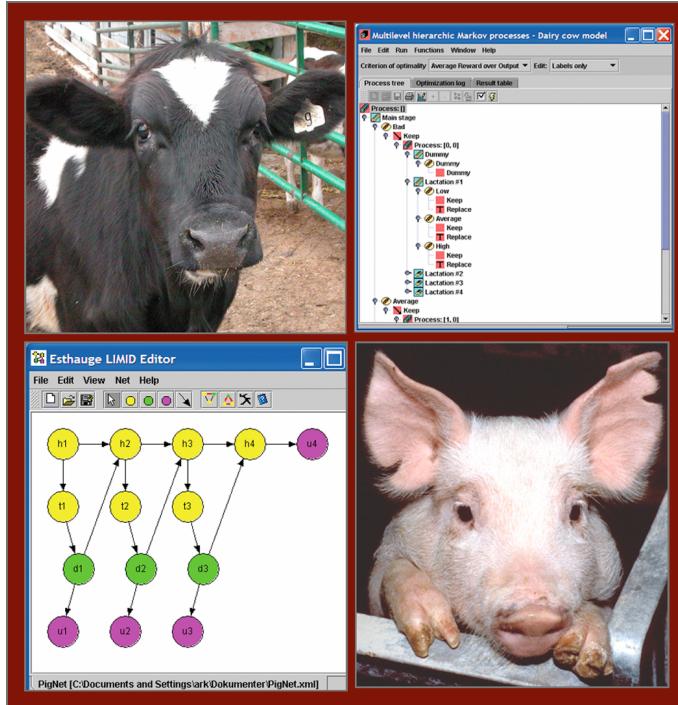


Herd Management Science



Preliminary edition

Compiled for the Advanced Herd Management course
KVL, August 28th - November 3rd, 2006

Anders Ringgaard Kristensen

Erik Jørgensen

Nils Toft

The Royal Veterinary and Agricultural University
Copenhagen, Denmark

Preface

In 1996 two of the authors organized a Nordic PhD course on *Planning and Control of Animal Production at Farm Level*. The third author, Nils Toft, attended the course as a PhD student. As part of the material for the course, 5 textbook notes were written and published as Dina Notes.

The notes were successfully used at the course in 1996, and afterwards they have served as important input to the Master level course on *Advanced Herd Management* given at the Royal Veterinary and Agricultural University. A couple of the notes have been slightly updated over the years, but basically the content has not changed.

Over the years, the authors have had the wish to collect and update the material in order to create an authoritative textbook covering more or less all aspects of herd management from the basic to the advanced level. This preliminary edition is the first attempt to realize this wish. Even though there is still a long way to go before the work is done, it is our hope that the preliminary version will turn out to be useful for the participants of the Advanced Herd Management course held at KVL in 2006.

The book is organized in two parts with basic herd management principles and classical theories in part I, and the more advanced methods in part II. Furthermore, the necessary mathematical and statistical theory is summarized in appendices for easy reference.

Part I has been written with a bachelor level course in mind, and the contents reflect what we think that any animal scientist should master, whereas Part II has been written for a more advanced level for graduate students who wish to specialize in Herd Management. Given the structural development in modern agriculture with ever increasing herd sizes, we expect that the need for an advanced textbook focusing directly on herd management will increase.

Copenhagen, August 2006

Anders Ringgaard Kristensen
Erik Jørgensen
Nils Toft

This book has a home page at URL: <http://www.dina.dk/~ark/book.htm>

Contents

I Basic concepts	15
1 Introduction	17
1.1 Definition of herd management science	17
1.2 The herd management problem	21
1.3 The management cycle	23
1.3.1 The elements of the cycle	23
1.3.2 Production monitoring and decision making	26
1.3.3 Levels	27
1.3.4 Time horizons	27
2 The role of models in herd management	29
2.1 The decision making process	29
2.2 Why models?	30
2.3 Knowledge representation	32
2.3.1 The current state of the unit	32
2.3.2 The production function	34
2.3.3 Other kinds of knowledge	35
2.3.4 Knowledge acquisition as a decision	36
3 Objectives of production: Farmer's preferences	37
3.1 Attributes of the utility function	37
3.2 Common attributes of farmers' utility functions	38
3.2.1 Monetary gain	38
3.2.2 Leisure time	39
3.2.3 Animal welfare	39
3.2.4 Working conditions	42
3.2.5 Environmental preservation	43
3.2.6 Personal prestige	43
3.2.7 Product quality	44
3.3 From attributes to utility	44
3.3.1 Aggregation of attributes, uncertainty	44
3.3.2 Single attribute situation: Risk	44
3.3.3 Multiple attribute situation	48

3.3.4	Operational representation of a farmer's utility function	52
4	Classical production theory	55
4.1	One to one relations	55
4.1.1	The production function	55
4.1.2	What to produce	56
4.1.3	How to produce	58
4.1.4	How much to produce	60
4.2	Multi-dimensional production models	62
4.3	Limitations of the neo-classical production theory	63
4.4	Costs	63
4.4.1	Variable costs and fixed costs: Gross margins	63
4.4.2	Opportunity costs and internal prices	64
4.4.3	Investments	67
4.5	Other classical techniques	68
4.5.1	Partial budgeting	68
4.5.2	Cost-benefit analysis	69
4.5.3	Decision tree analysis	70
4.6	Classical replacement theory	72
4.6.1	Replacement with identical assets	72
4.6.2	Optimal replacement using discounting	74
4.6.3	Technological or genetical improvement	75
4.6.4	Limitations and applications of the classical replacement theory	77
4.7	Exercises	79
4.7.1	Gross margin	79
4.7.2	Internal prices	80
4.7.3	Marginal considerations, I	81
4.7.4	Marginal considerations, II	82
4.7.5	Investments	82
5	Basic production monitoring	83
5.1	The framework	83
5.2	The elements of the monitoring process	85
5.2.1	Data recording	85
5.2.2	Database	86
5.2.3	Data processing	87
5.2.4	Report with key figures	89
	Types of key figures	89
	Key figures and their interrelations	90
	Properties of key figures	92
	Correctness	93
	Validity	94
	Precision	96

5.2.5	Analysis from a statistical point of view	97
General principles	97	
Continuous properties	98	
Categorical properties without observation error	102	
Categorical properties with observation error	104	
Concluding remarks	108	
5.2.6	Analysis from a utility point of view	108
5.2.7	Decision: Adjustment/no adjustment	111
When to adjust	111	
Where to adjust	111	
How to adjust	111	
5.3	Exercises	112
5.3.1	Milk yield	112
5.3.2	Daily gain in slaughter pigs	112
Valuation weighings	112	
Individual gains	113	
First in - first out	113	
Comparison of methods	113	
Numerical example	114	
5.3.3	Reproduction in dairy herds	115
5.3.4	Disease treatments	116
5.3.5	Validity, reproduction	117
5.3.6	Economic significance of days open	117
II	Advanced topics	119
6	From registration to information	121
6.1	Information as a production factor	121
6.2	Data sources: examples	122
6.2.1	The heat detection problem	122
6.2.2	The weighing problem	122
6.2.3	The activity problem	123
6.2.4	Miscellaneous	123
6.3	Basic concepts and definitions	123
6.3.1	Utility value of registrations	124
6.3.2	Entropy based value	127
6.4	Measurement errors	127
6.4.1	Registration failure	127
6.4.2	Measurement precision	128
6.4.3	Indirect measure	128
6.4.4	Bias	129
6.4.5	Rounding errors	130
6.4.6	Interval censoring	130

6.4.7	Categorical measurements	130
6.4.8	Enhanced models of categorical measurements	133
6.4.9	Pre selection	134
6.4.10	Registration of interventions	134
6.4.11	Missing registrations	135
6.4.12	Selection and censoring	135
6.4.13	Discussion	137
6.5	Data processing	137
6.5.1	Threshold level	138
6.5.2	Change point estimation	139
6.5.3	Significance test in experiments	139
6.6	Utility evaluation examples	140
6.6.1	Production control	140
6.6.2	Recording of sow specific litter size	142
6.6.3	Precision in weighing of slaughter pigs	144
6.7	Outlook	146
7	Dynamic production monitoring by classical methods	147
8	Dynamic production monitoring by state space models	149
9	Decisions and strategies: A survey of methods	151
9.1	Framework of the survey	151
9.2	Rule based expert systems	151
9.3	Linear programming with extensions	153
9.4	Dynamic programming and Markov decision processes	155
9.5	Probabilistic Expert systems: Bayesian networks	158
9.6	Decision graphs	159
9.7	Simulation	161
10	Modeling and sensitivity analysis by Linear Programming	167
10.1	Introduction	167
10.1.1	Phases of modeling	167
10.1.2	Formulating the problem	168
10.1.3	Constructing a Mathematical Model	168
10.1.4	Deriving a solution	169
10.1.5	Testing the model and the solution	169
10.1.6	Establishing control over the solution	169
10.1.7	Implementation	170
10.2	Using linear programming	170
10.2.1	The model setup	170
10.2.2	Proportionality	170
10.2.3	Additivity	171
10.2.4	Divisibility	171

10.2.5 Certainty	171
10.2.6 Circumventing the assumptions	171
10.2.7 Construction of constraints	172
10.2.8 Natural extensions of the LP-model	174
10.3 Sensitivity analysis	175
10.3.1 Shadow prices	175
10.3.2 Price adjustments	177
10.3.3 Capacity changes	178
10.3.4 Additional constraints	178
10.4 Concluding remarks	180
10.5 Exercise	180
11 Bayesian networks	183
12 Decision graphs: Potential use and current limitations	185
12.1 Introduction	185
12.2 From decision tree to influence diagram	186
12.3 From Bayesian network to influence diagram	187
12.4 From Dynamic Programming to influence diagrams	188
12.5 Examples	190
12.5.1 The two-sow problem	190
12.5.2 The registration problem for the whole cycle	191
12.5.3 The repeated measurement problem	194
12.5.4 Optimal timing of matings	195
12.5.5 The feed analysis problem	196
12.6 Outlook	198
13 Dynamic programming and Markov decision processes	199
13.1 Introduction	199
13.1.1 Historical development	199
13.1.2 Applications in animal production	201
13.2 Variability and cyclic production: Markov decision programming	202
13.2.1 Notation and terminology	203
13.2.2 A simple dairy cow replacement model	203
13.2.3 Criteria of optimality	205
Finite planning horizon	205
Infinite planning horizon	205
13.2.4 Optimization techniques in general Markov decision programming	207
Value iteration	207
The simple dairy model optimized by value iteration	208
Policy iteration	211
Linear programming	212
13.2.5 Discussion and applications	213

13.3	The curse of dimensionality: Hierarchical Markov processes	214
13.3.1	The curse of dimensionality	215
13.3.2	Notation and terminology	217
13.3.3	Optimization	220
13.3.4	Discussion and applications	221
13.3.5	The numerical example formulated and solved as a hierarchical Markov process	223
13.4	Uniformity: Bayesian updating	227
13.4.1	Principles of learning from observations	227
13.4.2	Applications	228
13.5	Herd restraints: Parameter iteration	228
13.6	General discussion	230
14	Monte Carlo simulation	233
14.1	Introduction	233
14.2	The simulation method	234
14.2.1	Systems approach	235
14.2.2	The purpose of simulation modeling	236
14.2.3	Dimensionality of simulation modeling	237
14.3	Why use the Monte Carlo Approach	238
14.3.1	The Deterministic Approach	238
14.3.2	The Markov chain approach	238
14.3.3	Conclusion	240
14.4	An object oriented model of a pig herd	240
14.4.1	General layout	241
14.4.2	Modeling the confinement	242
14.4.3	Modeling the animal and its physiological states	242
	State description	242
14.4.4	Modeling the measuring device	242
14.4.5	Modeling the belief management system	243
14.4.6	Modeling the decision support system	244
14.4.7	Modeling the manager and the staff	245
14.5	Representing model dynamics	245
14.5.1	Selection and effect of stepping method	246
14.5.2	Difference equations	246
14.5.3	Differential equations	247
	Solution of the equations	248
14.5.4	State transitions in continuous time	248
14.5.5	Competing events	249
14.5.6	Discussion	249
14.6	Stochastic specification	249
14.6.1	Example: use of results from an experiment	250
14.6.2	The principles	251
14.6.3	What happens if the a priori distribution is too vague	253

14.6.4	Initializing simulation model from herd data bases	253
14.6.5	Probability distributions	255
14.7	Random number generation	255
14.7.1	Generation of uniform random numbers	255
14.7.2	Sampling from probability distributions in general	256
14.7.3	The multivariate case	258
14.7.4	Metropolis-Hastings method	259
14.7.5	Sampling from a set of random deviates	259
14.8	Representing Decision Rules	260
14.9	Inference based on the model	261
14.9.1	Definition of sensitivity	261
14.9.2	Calibration of model and search for optimal strategies . . .	262
14.9.3	Stochastic search techniques	263
14.9.4	Variance reduction (Experimental planning)	264
14.10	Verification and model comparison	265
14.11	Advanced simulation methods	266
14.11.1	Importance Sampling	266
14.11.2	Sensitivity calculation	268
14.11.3	Calibration with observed values	268
14.11.4	Acceptance rejection sampling	269
14.12	Outlook	269
A	A short review of basic probability theory and distributions	289
A.1	The probability concept	289
A.1.1	Interpretation of probabilities	289
A.1.2	Experiments and associated concepts	290
A.1.3	Probability distributions on a sample space	290
A.1.4	Independence	291
A.1.5	Conditional probabilities and sum rule	292
A.1.6	Bayes' theorem	293
A.2	Discrete distributions	294
A.2.1	The binomial distribution	295
A.2.2	The hypergeometric distribution	296
A.2.3	The Poisson distribution	296
A.3	Continuous distributions	297
A.3.1	The uniform distribution	298
A.3.2	The normal distribution	298
A.3.3	The truncated normal distribution	300
A.3.4	The exponential distribution	300
A.4	Sampling from a distribution	301
A.5	Hypothesis testing	301
A.5.1	Normal distributions	301
A.5.2	Binomial distributions	302

B Selected topics from linear algebra	303
B.1 Real numbers	303
B.1.1 Elements of the real number system	303
B.1.2 Axioms for real numbers	303
B.1.3 Solving equations	304
B.2 Vectors and matrices	304
B.2.1 Definition, notation and examples	304
B.2.2 Special matrices, vectors	305
B.2.3 Addition operation	305
Definition	305
Associative and commutative laws	306
Additive identity: “Zero element”	306
Additive inverse	306
B.2.4 Multiplication operation	306
Definition	306
Laws for matrix multiplication	307
Vector multiplication	307
Multiplicative identity: “One element”	308
B.2.5 Other matrix operations	308
B.2.6 The multiplicative inverse of a quadratic matrix	309
B.2.7 Systems of linear equations	310
B.3 Linear regression	311
C “Advanced” topics from statistics	315
C.1 Covariance and correlation	315
C.2 Random vectors	315
C.3 Multivariate distributions	316
C.3.1 The multinomial distribution	316
C.3.2 The multivariate normal distribution	317
Definition	317
Conditional distribution of sub-vectors	317
C.4 Hyper distributions	318
C.4.1 Motivation	318
C.4.2 Conjugate families	319
Definition	319
Conjugate family of a binomial distribution	320
Conjugate family of a Poisson distribution	320
Other conjugate families	321
D Transition matrices for Chapter 13	323
D.1 Probabilities for the actions “Keep” and “Replace”	323

E Solutions to exercises	331
E.1 Solutions to exercises of Chapter 4	331
E.1.1 Gross margin	331
E.1.2 Internal prices	332
E.1.3 Marginal considerations, I	332
E.1.4 Marginal considerations, II	333
E.2 Exercises of Chapter 5	335
E.2.1 Milk yield	335
E.2.2 Daily gain in slaughter pigs	335
E.2.3 Reproduction in dairy herds	337
E.2.4 Disease treatments	338
E.2.5 Validity, reproduction	338
E.2.6 Economic significance of days open	339

Part I

Basic concepts

Chapter 1

Introduction

1.1 Definition of herd management science

Several points of view may be taken if we want to describe a livestock production unit. An animal nutritionist would focus on the individual animal and describe how feeds are transformed to meat, bones, tissues, skin, hair, embryos, milk, eggs, manure etc. A physiologist would further describe the roles of the various organs in this process and how the transformations are regulated by hormones. A biochemist would even describe the basic processes at molecular level.

A completely different point of view is taken if we look at the production unit from a global or national point of view. The individual production unit is regarded only as an arbitrary element of the whole livestock sector, which serves the purpose of supplying the population with food and clothing as well as manager of natural resources. A description of a production unit at this level would focus on its resource efficiency in food production and its sustainability from an environmental and animal welfare point of view.

Neither of these points of view are relevant to a herd management scientist even though several elements are the same. The herd management scientist also considers the transformation of feeds to meat, bones, tissues, skin, hair, embryos, milk, eggs and manure like the animal nutritionist, and he also regards the production as serving a purpose as we do at the global or national level. What differs, however, is the farmer. From the point of view of a herd management scientist, the farmer is in focus and the purpose of the production is to provide the farmer (and maybe his family) with as much welfare as possible. In this connection welfare is regarded as a very subjective concept and has to be defined in each individual case. The only relevant source to be used in the determination of the definition is the farmer himself.

The herd management scientist assumes that the farmer concurrently tries to organize the production in such a way that his welfare is maximized. In this process he has some options and he is subjected to some constraints. His options are to regulate the production in such a way that his welfare is maximized given the

constraints. The way in which he may regulate production is by deciding what factors he wants to use at what levels. A factor is something which is used in the production, i.e. the input of the transformation process. In livestock production, typical factors include buildings, animals, feeds, labor and medicine. During the production process, these factors are transformed into products which in this context include meat, offspring, milk, eggs, fur, wool etc. The only way a farmer is able to regulate the production - and thereby try to maximize his welfare - is by adjustments of these factors.

Understanding the factors and the way they affect production (i.e. the products and their amount and quality) is therefore essential in herd management. Understanding the constraints is, however, just as important. What the constraints do is actually to limit the possible welfare of the farmer. If there were no constraints any level of welfare could be achieved. In a real world situation the farmer faces many kinds of constraints. There are legal constraints regulating aspects like use of hormones and medicine in production, storing and use of manure as well as housing in general. He may also be restricted by production quotas. An other kind of constraints are of economic nature. The farmer only has a limited amount of capital at his disposal, and usually he has no influence on the prices of factors and products. Furthermore, he faces some physical constraints like the capacity of his farm buildings or the land belonging to his farm, and finally his own education and skills may restrict his potential welfare.

In general, constraints are not static in the long run: Legal regulations may be changed, quota systems may be abolished or changed, the farmer may increase or decrease his capital, extend his housing capacity (if he can afford it), buy more land or increase his mental capacity by training or education. In some cases (e.g. legal constraints) the changes are beyond the control of the farmer. In other cases (e.g. farm buildings or land) he may change the constraints in the long run, but has to accept them at the short run.

We are now ready to define herd management:

Definition 1 *Herd management* is a discipline serving the purpose of concurrently ensuring that the factors are combined in such a way that the welfare of the individual farmer is maximized subject to the constraints imposed on his production.



The general welfare of the farmer depends on many aspects like monetary gain (profit), leisure time, animal welfare, environmental sustainability etc. We shall denote these aspects influencing the farmer's welfare as attributes. It is assumed that the consequences of each possible combination of factors may be expressed by a finite number of such attributes and that a uniquely determined level of welfare is associated with any complete set of values of these attributes. The level of welfare associated with a combination of factors is called the utility value. Thus the purpose of herd management is to maximize the utility value. A function returning the utility value of a given set of attributes is called a utility function. In Chapter 3, the concept of utility is discussed more thoroughly.

The most important factors in livestock production include:

- farm buildings
- animals
- feeds
- labor
- medicine and general veterinary services
- management information systems and decision support systems
- energy

In order to be able to combine these factors in an optimal way it is necessary to know their influence on production. As concerns this knowledge, the herd management scientist depends on results from other fields like agricultural engineering, animal breeding, nutrition and preventive veterinary medicine. The knowledge may typically be expressed by a *production function*, f , which in general for a given *stage* (time interval), t , takes the form:

$$\mathbf{Y}_{s,t} = f_{s,i}(\mathbf{x}_{s,t}, \mathbf{x}_{s,t-1}, \dots, \mathbf{x}_{s,1}) + \mathbf{e}_t, \quad (1.1)$$

where $\mathbf{Y}_{s,t}$ is a vector of n products produced, $f_{s,i}$ is the production function, $\mathbf{x}_{s,t}$ is a vector of m factors used at stage t , and \mathbf{e}_t is a vector of n random terms. The function $f_{s,i}$ is valid for a given production unit s , which may be an animal, a group or pen, a section or the entire herd. The characteristics of the production unit may vary over time, but the set of observed characteristics at stage t are indicated by the *state* of the unit denoted as i . The state specification contains all relevant information concerning the production unit in question. If the function is defined at animal level the state might for instance contain information on the age of the animal, the health status, the stage of reproductive cycle, the production level etc. In some cases, it is also relevant to include information on the disease and/or production history of the animal (for instance the milk yield of *previous* lactation) in the state definition.

The total production \mathbf{Y}_t and factor consumption \mathbf{x}_t are calculated simply as

$$\mathbf{Y}_t = \sum_{s \in S} \mathbf{Y}_{s,t}, \quad (1.2)$$

and

$$\mathbf{x}_t = \sum_{s \in S} \mathbf{x}_{s,t}, \quad (1.3)$$

where S is the set off all production units s at the same level.

Eq. (1.1) illustrates that the production is only partly under the control of the manager, who decides the levels of the factors at various stages. The direct effects of the factors are expressed by the production function f , but the actual production also depends on a number of effects outside the control of the manager as for instance the weather conditions and a number of minor or major random events like infection by contagious diseases. These effects outside the control of the manager will appear as random variations which are expressed by \mathbf{e}_t . This is in agreement with the general experience in livestock production that even if exactly the same factor levels were used in two periods, the production would nevertheless differ between the periods.

An other important aspect illustrated by Eq. (1.1) is the *dynamic* nature of the herd management problem. The production at stage t not only depends on the factor levels at the present stage, but it may very well also depend on the factor levels at previous stages. In other words, the decisions made in the past will influence the present production. Obvious examples of such effects is the influence of feeding level on the production level of an individual animal. In dairy cattle, for instance, the milk yield of a cow is influenced by the feeding level during the rearing period, and in sows the litter size at weaning depends on the feeding level in the mating and gestation period.

The production \mathbf{Y}_t and factors \mathbf{x}_t used at a stage are assumed to influence the attributes describing the welfare of the farmer. We shall assume that k attributes are sufficient and necessary to describe the welfare. If we denote the values of these attributes at a specific stage t as $u_{1,t}, \dots, u_{k,t}$, we may logically assume that they are determined by the products and factors of the stage. In other words, we have:

$$\mathbf{u}_t = h(\mathbf{Y}_t, \mathbf{x}_t), \quad (1.4)$$

where $\mathbf{u}_t = (u_{1,t}, \dots, u_{k,t})'$ is the vector of attributes. We shall denote h as the *attribute function*. The over-all utility, U_N for N stages, t_1, \dots, t_N , may in turn be defined as a function of these attributes:

$$U_N = g(\mathbf{u}_{t_1}, \mathbf{u}_{t_2}, \dots, \mathbf{u}_{t_N}) \quad (1.5)$$

where g is the *utility function*, and N is the relevant time horizon. If we substitute Eq. (1.4) into Eq. (1.5), we arrive at:

$$U_N = g(h(\mathbf{Y}_{t_1}, \mathbf{x}_{t_1}), h(\mathbf{Y}_{t_2}, \mathbf{x}_{t_2}), \dots, h(\mathbf{Y}_{t_N}, \mathbf{x}_{t_N})). \quad (1.6)$$

From Eq. (1.6) we observe, that if we know the attribute function h and the utility function g and, furthermore, the production and factor consumption at all stages have been recorded, we are able to calculate the utility relating to any time interval in the past. Recalling the definition of herd management, it is more relevant to focus on the future utility derived from the production. The only way in which the farmer is able to influence the utility is by making decisions concerning the factors. Having made these decisions, the factor levels \mathbf{x}_t are known also for future

stages. The production levels \mathbf{Y}_t , however, are unknown for future stages. The production function may provide us with the expected level, but because of the random effects represented by \mathbf{e}_t of Eq. (1.1), the actual levels may very well deviate from the expected. An other source of random variation is the future state i of the production unit. This becomes clear if we substitute Eq. (1.1) into Eq. (1.5) (and for convenience assume that the production function is defined at herd level so that $s = S$):

$$U_N = g(h(f_{s,i_1}(\mathbf{x}_{t_1}, \dots, \mathbf{x}_1) + \mathbf{e}_{t_1}, \mathbf{x}_{t_1}), \dots, h(f_{s,i_N}(\mathbf{x}_{t_N}, \dots, \mathbf{x}_1) + \mathbf{e}_{t_N}, \mathbf{x}_{t_N})). \quad (1.7)$$

From Eq. (1.7) we conclude, that even if all functions (production function, attribute function and utility function) are known, and decisions concerning factors have been made, we are not able to calculate numerical values of the utility relating to a future period. If, however, the distributions of the random elements are known, the *distribution* of the future utility may be identified.

When the farmer makes decisions concerning the future use of factors he therefore does it with incomplete knowledge. If, however, the distribution of the possible outcomes is known he may still be able to make rational decisions as discussed in the following chapters. Such a situation is referred to as *decision making under risk*.

1.2 The herd management problem

Based on the definition of herd management (Definition 1) the herd management problem may be summarized as maximization of the farmer's utility U_N with respect to the use of factors $\mathbf{x}_1, \dots, \mathbf{x}_{t_N}$ subject to the constraints imposed on production.

From Eq. (1.7) and the discussion of Section 1.1 we may logically conclude that in order to be able to make rational decisions concerning a certain unit (animal, group, herd) he needs the following knowledge:

1. The state of the production unit at all stages of the planning horizon.
2. The production function(s) and the distribution(s) of the random term(s).
3. All attribute functions relevant to the farmer's preferences.
4. The utility function representing farmer's preferences.
5. All constraints of legal, economic, physical or personal kind.

As concerns item 1, it is obvious, that the future state of the production unit is not always known at the time of planning. The state represents all relevant information on the unit (i.e. the traits of the unit) and very often it varies at random

over time. It is therefore likely, that it is not possible to make decisions relating to all stages at the same time. Instead, we may choose a strategy (or policy as it is also some times denoted) which is a map (or function) assigning to each possible state a decision. In other words, having chosen a strategy we may at all stages observe the state of the unit and make the decision provided by the strategy.

The identification of an optimal strategy is complicated by the fact that decisions may also influence the future state of the unit as illustrated by the following example. Let us assume that the unit is an animal, that the relevant information (the state) is the weight, i , of the animal and what we have to decide is the feeding level, x , of the animal. It is obvious, that the optimal level of x will depend on the weight of the animal, i , but it is also obvious, that the feeding level, x , will influence the future weight, j , of the animal.

A thorough inspection of Eq. (1.7) will unveil that the decision made may also influence the future production, but, since the state i is assumed to contain all relevant information on the production unit, we may just include information on previous decisions in the definition of the state. Thus we do not need to consider that aspect further. Consider again the example of Section 1.1, where we noted that the feeding level of a heifer influences the future milk production as a dairy cow. If we denote the feeding level at various stages as x_1, x_2, \dots , the stage of first calving as k and the milk yield of stage n as y_n we might express the production function in the same manner as in Eq. (1.1):

$$y_{k+n} = f_{si}(x_{k+n}, \dots, x_1) + e_{k+n}, \quad (1.8)$$

where the state i for instance may be defined from combinations of traits like lactation number, stage of lactation and body weight. Thus the *state space* Ω_1 is defined as the set of all possible combinations of those individual traits, and an individual state $i \in \Omega_1$ is any combination of values representing the three traits. If, however, we redefine the state space to include also the feeding levels of previous stages (i.e. $\Omega_2 = \Omega_1 \cup \{x_1, \dots, x_{k+n-1}\}$) exactly the same information may be contained in the relation

$$y_{k+n} = f_{si}(x_{k+n}) + e_{k+n}, \quad (1.9)$$

where, now, $i \in \Omega_2$.

Eq. (1.9) illustrates that the state concept is very essential for the planning process. The identification of the state of the production unit is therefore a very important task. In order to be able to identify the state we have to perform some registrations concerning the production unit. It may some times happen that we are not able to observe the traits of the state space directly, but only indirectly through other related traits. The precision of our knowledge concerning the true state may therefore vary, but Bayesian updating methods are available for handling such situations with imperfect knowledge. Those aspects are dealt with in a later chapter.

We may now reformulate the information needs when choosing an optimal strategy:

Definition 2 *The information needs for making rational decisions include:*

1. The current state of the unit.
2. A production function describing the immediate production given stage, state and decision and the distribution of the possible random term(s).
3. The distribution of the future state given stage, state and decision.
4. All attribute functions relevant to the farmer's preferences.
5. The utility function representing farmer's preferences.
6. All constraints of legal, economic, physical or personal kind.

■

As concerns items 4 and 5, reference is made to Chapter 3.

1.3 The management cycle

1.3.1 The elements of the cycle

Herd management is a cyclic process as illustrated by Figure 1.1. The cycle is initiated by identification of the farmer's utility function as discussed in Chapter 3. Also the constraints have to be identified no matter whether they are of the legal, economic, physical or personal kind discussed in Section 1.1. The number of constraints will depend heavily on the time horizon considered. If the time horizon is short, the farmer faces more constraints of economic, physical and personal nature than when he is considering a long time horizon.

Having identified the farmer's utility function and the relevant constraints, one or more *goals* may be defined. It is very important not to confuse goals with the attributes of the utility function. The attributes represent basic preferences of the farmer, and they are in principle invariant, or - to be more precise - they only vary if the farmer's preferences change (for instance he may give higher priority to leisure time or working conditions as he becomes older). Goals, on the other hand, may change as the conditions change. They are derived from the farmer's preferences in combination with the constraints, and since we noted in Section 1.1 that, for instance, legal constraints may very well change over time, the same of course apply to goals. The purpose of goals is only to set up some targets which (if they are met) ensure that under the circumstances defined by farmer's preferences and the constraints the production will be successful. In practice, goals may be expressed as a certain level of production, a certain efficiency etc.

It should be noticed, that goals are often defined as a result of planning under a longer time horizon than the one considered. Traditionally, three different time

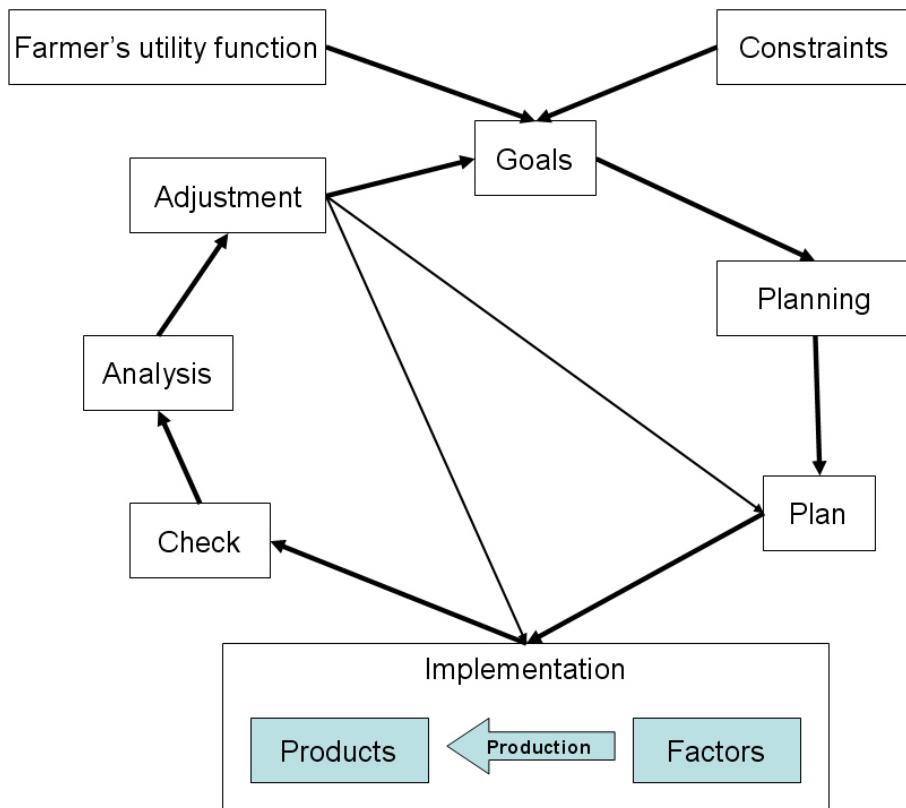


Figure 1.1: The elements of the management cycle.

horizons (levels) are distinguished. The *strategic* level refers to a long time horizon (several years), the *tactical* level refers to an intermediate time horizon (from a few months to a few years) and the operational level refers to a short time horizon (days or weeks). Thus, goals for the *operational* level are typically defined at the tactical level.

When the goals have been defined, the process of *planning* may be initiated. The result of the process is, of course, a *plan* for the production. A plan is a set of decided actions each concerning the future allotment of one or more factors. Alternative actions are evaluated on their expected utility as discussed in Chapter 3. Accordingly, the expected resulting production from each plan has to be known in order to be able to evaluate the utility (cf. Eq. (1.6)). So, what the plan actually contains is a detailed description of the factor allotment and the expected resulting production.

The next element of the management cycle is *implementation*. From a theoretical point of view this element is trivial (but certainly not from a practical). Implementation is just to carry out the actions described in the plan, and during the production process the factors are transformed into products.

During the production process, some *registrations* are performed. The registrations may refer to factors as well as products. Based on the registrations, some *key figures* (e.g. average number of piglets per farrowing, average milk yield per cow) describing the performance of production may be calculated. During the production *check* these calculated key figures are compared to the corresponding expected values known from the planning process.

The result of the comparison may either be that the production has passed off as expected or that one or more deviations are identified. In the first case, the production process is continued according to the plan. In the latter case, the deviations have to be *analyzed*. The purpose of the analysis is to investigate whether the deviations are significant from (a) a statistical point of view **and** (b) from a utility (often economic) point of view. Because of the random elements of the production function (e_t of Eq. (1.1)), and because of observation errors relating to the method of measurement it may very well happen that even a considerable observed deviation from a statistical point of view is insignificant. That depends on the magnitudes of the random elements and the sample size (for instance the number of animals). Even if a deviation *is* significant from a statistical point of view (because of small random variation and a large sample) it may still turn out to be insignificant from a utility point of view. In Chapter 5, the statistical analysis is discussed in more details.

If it is concluded during the analysis that a deviation is significant from a statistical point of view as well as a utility point of view, some kind of adjustment may be relevant. Depending on the nature of the deviation, the adjustment may refer to the goals, the plan or the implementation. If the deviation concerns the factor allotment, the implementation has to be adjusted. During the planning process certain factor levels were assumed, but through the check it was revealed that the

actual factor allotment was different. Accordingly, something went wrong during the implementation of the plan.

If the deviation concerns the output level (i.e. the products), the conclusion depends on whether or not a deviation concerning the factor levels was found simultaneously. In that case, the deviation in output level is probably only a result of the deviation in input level. Accordingly the adjustments should focus on the implementation process.

If, however, there is a deviation concerning output, but none concerning input, we really face a problem. During the planning process, we assumed that if we used the factors represented by the vector \mathbf{x} we could expect the production $E(\mathbf{Y})$. Now, the control process show that the actual factor allotment was \mathbf{x} , but the actual production was \mathbf{Y}' which differs significantly from $E(\mathbf{Y})$ both from a statistical and a utility point of view. If we consult Eq. (1.1), we have to conclude, that the only possible explanation is that we have used a deficient production function $f_{s,i}$ (in other words, the *validation* of the model used has been insufficient). The reason may be that the state i of the production unit s differed from what we assumed during the planning process, or it may simply be because of lacking knowledge on the true course of the production function. Under all circumstances, the situation calls for a new planning process where correct assumptions are made. Accordingly the adjustments refer to the plan. During the new planning process it may also turn up, that one or more goals are impossible to meet, and in that case they have to be adjusted.

Finally, it should be emphasized that if any constraints (legal, economic, physical or personal) changes, new goals have to be defined and new plans must be made.

1.3.2 Production monitoring and decision making

An other way of looking at the elements of herd management is to distinguish between *production monitoring* and *decision making*.

Production monitoring involves check and analysis of the production results. It is therefore based on results already obtained (and thus looking backwards). Decision making, on the other hand, is based on expected future results (and thus looking forwards). It involves adjustment of the goals, plans or implementation (which is actually something which is decided) and planning. Since it is based on expected future results, *prediction* plays an important role. The best basis for prediction is the results obtained so far in combination with external information. In other words, production monitoring is a necessary foundation for decision making. This fact illustrates how closely connected the elements of the management cycle actually are.

1.3.3 Levels

In herd management we face a hierarchy of decisions made at different levels with different time horizons. In this section, and the following section, we shall consider the implications of this hierarchy.

By *level* we mean the production unit considered. As discussed in Section 1.1, the unit may be an individual animal, a group or a pen, a section or even the entire herd. For instance, the decision to build a new barn is an example at herd level. Decisions concerning feeding are typically made at group or section level whereas decisions concerning culling or medical treatment may be made at animal level.

As long as decisions at different levels are mutually independent we may solve a problem at one level without considering the other level. Unfortunately it is very rare that such independence exists as the following example shall illustrate.

We want to build a new barn for our dairy cows, and two alternatives, a^1 and a^2 , are available. The two barns of course differ in several respects, but one of the differences is that barn a^1 allows for grouping of the cows whereas grouping is not possible in barn a^2 . This difference means that in order to make an optimal decision concerning the kind of barn to build, we also have to consider how to feed the animals in each of the two systems, because the option of grouping makes other feeding strategies possible than if all cows are housed in a single group. It then turns out that feeding strategy α_1 is optimal in barn a^1 whereas strategy α_2 is optimal in barn a^2 .

Unless α_1 has exactly the same utility consequences as α_2 , the differences must be taken into account when the optimal barn is chosen.

1.3.4 Time horizons

A general aspect of herd management under risk is that decisions have to be made without certainty about the future state of the production unit. The uncertainty increases with the time horizon of the decision, i.e. it is more prevalent at the tactical level than at the operational level. Having made a decision at the tactical level, the manager is restricted by the consequences for the duration of the time horizon. It may very well later turn out, that the actual state of the production unit differs from the expected state at the time of the decision, but the only way the manager may adjust to the new situation is by making decisions at the operational level. These decisions should be conditionally optimal given the tactical decision made and the current state of the production unit. In other words, the decisions at the operational level may be regarded as a way of adjustment to risk and in that way compensate for the incomplete knowledge on the future state of the production unit.

In general, it must be assumed that if decision a_1 is made at the tactical level, then strategy α_1 is optimal for decisions at the operational level (a strategy is defined as a set of decisions relating to the set of possible states of the production unit). On the other hand, if decision a_2 is made at the tactical level, then strategy

α_2 is optimal at the operational level. It will be an exception, if $\alpha_1 = \alpha_2$. In other words, it is not possible to choose an optimal decision a' at the tactical level, unless a conditionally optimal strategy α' has been determined at the operational level (conditional given the tactical decision).

In case of a management problem with limited time horizon (for instance the duration of the tactical decision considered) the mutual dependency between decisions at the tactical and operational level is not really a problem. We just have to determine optimal policies at the operational level given each of the alternative tactical decisions and, afterwards, to choose the tactical decision maximizing the objective function. A problem corresponding to this situation is discussed by Jensen (1995), who considered optimal mildew management policies in winter wheat under different nitrogen fertilization strategies. In that example, the decision at the tactical level is to choose a nitrogen fertilization strategy and the decisions at the operational level are to treat the crop for mildew. The time horizon is limited to the growing season just like the tactical decision. The problem was solved within the framework of a Bayesian network in combination with a usual backwards dynamic programming algorithm.

If, however, the time horizon is unknown or at least very long as it is typically the case in animal production, the situation is far more complicated. Examples of tactical decisions include mating of a female animal with a male animal of a certain quality or choosing a certain feeding level for an animal. Such decisions have (depending on the animal species and other circumstances) a time horizon of a few months, but unlike the mildew management problem, the time horizon of the production is not limited to a growing season or the like. Instead the production is continuous, which is often modeled by an infinite time horizon. In order to cope with such a situation, the decisions at the tactical and operational level have to be optimized simultaneously in order to ensure over-all optimality.

The terms "tactical" and "operational" are of course rather arbitrary. In general we have to deal with decisions at several levels having different time horizons.

Chapter 2

The role of models in herd management

2.1 The decision making process

From the very definition of herd management (Definition 1) it is clear that making decisions concerning what actions to take is the core element. But, before decisions can be made, the necessary knowledge as described in Section 1.2 must be collected. We may distinguish *general* knowledge from *context specific* knowledge.

By general knowledge, we refer to knowledge of the kind that can be found in textbooks of animal nutrition, animal breeding, agricultural engineering etc. Obvious examples are what feedstuffs are relevant for different groups of animals and how to evaluate the feedstuffs (i.e. what properties are relevant to consider). This general knowledge also includes the common forms of production functions, growth curves, lactation curves etc.

By context specific knowledge, we refer to knowledge that relates directly to the decision problem in question. This includes information about the current state of the unit and the effect of the state on the numerical part of the production function. The source of the context specific knowledge is observations done in production. In other words, there is a direct path from data observed in production to decisions on the basis of data. This path of information may be illustrated as in Figure 2.1. In the figure, we distinguish between data and information. The term “data” is used for the raw registrations whereas “information” is data which has been processed (and most often heavily reduced) to a form that can be used as a basis for decision making.

As an illustration of the path shown in Figure 2.1 we can consider the feeding decision in dairy cows. The data could be individual test day measurements of milk yield, the information could be average daily milk yield during the first 24 weeks of lactation, and the decision is the composition of the feed ration. In this example, the first processing (from data to information) involves interpolation of



Figure 2.1: The path from data to decision.

the individual lactation curves from test day yields and calculation of the daily average. The second processing (from information to decision) involves for instance calculation of a least costs ration by linear programming with constraints based on the average milk yield. Both processings are rather computationally demanding (but methodologically trivial) and illustrate the complexity of the decision making process.

As already mentioned, the decision process is complicated by the fact that most of the knowledge is associated with uncertainty. As an example of uncertainty of state, consider the state “pregnant” of a dairy cow. If the cow has not been inseminated, the probability that it is pregnant is zero. If the farmer knows that the cow was inseminated a week ago, he will assume it to be pregnant with a probability equal to the conception rate in the herd which we shall assume is 0.4. If, after 5 weeks, the cow has not shown heat, the probability of the cow being pregnant increases. If the heat detection rate of the herd is 0.5, the probability that the cow is pregnant becomes 0.7. If, furthermore, the cow is later tested for pregnancy (and the test is positive), the probability that it is pregnant increases even further, but it is still not equal to 1 because of the uncertainty related to pregnancy diagnoses. This example illustrates that what we want to know is the pregnancy state of the animal which we are not able to observe directly. Instead we observe the outcomes of events like heat detection and pregnancy diagnoses which increase our knowledge, but never totally remove uncertainty.

Also the relation between factors and products is associated with uncertainty. When we for instance feed a certain ration to a dairy herd we do not know the exact outcome measured as milk and growth. At best we know the expected outcome and a relevant distribution describing the likelihood of the possible deviations from the expected values.

Making rational decisions in herd management is therefore a very complex task where the farmer has to combine knowledge from many different sources and further take uncertainty into account. In this book, we shall introduce mathematical models as relevant tools for production monitoring and decision making in livestock herds.

2.2 Why models?

The traditional way of supporting farmers making decisions has been either through general norms, standards and recommendations or through advice given by experts

(agricultural consultants or veterinarians). The first method completely ignores the fact that production systems, farmers' preferences and constraints vary from case to case. Thus decisions based on general standards are typically non-optimal in the individual case. The strength of the expert is the ability to include knowledge on those individual conditions in the decision making process. The weakness of the expert on the other hand is the very limited capacity of humans to combine information from different sources in a consistent way. Furthermore, the ability of an expert to deal with uncertain information is usually not good.

The appearance of personal computers in the eighties soon lead to great expectations for using computers as a tool for decision support in agricultural production. The idea was to bring the knowledge of the experts directly to the dairy farmers by means of the so-called rule-based expert systems. McKinon and Lemmon (1985) concluded that expert systems "... are the ideal conduit of new knowledge from the agricultural scientists' laboratory to usage at the farm level ..." and "... once developed they can raise the performance of the average worker to the level of an expert". In other words, an expert system tries to mimic a human expert (and ultimately replace him). The philosophy behind is an almost blind belief in expert knowledge. As a management scientist realizing the complexity of decision making such a belief is very hard to share. Under all circumstances it may be concluded that an expert system will never be better than the experts it mimics.

A promising way to circumvent the shortcomings of general standards and expert advice is to construct a model representing the production unit and the decision problem. The main advantages of model based decision support include in the ideal situation the ability to take individual conditions into account, a concise framework for combination of information from different sources, direct representation of uncertainty and efficient search algorithms for determination of optimal decisions. Most models may be described by their *structure* representing general knowledge and their *numerical contents* representing context specific information.

Models may further contribute with extensive sensitivity analyses concerning optimal decisions, deviating conditions and parameter values. It is therefore believed by many herd management scientists that models may provide better decisions than experts, and that is indeed the reason for constructing and using them. Disadvantages mostly refer to the model building process which may be a very demanding task and to the operational use of the models which (because of the complexity of real world problems) may require computer calculations of vast dimensions.

In order to be used for decision support a model must represent all kinds of knowledge needed for making the decision in question. Furthermore, the precision of the knowledge ("level of uncertainty") has to be represented for a correct evaluation of alternative decisions. How to represent knowledge is the first important issue in model construction. A second and a third important issue are how to obtain the desired knowledge and what algorithms to use in search of optimal decisions.

2.3 Knowledge representation

Models are in general used for representation of all kinds of knowledge mentioned in Definition 2. They furthermore cover the processing of data into information and the processing of information into decisions as illustrated in Figure 2.1. In the following subsections we illustrate some principles of knowledge representation (including representation of the uncertainty).

2.3.1 The current state of the unit

Let us assume that the relevant unit is a dairy cow and we have to make a decision concerning for instance culling, insemination or medical treatment. Such decisions are all based on the expected future performance of the animal which in turn depends on the observed performance until now. The distinction between observed and future performance is important. Historical records are only relevant as a basis for prediction, since it is a well known fact that it is not possible to change the past. In other words, when we choose the traits to be represented in a model we want the best possible basis for prediction. Most often the state of an animal is modeled as a set of variables each representing a relevant trait like parity, stage of lactation, observed milk yield, days open, mastitis history etc. We shall refer to such variables as “state variables” included in the “state space”.

It is important to realize that the best variables to be used for prediction of future performance are often neither observed nor observable. As an example, consider the trait “milk yield”. The reason for including a measure of milk yield as a state variable is to be able to predict future milk yield. What we observe in practise is test day milk yield, but a more relevant variable to use for prediction would be some kind of abstract “milk yield capacity” of the cow in question. This is common knowledge to any dairy farmer who uses estimates like 305 days milk yield for evaluation of his cows, because individual test day milk yields are subject to large random fluctuations and, furthermore, depend very much on stage of lactation.

We therefore conclude that a variable measuring the milk yield capacity of the cow is relevant. An other obvious state variable is the pregnancy state of the cow. Neither of these variables are observable, but they are essential in the prediction of the future performance. In Figure 1 the relation between these unobservable state variables and a number of directly observable variables is illustrated. An ellipse in the figure is a random variable, and an arrow indicates a causal relation between two variables. Variables labeled by an asterisk are observable. Others are not.

The abstract milk yield capacity is the combined effect of genotype and permanent environment. Since the observed test day milk yields depend on the capacity there are arrows from the capacity to individual monthly test day records. But the over-all capacity is not the only effect influencing test day milk yield. Also temporary environmental factors (“Temp 1”, …, “Temp 6” in the figure) influence the milk yield on a specific day. It is reasonable to assume that short term environmen-

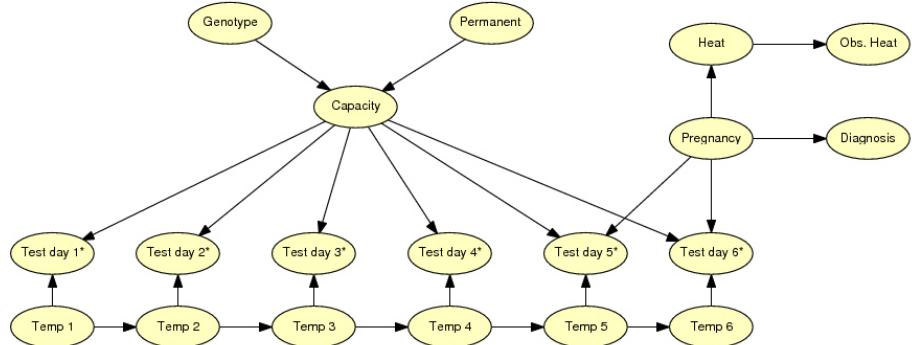


Figure 2.2: A conceptual model describing the milk yield, pregnancy and heat state of a dairy cow. Only variables marked by an asterisk are directly observable.

tal effects are auto-correlated which is illustrated by arrows from each effect to the following. In short term predictions, an estimate of the temporary environmental effect combined with an estimate of the over-all capacity is necessary. In long term predictions the capacity may be sufficient.

As concerns the pregnancy state of the cow we logically observe that pregnancy may influence test day milk yield (at least in the late part of lactation), but of course not the over-all capacity. Also for this variable we should notice that it is unobservable. We may look for heat and perform pregnancy diagnosis, but neither method is exact. It is very important in prediction, that the uncertainty is represented. Otherwise there is an obvious risk of biased estimates.

Until now we have only described the relations in general terms, but the network shown in Figure 2.2 may also be implemented numerically, so that each time a new test day milk yield is recorded, the capacity and the temporary effect are re-estimated. The tool for this kind of knowledge representation is Bayesian Networks as described by Jensen (2001). Bayesian Networks represent a very promising tool for knowledge representation and knowledge management in animal production models. Various software tools are available for construction and implementation of Bayesian Networks. There are already several examples of applications to dairy cattle. Dittmer (1997) uses the technique for prediction of calving dates; Hogeveen et al. (1994) as well as Vaarst et al. (1996) used Bayesian Networks for mastitis diagnosis; Rasmussen (1995) constructed a system for verification of parentage for Jersey cattle through blood type identification, and McKendrick et al. (2000a) used the method for diagnosis of tropical bovine diseases.

Although Bayesian networks are obvious to use in this context we should realize that other methods are also available. In particular Kalman filtering techniques have been used for estimation of unobservable variables. The whole setup shown

in Figure 2.2 (except for the pregnancy related variables) has been formulated by Goodall and Sprevak (1985). In their model, the over-all capacity was represented by the 3 parameters of the Wood lactation curve. Each time the test day milk yield was recorded the parameters and the temporary effect were re-estimated providing a good basis for prediction of future milk yield. Other examples of similar techniques are Thysen (1993a) who monitored and predicted bulk tank somatic cell counts; Thysen and Enevoldsen (1994) who monitored heat detection rates and pregnancy rates at herd level, and de Mol et al. (1999) who used the Kalman filer for analysis of sensor-based milk parlour data. Finally, Madsen et al. (2005) as well as Madsen and Kristensen (2005) used the method for monitoring the drinking pattern of growing pigs.

The strength of these Bayesian methods (Kalman filtering and Bayesian networks) is that uncertainty about the true level of important variables is directly represented. Furthermore, the methods provide a concise framework for combination of knowledge from different sources thus increasing the precision of knowledge as further observations are done.

2.3.2 The production function

As discussed in Chapter 1, the relation between factors and products is in usually represented by a production function taking the factor levels as arguments and returning the resulting production. As an example, we shall consider the relation between feeding and milk production in dairy herds. Assume that the cows are fed a ration consisting of n feedstuffs in the amounts x_1, x_2, \dots, x_n . In principle, the production function should take x_1, x_2, \dots, x_n as arguments, but our general knowledge tells us that it is more relevant to consider the amounts of *energy*, z_1 , *protein*, z_2 , and *fat*, z_3 . If the nutritional contents of the individual feedstuffs are known, we may easily calculate z_1, z_2 and z_3 from x_1, x_2, \dots, x_n . If, for instance, the energy concentration of feedstuff i is ρ_{1i} , the total energy content z_1 becomes

$$z_1 = \sum_{i=1}^n \rho_{1i} x_i. \quad (2.1)$$

A quadratic production function taking the energy, protein and fat contents of the feed ration as arguments was estimated already by Thysen (1983). If y is the milk yield, the relation is $y = f(z_1, z_2, z_3)$, where the production function f may be written as

$$\begin{aligned} f(z_1, z_2, z_3) &= c_1 z_1 + c_{11} z_1^2 \\ &+ c_2 z_2 + c_{22} z_2^2 \\ &+ c_3 z_3 + c_{33} z_3^2 \\ &+ c_{12} z_1 z_2 + c_{13} z_1 z_3 + c_{23} z_2 z_3. \end{aligned} \quad (2.2)$$

where the c s are constants. Since Eq. (2.2) is a deterministic function we conclude that the value returned for the milk yield y is only the expected value. In practise,

the actually observed milk yield may vary considerably for the same feed ration. The first step in order to represent this uncertainty is to include a random term e in the equation, so that the random milk yield Y becomes

$$Y = f(z_1, z_2, z_3) + e \quad (2.3)$$

where $E(Y) = y$. Usually a normal distribution may be assumed for the random term e . This representation is a considerable improvement. If the uncertainty is ignored, the influence of feeding on milk yield will be over-estimated and wrong decisions may be made. However, even Eq. (2.3) ignores a significant part of the uncertainty. It assumes implicitly that the true contents of energy, protein and fat are known. In practise that is not the case. Cows are fed a mixture of some kind of roughage (for example silage) and concentrates. In particular the nutritional contents of the roughage is uncertain. The values used as a basis for ration formulation may either be standard values from tables or they may origin from some kind of analysis of a sample of the roughage.

If we simplify the problem to feeding a mixture of silage, x_1 , and concentrate, x_2 , and only consider energy, the uncertainties may be illustrated as in Figure 2.3. Just like in Figure 2.2, each ellipse represents a state variable. Again, those marked by an asterisk are observable, others are not. The variable “Milk yield*” is at herd level and therefore directly observable. The variables “Silage true”, “Concentr.*”, “Silage obs.*” and “Ration” all refer to the energy content. In particular, “Silage true” and “Concentr.*” are the variables ρ_{11} and ρ_{12} in Eq. (2.1).

Since we do not know the true energy content of the silage, we neither know the true energy content of the full ration (the uncertainty concerning the concentrate is ignored). We have to rely on the observed energy content of the silage. The precision of the observed value will depend very much on the method used for observation. If we just use standard values for silage the precision will be low. If we include information on the dry matter content the precision will increase and a laboratory analysis will give even higher precision. The problem is similar to the pregnancy diagnosis problem shown in Figure 2.2: - Several methods are available each providing its own level of precision.

Just as Figure 2.2 may be implemented numerically using Bayesian networks, the same applies to Figure 2.3. In particular it should be noticed that the arrow from “Ration” to “Milk yield*” numerically represents the production function f and the random term e as specified in Eqs. (2.2) and (2.3). The “Herd size*” variable is just a scaling factor.

2.3.3 Other kinds of knowledge

As concerns the attribute functions and the utility function of the farmer, models also play an important role. Reference is made to Chapter 3 for a detailed discussion.

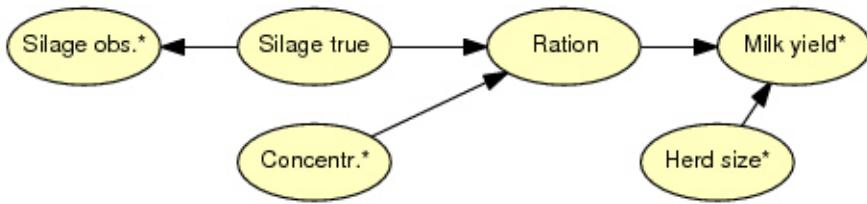


Figure 2.3: Relevant state variables describing the energy content and the resulting milk yield of a ration fed to dairy cows. Observable variables are marked by an asterisk.

For other kinds of knowledge (distribution of future states, constraints) the representation depends heavily on the modeling technique and reference is made to later chapters.

2.3.4 Knowledge acquisition as a decision

In this section we shall focus on acquisition of the context specific knowledge related to the state of the unit considered. General knowledge about relevant feed-stuffs, medical treatments etc. is not discussed. Knowledge about states is obtained through observations of the relevant unit. It is important to realize that to observe the unit is actually a decision. The purpose of observation is to improve knowledge, but on the other hand, observations are always associated with costs (either in money or time). In other words, making observations may be regarded as a production factor like animals and feedstuffs. Just like we want to determine an optimal feeding level we must try to determine an optimal observation level. Having identified what trait of the unit is of interest (for example “pregnancy state” or “energy content”), the decisions involved include at least

- what exactly to observe?
- how to observe?
- how often to observe?

These 3 questions may together be referred to as the *method*, which has already been briefly discussed in relation to the feedstuffs example of Figure 2.3. Since different methods have different precision and represent different costs, the choice of method is certainly not irrelevant. When making this kind of decision, the economic value of information is important, but unfortunately, very little attention has been paid to that matter in herd management research. An exception is a very interesting research conducted by Jørgensen (1985). His results showed remarkable diminishing returns to scale for increasing number of traits observed.

Chapter 3

Objectives of production: Farmer's preferences

3.1 Attributes of the utility function

A general characteristic of an attribute is that it directly influences the farmer's subjectively defined welfare and, therefore, is an element of the very purpose of production. This may be illustrated by a few examples. The average milk yield of the cows of a dairy herd is not an attribute of a utility function, because such a figure has no direct influence on the farmer's welfare. If, however, he is able to sell the milk under profitable conditions he will experience a monetary gain, which certainly may increase his welfare and, therefore, may be an attribute. In other words, the purpose of production from the farmer's point of view could never be to produce a certain amount of milk, but it could very well be to attain a certain level of monetary gain.

Animal welfare may in some cases be an attribute of the utility function. Whether or not it is in the individual case depends on the farmer's reasons for considering this aspect. An argument could be that animals at a high level of welfare probably also produce at a higher level and thereby increase the monetary gain. In that case, animal welfare is just considered as a shortcut to higher income, but it is not considered to be a quality by itself. Accordingly, it should not be considered to be an attribute. If, on the other hand, the farmer wants to increase animal welfare even if it, to some extent, decreases the levels of other attributes like monetary gain or leisure time then it is certainly relevant to consider it to be an attribute of the utility function.

This discussion also illustrates that attributes are individual. It is not possible to define a set of attributes that apply to all livestock farmers. In the following section, however, we shall take a look at some examples of typical attributes of farmers' utility functions.

3.2 Common attributes of farmers' utility functions

Typical attributes describing the welfare of a livestock farmer include:

- Monetary gain
- Leisure time
- Animal welfare
- Working conditions
- Environmental preservation
- Personal prestige
- Product quality

In this section, we shall discuss each of these attributes and focus on how to define relevant attribute functions in each case and how to compare contributions from different stages.

3.2.1 Monetary gain

It seems very unlikely that monetary gain should not be an attribute of all professional farmers' utility function. In many practical cases it is even the only one considered when decisions are made. Usually, it is very easy to define the attribute function of monetary gain. If we assume product prices p_{y_1}, \dots, p_{y_n} and factor prices p_{x_1}, \dots, p_{x_m} to be fixed and known, the partial attribute function simply becomes (assuming that monetary gain is the 1st attribute):

$$u_{1t} = h_1(\mathbf{Y}_t, \mathbf{x}_t) = \sum_{i=1}^n p_{y_i} Y_{ti} - \sum_{i=1}^m p_{x_i} x_{ti}. \quad (3.1)$$

Monetary gains from different stages are not directly comparable. If a certain amount of money is gained at the present stage, it is possible to invest it and earn interest, so that the amount has increased at the following stage. The usual way to account for that is by discounting, so that we deal with the *present value* of future monetary gains. If we assume all stages to be of equal length, and the interest rate to be constant, the present value of monetary gains from N stages is calculated as:

$$\nu_N = \sum_{t=1}^N \beta^{t-1} u_{1t}, \quad (3.2)$$

where $0 < \beta \leq 1$ is the *discount factor* usually calculated as

$$\beta = e^{-r}, \quad (3.3)$$

where r is the interest rate per stage. For further information on the theory and principles of discounting, reference is made to textbooks on economics.

3.2.2 Leisure time

Just like monetary gain, leisure time is probably an attribute of all farmers' utility functions. It is, however, slightly more complicated to represent it, because not only the total value, but also the distribution over time (day, week, year) is relevant. Since leisure time is the counterpart of work, it may therefore be relevant to split up the factor labor into several sub-factors like for instance work on weekdays from 6 am to 6 pm, x_1 , work on weekdays from 6 pm to 6 am, x_2 , work in weekends from 6 am to 6 pm, x_3 , work in weekends from 6 pm to 6 am, x_4 , etc. These sub-factors are assumed to refer to the farmer's own work, whereas the work by employees has to be represented by other sub-factors.

Using these sub-factors we may quite easily at any stage t calculate corresponding sub-attributes like total leisure time u_{2t1} , leisure time in weekends u_{2t2} , during day time u_{2t3} etc. If the stage length in days is denoted as D , these sub-attributes are objectively calculated as for instance total leisure time in hours (where the index j refers to different sub-periods):

$$u_{2t1} = 24D - \sum_j x_j, \quad (3.4)$$

and correspondingly for other sub-attributes. There exists no general over-all partial attribute function h_2 for leisure time, because the properties of that function depends on the farmer's individual (subjective) preferences. It seems however, reasonable to assume, that it may be defined in each individual case as a function of a number, v , of objectively calculated sub-attributes like those discussed, i.e.

$$u_{2t} = h_2(u_{2t1}, \dots, u_{2tv}). \quad (3.5)$$

As concerns the comparison of values relating to different stages, the sub-attributes are to some extent additive (e.g. total leisure time). In other cases it may be more relevant to look at the average value (e.g. work load in weekends) and/or the maximum value. If, for instance, x_3 and x_4 from the example above are both zero, it means that the farmer does not have to work at all during the weekend, which some farmers would give a high priority. Correspondingly, if $u_{2t1} = 24D$ for a stage, it means that the farmer is able to have a vacation.

3.2.3 Animal welfare

While monetary gain and leisure time are probably attributes of all farmers' utility functions (albeit that the relative weights may differ), the situation is different with animal welfare. There is no doubt that some farmers will include the attribute, but since the level of animal welfare has no direct consequences for the farmer's welfare it is probably ignored by others. It does not mean that the farmer necessarily ignores animal welfare, but the reason for considering it may only be for legal reasons or because it to some extend affects the production and consequently other attributes like monetary gain and/or labor. If that is the case, animal welfare

is only considered as far as to meet legal demands and not to affect the other attributes. In such a situation it should not be defined as a separate attribute of the utility function.

Recalling Section 3.2.1 we note that monetary gain is an attribute which may objectively be represented by a single numerical value. As concerns leisure time, we needed several numerical values (or sub-attributes) for an objective representation. If we want to express that attribute as a single numerical value it is unavoidable a subjective personal value derived from the preferences of that particular farmer. Another farmer would probably attach different weights to the various objective sub-attributes involved and, therefore, his over-all evaluation of the scenario would differ.

When it comes to an attribute like animal welfare, it is even a question whether objective sub-attributes exist. In general, the assessment of the very concept of animal welfare is a problem. The discussion of a proper definition of welfare is old as the examples collected below by Kilgour and Dalton (1984) will illustrate:

- *A wide term that embraces both the physical and mental wellbeing of the animal* (Brambell, 1965).
- *Existence in reasonable harmony with the environment, both from an etho-logical and physiological point of view* (Anonymous, 1977).
- *Welfare can be satisfied if three questions can be answered in the affirmative: (a) Are the animals producing normally? (b) Are they healthy and free from injury? (c) Is the animal's behaviour normal?* (Adler, 1978).
- *Welfare is a relative concept. Profit is a matter related to welfare and determining the relationship between welfare and profit is a scientific matter. The choice between welfare and profit is an ethical matter* (Brantas, 1978).
- *Considering that while an animal is producing protein without observable signs of pain, then it can be considered to be comfortable. Distress, strain, gross abuse and suffering are used to describe unfavorable circumstances ... the presence of raw flesh or heavy bruises would be classified this way, but callouses or hard skin caused by concrete floors and producing no pain would not be classed as distressing* (Randall, 1976).
- *On a general level it is a state of complete mental and physical health where the animal is in harmony with its environment. On an empirical level it may be measured by studying an animal in an environment which is assumed to be ideal, and then comparing it with an animal in the environment under investigation* (Hughes, 1976).
- *Handling animals in the least disturbing manner with full consideration for their normal species-specific behaviour requirements* (Kilgour, 1987).

None of the definitions listed are appropriate in this context. A more consistent approach seems to be to consider animal welfare analogously to human welfare as discussed by Sandøe and Simonsen (1992). If we accept this approach, the welfare of an animal may be described by a utility function involving a number of attributes each representing an element of animal welfare. The principles are exactly the same as with the farmer's utility function, but the attributes considered naturally differ. Animal welfare has nothing to do with monetary gain, environmental preservation or product quality. Instead, relevant attributes might be:

- Absence of hunger
- Absence of pain
- Thermal comfort
- Absence of fear
- Pleasure, joy

Just like the farmer's utility, animal welfare may be controlled by the dynamic allocation of factors to the production processes. Obviously, absence of hunger is ensured by an appropriate supply of feeds, but it is important to realize that what matters in relation to welfare is the subjective experience by the animal. It is certainly possible to feed animals in such a way, that all requirements for maintenance and production are satisfied, but the animal nevertheless feels hungry. Such a situation may occur if an animal is fed a very concentrated ration as it is typically the case with sows during the gestation period. Even though all physiological concerns are met, the welfare of the animal may be violated.

As concerns pain, many factors are involved: Physical injuries caused by inappropriate housing conditions may expose the animal to pain. The same applies of course to injuries caused by other animals in the flock. An other source of pain is disease, which may be prevented or cured by drugs, vitamins, minerals or feeding in general as well as by the housing conditions.

By thermal comfort we mean that the animal neither feels cold or too warm. This may be achieved in many different ways involving many different factors. If an animal feels cold, we may for instance install a heating system, reduce the ventilation, increase the number of animals per square meter, or supply the animal with straw for bedding.

It seems reasonable to assume that farm animals are able to experience fear. If we accept this assumption we have to consider how to prevent it. The two most important factors in this respect are probably labor and other animals. A negative influence from labor is prevented by appropriate daily routines and the influence from other animals is regulated through flock sizes and re-grouping.

As concerns the last attribute mentioned (pleasure, joy) it is a question whether it makes sense in relation to animals, but according to Sandøe and Simonsen (1992)

several researchers define this attribute and even attributes like satisfaction and expectation as elements of animal welfare.

Having identified the attributes is of course only half the way to an operational representation of animal welfare as a possible attribute of the farmer's utility function in relation to herd management. Since the attributes of animal welfare represent the animals' subjective experience, they cannot be measured directly. All that can be measured are physiological, behavioral, pathological and other objective parameters which may serve as evidence for the occurrence of the relevant subjective experiences. Sandøe and Simonsen (1992) therefore argue that animal welfare researchers have to find out which measurable parameters will serve as indicators of the occurrence of which experiences. This is an other example of a situation where the herd management scientist depends on results achieved in other research disciplines. What is actually measured are typically parameters like hormone levels, conflict and abnormal behavior, results of various sorts of choice tests conducted and disease incidences. Sandøe and Simonsen (1992) refer to Dawkins (1980) and Fraser and Broom (1990) for a more full and detailed account.

Reference is also made to Sandøe and Hurnik (1996) for a collection of articles discussing concepts, theories and methods of measurement in relation to the welfare of domestic animals.

An other problem is how to combine the conclusions regarding individual attributes into the aggregate notion of over-all animal welfare, which is the possible attribute of the farmer's utility function. Since, however, this problem is analogous to the problem of combining the various attributes of the farmer's utility function, reference is made to Section 3.2.2, where that issue is discussed.

3.2.4 Working conditions

Also the working conditions may affect the farmer's welfare. When working in barns, dust may contaminate the respiratory system and even damage the lungs. Correspondingly, very hard work (heavy burdens etc.) may harm the back. It is probably not possible to express the working conditions objectively by a single numerical attribute. It must, however, be assumed that if the general working conditions are so bad that they are a threat to the farmers permanent health condition improvements will have a very high priority in the utility function. When expressing the over-all value of the attribute over several stages it may therefore be relevant to define it as (assuming that the working conditions are the 4th attribute of the utility function):

$$u_{4N} = \min_t \{u_{4t}\}, \quad (3.6)$$

where u_{4t} , $1 \leq t \leq N$, is the resulting value at stage t of the attribute representing the working conditions. In words, Eq. (3.6) says that what matters to the farmer are the worst working conditions experienced during the N stages.

3.2.5 Environmental preservation

Just like some of the previously discussed attributes it may also be relevant initially to define a number of sub-attributes u_{5t1}, \dots, u_{5tv} each representing an aspect of environmental preservation. In most cases they may be objectively calculated as numerical figures even though the precision may vary. An example of such a sub-attribute is according to Jensen and Sørensen (1999) the net loss of nutrients (e.g. nitrogen) to the environment which may be calculated as

$$u_{5t1} = \sum_{i=1}^m c_{x_i} x_{ti} - \sum_{i=1}^n c_{Y_i} Y_{ti}, \quad (3.7)$$

where c_{x_i} and c_{Y_i} are the concentrations of the nutrient in question in the i th factor and the i th product, respectively. Other relevant sub-attributes mentioned by Jensen and Sørensen (1999) include the total consumption of energy and indicators of the risk of residues of pesticides in the environment.

When dealing with the leisure time attribute we concluded that no general over-all partial attribute function exists. For an individual farmer, however, we assumed that such a subjectively defined function exists as illustrated by Eq. (3.5). As concerns environmental preservation, the aggregate attribute function probably also has to be subjectively defined, but, in this case the reason is merely lack of knowledge on the relative importance of the sub-attributes in relation to environmental preservation.

When the attribute is evaluated over several stages, calculation of average values of the sub-attributes are most often the relevant method.

3.2.6 Personal prestige

Already by using the word "personal" we suggest that we are dealing with a subjectively defined attribute. In many cases, personal prestige has something to do with a high level of productivity (e.g. milk yield per dairy cow or litter size per sow), the presence of a factor of a particular kind (e.g. a milking parlor or a big tractor) implying that we are dealing with sub-attribute functions of the kind

$$u_{6t1} = \frac{Y_{ti}}{x_{tj}}, \quad (3.8)$$

(where Y_{ti} and x_{tj} are, for instance, the total milk yield and the number of cows at stage t) or the kind

$$u_{6t2} = \begin{cases} 1, & x_{ti} > 0 \\ 0, & x_{ti} = 0 \end{cases}. \quad (3.9)$$

Even though the list of sub-attributes is highly subjective, the calculation of their numerical values are, as illustrated, most often objective. Otherwise their worth as the basis of personal prestige would be lacking. Also the over-all partial attribute function h_6 is subjectively defined.

3.2.7 Product quality

All farmers are probably interested in product quality to the extent that it influences the price of the product (e.g. the hygienic status of milk). If, however, that is the only reason for considering product quality, it should not be defined as an attribute of the farmer's utility function. If the farmer is interested in product quality also if it is independent of (or even in conflict with) monetary gain and other attributes already included then it is relevant to consider it as an attribute of the utility function.

The relevant sub-attributes to consider will of course depend on the product, but typical examples when dealing with animal production are the hygienic status, the risk of residues of medicine or indicators of the nutritional value for humans (e.g. fat or protein percentage).

3.3 From attributes to utility

3.3.1 Aggregation of attributes, uncertainty

Having identified the attributes of a farmer's utility function we face the problem of how to combine these single attributes into an aggregate over-all utility of the farmer. The examples of the previous section clearly illustrate that typical attributes are very different in nature and measured in different units (e.g. monetary gain and animal welfare). When we use the attributes in a planning situation we furthermore face the problem that we are not able to calculate exact values relating to future stages as discussed in relation to Eq. (1.7). We have to consider the attributes (and thereby utilities) of future stages as random variables reflecting that we are producing under risk. In other words, the farmer has to make decisions based on distributions of attributes and utilities rather than fixed values.

3.3.2 Single attribute situation: Risk

In the most simple case the farmer's utility function only includes one attribute, i.e.

$$U_N = g(u_N), \quad (3.10)$$

where u_N is the aggregate value of the attribute over N stages. For convenience, we shall assume that the attribute is monetary gain, but in principle it could be any attribute. In a real situation it is hardly likely that a utility function only depends on one attribute, but the considerations below are also relevant if we consider a decision that only influences one attribute, so that the remaining attributes may be considered as fixed.

In order to illustrate the nature of a typical single-attribute utility function we shall initially consider a question: *How do we feel about earning an extra fixed amount of money?* We shall denote our present annual income as a , and the potential extra income as Δa (where $\Delta a > 0$). There is probably no doubt that we

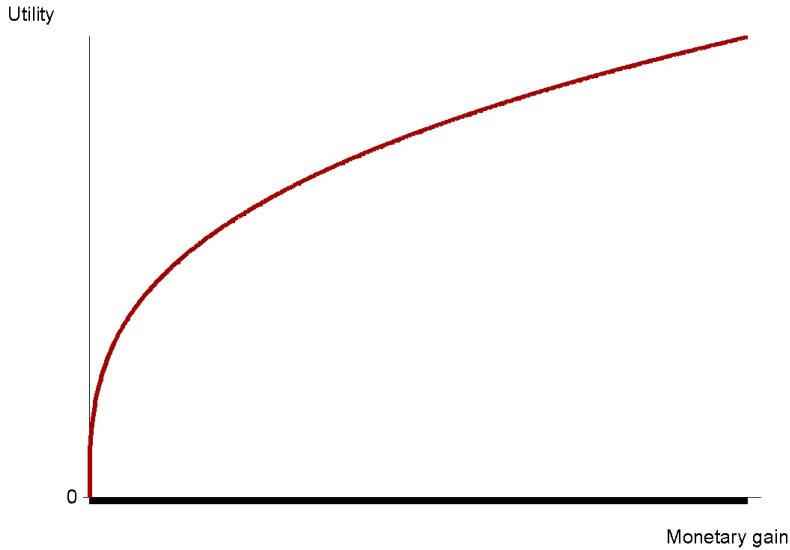


Figure 3.1: A typical single attribute utility function

would all prefer the total income $a + \Delta a$ to a no matter the actual values of a and Δa . In other words, the utility function g must have the property that

$$g(a + \Delta a) \geq g(a) \quad (3.11)$$

for $\Delta a > 0$.

In order to be able to answer the question more specifically, it is not sufficient information just to know that Δa is positive. We also have to know the value of not only Δa , but also of a . A simple example illustrates this. If our present annual income is a million (i.e. $a = 1,000,000$) we are probably rather indifferent to earning an extra amount of say ten thousand (i.e. $\Delta a = 10,000$). If, on the other hand, our present annual income is only fifty thousand ($a = 50,000$), we would regard the same option as a considerable improvement of our situation. In other words, the marginal utility of the same monetary gain depends heavily on the present level of income. This is illustrated in Figure 3.1 which shows the typical course of a single attribute utility function defined as

$$U = u^\alpha, \quad (3.12)$$

where, most often, $0 < \alpha < 1$. In Figure 3.1, $\alpha = \frac{1}{3}$. For a discussion of other common algebraic representations of utility functions, reference is made to Anderson et al. (1977).

Now, assume that we have to choose between two alternative actions. Since actions always relate to the allotment of factors, we shall denote them as \mathbf{x}^1 and \mathbf{x}^2 , respectively. The resulting production from these actions are correspondingly denoted as \mathbf{Y}^1 and \mathbf{Y}^2 . Applying Eqs. (1.1) and (3.1) we may calculate the monetary gain under action 1 as:

$$u_1^1 = \sum_{i=1}^n p_{yi} (f_i(\mathbf{x}^1) + e_i^1) - \sum_{i=1}^m p_{xi} x_i^1 = \sum_{i=1}^n p_{yi} f_i(\mathbf{x}^1) - \sum_{i=1}^m p_{xi} x_i^1 + \epsilon^1, \quad (3.13)$$

where the random variable ϵ^1 is defined as

$$\epsilon^1 = \sum_{i=1}^n p_{yi} e_i^1. \quad (3.14)$$

The monetary gain under action 2 is calculated correspondingly. Assuming fixed prices, the only random elements of the monetary gains are ϵ^1 and ϵ^2 . Without loss of generality, we may assume that $E(\epsilon^1) = E(\epsilon^2) = 0$. Eq. (3.13) may therefore be written as

$$u_1^1 = E(u_1^1) + \epsilon^1. \quad (3.15)$$

Let us assume that $E(u_1^1) = 100$ and $E(u_1^2) = 105$. If the outcome is known with certainty (i.e. $P(\epsilon^1 = 0) = P(\epsilon^2 = 0) = 0$), the choice between the two actions is easy. Using the utility function defined in Eq. (3.12) we may calculate the corresponding utilities as $U^1 = 100^{1/3} = 4.64$ and $U^2 = 105^{1/3} = 4.72$ implying that action 2 is preferred over action 1. This is certainly not surprising, and in fact we only needed the logically deduced Eq. (3.11) in order to arrive at the same conclusion.

If, however, risk is involved the situation is more complicated. By risk we mean that the random variables ϵ^1 and ϵ^2 have known distributions with variances greater than zero. A simple numerical example shall illustrate how this may influence the expected utility associated with each of the two alternative actions. In Table 3.1, the assumed distributions of ϵ^1 and ϵ^2 are specified, and the resulting expected utilities are calculated. It should be emphasized that the expected monetary gains under the two actions are still 100 and 105, respectively.

It may surprise to observe that even though the expected monetary gain under action 2 is higher than under action 1, the opposite applies to the expected utilities. Thus in the example we should indeed prefer action 1 over action 2. The explanation is simply that action 2 is more risky than action 1, and since the applied utility function only modestly rewards outcomes greater than the expected value but heavily punishes lower outcomes (cf. Figure 3.1), the expected utility of a risky action will always be lower than the expected utility of a less risky action having the same expected outcome.

Table 3.1: Probability distributions and expected utilities of two alternative actions.

Action 1				Action 2			
ϵ^1	u_1^1	$P(\epsilon^1)$	U^1	ϵ^2	u_1^2	$P(\epsilon^2)$	U^2
-10	90	0.25	4.48	-65	40	0.25	3.42
0	100	0.50	4.64	0	105	0.50	4.72
10	110	0.25	4.79	65	170	0.25	5.54
Expected utility, $E(U^1)$		4.64	Expected utility, $E(U^2)$		4.60		

The fact that maximization of expected utility is the relevant criterion for choosing between risky actions follows from the so-called expected utility theorem which is discussed by Anderson et al. (1977, p 65-69). The theorem may be deduced from only three perfectly reasonable axioms describing consistent human attitudes to risky choices. The same axioms also ensure the very existence of a unique subjective utility function.

A common numerical representation of the risk associated with an action is the variance of the outcomes. In the example we may directly calculate the variances from Table 1. Under action 1 the variance is $V^1 = 0.25 \times (-10)^2 + 0.50 \times 0^2 + 0.25 \times 10^2 = 50$. The corresponding variance under action 2 is $V^2 = 0.25 \times (-65)^2 + 0.50 \times 0^2 + 0.25 \times 65^2 = 2112.5$ indicating a far more risky action.

In the single attribute situation, it is only because of random variation (i.e. risk) that we have to consider the concept of utility and the shape of the function converting outcomes to utilities. In many cases it may therefore be more natural to use the expected monetary gain and the variance directly in the evaluation of risky actions. That approach is called (E,V) analysis. The idea is to draw so-called iso-utility curves in a diagram where the x axis is the variance and the y axis is the expected monetary gain. An iso-utility curve has the property, that all loci along the curve represent mean-variance combinations that yield the same level of utility. Any risky action with known expected gain and variance may be plotted as a locus in the diagram. By comparing the loci of the actions with the course of the iso-utility curves, the action representing the highest expected utility may be chosen.

In Figure 3.2, iso-utility curves representing 3 constant levels of utility are shown in an (E,V) diagram. If we plot the two actions from the numerical example of Table 3.1 in the diagram (recalling that Action 1 had the lowest variance and the lowest expected monetary gain) we easily see that action 1 represents the highest expected utility.

In the discussion of this section, we have implicitly assumed that the farmer in general regards risk as something that should be minimized. In other words we assume him to be risk averter. The concave utility curve of Figure 3.1 represents a risk averter. At least in theory, a farmer may be risk preferrer. In that case his utility function is convex. Also in that case, Eq. (3.12) may be used as an algebraic representation of the utility function. For a risk preferrer, however, the parameter

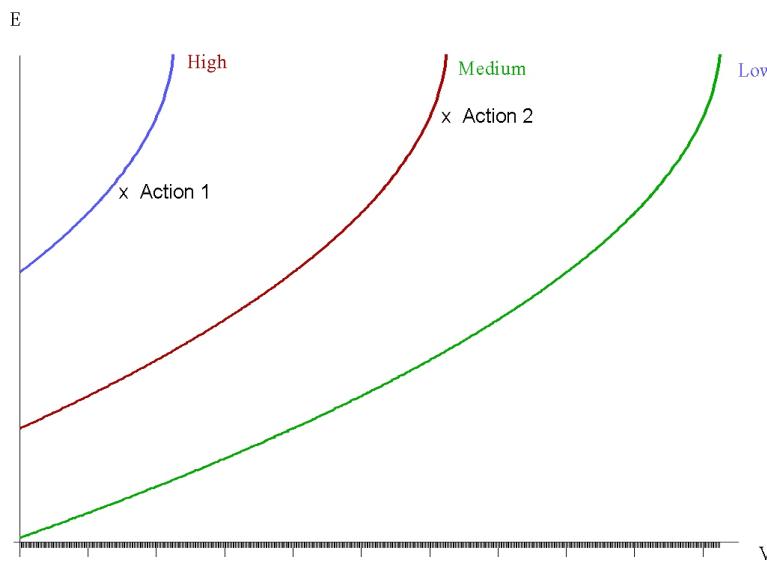


Figure 3.2: Iso-utility curves for 3 levels of utility. The two actions defined in Table 3.1 have been plotted in the diagram.

α has to be greater than 1. For a discussion of those aspects, reference is made to Anderson et al. (1977), whom we also refer to for a discussion of how to identify a farmer's utility function in practise.

3.3.3 Multiple attribute situation

In case that an action influences more than one attribute of the farmer's utility function we have to consider not only how each action affects each attribute, but also how much weight relatively the fulfilment of each attribute should be given. Animal breeding scientists face a completely analogous problem when they define breeding goals involving several traits as illustrated by a small example. Let us assume that the three traits milk yield, daily gain and fertility are relevant when dairy sires are selected for breeding. The relative breeding values for each trait may be estimated by usual methods, and we shall assume that we have to select one sire among the three candidates listed in Table 3.2.

Initially, we shall only consider the first four columns of Table 3.2. If we only had to choose between Sire 1 and Sire 2 there would be no problem since Sire 1 is superior in all three traits, but when we have to choose between Sire 1 and Sire 3 we face a problem, because Sire 3 is superior with respect to milk yield but inferior with respect to daily gain and fertility. There is no general solution to this problem, because the final choice will depend on the weight attached to milk yield relative to the weights attached to daily gain and fertility.

Table 3.2: Three sires and their relative breeding values for the traits *milk yield*, *daily gain* and *fertility*. Index 1 and Index 2 have been calculated using different economic weights.

Sire number	Milk yield	Daily gain	Fertility	Index 1	Index 2
1	110	107	104	107.75	107.00
2	109	106	103	106.75	106.00
3	114	103	102	108.25	105.50

The animal breeding scientist solves the problem by defining a so-called breeding index which is a linear combination of the individually estimated breeding values, i.e.

$$B_i = \sum_{j=1}^J c_j b_{ij} \quad (3.16)$$

where B_i is the breeding index of the i th animal, c_j is the weight attached to the j th trait, and b_{ij} is the estimated breeding value of the i th animal concerning the j th trait. If, for instance, we define the relative weight attached to milk yield to be 50% and the relative weights attached to daily gain and fertility to be 25% each (i.e. $c_1 = 0.50$ and $c_2 = c_3 = 0.25$) the breeding indexes of the three sires of the example become as shown under Index 1 in Table 3.2. As it appears from the table, Sire 3 should be chosen. If on the other hand, the relative weight attached to daily gain was 50% leaving 25% for each of the two remaining traits (i.e. $c_1 = c_3 = 0.25$ and $c_2 = 0.50$) the breeding indexes become as shown under Index 2 leaving Sire 1 as the superior one. This clearly illustrates that the relative importance attached to the traits determines the mutual ranking of the animals.

In herd management science we do not choose among Sires but among alternative actions representing different factor allotments. Neither do we evaluate alternatives on breeding values but on attributes of the farmer's utility function. Nevertheless, if in the example we replace *Sire* by *Action*, *trait* by *attribute* (*milk yield* by *monetary gain*, *daily gain* by *leisure time*, *fertility* by *animal welfare*) and *index* by *utility*, the example will still make sense. Also in herd management it is very likely that if we, for instance, rank actions on monetary gain the ranking changes if we use leisure time as our criterion of evaluation instead. Also the multi-attribute utility function may be defined analogously with Eq. (3.16). Using the notation of the previous sections, the utility function would just be a linear combination of the individual attribute functions:

$$U = \sum_{j=1}^J c_j u_j = \sum_{j=1}^J c_j h_j(\mathbf{Y}, \mathbf{X}), \quad (3.17)$$

where c_1, \dots, c_J are constant values which we may assume sum to 1 without loss of generality.

In the animal breeding example, Eq. (3.16) says that a higher breeding value concerning one trait may compensate for a lower value concerning other traits. The same is expressed for attributes like monetary gain, leisure time and animal welfare in Eq. (3.17). This is certainly a realistic property of a utility function, but, nevertheless, if the utility function is defined as in Eq. (3.17) we implicitly make some assumptions which by a closer look may seem unrealistic. In order to illustrate this we shall assume that only two attributes (monetary gain u_1 and leisure time u_2) are relevant for the management problem considered. Eq. (3.17) thus reduces to

$$U = cu_1 + (1 - c)u_2. \quad (3.18)$$

If we set U equal to some constant U' , Eq. (3.18) may be rearranged into

$$u_2 = \frac{U'}{1 - c} - \frac{c}{1 - c}u_1 \quad (3.19)$$

implying that all combinations of u_1 and u_2 yielding the same utility U' form a linear relationship in a diagram. In Figure 3.3, three such *iso-utility curves* representing a low, medium and high fixed level of utility U' are shown. The course of the curves illustrate that decreased leisure time may be compensated by increased monetary gain, and the linear relationship implies that the marginal rate of substitution between the two attributes is constant. In other words, a constant improvement Δu_1 of the monetary gain may compensate a constant reduction Δu_2 of leisure time *no matter whether the initial level of monetary gain or leisure time is high or low*. From Eq. (3.19) we see that the constant marginal rate of substitution is:

$$\frac{\Delta u_2}{\Delta u_1} = \frac{c}{1 - c}. \quad (3.20)$$

In Figure 3.4, iso-utility curves illustrating varying marginal rates of substitution between attributes are shown. Again, three fixed levels of utility are represented in the diagram. As shown in the figure, the marginal rate of substitution $\Delta u_{21}/\Delta u_1$ is very high if we are dealing with a poor farmer with plenty of leisure time whereas it is very low ($\Delta u_{22}/\Delta u_1$) in the opposite situation (a rich farmer working all the time).

A possible algebraic representation of a utility function having iso-utility curves like those of Figure 3.4 is:

$$U = u_1^\alpha u_2^\beta, \quad (3.21)$$

where (for the risk averter) $0 < \alpha < 1$ and $0 < \beta < 1$. It should be noticed, that if one of the attributes is left constant, Eq. (3.21) reduces to the single-attribute utility function of Eq. (3.12). For a discussion of other possible representations of a multi-attribute utility function, reference is made to Anderson et al. (1977).

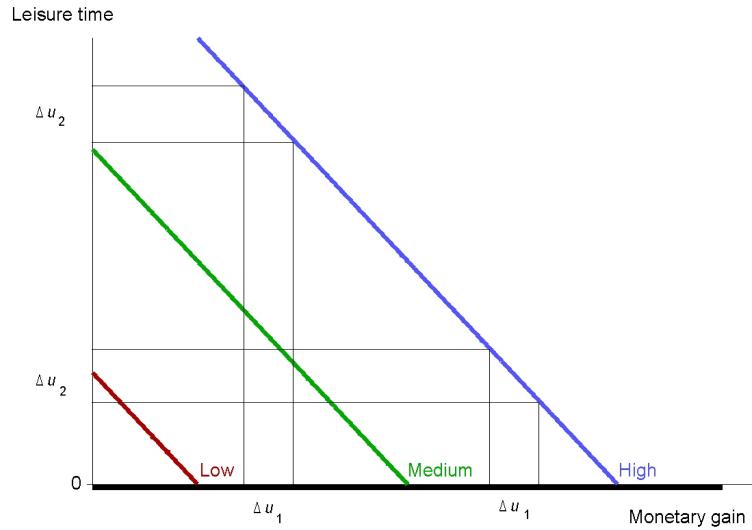


Figure 3.3: Iso-utility curves representing a low, medium and high fixed level of utility calculated according to Eq. (3.18). The marginal substitution rate $\Delta u_2/\Delta u_1$ is constant.

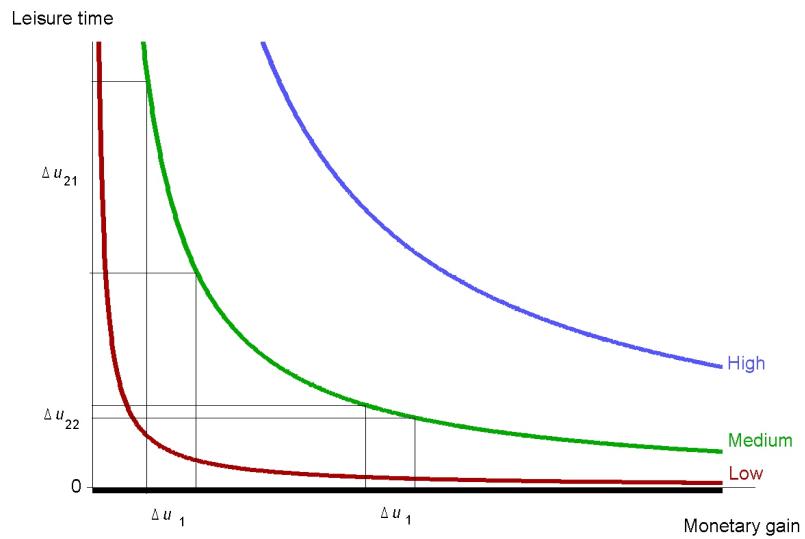


Figure 3.4: Iso-utility curves corresponding to the utility function of Eq. (3.21). We have $\Delta u_{21}/\Delta u_1 \gg \Delta u_{22}/\Delta u_1$.

3.3.4 Operational representation of a farmer's utility function

Tell me, how shall I feed my cows this winter? the farmer asks. The herd management scientist answers him back: *Let me know your utility function, and I shall tell you how to feed your cows.* From a theoretical point of view, the previous sections have illustrated that the herd management scientist is right. In order to choose an optimal action among a set of alternatives, we need to know the farmer's utility function. In other words we have to know what attributes he includes and how they are weighted in a specific situation. On the other hand, the farmer is off course not able to specify his utility function, so what do we do? Does the dialogue end here, or are we able to identify the utility function to an extent where it makes sense to make decisions?

Even though for instance Anderson et al. (1977) to some extend describe how single- and multi-attribute utility functions are identified in practice, it remains still a very difficult task and even the representation of relevant attribute functions for attributes like animal welfare and working conditions may be a problem.

On the other hand, we may argue that it is not always necessary to know everything about the utility function when an action is chosen. In several cases we may logically conclude that incomplete information is sufficient to choose an action:

- In some cases, an action only influences one attribute. If risk is involved, it is sufficient to know the single attribute utility function. In case of a decision not involving risk, even the attribute function is sufficient.
- Many actions (typically those with short time horizon) only marginally influence the attributes. It is therefore sufficient to know the local marginal rates of substitution between attributes. In Figure 3.4, this may be illustrated by plotting the current combination of monetary gain and leisure time and draw a straight line with a slope equal to the marginal rate of substitution through that locus. It is not claimed that this constant rate is universal, but for marginal changes the approximation may suffice. This means that the simple additive Eq. (3.17) is used as a local approximation to a theoretically correct (but unknown) utility function with varying substitution rates.
- The utility function may only vary little over a rather large range of actions. It may be sufficient to choose a satisfactory action rather than an optimal one. This is particularly the case if a farmer's single-attribute utility function has a course as illustrated in Figure 3.5. In that case it is important, that the value of the attribute in question is at least as high as some satisfactory level u^* . Lower levels are punished by the utility function, but higher values are not rewarded. The attribute in question in Figure 3.5 could for instance be animal welfare. Most farmers will probably try to improve it if it is very low, but having reached a certain level, they do not worry about it anymore. When they make decisions with long time horizon (for instance regarding housing facilities) they make some decisions to ensure a satisfactory level,

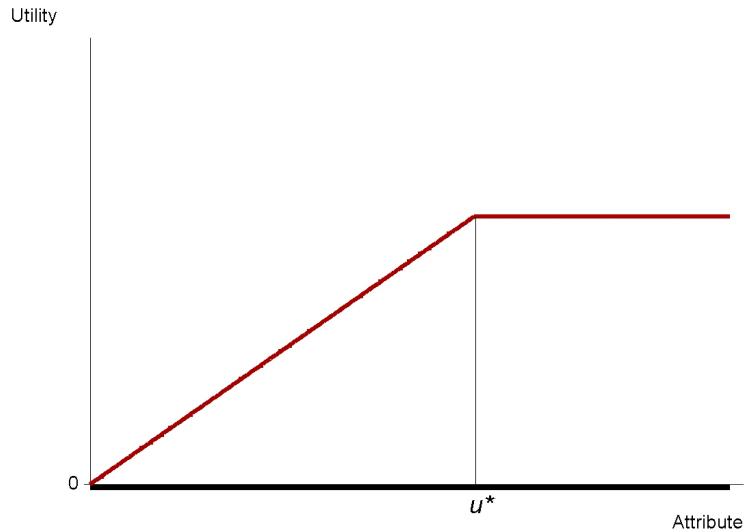


Figure 3.5: A single-attribute utility function reflecting the existence of a satisfactory level.

and afterwards they consider the matter of animal welfare to be out of concern.

An alternative to utility functions is the lexicographic utility concept which may be more operational in a real world situation. Instead of defining an aggregate multi-attribute utility function the farmer has to rank attributes from the most important to the least important. If four attributes are relevant the ranking could for instance be as follows:

1. Monetary gain (u_1)
2. Working conditions (u_2)
3. Animal welfare (u_3)
4. Leisure time (u_4)

In the pure form, the concept requires that actions are first evaluated entirely on the attributed given highest priority (in this case monetary gain), and Action 1 is preferred to Action 2 if and only if $u_1^1 > u_1^2$ (no matter the values of the other attributes). Only if $u_1^1 = u_1^2$, the actions are evaluated on the second attribute (working conditions) and, again, Action 1 is preferred to Action 2 if, and only if, $u_2^1 > u_2^2$ and so on. Only if $u_1^1 = u_1^2, u_2^1 = u_2^2$ and $u_3^1 = u_3^2$ the actions are evaluated on the fourth attribute (leisure time).

The lexicographic concept may also be combined with definition of satisfactory levels concerning the most important attributes and then maximizing the least important one subject to constraints on the others.

Finally, the obvious direct choice option should be mentioned. The farmer is just confronted with the multi-attribute consequences of the alternative actions. In other words, he is told that Action 1 represents a monetary gain of u_1^1 , working conditions expressed as u_2^1 , animal welfare at the level u_3^1 and leisure time at the level u_4^1 . Similar values concerning the other possible actions are given, and based on this information the farmer chooses the preferred action directly. This method has the obvious advantage that the advisor does not have to know the farmer's utility function, but only the relevant attributes.

Chapter 4

Classical production theory

The neo-classical production theory is a very general theory covering almost all aspects of production optimization. It is entirely based on a symbolic representation of the production function and use of general mathematical principles in the determination of optima. It is often said, that the neo-classical production basically deals with three problems:

1. What to produce.
2. How to produce.
3. How much to produce.

In this chapter we shall give a very simplified introduction to the theory by illustrating the basic marginalization principles behind the solution to the three problems. For detailed discussions and less restricted models, reference is made to textbooks on micro economics.

Furthermore, we shall discuss the concept of production costs by introducing key concepts like variable costs, fixed costs, gross margin, opportunity costs, internal prices and investments.

As an extension to the marginalization principles we shall introduce the concept of partial budgeting as an operational method to be used in regular planning of animal production.

Finally, we present a brief introduction to the classical replacement theory which is relevant for a wide range of decision problems in herd management.

4.1 One to one relations

4.1.1 The production function

In this section we shall use a modified version of the general production function as presented by Eq. (1.1) in Chapter 1. We shall use the form

$$F(\mathbf{y}, \mathbf{x}) = F(y_1, \dots, y_n, x_1, \dots, x_m) = 0, \quad (4.1)$$

where $\mathbf{y} = (y_1, \dots, y_n)'$ is a vector of n products, and $\mathbf{x} = (x_1, \dots, x_m)'$ is a vector of m factors. Compared to Eq. (1.1) it means that the dynamic aspect of the planning problem is ignored by skipping the dependence on factor use in previous stages, and that the uncertainty of the outcome as illustrated by the random term of Eq. (1.1) is ignored. The purpose of this simplification is to deduct some conditions for optimality that are of general interest in production planning.

The interpretation of the product levels y_1, \dots, y_n also differs from that of Eq. (1.1) in the sense that y_i now denotes the *maximum* achievable amount of the i th product given the values of $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n, x_1, \dots, x_m$. Thus, logically, products may be mutually substituted for given \mathbf{x} (i.e. by decreasing y_i , a higher level of y_j may be reached).

It should be noticed that Eq. (4.1) reflects an assumption about immediate adaptation of production output to changes in factors.

We shall investigate a number of one-to-one relations:

What to produce: A product-product relation where we determine the optimal combination of two products assuming fixed factor level.

How to produce: A factor-factor relation where we determine the optimal combination of two factors assuming fixed production level.

How much to produce: A factor-product relation where we decide on the optimal level of one factor used for production of one product.

4.1.2 What to produce

Assume that we only produce two products, i.e. that $\mathbf{y} = (y_1, y_2)'$, and a fixed factor input $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_m)'$ is used in production. The production function (4.1) then reduces to

$$F(y_1, y_2, \bar{\mathbf{x}}) = 0, \quad (4.2)$$

Now, let y_1 vary in small steps from 0 to y_1^+ , where y_1^+ is the maximum production possible for factor input $\bar{\mathbf{x}}$. For each value of y_1 in the interval, we solve the equation (4.2) with respect to y_2 and plot the result in a diagram with y_1 along the first axis and the corresponding value of y_2 along the second axis. If, for instance, $y_1^+ = 10$, the result could be as shown in Figure 4.1(a). We shall refer to the curve as an *iso-factor* curve, because points along the curve represent product combinations that can be achieved using the same factor input.

The curve of the figure shows possible combinations of y_1 and y_2 for given, fixed, factor input $\bar{\mathbf{x}}$. If the price of product 1 is p_1 , and the price of product 2 is p_2 , the total profit from production is simply $u' = p_1 y_1 + p_2 y_2 - p_{\bar{\mathbf{x}}}$, where $p_{\bar{\mathbf{x}}}$ is the total price of the input factors. Assuming that we wish to maximize income

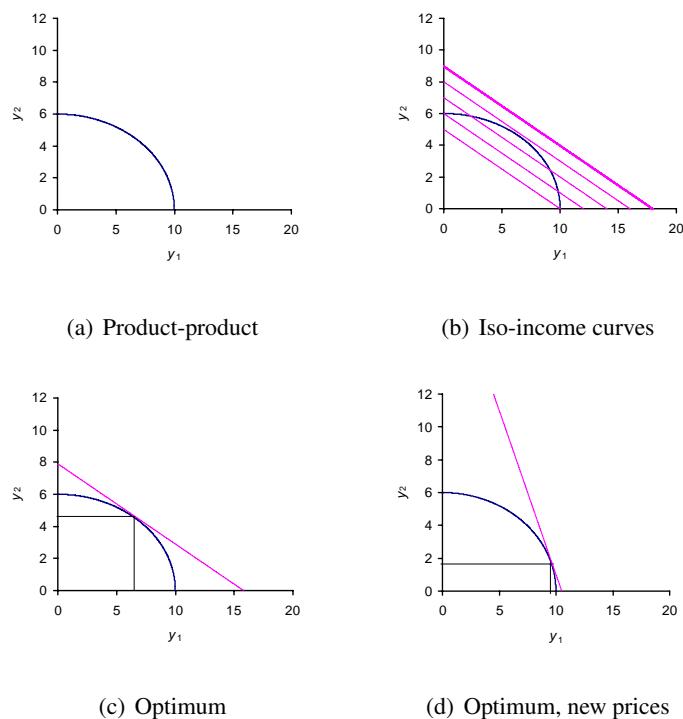


Figure 4.1: Optimization of product-product relation for given factor input \bar{x}

from production, we easily see that we can ignore the price of the factors (since it is fixed) and only look at the total income u (if we maximize u we also maximize u'):

$$u = p_1 y_1 + p_2 y_2. \quad (4.3)$$

Putting $u = \bar{u}$ and solving for y_2 , Eq. (4.3) may be rearranged to

$$y_2 = \frac{\bar{u}}{p_2} - \frac{p_1}{p_2} y_1. \quad (4.4)$$

It should be noticed, that all combinations of y_1 and y_2 satisfying Eq. (4.4) will represent the same income level, \bar{u} . Those combinations form a straight line with slope $-\frac{p_1}{p_2}$ in a diagram, and in Figure 4.1(b), five such *iso-income* lines, each representing a fixed income level, have been added to the iso-factor curve.

Since the iso-factor curve is the upper border of the achievable product combinations, we wish to find the combination maximizing the income. This is done by finding the iso-income line which has the highest possible income u in the set bounded by the iso-factor curve. Figure 4.1(c) illustrates that the optimal product combination is found on the border of the permitted area and has the property that the marginal substitution rate between y_1 and y_2 equals the slope of the iso-income line. In other words, the optimal combination of products is found where

$$\frac{dy_2}{dy_1} = -\frac{p_1}{p_2}. \quad (4.5)$$

Eq. (4.5) illustrates that the optimal combination (for given production function) is entirely determined by the product prices. Thus, new prices will lead to another optimal product combination as shown in Figure 4.1(d).

4.1.3 How to produce

The treatment of this problem has many similarities with the previous problem. The difference is that we now consider the optimal combination of two factors for production of a fixed product output. In other words, we have $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_n)'$, and a variable combination of two factors $\mathbf{x} = (x_1, x_2)'$ is used in production. The production function (4.1) then reduces to

$$F(\bar{\mathbf{y}}, x_1, x_2) = 0, \quad (4.6)$$

Now, let x_1 vary in small steps from 0 to x_1^+ , where x_1^+ is the consumption of factor 1 if the production is based entirely on that factor. For each value of x_1 in the interval, we solve the equation (4.6) with respect to x_2 and plot the result in a diagram with x_1 along the first axis and the corresponding value of x_2 along the second axis. If, for instance, $x_1^+ = 10$, the result could be as shown in Figure 4.2(a). We shall refer to the curve as an *iso-product* curve, because points along the curve represent factor combinations that can be used to produce the same output.

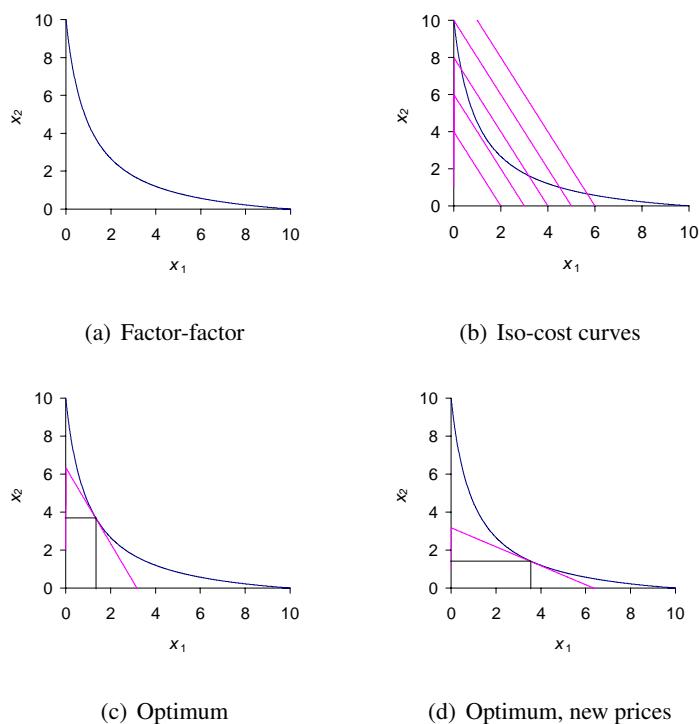


Figure 4.2: Optimization of factor-factor relation for given output \bar{y}

The curve of the figure show possible combinations of x_1 and x_2 that may be used in the production of a given, fixed, product output \bar{y} . If the price of factor 1 is p_1 , and the price of factor 2 is p_2 , the total profit from production is simply $u' = p_{\bar{y}} - (p_1x_1 + p_2x_2)$, where $p_{\bar{y}}$ is the total market value of the product output. Assuming that we wish to maximize income from production, we easily see that we can ignore the market value of the products (since it is fixed) and only look at the total costs c (if we minimize c we maximize u'):

$$c = p_1x_1 + p_2x_2. \quad (4.7)$$

Putting $c = \bar{c}$ and solving for x_2 , Eq. (4.7) may be rearranged to

$$x_2 = \frac{\bar{c}}{p_2} - \frac{p_1}{p_2}x_1. \quad (4.8)$$

It should be noticed, that all combinations of x_1 and x_2 satisfying Eq. (4.8) will all represent the same cost level, \bar{c} . Those combinations form a straight line with slope $-\frac{p_1}{p_2}$ in a diagram, and in Figure 4.2(b), five such *iso-cost* lines, each representing a fixed cost level, have been added to the iso-product curve.

Since the iso-product curve is the lower border of the possible factor combinations, we wish to find the combination minimizing the cost. This is done by finding the iso-cost line which has the lowest possible cost c in the set bounded by the iso-product curve. Figure 4.2(c) illustrates that the optimal factor combination is found on the border of the permitted area and has the property that the marginal substitution rate between x_1 and x_2 equals the slope of the iso-cost line. In other words, the optimal combination of factors is found where

$$\frac{dx_2}{dx_1} = -\frac{p_1}{p_2}. \quad (4.9)$$

Eq. (4.9) illustrates that the optimal combination (for given production function) is entirely determined by the factor prices. Thus, new prices will lead to another optimal factor combination as shown in Figure 4.2(d).

4.1.4 How much to produce

Here we assume that one product is produced by use of only one factor. The purpose of the optimization is to find the optimal number of units to produce by adjusting the input level of the factor. The production function (4.1) then reduces to

$$F(y, x) = 0, \quad (4.10)$$

or, in a more convenient form

$$y = f(x), \quad (4.11)$$

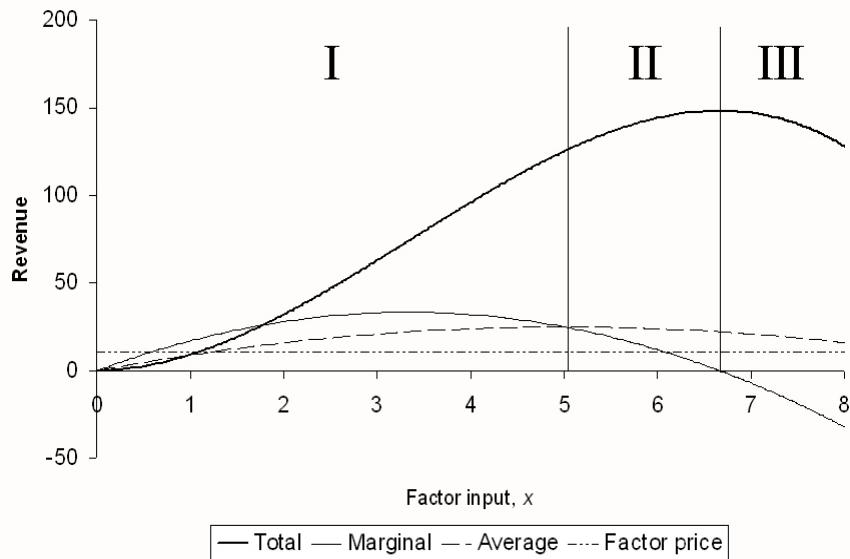


Figure 4.3: Total, average and marginal revenue as functions of factor level. The constant factor price is also shown. The optimal factor level is found in Region II, where the marginal revenue equals the factor price (around $x = 6$).

where f is a function, so that for all x , $F(f(x), x) = 0$. It should be noticed, that the convenient form (4.11) resembles the form used in (1.1). If p_y is the product price, the *total revenue* from production is $t(x) = p_y y = p_y f(x)$. In Figure 4.3, a typical “S-shaped” total revenue curve is shown. The course of the curve reflects that, initially, the efficiency of production will typically be increasing with the scale (factor input level) of production. At some point (between $x = 3$ and $x = 4$ in the figure) the growth of the total revenue curve will slow down, typically because of some kind of capacity constraint. In animal production such capacity constraints could for instance be imposed by the housing system or the biological production capacity of the animals. At some factor level (between $x = 6$ and $x = 7$ in the figure) the absolute maximum capacity is reached. Increasing factor input further will even decrease total revenue.

Referring to Figure 4.3, we can logically conclude that the optimal factor level is not in region III. It would certainly not be rational to increase factor input to a level where a higher total revenue could be achieved with a lower factor level (in Region III we have passed maximum production). We may also logically exclude Region I, even though it requires a little further explanation. We shall define the *average revenue*, $a(x)$ per factor unit as

$$a(x) = \frac{t(x)}{x} = p_y \frac{f(x)}{x}, \quad (4.12)$$

and the *marginal revenue*, $m(x)$, as

$$m(x) = t'(x) = p_y f'(x). \quad (4.13)$$

In Figure 4.3 both curves are plotted together with the total revenue, $t(x)$. The marginal revenue expresses how much the total revenue increases if we increase the factor input level by 1. It is obvious that as long as adding an additional factor unit will increase total revenue by more than the average revenue so far, it will be beneficial to do it. As it is seen in the figure, Region I is characterized by exactly that property (that the marginal revenue is higher than the average). The optimal factor input level must therefore be higher. Thus, optimum must be found in Region II.

It is rather simple to show mathematically exactly where the optimum is. We can calculate the profit, u , from production as

$$u = t(x) - p_x x = p_y y - p_x x, \quad (4.14)$$

where p_x is the factor price. Again, we wish to maximize profit from production, so the problem reduces to finding the value of x maximizing u . According to usual mathematical theory, a maximum is determined by differentiation and solving the equation where the differentiated function is put equal to zero:

$$\begin{aligned} \frac{du}{dx} &= 0 \Leftrightarrow \\ \frac{d}{dx}(p_y y - p_x x) &= 0 \Leftrightarrow \\ \frac{d}{dx}(p_y f(x) - p_x x) &= 0 \Leftrightarrow \\ p_y f'(x) - p_x &= 0 \Leftrightarrow \\ m(x) &= p_x \end{aligned} \quad (4.15)$$

In words, Eq. (4.15) expresses that *the factor input level should be increased until the marginal value of an additional factor unit equals the unit price of the factor*. In Figure 4.3 the optimum is therefore found in Region II, exactly where the marginal revenue is equal to the factor price.

4.2 Multi-dimensional production models

If the full multi-dimensional production function (4.1) is used, the simple marginal considerations of Section 4.1 will not suffice anymore. Instead, a more generalized approach known as Lagrange's method based on partial differentiation and solving of a set of linear equations must be used. We shall not discuss the method further here, but refer the interested reader to a textbook of economics. For an applied example related to milk production under a quota, reference is made to Rasmussen and Nielsen (1985).

4.3 Limitations of the neo-classical production theory

The neo-classic production theory provides us with a general framework for production optimization, and the marginalization criteria derived from it are applicable also in many practical on-farm situations.

Nevertheless, the theory has several serious limitations for use in a herd management context. We can summarize the limitations as follows:

Uncertainty: The neo-classical production theory assumes full certainty about the production outcome and the state of the factors. As discussed in details in Chapter 2, uncertainty is an inherit property of the decision making process in herd management.

Dynamics: The neo-classic production model is static. It assumes implicitly that factors are purchased, products generated and sold within the same time interval. In herd management, many decisions made at an earlier stage have consequences for several months and years as discussed in Chapter 1.

Adaptation: It is assumed that any change in factor input results in an immediate (and full) response in output level. In a biologically based production, it is reasonable to assume some inertness in the response to changes (for instance changes in feeding).

In order to cope with those limitations, more advanced models must be used. In the second part of this book, such models will be introduced.

4.4 Costs

It is important to distinguish between costs and payments. A payment takes place at a certain *moment* (date, hours) when money is actually transferred to someone else. As opposed to that, costs are related to a *period*. The costs relating to a period is simply the value in money of the factors used in production during the period. It is of no relevance for the costs when the factors are actually paid for. It may very well be in an other period, or - for that matter - never. The costs remain the same.

4.4.1 Variable costs and fixed costs: Gross margins

An important distinction is made between *variable costs* and *fixed costs* (also referred to as *capacity costs*). By variable costs we refer to costs that vary (more or less) proportional to the scale of production. Obvious examples of variable costs are feed costs, costs of bedding material, insemination costs and medicine costs. If production is doubled, we also expect those costs to be (approximately) doubled.

By fixed costs we refer to costs that do not vary with the size of production (at least not in the short run). Examples of such costs include housing costs and salary to permanently employed staff. No matter whether we decide to reduce

the production output to 50%, the costs related to farm buildings equipment and staff remain the same. The examples also illustrate that there is no distinct border between variable and fixed costs. It is, for example, obvious that costs related to staff salary are less “fixed” than those related to the farm buildings: A herdsman may be fired at a certain notice, but even if we demolish the farm buildings, we will have to continue the payments on the loan.

Because the farmer cannot really do anything to reduce the fixed costs in the short run, the operational and tactical management will focus only on the variable costs which are under direct control even in the short run. It is obvious that it is not sufficient just to try minimizing the variable costs (that would be easy: if production is reduced to zero the variable costs will also be zero). As illustrated in Figure 4.3, the variable costs should be increased to the level where the marginal revenue equals the marginal costs if profit is to be maximized.

In order to follow the economic efficiency of production the farmer typically calculates the *gross margin* (as defined in Definition 3) for the various production processes on the farm. The processes may for instance be the production based on the sows, the slaughter pigs and the crops.

Definition 3 The *gross margin* of a production process is defined for a specified period as the total revenue generated by the process during the period minus the associated variable costs. The gross margin is also known as the *contribution margin*.

The advantage of the gross margin concept is that it enables the farmer to compare over time, over processes and with other farmers. The comparison over processes is important because if the farmer finds the gross margin from slaughter pigs too low, he might decide skipping slaughter pig production and use the facilities for something more profitable (e.g. expand the sow herd).

The reason for using the gross margin (instead of net profit) for comparison between farmers is that the fixed costs vary very much from farm to farm depending on the housing facilities and many other circumstances. Thus, the net profit is simply not comparable over farms. The variable costs are more likely to be comparable. Accordingly, it makes sense to use gross margin as a measure of economic efficiency between farms.

4.4.2 Opportunity costs and internal prices

In many cases, the interpretation of the costs are straight forward: A certain amount of a factor has been used during the period. Each unit has a well-defined price, because the factor is bought at the market, and the costs are the number of units multiplied by the unit price.

There are, however, many cases where more indirect considerations are needed. Examples include:

- Feeds may consist of home-grown crops.

- A slaughter pig production may be based on piglets from the farmer's own sow herd.
- The value of the animals represents a considerable amount of money.

It is obvious that the costs related to the items above should be included in the gross margin, and that is exactly where the concept of opportunity costs, as defined below, is relevant.

Definition 4 The *opportunity cost* of a factor used in production is the value of the factor if it was used in the best possible alternative way. The name of the concept refers to the fact that this opportunity is lost.

When the farmer for instance calculates the gross margin for slaughter pig production, he should include the costs of the piglets no matter whether he buys them at the market, or he “buys” them from his own sow production. This is an example of the opportunity cost concept. By using the piglets in his own slaughter pig production, the farmer loses the opportunity of selling the piglets at the market. The *internal price* of the piglets to be used when calculating the gross margin of the slaughter pigs is simply the value of the piglets on the market. This illustrates that if we are dealing with a factor commonly exchanged over a market, the internal price is simply the market value.

In other cases the calculation of internal prices is more complicated, and when the internal price is used in planning, the value may even depend on the time horizon as the following example will illustrate.

Example 1 Internal production price of roughage.

Dairy cows are typically fed a ration consisting of some kind of mix of roughage and concentrates. The roughage can either be based on home-grown crops, or it can (at least to some extend) be based on industrial by-products bought at the market. The decision about the ingredients of the feed ration is based on the nutritional contents and the price.

The price of industrial by-products is well-defined and easy to get from market information, but what about the internal price of home-grown roughage? The answer depends on the time horizon.

Before the growth season: At this stage the dairy farmer still has the option *not* to grow the roughage and instead base his feeding on industrial by-products. Such a decision will allow him to use the land intended for the roughage for other crops to be sold at the market. If he chooses to grow roughage he loses that opportunity. The correct internal price must reflect that fact, and accordingly, it is defined as follows:

$$\begin{array}{rcl}
 + & \text{Direct variable costs of roughage} & d \\
 + & \text{Gross margin of alternative crop to be sold} & g \\
 \hline
 = & \text{Internal production price of roughage} & p
 \end{array}$$

The direct variable costs include seeds, fertilizer, chemicals, and costs of fuel, cultivation and machinery. The gross margin of the alternative crop includes the market value of the crop less the variable costs associated with production (according to the standard definition of the gross margin concept as given in Definition 3). For a numerical example, reference is made to Exercise 4.7.2.

The internal production price defined in this way is directly comparable with the price of feedstuffs available at the market. When planning next year's winter feeding, it is therefore the internal production price that should be used for home grown roughage.

During the growth season: When the roughage is already growing in the fields, the option of growing an alternative crop to be sold at the market is no longer available. Any consideration of future feeding should therefore be based on a price setting reflecting the absence of that opportunity. Accordingly, the internal price will be lower, because the lost gross margin of an alternative crop should now be ignored (the option is no longer available, and you can't change the past).

The internal price of the roughage in this situation therefore *only* consists of the variable costs *from now*. Variable costs until now (e.g. seeds and fertilizer) should be ignored, because they are already facts that can't be changed no matter what the farmer does. The internal price of a roughage already growing in the fields is therefore typically very low. The low price reflects the fact that when the decision to grow the roughage has already been made, it is in most cases unrealistic to replace it with feedstuffs bought at the market.

It must be emphasized that these considerations assume that there is no market for the roughage. If, on the other hand, there is a market, the price used in planning should be the market value of the roughage.

After harvest: When the roughage has been harvested, all previous costs are of no relevance for the internal price. Those costs have already been paid, and no matter what we do, the money is spent, and it cannot be regained.

If there is a market for the roughage, the price used in feed planning should reflect the market value. In many cases, however, there is not really a market for roughage, and accordingly it is realistic to define the internal price to be zero, simply because the roughage has no alternative value. When it is already available in the rick or the silo, it is unrealistic not to use it for feeding.

The only relevant question is which category of animals to feed it to (the heifers, the high yielding dairy cows, the dry cows, etc.) The answer to that question should be based on where the marginal contribution to the gross margin is highest.

4.4.3 Investments

The unit costs of factors used entirely during one time stage are rather easy to understand. We use a certain amount of a feedstuff during the stage, and the unit price is known either from the market or from calculation of internal prices.

When it comes to factors like farm buildings or inventory the cost associated with a given stage (e.g. a year) is more difficult to assess, simply because we use the building or inventory for several years. We typically know the total price of e.g. a building when it was built or purchased, but that value may be very high, and it would not be “fair” to charge the full amount to the year of construction. Intuitively, it seems reasonable to disperse the cost over several years.

When dispersing the investment, the key question is for how long we can use the asset in production. If it is a farm building it may for instance be 20 years, and less for inventory. Let I be the initial investment (the price of the asset at purchase) and n be the number of years that we expect to use the asset. It would be tempting just to calculate the annual costs, c , as $c = I/n$, but from an opportunity cost point of view this is *not* correct. Usually, I represents much money, and by using it for investment in an asset means that the farmer loses the opportunity of investing the money in an alternative way (he might for instance buy stocks or bonds).

The lost returns (typically in the form of interests) from the alternative investment should also be included in the annual costs of the investment. In economics, there are standard methods for calculation of annual costs of investments. If I , c and n are defined like before, and furthermore r is the interest rate of the alternative investment, the annual costs are calculated as

$$c = I \frac{r}{1 - (1 + r)^{-n}}. \quad (4.16)$$

It should be noticed that the annual costs calculated this way includes the *depreciation* (loss in value) as well as the *interest costs*.

Example 2 Investment in new farm building.

A pig farmer decides to build a new farrowing department. The initial price of the investment is $I = 1,000,000$, the interest rate is $r = 0.05$ (i.e. 5%), and the building is expected to be used for $n = 20$ years. According to Eq. (4.16), the annual cost, c , is

$$c = I \frac{r}{1 - (1 + r)^{-n}} = 1,000,000 \times \frac{0.05}{1 - (1 + 0.05)^{-20}} = 80,243$$

The annual costs of the investment will be part of the fixed costs of the production.

■

The investments discussed so far refer to assets that are purchased at a certain moment, used over a number of years, and then discarded. In animal production there is, however, also another kind of investment. For an intensive livestock farm, the value of the stock of animals is considerable. At some time in the past, the farmer has invested in the animals, but since animals are regularly replaced, the stock as such, is not depreciated. It should, however, be taken into account that by investing his money in the herd, the farmer loses the opportunity of investing his money in an alternative way. Thus, according to the opportunity cost concept, the lost returns (interests) should be charged to the variable costs of production.

Because there is no depreciation involved, the annual costs of this kind of investment is just calculated as $c = Ir$.

4.5 Other classical techniques

A number of techniques for simple comparison of pre-defined alternatives have been described in literature. Despite their simplicity they are often used under practical conditions for what could be denoted as “quick and dirty” analysis, in cases where only few alternatives are considered. In this section a very short introduction to the methods known as *partial budgeting*, *cost-benefit analysis* and *decision tree analysis* is given. For further details reference is made Dijkhuizen and Morris (1997). The following description relies heavily on Dijkhuizen et al. (1995).

4.5.1 Partial budgeting

In cases where a static analysis suffice, and uncertainty only plays a minor role, partial budgeting could be the method of choice. It is simply a quantification of the economic consequences of a specific change in farm production, and it requires only a limited data collection. It is particularly useful for analysis of relatively small changes within the production process. The general format is sketched in Definition 5. The advantage of the method is that only marginal changes need being included in the calculations. The principles are illustrated in Example 3.

Definition 5 A *partial budget* is used to evaluate the consequences of a (minor) change in production. It consists of the following four sections:

- + Additional returns realized from the change
 - + Reduced costs as a result of the change
 - Returns foregone as a consequence of the change
 - Extra costs as a result of the change
-
- = Result

If the result is a positive number, the change should be adopted. A negative result means that the change is not profitable.

Example 3 Growing roughage or buying industrial by-products

At present a dairy farmer uses a certain industrial by-product for feeding. He considers whether or not to replace the by-product by home-grown roughage. From a nutritional point of view the by-product and the roughage are equivalent. A consequence of the change will be that the farmer must give up growing an other crop sold at the market.

The decision could be based on the following partial budget which should be at annual level:

- | | |
|-------|--|
| + | (There are no additional returns) |
| + | Reduced costs (I): The price of the by-product |
| + | Reduced costs (II): Variable costs of growing current crop |
| - | Returns foregone: Market value of current crop |
| - | Extra costs: Variable costs of growing the roughage |
| <hr/> | = Result |

If the result is a positive number, the by-product should be replaced by the roughage. The example illustrates that the previously defined internal production price of Example 1 is actually based on partial budgeting principles.

4.5.2 Cost-benefit analysis

Cost-benefit analysis may be considered as an extended version of partial budgeting, where we are dealing with a longer time horizon involving proposed courses of action. Since the time at which costs or benefits occur may differ between the alternatives, it is important that the future values are discounted to *present values*. Like in Eq. (3.2) we shall use discounting to account for time differences. The present value p^0 (at time 0) of a benefit (or cost) p earned at time t is simply $p^0 = e^{-rt}p$, where r is the interest rate. The reason for discounting is the time preference of money. A benefit earned in one year has less value today, because of potential interest yields.

The interest rate used in cost-benefit analysis is called the *discount rate* because it makes future values smaller than present values. The higher the discount rate the more a program with high initial costs and a low level of benefits over a longer period of time will be penalized.

The evaluation of the alternatives is exactly as with partial budgeting except for the important difference that all future costs and benefits are discounted to present values before the final result is calculated. Refer to Example 4 for an illustration of the method.

Example 4 Cost-benefit analysis (Dijkhuizen et al., 1995)

A course of actions is assumed over a 4 year period to give rise to the un-discounted

future costs and benefits shown in the table below. The corresponding discounted values are also shown (the discount rate is around 5%).

Year	Discount factor	Un-discounted		Discounted	
		Costs	Benefits	Costs	Benefits
1	0.95	28	0	26.6	0.0
2	0.91	15	10	13.7	9.1
3	0.86	10	20	8.6	17.2
4	0.82	0	25	0.0	20.5
Total		53	55	48.9	46.8

Evaluated directly on the sum of un-discounted values, the benefits are expected to be higher than the costs (implying a profitable project), but when the time differences are taken into account, the discounted costs are higher than the benefits. In other words, the project is *not profitable* under the interest rate used.

■

4.5.3 Decision tree analysis

If uncertainty about the possible outcomes plays an important role for the alternatives considered, a decision tree analysis could be the right choice. Due to its simplicity and the ability to handle uncertainty, it is probably one of the most frequently used techniques of decision analysis. It is in particular very popular in the veterinary community.

Decision trees belong to a large group of graphical models which also includes influence diagrams and so-called LIMIDs which will be described in a later chapter. The decision tree is probably the most simple member of the group. It is defined as a graphical method of expressing in chronological order, the alternative actions available to the decision maker and the choices determined by chance (i.e. the uncertain outcomes).

Decisions are represented by squares (decision nodes), and chance events with uncertain outcome are represented by circles (chance nodes). The branches following each decision node must be exhaustive, so that there is one and only one branch for each decision alternative. After each chance node there is a certain probability p_i that event i occurs. The probabilities following a chance node must add up to 1.00. Each decision alternative and each chance outcome may be associated with a reward (negative or positive). These rewards are written on the relevant branches of the tree.

The optimal decision alternatives are found by successively “pruning” branches starting with the leaves. The branches originating from a chance node are pruned by calculation of the expected rewards of each branch, and the expected value is added to the reward of the branch ending in the node. The branches originating from a decision node are pruned by picking the value from the branch having the

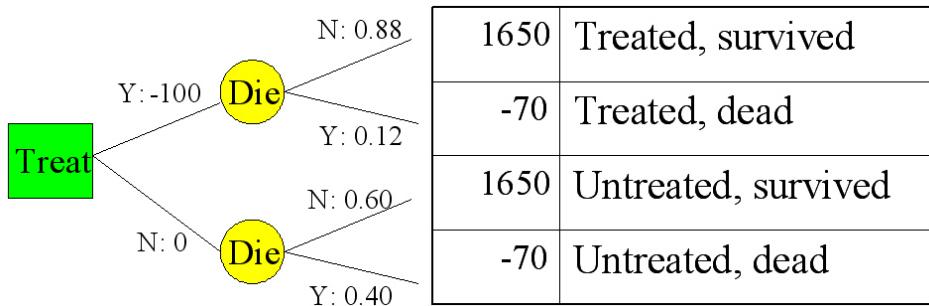


Figure 4.4: A decision tree for the treatment decision described in Example 5. The rewards of the decision alternatives are written along the corresponding branches, and the rewards of the chance outcomes are written in the table to the right of the leafs.

highest reward and adding it to the reward of the branch ending in the node. Obviously, the decision alternative having the highest reward is the optimal. For an illustration, reference is made to Example 5.

Example 5 To treat, or not to treat ...

A calf is suffering from a certain disease, and the veterinarian must decide whether or not to treat it. The value of the calf, if it recovers, is 1650 DKK. If it dies, the cost of getting rid of it is 70 DKK, and the cost of the treatment is 100 DKK. If the calf is treated, the probability of recovery is 0.88, whereas it is only 0.60 if left untreated. A decision tree representation of the decision problem is shown in Figure 4.4.

The pruning of the tree has the following stages:

Pruning of branches from upper “Die” node: This is a chance node, so we must calculate the expected reward, r_1 , as $r_1 = 0.88 \times 1650 + 0.12 \times (-70) = 1443.60$. This value is added to the reward of the branch ending in the node. The revised reward of that branch is thus $-100 + 1443.60 = 1343.60$.

Pruning of branches from lower “Die” node: This is also a chance node, so we must again calculate the expected reward, r_2 , as $r_2 = 0.60 \times 1650 + 0.40 \times (-70) = 962$. This value is added to the reward of the branch ending in the node. The revised reward of that branch is thus $0 + 962 = 962$.

Pruning of branches from the “Treat” node: This is a decision node, so we must select the branch having the highest reward. Since $1343.60 > 962$ we see that it is the upper branch representing the “Treat = Yes” decision alternative, which, accordingly, is the optimal decision.

■

4.6 Classical replacement theory

In economics, the term *replacement* covers the decision to take an existing “asset” out of production and instead use a new one with essentially the same functions. The asset can be any factor being used over several stages. Examples include farm buildings, inventory, sows, dairy cows, slaughter pigs etc.

Usually, we regard the present asset as a single item in a chain of assets successively replacing each other. If the time horizon of production is known and fixed, the number of assets in the chain will be final. In many cases, however, the time for termination of production will be unknown - it is just known to be “far” away in the future. It may then be relevant, as an abstraction, to consider the present asset as the first in an infinite chain of assets. The infinite chain assumption has several computational advantages as we shall see later.

The optimal time for replacement may be determined by the technical properties of the asset, simply because it is not able to produce anymore. A trivial example is light bulbs that typically produce light at a constant level, but suddenly fail. Another example could be a sow or a dairy cow that dies. Most often, however, the asset is still able to produce (at least to some extent). It is therefore replaced because the new asset is expected to contribute more to the total utility value of production.

In the following subsections, the classical replacement theory is briefly introduced. One of the first to describe the replacement problem systematically was Preinrich (1940), but later several textbooks have described the concepts, so for a more detailed description, reference is made to such books (e.g. Rapp, 1974).

4.6.1 Replacement with identical assets

Initially, we shall assume that we are dealing with an infinite chain of *identical* assets, a_0, a_1, \dots , where a_0 is the present asset, a_1 is the next etc. Our problem is to define the right moments t_0, t_1, \dots for replacement of each of these assets. Thus, if t_0, t_1, \dots are optimal, $\delta_n = t_n - t_{n-1}$ is the optimal lifetime of asset n .

By “identical” we mean that the new asset is exactly as the existing one. In other words, we assume that there is no technological or genetical improvement over time. The assumption of identical assets logically implies that the optimal lifetime intervals of all assets are the same, i.e. that there exists an optimal lifetime, δ , so that $\delta = \delta_0 = \delta_1 = \dots$. Accordingly, the problem reduces to the determination of the optimal lifetime, δ .

For any of the assets we assume that the marginal profit at time t (when inserted into production at time 0) is described by the function $m(t)$. The total profit from an asset (bought at the initial price A) over its lifetime, $\pi(\delta)$, is therefore calculated as

$$\pi(\delta) = \int_0^\delta m(t)dt - A, \quad (4.17)$$

with the direct (and obvious) implication that $\pi'(\delta) = m(\delta)$. The asset may also have a *salvage value* at time t , but for convenience, we assume that the changes in salvage value are included in $m(t)$ (if the asset is an animal, the salvage value is the slaughter value).

We assume that the overall aim of production is to maximize profit over time. Since we assume an infinite chain of assets, we cannot look at the total profit, because it is probably not bounded (an infinite chain is expected to give rise to an infinite profit). In other words, it does not make sense to maximize $\sum_{n=0}^{\infty} \pi(\delta)$. A logical alternative to total profit is to maximize average profit per time unit, $a(\delta)$. Thus, we try to identify the lifetime δ maximizing

$$a(\delta) = \frac{\pi(\delta)}{\delta}. \quad (4.18)$$

According to standard mathematical principles we find a maximum of a function by differentiation and putting equal to zero. In other words, we wish to find a δ satisfying:

$$\begin{aligned} a'(\delta) &= 0 &\Leftrightarrow \\ \frac{d}{d\delta} \frac{\pi(\delta)}{\delta} &= 0 &\Leftrightarrow \\ \frac{\pi'(\delta)}{\delta} - \frac{\pi(\delta)}{\delta^2} &= 0 &\Leftrightarrow \\ \pi'(\delta) &= \frac{\pi(\delta)}{\delta} &\Leftrightarrow \\ m(\delta) &= a(\delta) &. \end{aligned} \quad (4.19)$$

From Eq. (4.19) it is seen that *the optimal lifetime is found where the marginal profit equals the average profit*.

Intuitively, this makes good sense: As long as the marginal profit of keeping the asset is higher than the average profit, it should be kept. As long as the benefit of keeping the asset for an additional time unit is higher than the average, keeping it will logically increase the average profit per time unit.

Example 6 Sow replacement

Application of the general replacement theory for sows is straight forward. Based on data from a sow herd, Toft and Jørgensen (2002) showed that the litter size profile of sows (average number piglets $p(n)$ for parity n) in a given herd may be described by the model

$$p(n) = -\theta_1 \exp(-(n^2 - 1)\theta_2) + \theta_3 - \theta_4 n, \quad (4.20)$$

where $\theta_1, \theta_2, \theta_3$ and θ_4 are parameters to be estimated from herd data. The function $p(n)$ together with assumptions regarding mortality, slaughter value of sows (as a function of age), feed consumption and prices determine the marginal and average profit curves, and under realistic assumptions, the results are as shown in Figure 4.5.

As it is seen from the figure, the optimal replacement time is after 7th farrowing.

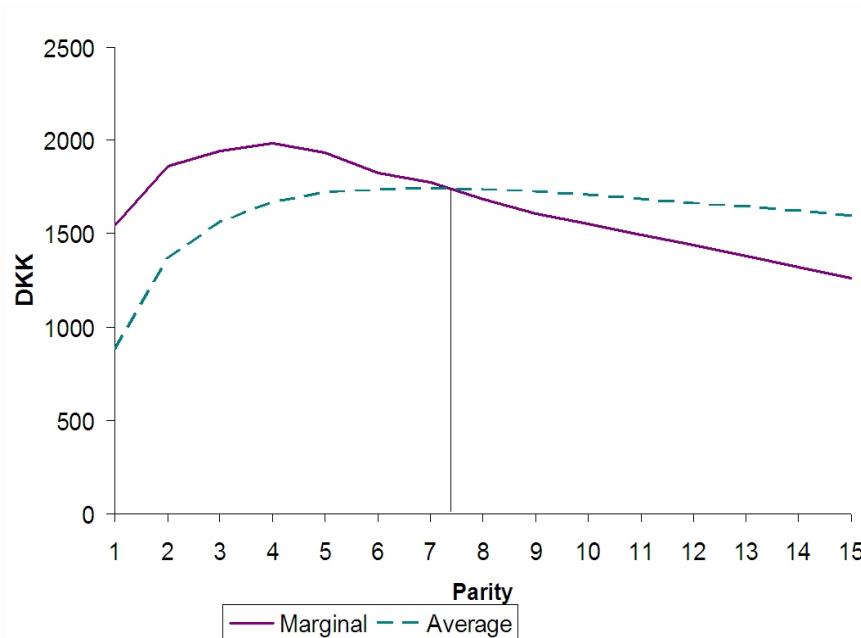


Figure 4.5: Marginal and average profit (DKK) for a sow. The optimal lifetime, δ , is found at the intersection of the two curves. Under the conditions used, the sow should be replaced after 7th farrowing.

■

Even though Figure 4.5 refers to sows, the course of the marginal and average curve is typical for many replacement problems. The initial low values for both curves are due to the initial replacement costs (the difference between the market price of the new asset and the salvage value). The growth of both curves reflects that a new asset typically has a high productivity and low maintenance costs. After some time the productivity decreases, and the maintenance costs increase. Accordingly, the marginal profit (and with some delay also the average profit) starts decreasing.

4.6.2 Optimal replacement using discounting

In the previous subsection we used a criterion maximizing average net profit per time unit. In economics, however, a more usual criterion of optimality is to maximize the present value of expected future profits. As described in a previous section, the present value p^0 (at time 0) of a profit p earned at time t is simply $p^0 = e^{-rt}p$, where r is the interest rate. Since, for $r > 0$, we have, $0 < e^{-rt} < 1$, the infinite sum of present values will actually converge to a fixed value.

The present value of the total profit over the lifetime of an asset at the time of insertion (time 0) is

$$\pi^0(\delta) = \int_0^\delta e^{-rt} m(t) dt - A, \quad (4.21)$$

implying that $\pi^{0'}(\delta) = e^{-rt}m(t)$. Denote as $c(r, \delta)$ a constant profit stream (money per time unit) resulting in exactly the same present value over δ time units. Clearly, $c(r, \delta)$ must satisfy

$$\pi^0(\delta) = \int_0^\delta e^{-rt} c(r, \delta) dt = c(r, \delta) \int_0^\delta e^{-rt} dt. \quad (4.22)$$

Solving Eq. (4.22) for $c(r, \delta)$ yields (by solution of the integral)

$$c(r, \delta) = \frac{\pi^0(\delta)}{\int_0^\delta e^{-rt} dt} = \frac{r\pi^0(\delta)}{1 - e^{-r\delta}}. \quad (4.23)$$

Now, let $p^0(\delta)$ be the total present value of the entire chain calculated at the time of insertion of asset 0. We have

$$p^0(\delta) = \sum_{n=0}^{\infty} e^{-r\delta n} \pi^0(\delta) = \pi^0(\delta) \sum_{n=0}^{\infty} (e^{-r\delta})^n = \frac{\pi^0(\delta)}{1 - e^{-r\delta}}. \quad (4.24)$$

The last equal sign of Eq. (4.24) uses the well known mathematical fact that for $0 < q < 1$ we have $\sum_{i=0}^{\infty} q^i = 1/(1 - q)$. In order to decide on an optimal lifetime, δ , we differentiate and put equal to zero:

$$\begin{aligned} p^{0'}(\delta) &= 0 && \Leftrightarrow \\ \frac{d}{d\delta} \frac{\pi^0(\delta)}{1 - e^{-r\delta}} &= 0 && \Leftrightarrow \\ \frac{\pi^{0'}(\delta)(1 - e^{-r\delta}) - \pi^0(\delta)re^{-r\delta}}{(1 - e^{-r\delta})^2} &= 0 && \Leftrightarrow \\ \pi^{0'}(\delta)(1 - e^{-r\delta}) - \pi^0(\delta)re^{-r\delta} &= 0 && \Leftrightarrow \\ \pi^{0'}(\delta) &= \frac{\pi^0(\delta)re^{-r\delta}}{1 - e^{-r\delta}} && \Leftrightarrow \\ e^{-r\delta}m(\delta) &= c(r, \delta)e^{-r\delta} && \Leftrightarrow \\ m(\delta) &= c(r, \delta) && . \end{aligned} \quad (4.25)$$

Since $c(r, \delta)$ is a kind of generalized average profit per time unit, we see that also when we use discounting, the optimal replacement time δ is found where the marginal profit equals average profit.

4.6.3 Technological or genetical improvement

In this subsection, we will not assume anymore that the assets of the infinite chain are completely identical. Instead we shall assume that future assets in some sense are better than the present one. This is a rather realistic assumptions due to technological or genetical improvements over time.

In order to solve the problem under such conditions, we need to make an assumption concerning the nature of the improvement. We shall denote the marginal and total profit function of the n th asset in the chain as $m_n(t)$ and $\pi_n(\delta)$, respectively. A *linear* improvement means that the marginal profit function (where $t = 0$ is the time of insertion) is of the form

$$m_n(t) = \alpha\delta n + m_0(t), \quad (4.26)$$

where α is the increase per time unit. An *exponential* improvement leads to the marginal profit function

$$m_n(t) = e^{\beta\delta n}m_0(t), \quad (4.27)$$

where β is the relative increase per time unit. Thus, $\beta = 0.01$ implies that the marginal profit (from *new assets*) increases by 1 percent per time unit. In the following we shall investigate the exponential form and its consequences. It follows directly from (4.27) that

$$\pi_n^0(\delta) = e^{\beta\delta n}\pi_0^0(\delta). \quad (4.28)$$

Our criterion for determination of the optimal lifetime is now (like the previous subsection) that we wish to maximize the present value $p^0(\delta)$ of the entire chain, where

$$p^0(\delta) = \sum_{n=0}^{\infty} e^{-r\delta n}\pi_n^0(\delta) = \sum_{n=0}^{\infty} e^{-r\delta n}e^{\beta\delta n}\pi_0^0(\delta) = \pi_0^0(\delta) \sum_{n=0}^{\infty} e^{(\beta-r)\delta n}. \quad (4.29)$$

By use of the rule $\sum_{i=0}^{\infty} q^i = 1/(1 - q)$ we end up with

$$p^0(\delta) = \frac{\pi_0^0(\delta)}{1 - e^{(\beta-r)\delta}}. \quad (4.30)$$

In order to decide on an optimal lifetime, δ , we differentiate and put equal to zero:

$$\begin{aligned} p^0'(\delta) &= 0 && \Leftrightarrow \\ \frac{d}{d\delta} \frac{\pi_0^0(\delta)}{1 - e^{(\beta-r)\delta}} &= 0 && \Leftrightarrow \\ \frac{\pi_0^0'(\delta)(1 - e^{(\beta-r)\delta}) + \pi_0^0(\delta)(\beta - r)e^{(\beta-r)\delta}}{(1 - e^{(\beta-r)\delta})^2} &= 0 && \Leftrightarrow \\ \pi_0^0'(\delta)(1 - e^{(\beta-r)\delta}) + \pi_0^0(\delta)(\beta - r)e^{(\beta-r)\delta} &= 0 && \Leftrightarrow \\ \pi_0^0'(\delta) &= -\frac{\pi_0^0(\delta)(\beta - r)e^{(\beta-r)\delta}}{1 - e^{(\beta-r)\delta}} && \Leftrightarrow \\ e^{-r\delta}m_0(\delta) &= c(r - \beta, \delta)e^{(\beta-r)\delta} && \Leftrightarrow \\ m_0(\delta) &= c(r - \beta, \delta)e^{\beta\delta} && . \end{aligned} \quad (4.31)$$

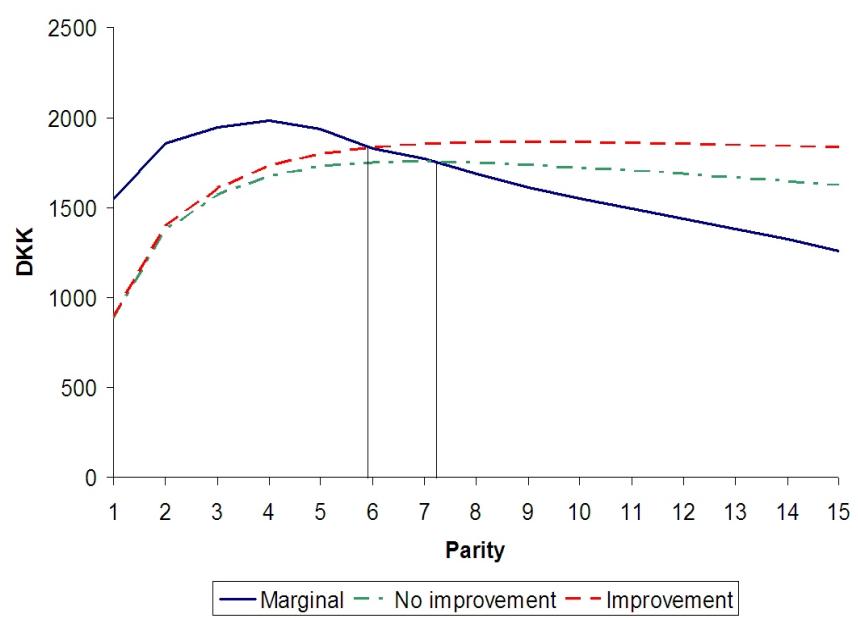


Figure 4.6: Optimal replacement time for sows with and without genetical improvement. The curve marked as “No improvement” is the function $c(r, \delta)$ as defined in Eq. (4.23), whereas the one marked as “Improvement” is the function $c(r - \beta, \delta)e^{\beta\delta}$ as shown in Eq. (4.31). Genetical improvement implies earlier replacement.

Under usual conditions, $c(r - \beta, \delta)e^{\beta\delta} > c(r, \delta)$ implying that the optimal lifetime of the present asset is lower when technological or genetical improvements are assumed.

Example 7 Sow replacement with genetical improvement

In Figure 4.6 the discounting criterion is used for the sow replacement problem introduced in Example 6. An annual interest rate of 5% is assumed. The figure shows the optimal replacement time without genetical improvement compared to a situation with an annual genetical improvement of 1%.

As it is seen from the figure, the optimal replacement time under the discounting criterion is the same as the one in Figure 4.5 (after 7th farrowing) as long as genetical improvement is ignored. Adding a 1% annual genetical improvement shifts the optimal replacement time to the left.

■

4.6.4 Limitations and applications of the classical replacement theory

The classical replacement theory provides us with a general framework for dealing with optimal lifetime of assets. It has, also under practical farm conditions, many

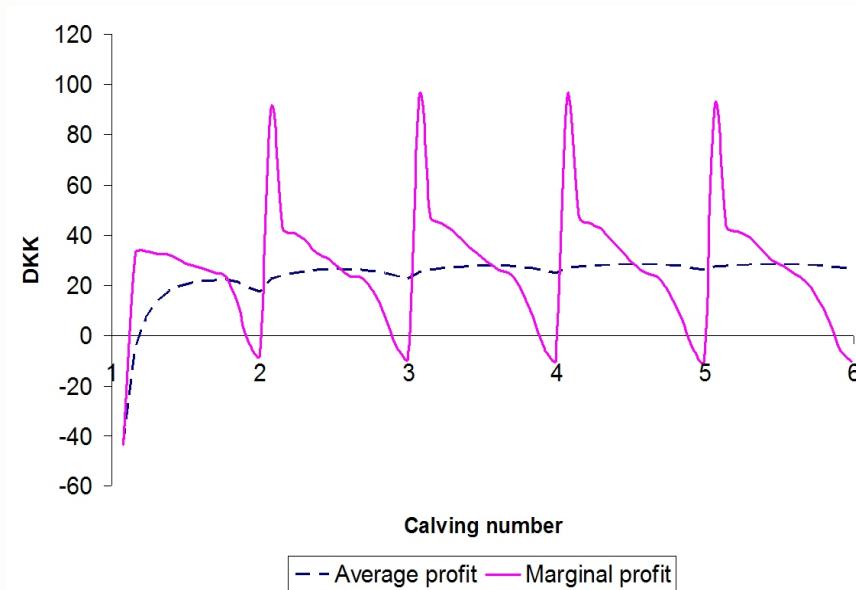


Figure 4.7: Average and marginal profit of a dairy cow over parities. There is an intersection between the curves in each lactation.

areas of application. It is well suited for all kinds of machinery, and, to some extent, optimal slaughter time for broilers, bull calves and slaughter pigs. It has also some relevance for the sow replacement problem, and (within lactation) the dairy cow replacement problem.

In general, the main limitations may be summarized as:

Uncertainty: The classical replacement theory assumes full certainty about the marginal profit function, the investment costs and all prices. As discussed in details in Chapter 2, uncertainty is an inherit property of the decision making process in herd management. The uncertainty is partly a consequence of imperfect knowledge, and partly of random variation.

Uniqueness: The general theory implicitly assumes that the marginal and average profit functions are as shown in Figure 4.5 with a uniquely determined intersection. For several applications the intersection is not unique. This is, for instance, the situation in dairy cows, where the average and marginal profits are as shown in Figure 4.7.

Availability: The theory assumes that a new asset for replacement is always available.

For several of the areas mentioned, the limitations are not so serious that the method cannot be used. When dealing with the animal replacement problem, several characteristics of the problem are problematic in relation to the theory as described already by Ben-Ari et al. (1983). Those characteristics are:

Uniformity: The traits of animals are difficult to define and measure. Therefore, the true state of the factor (the animal) is typically not known with certainty.

Variability: The traits of animals vary considerably among animals as well as over time for the same animal.

Reproductive cycle: For some animals (in particular dairy cows), the reproductive cycle makes the replacement problem two dimensional. It must be decided in which lactation to replace the animal *and* at what stage inside lactation.

Herd constraints: Often the replacement decision for an individual animal is not independent of decisions made for other animals due to herd constraints. Obvious examples include:

Availability of replacement animals: Most often there is a limited supply of e.g. heifers or gilts for replacement.

Production quotas: If for instance a dairy herd produces under a milk quota, the decision to cull one cow influence the ability of the whole herd to meet the quota.

Batch production: In slaughter pigs, for instance, no new piglets are inserted into the pen before it is empty. The principle may even be applied on section level. The slaughter decision therefore has two dimensions: When to slaughter individual pigs, and when to terminate the batch by sending the remaining pigs to the slaughterhouse.

Due to the characteristics listed, the standard method for dealing with the animal replacement problem has become dynamic programming (also known as Markov decision processes) in recent years. For a survey of such techniques, reference is made to Kristensen (1994).

Despite the limitations, there are (in particular in older research) many examples of application of the classical replacement theory in animals. For a survey in relation to dairy cows, reference is made to van Arendonk (1984). In sows, Dijkhuizen et al. (1986) has used the technique.

4.7 Exercises

4.7.1 Gross margin

A dairy farmer has housing facilities for 110 dairy cows, and the size of his milk quota is 650,000 kg milk. During the last quota year he had on average 100 dairy cows. His total milk production amounted to 680,000 kg milk (delivered to the dairy). During the year he sold 45 dairy cows and received total a payment of 320,000 DKK. A total of 110 healthy calves were born and 46 heifers were bought/transferred.

For feeding of the cows the farmer used 213,100 SFU of concentrates and 340,000 SFU of roughage. Various costs for bedding, breeding, veterinary services etc amounted to 113,400 DKK. A permanently employed herdsman has been paid 300.000 DKK whereas interests and depreciation costs of the housing facilities were 300,000 DKK.

Relevant prices were:

Milk, DKK per kg,	2.75
Calves, DKK per calf,	1250.00
Heifers, DKK per animal,	8000.00
Concentrates, DKK per SFU,	1.95
Roughage, DKK per SFU,	1.25
Quota penalty, DKK per kg mil above the limit,	2.70

- a Calculate the gross margin (net returns to housing, labor and management) per cow per year
- b Calculate the gross margin (net returns to housing, labor and management) per kg quota milk
- c Which of the two gross margins is relevant in the herd in question?
- d What is meant by “gross margin”? What is the difference between gross margin and net profit?

4.7.2 Internal prices

A dairy farmer considers whether he shall grow beets for winter feeding or he shall buy industrial by-products instead. If he chooses *not* to grow beets he intends to grow spring barley on the field in question instead. In that case, the barley will be sold. The farmer has made the following calculation of the costs and revenues of growing beets and barley (all figures are per hectare):

	Barley	Beets
Sowing costs, DKK	427	576
Fertilizer, DKK	889	2291
Chemicals, DKK	260	1270
Drying costs, DKK	358	0
Plastic, DKK	0	43
Fuel, cultivation, machinery, DKK	4060	6520
Yield, hkg grain	64	0
Yield, hkg straw	40	0
Yield, SFU	0	12,500
Value of grain, DKK per hkg	124	-
Value of straw, DKK per hkg	20	-

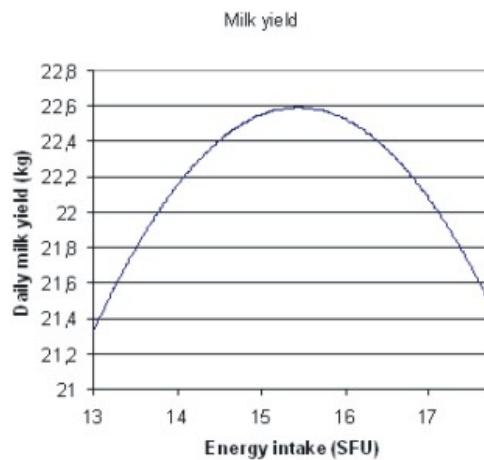


Figure 4.8: Milk yield as a function of energy intake (fixed level of protein and fat).

- a Calculate the gross margin of spring barley.
- b Calculate the internal production price for beets.
- c Shall the farmer grow beets if industrial by-products with exactly the same nutritional content can be purchased for 1.00 DKK?

4.7.3 Marginal considerations, I

At fixed level of crude protein and fat, the relation between energy intake (SFU) and milk yield of dairy cows was found to be

$$Y = -27.94 + 6.546X - 0.212X^2, \quad (4.32)$$

where Y is daily milk yield and X is the daily energy intake. The function is shown in Figure 4.8.

It is assumed that there is no limited quota.

- a Indicate directly in Figure 4.8 the marginal increase in milk yield if the energy intake is increased from 14 to 15 SFU.
- b Calculate the marginal increase in milk yield at the exact energy intake $X = 14$ SFU. Illustrate in Figure 4.8.
- c Determine the *optimal* energy intake under the assumption that the price of an *additional* energy unit is 1.60 DKK and the price of a kg milk is 2.60 DKK.

- d What determines the optimal energy intake, under the assumption that the *production function* (the relation between the factor X and the product Y) is as shown in Figure 4.8?
- e Why is it necessary to assume that there is no quota?

4.7.4 Marginal considerations, II

Bull calves fed ad libitum are assumed to consume the following number of energy units (SFU) per day:

$$X = 3.56 \ln(W) - 13.61, \quad (4.33)$$

where X is daily energy intake (SFU) and W is the weight in kg. The corresponding daily gain G (kg per day) is assumed to be:

$$G = \frac{X}{2.17 \exp(0.00256W)} \quad (4.34)$$

It is assumed that the housing capacity is “unlimited”, that the slaughter price is 12.00 DKK per kg live weight and the feed price is 1.50 DKK per SFU.

- a Describe how to find the optimal slaughter weight of bull calves. Use for instance a spreadsheet to determine it.
- b Why is it necessary to assume that the housing capacity is “unlimited”?

4.7.5 Investments

A pig farmer buys new equipment for 500,000 DKK. It is expected to be used for the next 10 years and the interest rate is 12%.

- a What is the annual cost of the investment?
- b In what way does it matter whether the farmer:
- already has the money in the bank?
 - has to borrow the money in the bank?
 - borrows the money from his rich uncle?
- c What does depreciation mean?

Chapter 5

Basic production monitoring

This chapter presents the theory behind traditional production monitoring systems based on production reports presenting a number of selected key figures. Later chapters will introduce more advanced methods for production monitoring.

5.1 The framework

With reference to the herd management cycle shown in Figure 1.1, production monitoring covers the steps “Check” and “Analysis”. In Figure 5.1, the elements relating to monitoring are shown in an expanded version.

The “Check” is basically a comparison of the results achieved in production with some predefined *target values*. The target values are a result of the planning process, and they are in principle of the same nature as the “goals” shown in Figure 1.1. The only difference (if any) is that the targets are very specific and directly comparable with the key figures calculated from a (more or less continuous) flow of registrations of events relating to the factors or the products. Examples of key figures (and corresponding targets) are conception rate, average daily gain, average daily milk production, average litter size etc.

The result of the check is binary: Either is the result at least as good as the target value or it is not. In case of a negative outcome (the result is not as good as the target value), an *analysis* should be carried out. It consists of two elements:

1. Is the deviation significant from a statistical point of view?
2. Is the deviation significant from a utility point of view?

If the answer to both questions is “yes”, an adjustment of the production may be relevant.

The elements of the monitoring process may be listed as:

- Data recording
- Database

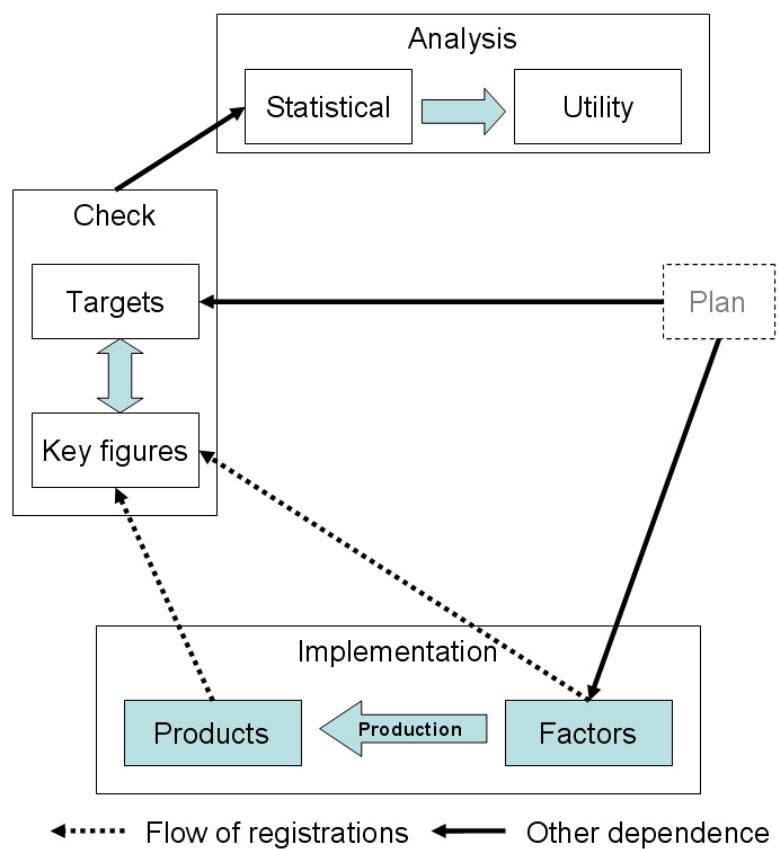


Figure 5.1: Expanded “monitoring part” of the management cycle.

- Data processing
- Report with key figures
- Analysis
 - Statistical
 - Utility
- Decision: Adjustment/no adjustment

In modern livestock production the basic production monitoring is most often supported by some kind of computer program (often called a *Management Information System* or just “MIS”). Such system will typically have an interface for entering records (which may be partly automated) in a database, and it performs the necessary data processing and presents a report with key figures. Sometimes the report also contains relevant target values and even results of a statistical analysis. The utility evaluation, however, is usually not part of the system. In the following subsections we shall discuss the elements of the monitoring process further.

5.2 The elements of the monitoring process

5.2.1 Data recording

Recording (data registration) is usually related to *events* in production. An event is when “something” happens. Examples of events in a sow herd include mating, farrowing, weaning, moving etc. In a dairy herd, events also include for instance test of milk production at cow level, and, in automatic milking systems, we may even consider milking as an event with associated recording. A precise definition of a *record* (the result of a data registration) is shown in Definition 6 together with an actual example of a record registered in a sow herd (Example 8).

Definition 6 A *record*, ρ , is defined by the following properties:

Event, $\rho.e$: What does the record relate to?

Identification, $\rho.i$: The unique number of the animal(s)/section/batch in question.

Registration level, $\rho.l$: Is it on animal/pen/section/batch/herd level?

Time, $\rho.t$: Date (or date and time).

Property, $\rho.p$: What is measured?

Value, $\rho.v$: The numerical or categorical value actually measured.



Example 8 An actual record from a sow herd.

Property	Value
Event	$\rho.e = \text{"Farrowing"}$
Identification	$\rho.i = \text{"Sow \#1234"}$
Registration level	$\rho.l = \text{"Animal level"}$
Time	$\rho.t = \text{"February 1st, 2006"}$
Property	$\rho.p = \text{"Live-born piglets"}$
Value	$\rho.v = 12$

The traditional registration method is paper and pencil, but in recent years also different kinds of electronic equipment has become available for data recording.

5.2.2 Database

The individual records collected over time are typically entered into a database. A (relational) database is organized as a collection of interrelated *tables*. Proper design of a database (i.e. what tables to define and how to interrelate them) is not a trivial task, but, for the ordinary user it is not necessary to master the art of database design. Readers who are interested in the subject may consult a textbook as for instance Watson (2002).

Even though most MISs provide automatically generated production reports with key figures it is important to know the principles behind the use of a database. Such knowledge will enable the user to make more detailed and/or more specialized analyses of the production than those provided by the MIS.

The communication with the database is usually done by use of the Structured Query Language (SQL) which is a common standard for describing and querying databases. It is available with many commercial relational database products, including DB2, Oracle and Microsoft Access (Watson, 2002). Most MISs rely on a standard commercial database, implying that it can be accessed by usual statistical software tools like SAS or R. Occasionally it happens that the manufacturer of the MIS has blocked the database for external communication by password protecting it. Such limitations should *not* be accepted by farmers.

A typical way of organizing the database is by collecting records relating to the same event in a table. Thus, the database typically has a table for each event considered. Example 9 shows a table for the event “Farrowing” in a sow herd. Typically each property of a record (cf. Definition 6) either corresponds to a column in the table, or, it is a general property of the table as such. In the example, there are columns for “Identification” (Sow #), “Time” (Date) and “Property” (Live-born), whereas “Event” and “Registration level” are general properties of the table (it is a table for the event “Farrowing” and all records are at animal level).

Example 9 A table in the database of a sow herd.

Sow #	Date	Live-born	Stillborn	Course
1234	15 January 2006	12	2	Easy
678	16 January 2006	9	4	Diff.
1001	18 January 2006	14	1	Easy
...
...
...

It should be noticed that each row in the table actually contains information about 3 different records, but, all relating to the same event (“Farrowing”). The example assumes that 3 properties (cf. Definition 6) are measured for the event “Farrowing”: Number of live-born piglets, number of still-born piglets, and the course of the farrowing as a categorical value (either “Easy” or “Difficult”). Since the number of total born piglets is a simple sum of live-born and stillborn, a separate column for “Total born” will typically be omitted in order to avoid redundant information in the database.

The database as such may be regarded as a collection of tables, where each table typically relates to an event.

5.2.3 Data processing

The data processing dealt with here corresponds to the first processing shown in Figure 2.1. In this chapter the description will be kept at a basic level in agreement with traditional monitoring systems. A description of the path from data to information at a more advanced formal level will be given in Chapter 6.

Referring to Figure 2.1, the information resulting from the processing in the traditional basic setup is a number of selected key figures characterizing the efficiency of production. The final decision at the end of the path is whether or not to make adjustments (and what kind of adjustments).

The purpose of the processing is *data reduction*. In principle all information about the herd is already present in the database, but nobody will be able to interpret it unless it is reduced to a dimension that can be over viewed by the decision maker.

The key figures will typically refer to a specific period, i.e. the most recent month, quarter or year. What the MIS actually does in order to obtain the key figures is to ask questions to the database (using the SQL language). The answers are then further processed into simple or more complex key figures. In Table 5.1 the calculation of a few selected key figures for a sow herd is shown. It is obvious, that if registrations of relevant events like farrowing, mating, and pregnancy diagnosis have been performed, the answers to the questions are easily found in the database. Readers, who are interested in the syntax of the SQL language, used for asking questions, may for instance consult Watson (2002).

In some cases, the questioning of the database and the subsequent processing may be more complex. As an example, we shall use the key figure “Average daily

Table 5.1: Examples of simple key figures for a sow herd. The key figures are assumed to refer to a period of w weeks.

Key figure	Question asked to the database	Answer	Formula
Farrowings per week	How many farrowings were there in the period?	f	f/w
Matings per week	How many matings were there in the period?	m	m/w
Live-born piglets per litter	How many piglets were born in the period? How many farrowings were there in the period?	p f	p/f
Conception rate	How many matings were there in the period? How many positive pregnancy diagnoses were there in the period?	m d	d/m

gain” in slaughter pigs. In commercial production systems, slaughter pigs are not individually identified. The straightforward calculation of individual gain from insertion to slaughter followed by calculation of the average value per day can therefore not be used. Instead an indirect method based on valuation weighing is often used. The principles are shown in Example 10.

Example 10 Calculation of average daily gain in slaughter pigs.

When the pigs are not individually identified the *total* gain for a certain period may be calculated as follows:

+	Total weight, slaughtered pigs	w_s
+	Total weight, dead pigs	w_d
+	Valuation weight, end of period	v_e
-	Total weight, piglets inserted	w_i
-	Valuation weight, beginning of period	v_b
=	Total gain during the period	g_t

By “valuation weight” we mean the total weight of all slaughter pigs being present in the herd on a particular day. In order to arrive at the average daily gain, the total gain g_t must be divided by the total number of days in feed, d , in the period. Thus, the average daily gain g is $g = g_t/d$.

In order to retrieve the values used for calculation of g_t we must have tables in the database for the events “Slaughtering”, “Death”, “Insertion” and “Valuation”. Based on the tables for slaughtering, death and insertion we may at any day of the period calculate the number of pigs present in the herd. The sum of those numbers over all days of the period will provide us with the number of days in feed, d . ■

It should be emphasized, that for any rational interpretation of the key figures, it is necessary to know exactly how it is calculated, and what records it is based on. Such information is usually given in the documentation provided by the manufacturer of the MIS. If that is not the case, the farmer shouldn't buy it!

5.2.4 Report with key figures

If a MIS is used for the monitoring process, the key figures shown in the report are often more or less predefined even though some systems allow for selection of figures to be presented and thus provide a limited degree of flexibility.

Usually the report contains (too) many key figures and it may even be a problem for the farmer to choose the most relevant for closer evaluation. In the following subsections, the principles for choosing the best key figures and for interpretation will be described.

Types of key figures

The purpose of key figures is to provide the farmer with a condensed set of information expressing the current state of the production. Many different kinds of key figures each expressing certain aspects of production can be defined. In general, we distinguish *economical* key figures from *technical* key figures. Whereas the first are expressed in money, the latter are expressed in product or factor units (or ratios). In this book, we shall mainly discuss technical key figures. Typically they attempt to express one of the following basic properties:

Productivity: Such key figures express the average production level of a product per time unit. Let $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{nt})'$ be the vector of n products produced at stage t . If we wish to express the productivity of product i in the period from stage $t - k$ to stage t , we may simply calculate the productivity π_{it} as

$$\pi_{it} = \frac{1}{k} \sum_{\tau=t-k}^t Y_{i\tau} \quad (5.1)$$

Typical examples of key figures expressing productivity include “piglets produced per week”, “average daily milk production” etc.

Factor consumption: This is just the counterpart of productivity. Let $\mathbf{x}_t = (x_{1t}, \dots, x_{mt})'$ be the vector of m factors consumed at stage t . The average consumption c_{it} of factor i in the period from stage $t - k$ to stage t , is calculated as

$$c_{it} = \frac{1}{k} \sum_{\tau=t-k}^t x_{i\tau} \quad (5.2)$$

Even though key figures in this category are rare, examples could be “feed consumption per month”, “average number of cows” etc.

Technical efficiency: Technical efficiency expresses the ratio of a product to a factor. Using the same notation as before, the key figure κ_{ijt} expressing the efficiency of the production of product i relative to the consumption of factor j is calculated as

$$\kappa_{ijt} = \frac{\pi_{it}}{c_{jt}} = \frac{\sum_{\tau=t-k}^t Y_{i\tau}}{\sum_{\tau=t-k}^t x_{i\tau}} \quad (5.3)$$

Key figures expressing technical efficiency are probably the most commonly applied and they include so well known properties as “milk yield per cow per year”, “piglets per sow per year”, “gain per feed unit” in slaughter pigs/calves etc. If we extend the definitions of factors and products to cover also input and output of sub-areas of the production process, even key figures as for instance “conception rate” (factor = insemination, product = pregnancy) fall into this category.

The only economical key figures we shall briefly discuss here are those expressing gross margins. They may be calculated as total figures, or relative to a main factor or a main product. In dairy herds, for instance, we may calculate gross margin per cow per year and/or gross margin per kg of milk produced. In general, relative values should be calculated relative to the most limiting constraint. Thus, if the dairy farmer produces under a quota, the most relevant figure is gross margin per kg of milk (if the the quota is the most limiting constraint). In the absence of a quota, it is more relevant to consider gross margin per cow per year, because in that case, the housing capacity is assumed to be the most limiting constraint.

Key figures and their interrelations

Even though the production report may contain many different key figures, it is important to realize that many are heavily correlated. In order to illustrate the interrelations, a collection of typical technical key figures for a sow herd are shown in Figure 5.2. They are all examples of commonly defined properties used for characterizing a sow herd.

A sow herd has one well-defined main product which is piglets (sold on the market or kept for fattened by the farmer himself). Therefore, the over-all productivity of the sow herd is expressed well by the key figure “piglets sold per year”. This key figure is shown at the top of the diagram, and with that as a starting point, a tree of key figures has been created by simple logical reasoning. The principle is that an arrow between two key figures indicates a causal influence. Starting from the top, the figure illustrates that the number of piglets sold is influenced (in this case determined) by the number sold per sow (an efficiency key figure) and the number of sows (a factor consumption key figure) and so on downwards. In general, all key figures below a certain position in the tree have an explanatory effect on the value at the position in question.

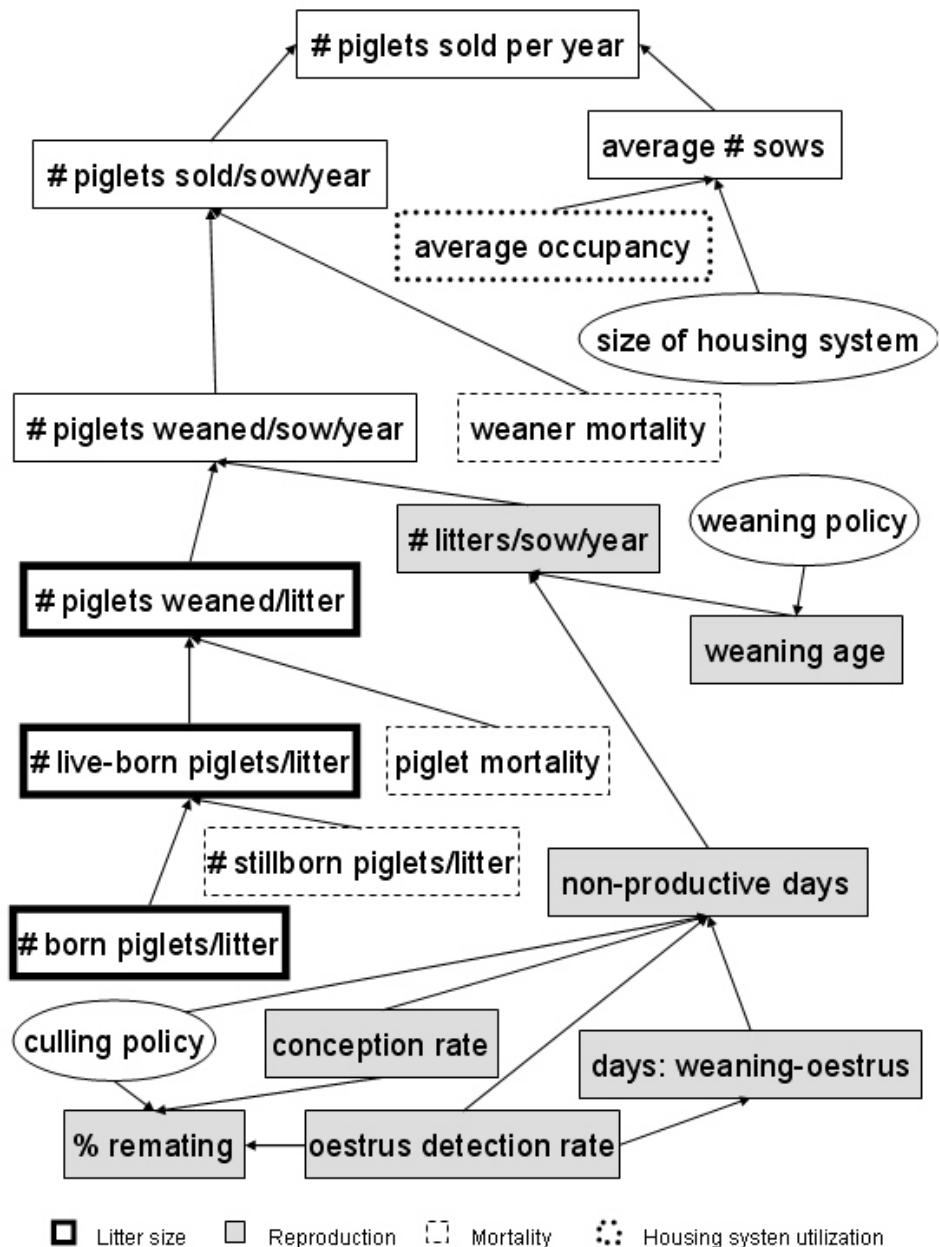


Figure 5.2: Typical key figures (rectangles) of a sow herd. Ellipses indicate decisions made by the manager. The arrows indicate interdependencies.

The figure illustrates that the over-all productivity in a sow herd may be explained by a large number of effects that basically can be split up into 4 distinct groups as the symbols indicate. The groups are:

- Litter size productivity.
- (Foetus and) piglet mortality.
- Reproductive efficiency.
- Housing system utilization efficiency.

Whereas the various key figures for litter size productivity are (logically) heavily inter-correlated, the 3 key figures for mortality are not logically inter-correlated (i.e. there are no arrows between them). They may, however, often be *biologically* inter-correlated in the sense that if a herd has a high piglet mortality, it is probably because of bad management, disease pressure etc. It is therefore likely that the same applies to weaners with a corresponding high mortality as a consequence.

As concerns the key figures for reproductive efficiency, the “root” of the subtree covering those figures, “litters per sow per year” represents a kind of total productivity concerning reproduction. The figures closer to the leaves of the tree accordingly represent explanations for the over-all reproductive efficiency. Proceeding to the leaves, the overall efficiency is split up into properties relating to the sows (“conception rate” and “days from weaning to oestrus”) and one relating to the staff (“oestrus detection rate”).

In the selection of key figures to monitor with particular care, the general advice is to select a few that are not (or only slightly) correlated. For the productivity in a sow herd it would be natural to choose one from each of the 4 groups identified in Figure 5.2.

It should be emphasized that the principles of Figure 5.2 where very aggregated properties (on the top) are explained by more specific underlying properties may be applied to any livestock production and to some extend also for key figures measuring factor consumption. It is left to the reader as an exercise to construct a similar figure for productivity in a dairy herd.

Properties of key figures

In the interpretation of calculated key figures, at least three properties are relevant:

Correctness: This property deals with the observation that the data registrations behind it must all be correct. From a theoretical point of view, the observation is trivial, but, from a practical point of view it is certainly not. Numerous errors may occur under the registration of data: The identification may be wrong (an observed insemination or farrowing may be ascribed to a wrong sow), registrations may be lost or forgotten, values may be misread etc.

Validity: The validity of a key figure expresses to what extent it gives us the desired information. As discussed in Section 2.3.1 we often want to know something that is not directly observable. We are therefore forced to observe or calculate something that is *related* to the desired property. It is, however, not the same. If it is “close” to the desired property, the validity of the key figure is “high”. If it is only slightly related to the desired property, the validity is “low”. These vague terms illustrate that we are usually not able to calculate a numerical value for the validity, but, only use a crude classification. It must be emphasized that the validity is not an absolute property of a key figure. It only makes sense when relating the key figure to a well defined desired information. In other words, the same key figure may have “low” validity in relation to one desired information and high in relation to another kind of information.

Precision: The precision expresses the certainty by which the key figure has been determined. We may regard the calculated value as an estimate (in statistical sense) of a true underlying value. A formal definition is given in Definition 7.

Definition 7 *Precision of a key figure:* Let the key figure κ be an estimate of an underlying true value θ estimated on the data set D using a known estimation technique resulting in an unbiased estimate (i.e. $E(\kappa) = \theta$) with the standard deviation σ so that $V(\kappa) = \sigma^2$. The precision of the key figure τ is then

$$\tau = \frac{1}{\sigma^2}. \quad (5.4)$$

■

In other words, the precision of a key figure is an absolute numerical value. It is one of the most important properties at all to consider when we interpret key figures. From the definition it is clear, that if a large data set D is available (i.e. it is a big herd), the precision will typically be high as we shall see later.

Even though Definition 7 is the correct and formal, we will often use the standard deviation σ as a direct expression for the precision. In that case, low values of σ correspond to high precision and vice versa.

Correctness

It is in practise very difficult to actually guarantee that the registrations behind a key figure are correct. This is simple a consequence of the fact that errors may occur in so many different ways that it is more or less impossible to foresee all of them.

What a MIS typically does in practise is to run some automatic checks of records. Examples of such checks are given in Example 11. As it is seen, such verification checks are often trivial, but nevertheless, they are very useful. The strength is that records that are obviously wrong are identified, but there is certainly no guarantee that records passing the tests are correct. If they are wrong, but not unusual, they will pass. To decide the acceptance limits of the checks is a tradeoff between the wish to identify wrong values and to avoid discarding records that are actually correct (even though they may be extreme). It is, for instance, not trivial to decide on an absolute maximum for a litter size to be accepted.

Example 11 Typical verification checks for a data record ρ (with fields as in Definition 6) from a sow herd.

Event, $\rho.e$	Property, $\rho.p$	Verification check
Farrowing	Litter size	$0 \leq \rho.v < 25$
Farrowing	(Any)	A mating record exists for the same sow
Mating (sow)	(Any)	A weaning record exists for the same sow
Weaning	Weight of piglets	$5n < \rho.v < 10n$, (n = number of piglets)

Another kind of verification check typically done is to create a kind of balance accounting for all animals present in the herd. If a valuation has been carried out on a certain historic date, and records of all animals born, bought, sold or dead are kept, a balance as illustrated in Example 12 may be created.

Example 12 Balance accounting for all animals in a slaughter pig unit.

+ Slaughter pigs in the unit, previous valuation day	n_p
+ Weaners bought since previous valuation day	n_b
- Dead slaughter pigs since previous valuation day	n_d
- Sold slaughter pigs since previous valuation day	n_s
= Slaughter pigs present in the unit today	n_t

The number of pigs present in the unit (determined by counting) must match the calculated number n_t in order for the records to be correct.

Again, even though the balance check of Example 12 is extremely useful in practise, passing it still doesn't guarantee that the records are correct. Passing is a *necessary* condition for correct records, but it is certainly not a *sufficient* condition. Nevertheless it is a simple, but effective, method for checking that "all" events have been registered.

To sum up, several verification checks can be defined, but only care with registrations can ensure that the records behind a key figure are correct.

Validity

Of the three basic properties of key figures, validity is probably the most difficult to understand. In order to improve the understanding of the concept, we shall take

a look at few examples (more examples can be found in the exercises of Section 5.3).

Suppose that we wish to describe the reproductive efficiency of a sow herd by one single key figure. The question is then which one we should choose. Referring to Figure 5.2, it is obvious that we should search among those shown on a grey background. Let us take a look at each of them one by one and assess their validity:

Litters per sow per year: If what we really wish to know is something about the over-all reproductive efficiency of the sow herd, this figure has a rather high validity. It is, however, also influenced by the weaning and culling policy of the pig farmer. Since the weaning and culling policy has nothing to do with the reproductive efficiency, it decreases the validity. If these policies are constant over time, changes in the value of the key figure reflect changes in reproductive efficiency. Thus for within farm use, the validity is high, but if key figures for different farms are compared, the validity is lower, because the weaning and culling policies probably vary among farms. Even for within-farm use, the key figure is rather unspecific. If the value decreases, it is not possible to know whether it is because of, for instance, low conception rate or low oestrus detection rate.

Weaning age: This is almost entirely a consequence of the weaning policy of the farm. Thus, the validity as an expression for reproductive efficiency is extremely low.

Non-productive days: This key figure is very often used as an expression for reproductive efficiency. It has a higher validity than “litters per sow per year”, because it does not depend logically (but slightly biologically) on the weaning policy. There is, however, still a marked influence of the culling policy (if re-maters are always culled, the number of non-productive days will decrease).

Days from weaning to oestrus: This is a rather specific key figure attempting to measure a property of the sows in the herd. Regarded as an expression for the state of the sows, the validity is decreased by the fact, that what is measured is *detected* oestrus. The value is therefore also influenced by the oestrus detection rate which is a property of the farmer.

Conception rate: An attempt to measure the ability of the sows to conceive. The value also expresses the efficiency of the inseminations (timing, stimulation, procedure) thus reducing the validity as an expression for a property of the sows.

Percent re-mating: A very composite key figure combining effects of conception rate, oestrus detection rate and culling policy. The validity (for almost any purpose) is therefore low.

Oestrus detection rate: An attempt to measure the ability of the staff to detect heat. It is not used very often in pig herds (whereas it is commonly applied in dairy herds). Even a very small probability of false positive detections may change the value dramatically.

The examples concerning sow herd reproduction illustrate that if a key figure is systematically influenced by aspects with no (or only little) relevance for the information we wish to obtain, the validity of the key figure as source of information is decreased. Thus, again, it must be emphasized that the concept of validity only makes sense in relation to specific information we request.

Precision

The standard deviation of a key figure expresses the degree of certainty of the calculated value. The uncertainty has two sources:

The sample uncertainty: This uncertainty expresses the natural variation (often referred to as the *biological variation*). It corresponds to the trivial observation that even though the production has been conducted in exactly the same way (i.e. exactly the same factors have been used in the same amounts) in two periods, we nevertheless, don't expect exactly the same result simply because of the random nature of (biological) production. Mathematically, the sample uncertainty is a consequence of the random term e_t of the production function (1.1) which is *not* under the control of the farmer.

The observation uncertainty: Depending on the registration methods used in order to obtain the records, a measurement error may be involved.

The effects of the two sources of uncertainty may be expressed by the following model:

$$\kappa = \theta + e_s + e_o, \quad (5.5)$$

where κ is the calculated key figure, θ is a true underlying value (which we try to estimate), e_s is the sample error, and e_o is the observation error.

Assuming the model (5.5), the variance of the estimate is $\sigma^2 = V(e_s + e_o)$. If the two error terms are independent (which they usually are), we have $\sigma^2 = V(e_s) + V(e_o)$. In other words, the key to the precision is to calculate the variances of e_s and e_o . In many cases it is realistic to assume both error terms to be normally distributed, i.e., $e_s \sim \mathcal{N}(0, \sigma_s^2)$ and $e_o \sim \mathcal{N}(0, \sigma_o^2)$. This is often a reasonable assumption for continuous properties like milk yield. In other situations, the variance of the key figure is the result of binomial variation, which is the case when the key figure measures a binary output as for instance with conception rate and heat detection rate. Also other distributions as for instance the poisson distribution may be used.

The next section contains a number of practical examples of how to calculate the precision.

5.2.5 Analysis from a statistical point of view

General principles

The way in which the key figures are actually presented depends on the MIS used in the herd. A typical way is, however, a table where each key figure is shown in a separate row containing the name of the key figure, the calculated value (perhaps for several periods) and sometimes also a target value. An example of such a row in a report for slaughter pigs is shown in Example 13.

Example 13 A row of a typical production report for slaughter pigs

Property	Quarter t-3	Quarter t-2	Quarter t-1	Quarter t	Target
...
Daily gain (g)	988	934	896	925	950
...

In the example, the target value for daily gain was 950 g, but the achieved value for the present quarter (time t) was only 925 g. The basic question is therefore: *Should we be worried?* It is easily seen that the calculated gain is 25 g less than the expected target, but, could it be a simple consequence of random biological variation and/or observation errors?

In statistical terms, the basic question is whether the calculated key figure κ could be an observation from a distribution having the mean θ' (where θ' is the target value) and the variance $\sigma^2 = V(e_s) + V(e_o)$. A formal way of testing this is to test the null hypothesis

$$H_0 : \theta = \theta' \quad (5.6)$$

Often, we are only worried in case of “negative” deviations from the target value. It may then be relevant to use a one-sided test of the type $H_0 : \theta \geq \theta'$ instead of (5.6). Under all circumstances, if the null hypothesis is rejected, we should be worried. The deviation of the calculated value from the target is in that case *not* likely to be the result of random biological variation and/or observation errors.

The practical setup of the test depends on the underlying distributions implying that it may be relevant to distinguish between continuous and categorical properties. In both cases, the most difficult part is to calculate the precision.

Even though the practical setup differs, the procedure under all circumstances involves the following steps:

1. Identification of the formula(s) used for calculation of the key figure from records.
2. Identification of the distributions of the data.
3. Identification of the distribution of the key figure under the null hypothesis.

Table 5.2: Average daily milk yield (ECM) during the first 24 weeks of the lactation in a herd.

Parity	No. of cows	Key figure	Target
1	10	20.4	23.5
≥ 2	12	26.0	27.8

4. Calculation of confidence limits for the key figure based on the precision.
5. Acceptance or rejection of the null hypothesis.

Referring to the “key figure tree” of Figure 5.2, the procedure will typically be to focus on the key figures in the “leaves” of the tree. Those close to the root are very aggregated values, where in particular identification of the distribution is difficult.

Continuous properties

We shall illustrate the procedure for the statistical evaluation through two examples concerning continuous properties: Average daily milk yield of dairy cows and average daily gain in slaughter calves.

Example 14 Average daily milk yield

In the Danish management information system for dairy cows, the average daily milk yield of the cows during the first 24 weeks of lactation is calculated for first lactation cows and other cows, respectively. In the system, the farmer may compare the results to his own targets (expected values). The results in a herd were as shown in Table 5.2. We observe, that the calculated result for all parities is lower than expected. The question is now, whether or not the deviation is significant from a statistical point of view.

Following the procedure described in last section we notice and conclude as follows:

Calculation of the key figure from data: The basic records behind the key figure are test day milk yields (including protein and fat content) at cow level. Using a standard method (which we shall not discuss further here) the individual test day milk yields are processed into individual cumulated milk yields in energy corrected milk (ECM) over the first 24 weeks of the lactation. We shall denote the calculated cumulated milk yield (ECM) of cow i as $\hat{Y}_{24,i}$ for $i = 1, \dots, n$ cows. The “hat” over the symbol $\hat{Y}_{24,i}$ is used to indicate that it is *not* the true cumulated milk yield, which we accordingly could denote as $Y_{24,i}$. The key figure κ , average daily milk yield (ECM) during the first 24 weeks of the lactation, is then calculated as

$$\kappa = \frac{1}{(7 \times 24)n} \sum_{i=1}^n \hat{Y}_{24,i}. \quad (5.7)$$

Distribution of the data: The basic data records here are actually test day milk yields (with protein and fat contents). The cumulated milk yield $\hat{Y}_{24,i}$ is therefore *not* directly observed. The use of a calculated value $\hat{Y}_{24,i}$ from test day records, implies that there is an observation error e_{oi} involved ($e_{oi} = \hat{Y}_{24,i} - Y_{24,i}$). The variance of the observation error depends on the number of test day records behind the calculated value (in case of weekly test day records, the variance of the observation error is smaller than if we have only monthly test day records). In a general approach, we should take the variance of the observation error into account when estimating the within-herd distribution of $\hat{Y}_{24,i}$, but in this particular case, Kristensen (1986) has estimated the within-herd variance of $\hat{Y}_{24,i}$ directly. He estimated the standard deviation of cumulated milk yield (ECM, first 24 weeks) between cows in the same herd as 490 kg milk for first lactation and 630 kg milk for other lactations. Since these standard deviations have been estimated directly on *calculated* cumulated milk yields, they actually represent the standard deviation of $\hat{Y}_{24,i}$ directly. Thus, the variance originating from the observation error is already taken into account. For a continuous property like cumulated milk yield, it is natural to assume a normal distribution, so we can conclude that $\hat{Y}_{24,i} \sim \mathcal{N}(\theta_1, 490^2)$ for first lactation, and $\hat{Y}_{24,i} \sim \mathcal{N}(\theta_2, 630^2)$ for older cows.

Distribution of the key figure: Using standard rules from statistics, the key figure as calculated from Eq. (5.7) is also normally distributed with mean θ_1 and θ_2 for first lactation and older cows, respectively. The standard deviation of the key figure is then (again according to standard rules) calculated as (for 1st lactation cows) $490/7/24/\sqrt{n}$ and (for older cows) $630/7/24/\sqrt{n}$. According to Table 5.2, there were 10 first lactation cows and 12 older cows. In other words, $n = 10$ and $n = 12$, and the standard deviation of the key figure for 1st lactation is therefore $490/7/24/\sqrt{10} = 0.92$. For other cows, the standard deviation is $630/7/24/\sqrt{12} = 1.08$. It should be noticed, that the standard deviation is a direct measure of the precision as it is easily seen from Definition 7.

Confidence limits: As a rule of thumb for normally distributed data, a deviation is significant if it exceeds the standard deviation multiplied by 2, so in this case, acceptable results would be $23.5 - 2 \times 0.92 = 21.7$ and $27.8 - 2 \times 1.08 = 25.6$.

Acceptance or rejection: Since the calculated key figure concerning first lactation cows is clearly below the deduced acceptable value, we conclude that the deviation is significant from a statistical point of view. On the other hand,

Table 5.3: Results used for calculation of average daily gain in a herd. Refer also to Example 10.

Total live weight of 25 calves delivered, kg	w_s	12,500
Valuation weight at the end of the period, kg	v_e	21,159
Total weight of 20 calves inserted, kg	w_i	-950
Valuation weight at the beginning of the period, kg	v_b	-23,441
Total gain during the period, kg	g_t	9,268
Total number of days in feed	d	8,827
Daily gain, g	$g = g_t/d$	1,050

the result concerning other cows is higher than the acceptable value, and accordingly we have to accept the result as a possible consequence of usual random variation. In other words, the result concerning those cows does not call for any adjustments.

■

Example 15 Daily gain of slaughter calves

Usually, total gain, y_t , for bull calves or slaughter pigs over a certain period is calculated as shown in Example 10 by use of valuation weighings and total days in feed. We shall assume that the standard deviation in true daily gain between calves of the same herd amounts to 200 g.

Calculation of the key figure from data: Based on the feeding and the housing conditions the goal for daily gain has been set to 1150 g in a herd with approximately 100 bull calves. The actual results registered in the herd for a 3 month period appear from Table 5.3 (no dead calves in the period).

Distribution of the data: The records behind the key figure are those shown in Table 5.3. It is reasonable to assume that all weight observations are drawn from normal distributions. The values of w_s and w_i are assumed to be sums of individual weighing records at animal level. We shall assume that the uncertainty of these values is negligible so that there is no observation error involved. Also the number of days in feed is assumed to be observed without error. The valuation weights, on the other hand, may be determined in many different ways. The observation error on these figures depends heavily on the method used. We shall assume that the standard deviation of the observation error (on v_e and v_b) is zero if all animals (approximately 100) have actually been weighed. If the total weight is calculated by weighing only representative calves, the corresponding standard deviation is assumed to be 5%. Finally, if the valuation weights are just assessed by visual observations, the standard deviation is assumed to be 10%.

Distribution of the key figure: Given the assumptions for the records, the key figure itself will also be normally distributed. The standard deviation will according to the general model (5.5) be calculated from the standard deviation, σ_s , of the sample error and the standard deviation, σ_o , of the observation error.

The value of σ_s is based on the assumption that the standard deviation of individual daily gains within a herd is 200 g. For the average of 100 calves, it is therefore (according to standard rules from statistics) $\sigma_s = 200/\sqrt{100} = 20$ g.

If the method used for valuation is weighing of all animals, there is no observation error implying that $\sigma_o = 0$. If only a representative fraction of the calves is actually weighed, σ_o was 5%, or in other words $\sigma_o = \sqrt{(21,159 \times 0.05)^2 + (23,441 \times 0.05)^2} = 1,579$ g. If only visual observations are used, we have $\sigma_o = \sqrt{(21,159 \times 0.10)^2 + (23,441 \times 0.10)^2} = 3,158$ g.

For the total standard deviation of the key figure, $\sigma = \sqrt{\sigma_e^2 + \sigma_o^2}$, we have $\sigma = \sqrt{20^2 + 0^2} = 20$ g, $\sigma = \sqrt{20^2 + 1,579^2} = 180$ g, and $\sigma = \sqrt{20^2 + 3,158^2} = 359$ g, respectively, for the three methods.

Depending on the observation method, the distribution of the key figure is under the null hypothesis $H_0 : \theta = 1,150$ g:

Weighing of all animals: $\kappa \sim \mathcal{N}(1150, 20^2)$

Weighing of a representative fraction: $\kappa \sim \mathcal{N}(1150, 180^2)$

Visual weight estimation: $\kappa \sim \mathcal{N}(1150, 359^2)$

Confidence limits: As in the previous example, we use the rule of thumb for normally distributed data, that a deviation is significant if it exceeds the standard deviation multiplied by 2. Thus, we obtain the following acceptable lower limits for daily gain under the null hypothesis:

Weighing of all animals: $1150 - 2 \times 20 = 1110$ g.

Weighing of a representative fraction: $1150 - 2 \times 180 = 790$ g.

Visual weight estimation: $1150 - 2 \times 359 = 432$ g.

Acceptance or rejection: The calculated daily gain was 1050 g. In other words, we conclude that if all animals have been weighed, the deviation is significant from a statistical point of view. If, however, only a fraction has been weighed, the deviation may very well be a result of random fluctuation. Neither is the deviation significant if the weight has been estimated visually.



Categorical properties without observation error

The previous examples referred to variables which might be regarded as at least approximately normally distributed. We now turn to examples involving categorical data, where other distributions are involved. We will first focus on examples without observation error.

Example 16 Reproduction in a dairy herd

The key figure considered here is the conception rate. Initially we shall assume that heat and pregnancy is observed without errors. The value 0.4 was observed in a herd, where the expected target value was 0.5.

Calculation of the key figure from data: The calculation is very simple. The number of inseminations N in a period is counted, and the number of cows n , where $0 \leq n \leq N$, conceiving is counted as well. The key figure κ is then just calculated as

$$\kappa = \frac{n}{N}. \quad (5.8)$$

Distribution of the data: For given N cows inseminated, the number n of cows conceiving will be binomially distributed with parameters N and θ , where θ is the conception rate.

Distribution of the key figure: We shall not use the distribution of κ for the further analysis, which instead will be based on the binomial distribution of n . The analysis assumes that N is known. Under the null hypothesis $H_0 : \theta = 0.5$, we have that $n \sim \mathcal{B}(N, 0.5)$.

Confidence limits: The lower acceptable limit is calculated under the null hypothesis, a chosen confidence level (e.g. 5%) and the parameters N and θ . Let $F_{N,\theta}(n)$ be the cumulative probability function of the binomial distribution with parameters N and θ . The lower acceptable limit n_a for n is then determined as the 0.05 quantile of $F_{N,\theta}(n)$, i.e.

$$n_a = F_{N,\theta}^{-1}(0.05). \quad (5.9)$$

If, for instance, 50 cows were inseminated (i.e. $N = 50$) we get $n_a = F_{50,0.50}^{-1}(0.05) \approx 19^1$.

Acceptance or rejection: Since the observed rate was 0.40 it means that out of the 50 cows inseminated, 20 cows actually conceived. Since the observed value is higher than the lower acceptable limit, we conclude that the deviation from the target may very well be the result of random variation.

¹Most statistical software packages come with built-in implementation of the cumulative (and sometimes even the inverse cumulative) distribution functions of standard distributions like the binomial, the Poisson and the hypergeometric distributions. Alternatively, the quantiles may be found in statistical tables.

Example 17 Disease cases

Often, a production report generated by a standard MIS contains key figures representing simple counts of disease cases. In this example we assume that individual disease cases are observed with certainty.

A dairy herd observed for a period 11 cases of retained placenta. The expected target value was 7 cases, and, as usual, we wish to know whether we should be worried from a statistical point of view.

Calculation of the key figure from data: The key figure κ is a simple count of the number of cases in a given period.

Distribution of the data: Because of the simplicity of the key figure, we address its distribution directly.

Distribution of the key figure: If the herd size and the disease incidence is constant, the number of cases of a specific disease may be represented by a Poisson distribution provided that individual cases appear independently of each others (which is not the case with contagious diseases). The actual number of cases may therefore be compared to the target (or expected result) using probabilities calculated from a Poisson distribution taking the defined target as its parameter. In other words, under the null hypothesis $H_0 : \theta = 7$, the key figure κ is Poisson distributed with expected value 7, i.e. $\kappa \sim \mathcal{P}(7)$.

Confidence limits: The upper acceptable limit is calculated under the null hypothesis, a chosen confidence level (e.g. 5%) and the parameter $\theta = 7$. Let $F_\theta(\kappa)$ be the cumulative probability function of the Poisson distribution with parameter θ . The upper acceptable limit κ_a for κ is then determined as the 0.95 quantile of $F_\theta(\kappa)$, i.e.

$$n_a = F_\theta^{-1}(0.95). \quad (5.10)$$

For $\theta = 7$, we get $n_a = F_\theta^{-1}(0.95) \approx 11$.

Acceptance or rejection: Since the observed number was 11, which is equal to the upper acceptable limit, we conclude that the observed deviation does not exceed what can be explained by random fluctuations, even though it is not likely (the deviation is significant at a 6% confidence limit).

Example 18 Disease prevalence

In systematic herd health programs, the veterinarian often estimates the prevalence of a disease. The typical method used is clinical examination of a representative sample of the animals.

In a sow herd with $N = 500$ sows, $n = 50$ sows were drawn at random for clinical examination. The observed prevalence was $\kappa = 0.10$, where the target value defined for the herd was $\theta = 0.05$. Initially we make the (often unrealistic) assumption that the veterinarian is able to distinguish with certainty between healthy and diseased animals.

Calculation of the key figure from data: The basic data material behind the key figure is n^+ positive diagnoses and n^- negative, where $n^+ + n^- = n$. The observed prevalence is simply $\kappa = n^+/n$.

Distribution of the data: Because of the simplicity of the key figure, we address its distribution directly.

Distribution of the key figure: Under the null hypothesis $H_0 : \theta = 0.05$, the number of positive diagnoses n^+ has a hypergeometric distribution with parameters $N = 500$, $n = 50$ and $D = N\theta = 500 \times 0.05 = 25$. In other words, $n^+ \sim \mathcal{H}(25, 500, 50)$.

Confidence limits: The upper acceptable limit is calculated under the null hypothesis, a chosen confidence level (e.g. 5%) and the parameter $\theta = 0.05$. Let $F_{N,n,\theta}(n^+)$ be the cumulative probability function of the hypergeometric distribution with parameters N, n and θ . The upper acceptable limit n_a^+ for n^+ is then determined as the 0.95 quantile of $F_{N,n,\theta}(n^+)$, i.e.

$$n_a^+ = F_{N,n,\theta}^{-1}(0.95). \quad (5.11)$$

For $N = 500$, $n = 50$ and $\theta = 0.05$, we get $n_a^+ = F_{N,n,\theta}^{-1}(0.95) \approx 4$.

Acceptance or rejection: Since the observed prevalence was $\kappa = 0.10$, it means that 5 out of the 50 examined sows were positive. Since this number is bigger than the upper acceptable limit n_a^+ , we conclude that it is not likely that the true prevalence is $\theta = 0.05$.

■

Categorical properties with observation error

In cases where we basically deal with binary data (pregnant versus not pregnant, diseased versus healthy etc.), the classification of a unit (often an animal) into one of the two observational classes is often based on some kind of test. Depending on the property observed, the nature of the test will vary considerably. Examples include:

Test result	True state		Sum
	+	-	
+	n^{++}	n^{+-}	$N^{+\bullet}$
-	n^{-+}	n^{--}	$N^{-\bullet}$
Sum	$N^{\bullet+}$	$N^{\bullet-}$	N

Figure 5.3: Sensitivity and specificity of a test. The *sensitivity*, s , is $s = n^{++}/N^{\bullet+}$, and the *specificity*, ρ , is $\rho = n^{--}/N^{\bullet-}$. See the text for further explanation.

- Oestrus detection performed by the farmer.
- Diagnosis based on clinical examination.
- Pregnancy diagnosis based on scanning.
- Pregnancy diagnosis based on examination by a veterinarian.
- Diagnosis based on a laboratory test.

Even though the tests are very different they also have properties in common. The exactness of the test is traditionally expressed by its *sensitivity* and *specificity*. The simplest way to look at these concepts is as illustrated in Figure 5.3.

The symbols “+” and “-” for the true state indicates whether or not the condition (e.g. a disease) is actually present. For the test result the symbols indicate the outcome of the test. The symbol n^{++} thus indicates the number of animals having a positive test outcome *and* are actually in possession of the condition (accordingly for the other symbols). The sensitivity is defined as the probability of a positive outcome of the test given that the condition is actually present. The specificity is defined as the probability of a negative outcome given that the condition is *not* present.

The sensitivity and specificity are inherent properties of a given test. In some cases, where the test is very standardized (e.g. laboratory tests or, in general, tests based on objective measurements), the sensitivity and specificity are known (more or less) constant values. In other cases (e.g. clinical examination) the sensitivity and specificity will vary depending on the person performing the test. Furthermore, it will probably vary from case to case and from day to day.

Now, assume that the true prevalence of some relevant condition (e.g. a disease) in a herd is θ (where $0 < \theta < 1$). If N animals are tested under the sensitivity, s , and the specificity, ρ , the number of animals tested positive $N^{+\bullet}$ will be binomially distributed with probability parameter p , where

$$p = s\theta + (1 - \rho)(1 - \theta). \quad (5.12)$$

It is easily seen, that if we have a perfect test with $s = \rho = 1$, then $p = \theta$. In case of an imperfect test, the probability parameter p will differ from θ . This

Table 5.4: Probability, p , of positive test outcome under different sensitivities and specificities and a constant true prevalence of $\theta = 0.10$.

Specificity	Sensitivity				
	0.6	0.7	0.8	0.9	1.0
0.6	0,42	0,43	0,44	0,45	0,46
0.7	0,33	0,34	0,35	0,36	0,37
0.8	0,24	0,25	0,26	0,27	0,28
0.9	0,15	0,16	0,17	0,18	0,19
1.0	0,06	0,07	0,08	0,09	0,10

is illustrated in Table 5.4 for a constant true prevalence ($\theta = 0.10$) and varying sensitivities and specificities of the test. As it is seen in the table, the probability of positive test outcomes may deviate considerably from the prevalence even for high sensitivities and specificities.

Having observed $N^{+•}$ positive outcomes out of N animals tested, the ratio $N^{+•}/N$ will (according to standard statistical theory) be an unbiased estimate of p , but, as Table 5.4 illustrates, it is certainly not an unbiased estimate of the true underlying prevalence θ . If, however, a standardized test with constant and known sensitivity and specificity has been used, we may just solve Eq. (5.12) for θ , i.e.

$$\theta = \frac{p + \rho - 1}{s + \rho - 1}. \quad (5.13)$$

With known sensitivity and specificity the analysis is therefore only slightly complicated compared to the examples of the previous section, because there is a one-to-one correspondence between p and θ . We shall illustrate the slightly modified procedure through an example.

Example 19 Conception rate with uncertain pregnancy diagnosis

Again we consider the conception rate in a dairy herd. After 6 weeks (from insemination), a pregnancy test is performed. We assume that the sensitivity is 0.90 and the specificity is 0.95. Like Example 16, we assume that the directly calculated value is 0.4. The expected target value for the *true* conception rate is 0.5.

Calculation of the key figure from data: The calculation is very simple. The number of inseminations N in a period is counted, and the number of cows $N^{+•}$ having a positive outcome of the pregnancy test is counted as well. The key figure κ is then just calculated as

$$\kappa = \frac{N^{+•}}{N}. \quad (5.14)$$

Distribution of the data: For given N cows inseminated, the number $N^{+•}$ of cows with positive pregnancy diagnosis will be binomially distributed with parameters N and p , where p is calculated as in Eq. (5.12).

Distribution of the key figure: The null hypothesis $H_0 : \theta = 0.5$ refers to the true underlying conception rate. With the assumptions made for sensitivity and specificity, a true conception rate of 0.5 corresponds to a probability of positive pregnancy test of $p = 0.9 \times 0.5 + (1 - 0.95) \times (1 - 0.5) = 0.475$ according to (5.12). Thus, $N^{+•} \sim \mathcal{N}(N, 0.475)$.

Confidence limits: The lower acceptable limit is calculated under the null hypothesis, a chosen confidence level (e.g. 5%) and the parameters N and p . Let $F_{N,p}(n)$ be the cumulative probability function of the binomial distribution with parameters N and p . The lower acceptable limit n_a for n is then determined as the 0.05 quantile of $F_{N,p}(n)$ as shown in Eq. (5.9). If, for instance, 50 cows were inseminated (i.e. $N = 50$) we get $n_a = F_{50,0.475}^{-1}(0.05) \approx 17$.

Acceptance or rejection: Since the observed rate was 0.40 it means that out of the 50 cows inseminated, 20 cows were tested positive. Since the observed value is higher than the lower acceptable limit, we conclude that the deviation from the target may very well be the result of random variation.

As it is seen from Example 19, observation uncertainty with binary data is easily taken into account if the sensitivity and specificity of the observational test is known. If, on the other hand, the sensitivity and specificity of the test are unknown, the situation is far more complicated.

If it makes sense to consider the values as drawn from known hyper distributions, the distribution of the observed key figure may be assessed numerically, but, very often it has no closed form. The analysis becomes therefore far more complicated, and no simple guidelines can be given.

If nothing (or very little) is known about the sensitivity and specificity, we can only conclude that the validity of key figures based on such data material is very low as Table 5.4 clearly suggests. An obvious example is simple counts of veterinary treatments for a specific disease in a period. For an animal to be treated, at least two conditions must be fulfilled:

1. The farmer suspects the animal to be diseased and calls for the veterinarian.
2. The veterinarian makes a positive diagnosis.

Thus, in fact two tests are performed: One by the farmer and one by the veterinarian. It is obvious that the sensitivities and specificities of those tests are unknown (and probably varying from case to case). Thus, the validity of the number of treatments as an indicator for the disease prevalence is very low.

Concluding remarks

It should be emphasized that the kind of statistical “tests” illustrated in the previous sections are only of indicative nature. They may not be confused with tests performed on data from controlled experiments. The purpose is only to provide a rough estimate of the significance of an observed deviation. The message is, that a result calculated from actual registrations is often associated with a rather big uncertainty as illustrated in the examples. In many cases it is possible to “estimate” that uncertainty rather precisely, but there is no correct level of significance to use in the evaluation of results. In a true situation the relevant level will also depend on the significance from a utility point of view.

5.2.6 Analysis from a utility point of view

The evaluation of deviations from a utility point of view is just as important as the statistical analysis. It may very well happen that a deviation is significant from a statistical point of view (where significance is merely a question of sample size), but insignificant from a utility point of view (and vice versa). Adjustments are of course only relevant if the deviations observed are significant from both points of view.

The evaluation from a utility point of view is complicated by the dynamics of production. If, for instance, the average number of days open in a dairy herd is higher than expected (and the statistical analysis showed that it was a significant deviation) we have to consider what the deviation means for the farmer’s utility. The direct consequences for production of a cow conceiving 16 weeks after calving (b) instead of 12 weeks after calving (a) are illustrated in Figure 5.4.

As it appears from the figure, the direct consequences include:

1. Next calving is delayed by 4 weeks.
2. The milk yield towards the end of the lactation is slightly higher.
3. The number of days in milk is increased by 4 weeks.
4. The milk lactation curve of next lactation is shifted 4 weeks to the right.

Expressed numerically, the consequences might for instance be as indicated in Table 5.5, where the economical net returns are calculated initially on a lactation basis and afterwards normalized to annual figures per cow.

The economic value per cow per year of a decrease in the number of days open from 16 to 12 weeks appears to be $13,637 - 13,054 = 584$. The *annual* difference in number of days open per cow is $16 \times 7 \times 365/364 - 12 \times 7 \times 365/392 = 20$ days. Assuming linearity, the cost of one additional day open is therefore $584/20 = 29$.

Assume now, that in a herd with 100 cows, the average number of days open has been calculated to 105 whereas the target is 90. It would be very tempting to

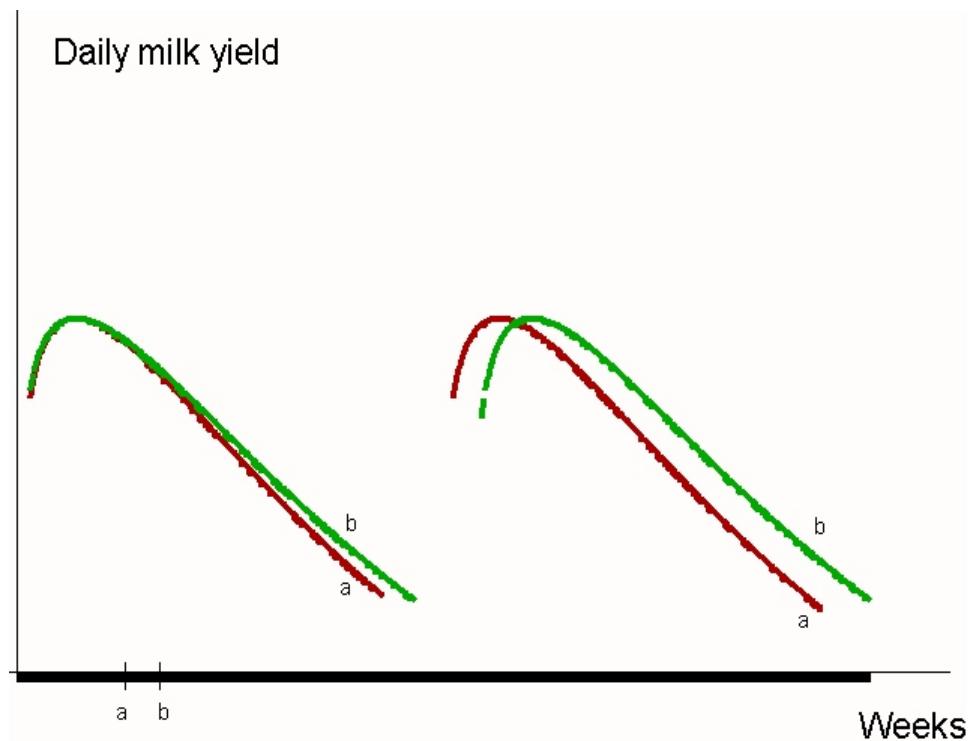


Figure 5.4: Daily milk yield during two lactations for a dairy cow conceiving 12 weeks after calving (a) compared to a cow conceiving 16 weeks after calving (b).

Table 5.5: Comparison of lactation results for a cow conceiving after 12 weeks with those of a cow conceiving after 16 weeks.

Property	Price DKK	12 weeks		16 weeks	
		Amount	Value	Amount	Value
Total milk yield, lactation	2.10	6000	12600	6200	13020
Number of calves born	1000	1	1000	1	1000
Total economic value, lactation			13600		14020
Calving interval, days		364		392	
Total economic value, year			13637		13054

claim that the total economic loss caused by this deviation is $(105 - 90) \times 29 \times 100 = 43,500$, but nevertheless, it would be a serious mistake. In order to reveal why, we shall consider the implicit assumptions behind the calculations above. Those assumptions are:

- The marginal value of an improvement is assumed to be constant (i.e. independent of the initial level). In other words, a decrease from 200 days open to 199 days is assumed to be of the same value as a decrease from 30 to 29 days. Such an assumption is certainly not realistic.
- All cows are assumed to be identical in the sense that they all conceive exactly after 12 (respectively 16) weeks. The real situation is very different. Even though an average result is for instance 105 days, individual results vary considerably among the cows of a herd. Thus, if the assumption of constant marginal value does not hold, we need to know the distribution of individual results in order to be able to estimate the economic consequences at herd level.
- All cows are assumed to be identical concerning milk yield. From the calculations it is obvious that a different level of milk yield would give us an other result.
- Conception is assumed to be independent of milk yield. In the real world, it is a well known problem that the conception rate of a high yielding cow is lower than that of a low yielding. It is therefore reasonable to assume that in particular the high yielding cows are those that cause a high average number of days open.
- It is assumed that no cows are replaced. In practice annual replacement rates of 30 to 50% are usual. It is obvious, that if a cow is replaced, the effects of delayed calving and delayed new lactation are irrelevant.
- The farmer's management is assumed not to interact with the observed result. The farmer probably has some policy concerning for instance replacement. If a cow is high yielding, he accepts more days open than if it is low yielding. If the number of days open grows, he probably also increases his replacement rate.
- It is assumed that there is no milk quota. In case of a milk quota, the marginal value of an increased milk yield is much lower than assumed in the example. The reason is that if a higher milk yield is obtained (for instance because of improvements in reproduction) the farmer has to decrease the number of cows in order not to violate the quota (refer to Kristensen and Thysen, 1991, for a discussion). Presence of a milk quota may reduce the costs of an additional day open by 60-70% as shown by Sørensen and Enevoldsen (1991).

These considerations clearly illustrate that calculations as those of the example are of little value in the evaluation of results obtained in production. It is very important that random variation, interactions, management policies and restraints are taken into account in such calculations. The most obvious tool to apply is certainly simulation, but also some of the other tools discussed in Chapter 9 may be used in some cases. Evaluations from a utility point of view remains, nevertheless a true art.

5.2.7 Decision: Adjustment/no adjustment

Since the only rational purpose of production monitoring is to provide a better basis for planning of the future production by adjustment of one or more factors, the final result of the monitoring process is of course to draw a conclusion concerning whether or not to make any adjustment.

When to adjust

From the discussion of the previous sections it is hopefully clear that we should make adjustments if we have found a deviation which is significant from a statistical *as well as* a utility point of view. The reason for this is that adjustment almost always are associated with costs (either in money or work). If adjustments were costless we should intervene no matter the size and significance of a (negative) deviation.

Where to adjust

Where to adjust depends on the deviations found. With reference to Figure 1.1, adjustments may be made of the goals, the plan or the implementation.

If there is a deviation in productivity key figures and corresponding deviations in factor consumption key figures (so that the deviations in factor consumption explain the deviations in production), it is the implementation that should be adjusted. We simply have a case where the plans have not been followed in practise.

If, on the other hand, there is a deviation in productivity, but none in factor consumption, the logically explanation is that the planning has been deficient. The results (factor consumption) show that we actually did as planned, but nevertheless, the production deviates from what we expected. The only logical explanation is that the knowledge used in planning has been insufficient. The adjustment may in that case result in new plans and/or adjusted goals.

How to adjust

When it has been decided that adjustments should be made, we have actually completed a full rotation of the management cycle of Figure 1.1. In other words, we are back in planning, and the general principles described in Chapter 4 are applied.

5.3 Exercises

5.3.1 Milk yield

One of the key figures provided by the official management information system of the “Danish Cattle Federation” is “kg EKM” (Energy Corrected Milk) per day for the cows having completed the period with fixed feeding level. The figure is calculated separately for 1st parity cows and other cows.

- Evaluate the validity of the key figure as an estimate for the true milk production capacity of the herd under the circumstances in question. Compare with the validity of other key figures like for instance the 305 days yield, the lactation yield and the average milk yield per cow per year.
- Evaluate the precision of the key figure.
- Evaluate the economic importance of a deviation from the goal.

According to the latest report from the management information system a herd has achieved the following results concerning milk yield:

Daily milk yield, period with fixed feeding level	Target	Result
1st parity cows	26.3	23.9
Other cows	29.6	28.1

The herd is fed a fixed level of concentrates during the first 24 weeks of the lactation. Kristensen (1986) estimated the standard deviation of *cumulated* milk yield during the first 24 weeks inside a herd to be around 490 kg milk in first lactation and around 630 kg milk for older cows. The number of cows behind the calculated average daily milk yield is not indicated in the production report, but a manual check of calving dates reveals that the figure for first lactation is based on 11 observations whereas the figure for other cows is based on 15 observations.

- Advise the farmer concerning the results achieved. Your advice must be based on the statistical significance of the deviation as well as the economic importance.

5.3.2 Daily gain in slaughter pigs

A very commonly applied key figure is *average daily gain* usually measured as gram per day. The figure may be calculated in several different ways.

Valuation weighings

The principles behind the calculation may be illustrated slightly simplified² as follows (cf. Example 10):

²Ignoring dead pigs.

+ Live weight of all pigs delivered during the period	xxxx
+ Valuation weight at the end of the period	xxxx
- Weight of pigs inserted during the period	xxxx
- Valuation weight at the beginning of the period	xxxx
Total gain during the period	yyyy

In order to calculate average daily gain the total gain must be divided by the number of feeding days (a feeding day is defined as the presence of one pig for one day).

- a. What exactly should be registered in the herd in order to calculate daily gain as described?

Individual gains

The principles behind this method may be illustrated as follows:

+ Live weight at delivery, pig <i>i</i>	xxxx
- Weight at insertion, pig <i>i</i>	xxxx
Total gain of pig <i>i</i>	yyyy

Daily gain for pig *i* is calculated as total gain for pig *i* divided by the number of feeding days for pig *i*. Average daily gain for the herd is then calculated as the average over all pigs.

- b. What exactly should be registered in the herd in order to calculate daily gain as described?

First in - first out

The principles behind this method may be illustrated as follows:

+ Total live weight at delivery	xxxx
- Reconstructed weight at insertion	xxxx
Total gain of group	yyyy

Average daily gain is then calculated as total gain divided by the number of feeding days.

- c. What exactly should be registered in the herd in order to calculate daily gain as described?

Comparison of methods

- d. Discuss advantages and disadvantages of the three methods (e.g. precision, work load, costs).

Numerical example

Assume that a herd has $R = 125$ pens each with a capacity of $N_i = 10$ pigs ($i = 1, \dots, R$). Standard deviation of live weights inside a pen is assumed to be $\sigma_i = 5$ kg for all pens ($i = 1, \dots, R$). Total standard deviation between all pigs in the herd is assumed to be $\sigma = 15$ kg. The following data have been registered for a 90 days period where the facilities have been full all of the time:

+ Live weight of all pigs delivered during the period	130,419
+ Valuation weight at the end of the period	75,952
- Weight of pigs inserted during the period	39,502
- Valuation weight at the beginning of the period	77,263
Total gain during the period	89,606

- e. Calculate average daily gain for the period.

The precision of the calculated daily gain κ depends on the method used for the determination. We may consider κ as an estimate of an unobservable latent true value θ as expressed by Eq. (5.5), where e_s is the *sample error* representing the random deviation of the pigs currently being present in the unit, and e_o is the *observation error* which depends heavily on the measurement method.

In the following it is assumed that the weight at insertion and the live weight at delivery are known with certainty (i.e. without observation error). Thus, the observation error only reflects the valuation weights. If all pigs are weighed on a scale, the observation error may be zero. Most often, however, only a minor fraction is actually weighed. Since the weight of the fraction may deviate from the weight of the whole herd, we therefore introduce a *fraction error*³ with a standard deviation σ_v . Observation of weights of the fraction may further be subject to a measurement error with a standard deviation of σ_m . The standard deviation of the total observation error e_o therefore depends on the values of σ_v and σ_m . The following table shows the influence of the standard deviations mentioned:

Measured trait	Value	Frac.	Meas.
+ Live weight of all pigs delivered	130,419	0	0
+ Valuation weight at the end	75,952	σ_v	σ_m
- Weight of pigs inserted	39,502	0	0
- Valuation weight at the beginning	77,263	σ_v	σ_m
Total gain during the period	89,606	$\sqrt{2\sigma_v^2}$	$\sqrt{2\sigma_m^2}$

In other words, the standard deviation of the total gain is $\sqrt{2(\sigma_v^2 + \sigma_m^2)}$. The exact numerical values depend heavily on the measurement method used for determination of the valuation weight. We shall consider the following options:

³The usual statistical term would be “sample error”, but since we already consider the present pigs of the herd as a random sample of an infinite population of pigs with a daily gain of κ we use the term “fraction error” in order to avoid ambiguity.

1. Weighing of *all pigs*: $\sigma_m = \sigma_v = 0$.
2. Weighing of a randomly selected pig from each pen: $\sigma_m = 0$ and $\sigma_v = s$, where s is calculated from Eq. (5.15) with $R = 125$, $N_i = 10$, $\sigma_i = 5$, $n_i = 1$.
3. Weighing of 125 randomly selected pigs: $\sigma_m = 0$ and $\sigma_v = s$, where s is calculated from Eq. (5.15) with $R = 1$, $N_i = 1250$, $\sigma_i = 15$, $n_i = 125$.
4. Visual assessment: $\sigma_m = 3\%$ and $\sigma_v = 0$.

$$s^2 = RN_i^2 \frac{\sigma_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right). \quad (5.15)$$

The standard deviation of the observation error e_o for *average daily gain* is just the standard deviation of the observation error for total gain divided by the number of feeding days.

If it is assumed that the standard deviation within the herd between individual pigs is 110 g the standard deviation of the sample error e_s becomes $\frac{110}{\sqrt{1250}} = 3.1$ g with a herd size of 1250.

- f. Calculate the standard deviation of the calculated daily gain under each of the 4 observation methods described above.

According to the production plan of the herd, the expected daily gain is 820 g.

- g. Do you recommend the farmer to interfere (assuming each of the 4 observation methods)?

The farmer informs you that method 2 has been used in the herd.

- h. Describe how to assess the economic consequences of the lower daily gain.

5.3.3 Reproduction in dairy herds

The management information system of a dairy herd shows that the pregnancy percent during the last 3 months has been 40% whereas the expected value was 50%.

Advice the farmer about the result taking the economic significance as well as the precision under the assumption, that the result is based on:

- a. 10 observations.
- b. 20 observations.
- c. 50 observations.

Hint: The number of inseminations leading to pregnancy is binomially distributed (n, p) , where n is the number of observations, and p is the probability of pregnancy. Calculate the probability of observing *at least* as many pregnancies as observed under the (null-) hypothesis $p = 0.50$.

- d. Comment on the results taking typical Danish herd sizes into account.

The dairy farmer feels that he has had reproductive problems in the herd for quite some time. He therefore orders (from the central management information system) a list of cows which during the last year have had at least 6 inseminations. The reason is that he suspects a few “problem cows” to be the primary reason for the dissatisfying result. It turns out that the list holds 5 cows. Advise the farmer based on the observed result and the precision under the assumption that the 5 cows on the list are found among:

- e. 20 cows.
- f. 100 cows
- g. 200 cows.

Hint: Calculate (under the hypothesis $p = 0.50$) the probability q of a cow having at least 5 unsuccessful inseminations. The number of cows with at least 6 inseminations is then binomially distributed (n, q) , where $n = 20, 100, 200$, respectively. Calculate the probability of observing at least 5 cows.

5.3.4 Disease treatments

A dairy herd has during the last year had the following number of treatments for footrot:

October	0
November	0
December	0
January	0
February	0
March	1
April	1
May	0
June	1
July	0
August	3
September	0
Total 12 months	6

The goal in the herd is 2 cases per year as a maximum. The cows are on pasture from May 1st till September 30th.

- a. Advise the farmer based on the results and there temporal distribution. Your advise must be based on the precision as well as the significance from an economic and an animal welfare point of view.
- b. How many cases is “enough” for concluding that the deviation from the goal is significant from a statistical point of view.

Hint: At constant herd size and incidence rate the number of cases will be Poisson distributed (Enevoldsen, 1993, p. 46-49).

- c. What is *in general* your opinion about monthly records of disease treatments and fluctuations from month to month?
- d. Discuss *in general* the validity of the number of disease treatments as an indicator for health in a herd..

5.3.5 Validity, reproduction

A dairy farmer has been told that several of his neighbors achieve 1.3 calvings per cow per year. The farmer has read in “Farmers’ weekly” that high economic gross margin relies on efficiency in reproduction and, furthermore, he hates to fall short of his neighbors he decides to go for 1.30 calvings per cow per year in his own herd.

His present results are 1.10 calvings per cow per year, 360 days calving interval and an annual replacement percent of 45. He uses all of his milk quota.

You are a consultant and the farmer asks you to advise him. What will you tell him?

Hint: Consider what influences the number of calvings per cow per year.

5.3.6 Economic significance of days open

A simulation study by Esslemont showed that at a herd level of milk yield around 6,500 kg (305 days’ yield) the milk yield per cow per year increases by 255 kg from 6,460 to 6,715 if the average calving interval is decreased from 14 to 13 months.

The economic gross margin *per cow per year* (D) is assumed to relate to milk yield per cow per year Y approximately as follows:

$$D = c + bY, \quad (5.16)$$

where c is a constant reflecting the “basic costs” of having a cow (i.e. feed for maintenance, value of calf, replacement costs etc), and b is a constant reflecting the marginal value of 1 kg extra milk. The numerical value of b equals the milk price minus the price of the feed used for production of 1 kg milk.

We assume that $c = -4,000$ DKK per cow per year, $b = 2.10$ DKK per kg milk, the annual replacement percent is 40 and the value of a calf is 1,200 DKK. What are the costs of one additional day open if:

- a. There is no milk quota?
- b. There is a 600.000 kg milk quota?

Hint: Start by calculating the marginal value of one kg milk per cow per year in each of the two situations.

Part II

Advanced topics

Chapter 6

From registration to information

6.1 Information as a production factor

With the advent of electronic equipment a whole new range of registrations becomes possible. The possibilities comprise electronic identification; automatic weighing; temperature- and activity measurements; and several registrations via video recordings using image analysis techniques (Van der Stuyft et al., 1991). The general attitude in the literature towards these new registrations is that they will improve, either income of the producer; welfare of the animal; reduce the environmental impact of animal production; or help to fulfil consumer demands for quality certification. However, the possible benefits of these registrations are not directly estimated.

With the traditional manual registration principles as described in Chapter 5, it seems that the choice of registrations has a large arbitrary element. The registration detail in animal production herds differ widely between species and production system. For example, in sow herds the registered traits comprise, e.g. litter size and event dates (mating, farrowing, etc.). It is generally accepted that these traditional traits are useful for decision purposes. In contrast, few traits are measured in slaughter pig production. The reason for this difference in registration detail is not clear. Again the benefits of the registrations are assumed rather than estimated. It seems that the choice of registrations has a large arbitrary element.

In the authors' opinion it is important to treat information as a production factor in line with, e.g feed. We need to define the quality and value of information, just as we define quality and value of feed stuffs. The value of feed stuffs is measured by the effect on the output, i.e. daily gain, feed conversion and meat quality. The value of information should be measured similarly on the output, i.e. the improvement in decisions. The value of information is dependent on its applicability in the decision process that is to say, to what extent the information helps the producers to reach their overall goal. Each decision in the herd has its own information needs. The value of information thus depends on the decision context, like the value of feed stuffs depends on whether they are used for sows or slaughter pigs.

Statistical decision theory using Bayesian techniques give the necessary theoretical tools for measuring improvement in decisions. However, a large effort is needed in defining and categorizing the decisions made in animal production, and in estimating the necessary probability distribution of the relevant traits.

The purpose of this chapter is to present a framework for the evaluation of registrations.

6.2 Data sources: examples

Currently, the possibilities for obtaining registrations from the herd are rapidly increasing. Electronic equipment already in use in feeding equipment and climate regulation has obvious potentials for use in connection with systems for decision support. A few examples will suffice.

6.2.1 The heat detection problem

Heat detection plays a very important role in the managing of a herd with reproductive animals. First of all, if the reproduction cycle is to be started, it is essential that the manager (or the optional male animal in the group) can detect the animals in oestrus at the correct time, in order to proceed with the mating. The detection and subsequent decision concerning mating have a very short time scale, i.e. a few hours. In addition heat detection is the best pregnancy diagnosis available, i.e. the most precise and the best timed. Several physiological and behavioural traits change during oestrus in the animal. Therefore, the idea of automatic detection of changes in these traits is straightforward.

Several methods have been studied (and advocated). Measurement of changes in the conductivity in the vaginal mucus, use of a teaser male, pedometer measurement of movement, temperature measurement, conductivity of milk, visit to the boar pen. Bressers et al. (1991, 1994); Thompson et al. (1995); Scholten et al. (1995). Reviews of methods for automatic heat detection in sows and dairy cows are given by Cornou (2006) and Firk et al. (2002), respectively.

6.2.2 The weighing problem

In experiments the growth of the animals has always been followed very closely. In production management as well, the growth has been felt to be an important trait to monitor. At least with the Danish pricing system, the carcass weights of the delivered animals have large impact on the carcass value, and thus the income. It has also been proposed that daily gain is a good early indicator of other problems in the herd such as disease, reduced feed quality etc. Therefore, large efforts have been put into (semi)automatic weighing equipment, ranging from simple equipment for girth measurement (a string) to automatic weights in each pen together with electronic identification. Weighing of the whole animal or only part of the animal near the feed through, weighing on perching equipment in poultry, and even

video based weighing based on automatic detection of body area. Turner et al. (1983); Anderson and Weeks (1989); Slader and Gregory (1988); Marchant and Schofield (1993); Ramaekers et al. (1995); Brandl and Jørgensen (1996) are only a few of the references.

6.2.3 The activity problem

Measurements of activity have already been mentioned in connection with the heat detection problem. Several other possibilities for use of activity measure. Before farrowing sows change behavioral pattern and the resulting shift can warn the manager to be ready for farrowing supervision. Decreased activity of individuals can be seen as an indicator of disease. Pigs change resting behavior according to the micro climate in their surroundings, if the temperature is too low they huddle and lay on their bellies, while they lay apart with legs outstretched in side position if it is too warm. Video monitoring can detect these position shifts and be used for climate control in the herd, i.e. the temperature is controlled using the pigs' own conception of temperature, rather than a thermometer. An activity problem on a far larger scale, is the problem of free ranging beef cattle. Simple identification equipment, in connection with satellite transmission can position the animals on the range. Van der Stuyft et al. (1991) is a key reference.

6.2.4 Miscellaneous

This is but a few of the suggested use of automatic registrations. Mastitis detection, hormonal analysis could be added to the list. Even measurement of the number of coughs in the herd has been studied (Dijkhuizen et al., 1989). Clearly, the possibilities are enormous.

6.3 Basic concepts and definitions

A noteworthy aspect of the examples in the preceding section, is that many of the techniques are suggested to solve identical problems, and in addition, problems that are already being solved in everyday. This is in close correspondence with e.g. feed composition problems or any other of the problems concerning optimal adjustment of input factors. Clearly, the use of registrations and observations in the herd is an input factor.

If we discuss selecting the optimal feed composition, we can describe each possible feed ingredient based on quantity and quality. A similar description is usually not given for the registration factor but in the following we will attempt to sketch possibilities for such a description. Jensen (1996, Section 5.5) discusses some of the concepts behind the possible value of information, if the correct data processing is used and the correct actions are taken. He presents a theorem (Theorem 5.4) that (if the value function is a convex function then) *the expected benefit of performing a test is never negative.*

This is only true if assumptions concerning the decision makers action is fulfilled. In this exposition, where we have a slightly different angle, examples to the contrary will be given. But, let us start the discussion with a modification of Theorem 5.4 in Jensen (1996). Thus, here we assume that,

- A registration has only value if it leads to an action.
- The value of a registration is that it reduces the uncertainty.

The value of a registration is thus defined from the actual action chosen, rather than the optimal action as assumed by Jensen (1996). Thus a negative value of the registration is possible if the decision maker reacts in a suboptimal manner.

The first item leads furthermore to the distinction between data and information as introduced in Chapter 2. As described in Section 2.1 and illustrated in Figure 2.1, the distinction between data and information, is that information can be used as basis for a decision, while data cannot. *Data is processed into information.*

More formally, a *registration* is simply a record, λ as defined in Definition 6, *Data* is defined as a set of registrations $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_k\}$ and *data processing* is a function of data Ψ , such that the information $I = \Psi(\Lambda)$. We have that $\Lambda \in \mathbb{R}^k$ and $I \in \mathbb{R}^l$, where usually $l \ll k$. Furthermore, a *decision strategy* is a map that assigns a decision to every possible state of the information set, i.e., $I \mapsto \Theta$.

Note that the data processing, Ψ , implies a reduction in data and thus only refer to a specific set of decisions. The standard assumption concerning the error terms in statistical models that they should be independent and random, is to prevent us from throwing information away. If the assumption is fulfilled, no information is lost.

From a decision support point of view we need not be that strict. The processing should only protect against throwing information away that might lead to an other action. Thus, no standard processing of data can be found, but it should be designed to reflect the actual decisions being made. In other words, many different data processing methods (corresponding to different functions, Ψ_1, Ψ_2, \dots) may be defined. In general, we prefer the processing Ψ_i leading to the best decisions, and by the best decisions we mean those having the highest expected utility value.

6.3.1 Utility value of registrations

The value of a registration can now be defined. It is simply defined as the expected utility based on actions when the registrations are made minus the expected utility when the registration is not made. The value of the registration, λ , thus depends on the set of data, Λ , it is part of, the data processing, Ψ , and the decision strategy, $I \mapsto \Theta$. The evaluation is illustrated by the following example.

Example 20 The two-sow problem

The two-sow problem is a problem of selection between sows. A pig producer has bought two sows (A and B) with the same expected farrowing date. One of

the sows has to be culled because there is only room for one sow in the farrowing department. The litter size of a sow is correlated with the litter size in her previous litter. Information about the litter size in the previous litter can be bought from the farmer, who sold the two sows. How much should he be willing to pay for the information, and which sow should he keep?

The farmer knows from experience that the state of nature, i.e. the relation between litter size in first and second litter can be described by the following bivariate normal distribution,

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 8.5 \\ 10.0 \end{pmatrix}, \begin{pmatrix} 6.25 & 1.25 \\ 1.25 & 6.25 \end{pmatrix} \right). \quad (6.1)$$

The actions that the farmer has to select between are *Select A* and *Select B*.

If he does not buy the information about previous litter size, he would select a sow at random and each of the actions would have equal probability. The litter size would follow the marginal distribution for parity 2 in the formula above, i.e. the expected number of piglets is 10.0.

If he buys the information (Y_{1A} , Y_{1B}) his decision strategy would probably be to assign action *Select A* to the outcome $\hat{Y}_{2A} > \hat{Y}_{2B}$, where \hat{Y}_{2i} is the predicted litter size in the second litter and *Select B* to the outcome $\hat{Y}_{2A} < \hat{Y}_{2B}$ and to select action at random if $\hat{Y}_{2A} = \hat{Y}_{2B}$.

The data consist of two registrations with

Event: Farrowing

Identification: i , where $i = A$ or $i = B$ for Sow A and B, respectively.

Registration level: Individual sow

Time: Date of farrowing

Property: Live born piglets

Value: Y_{1i}

We process the data as shown below in order to obtain the set of predicted litter sizes \hat{Y}_{2A} , \hat{Y}_{2B} and we have decided how to map this information into actions. We assume that the utility depends linearly on litter size, and therefore we can measure the utility in the same scale as litter size. The expected utility value of the registrations can now be found.

The conditional distribution of Y_{2i} for given value of Y_{1i} is

$$Y_{2i} \sim \mathcal{N}(10.0 + 0.2(Y_{1i} - 8.5), 6.0) \quad (6.2)$$

i.e., $\hat{Y}_{2i} = 10.0 + 0.2(Y_{1i} - 8.5)$. Note that we could have used information concerning the variance of the predicted level as well. As the variance is the same

and in addition the utility is linear, the variance does not influence our decisions, and accordingly has no value, and can be discarded.

For each possible level of litter size for sow A we know the probability from the prior distribution, furthermore the probability of each of the outcomes relevant for the decision can be found. As the problem is symmetric for the two animals we can simply calculate the expected litter size as the expected litter size of sow A if she is kept.

Now, define X as the superiority of Sow A over Sow B for *first* parity, i.e.

$$X = Y_{1A} - Y_{1B}$$

It follows directly, that $X \sim \mathcal{N}(0, 12.50)$, and $\text{Cov}(X, Y_{2A}) = \text{Cov}((Y_{1A} - Y_{1B}), Y_{2A}) = \text{Cov}(Y_{1A}, Y_{2A}) = 1.25$ (because the two sows are independent). It is clear that $\hat{Y}_{2A} > \hat{Y}_{2B} \Leftrightarrow X > 0$, so what we need is the expected litter size of Sow A at second parity given that it was superior to Sow B at first parity, i.e. we must calculate $E(Y_{2A} | X > 0)$.

First, we observe by applying standard rules (cf. Eqs. (C.4) and (C.5)), that $(Y_{2A} | X) \sim \mathcal{N}(10.0 + 1.25/12.50X, 6.25 - 1.25^2/12.50)$, or in other words,

$$(Y_{2A} | X) \sim \mathcal{N}(10.0 + 0.1X, 6.125).$$

We are now ready to calculate $E(Y_{2A} | X > 0)$ using the rules for truncated normal distributions (Appendix A.3.3):

$$\begin{aligned} E(Y_{2A} | X > 0) &= 10 + 0.1E(X | X > 0) \\ &= 10 + 0.1 \left(0 + \sqrt{12.50} \frac{\phi(0)}{1 - \Phi(0)} \right) \\ &= 10 + 0.1 \left(3.54 \frac{\frac{1}{\sqrt{2\pi}}}{0.50} \right) \\ &= 10.28. \end{aligned} \tag{6.3}$$

Thus, the expected litter size of the sow preferred (in this case the one performing best at first parity) is 10.28 piglets implying that the *value of the observations* of first litter size results is the value of $10.28 - 10.00 = 0.28$ piglets. ■

Note that in Example 20, we have considered the problem for a fixed observation strategy. In fact we have a sequential decision problem. What is the information value of only one observation? Depending on the first observation when should we buy the next observation. E.g. if the observed value for A is very high (or very low), it is not likely that we can improve the decision by obtaining an observation from sow B . We have also based our estimated value on a known state of nature (6.1). In general, only an estimate of the state of nature is available. This

will reduce the precision in the information and reduce the value of the information, i.e. it is actually lower than the 0.28 specified.

Naturally utility evaluation of registrations is a very time-consuming task and is usually not possible directly. In section 6.6 this approach towards utility evaluation will be used on more relevant examples. Now we will continue with other measures of value and with some attributes of registrations that allows us to evaluate the registration without considering the decisions involved

6.3.2 Entropy based value

If we wish to record a variable it seems reasonable to select the variable that will reduce the uncertainty most.

A measure of the uncertainty that is often used, is Shannon's entropy or information value, as it is also called. The entropy has important applications within the areas optimal coding etc. For instance, the many "zip", "pack" procedures originate with this concept. For a discrete variable taking values in the set $S = \{a_1, \dots, a_k\}$ the entropy is simply

$$\text{ent} = - \sum_{i=1}^k p_i \log_2(p_i) \quad (6.4)$$

where p_i is the probability of observing the value $a_i \in S$. (The use of \log_2 is not essential, but it is the generally accepted logarithm function to use, because in that case ent becomes the number of yes/no questions needed to find the value of the variable.) The formula can be extended to continuous variables. Applying (6.4) to the sow problem of Example 20 for litter size in first litter (with $S = \{1, \dots, 19\}$) is 3.37, while if the litter size was randomly distributed over the set, S , the entropy would have been 4.25

For a further discussion on myopic (nearsighted) measures of value of information see Jensen (1996).

6.4 Measurement errors

Another set of attributes of registration concerns the measurement error. In the following different types of measurement errors are discussed.

6.4.1 Registration failure

The first type of the error that can happen in the measurement process is registration failure. Registration failure refers to the situation where a registration has gone totally wrong, e.g. because the record is associated with a wrong identification, a measurement device is read incorrectly, or a wrong number is typed into the database. Such a registration is also called an outlier. Errors of this kind influence the *correctness* of a key figure based on the registrations as defined in Section 5.2.4.

Very often error detection procedures as described in Section 5.2.4 will detect outliers that are very far from the expected value, while plausible errors will pass undetected. A natural model of the outlier is that it is uniformly distributed on the interval $\mu - \kappa, \mu + \kappa$ where μ is the expected level and κ is the limit of the error detection procedure, e.g. $\kappa = 1.96$ units of standard deviation. In some data filtering methods (e.g. Thysen, 1993a) this outlier detection can be handled automatically, if an estimate of the general error rate, p_e , is available, i.e. that we know how often such an error can occur.

The process of securing that a registration has not failed is sometimes called *verification* of the measurement.

6.4.2 Measurement precision

The simplest error type that can occur is a random difference between the real value, X , and the observed, Y . The magnitude of this error is quantified by the precision using the model,

$$Y = X + \varepsilon \quad (6.5)$$

where ε is a random and independent error with expectation 0, often it is relevant to assume $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2) = \mathcal{N}(0, 1/\tau_\varepsilon)$, where τ_ε is defined as the precision, i.e., $\tau_\varepsilon = 1/\sigma_\varepsilon^2$. We will use this definition of the term precision, even for variables that are not normal distributed. A measurement error of this kind influences the precision of a key figure as defined in Section 5.2.4.

If we take the average of k independent measurements, the precision of the average is $k\tau$. Note that the independency between measurement is important, if the relation is to hold.

Some times the term *repeatability*, t , is used for measuring the quality of a registration. Repeatability is most often defined as the correlation between two measurements on the same subject, i.e., in this case:

$$t = \frac{\text{Cov}(Y_i, Y_j)}{\text{Var}(Y_i)} = \frac{\text{Var}(X)}{\text{Var}(X) + \sigma_\varepsilon^2} \quad (6.6)$$

When the precision increases, the repeatability approaches one. Note, however that the variance of the subjects measured will have a large impact on the *repeatability* of the measurements, while the *precision* is not influenced.

6.4.3 Indirect measure

Often the measurement taken is only an indirect measure of the true value. Most measurements are in fact indirect, such as hormone analysis etc. Measurement errors of this type influence the *validity* of a key figure as defined in Section 5.2.4. The value that we want to measure may even be some abstract phenomenon.

A simple case is the trait that we measure is not constant but varies around some value

$$Y = X_0 + \varepsilon_x + \varepsilon.$$

Here we have introduced an additional error term. We still have our “old” measurement error but in addition we have an error term that represents how good our indirect measure is. If ε_x has a low variance, the two variables will be highly correlated and we can obtain a good measure of the underlying trait. Repeated measurements will not generally solve the problem of low precision. Repeated measures improve our precision in $X_0 + \varepsilon_x$. A good example is pregnancy testing using ultra sound. Some times the pregnancy tester will react to something different than a pregnant uterus. If we repeat the testing, we will repeat the failure. Only, if the measurements are independent both with respect to ε_x and ε can we improve the precision by repeated measurement.

This case of indirect measure is in general too simple. Instead, if X is what we want to know, but Y is what we measure, a more general model would be,

$$\begin{aligned} Y &= f(X) + \varepsilon \\ Y &\approx X_0 + \beta_1(X - X_0) + \frac{1}{2}\beta_2(X - X_0)^2 + \varepsilon, \end{aligned}$$

where the approximation is based on a second order Taylor series. If the measurement is properly calibrated and developed, the relation will be very close to linear and the second order term will vanish. On average the measurement would give the correct result and we can treat the deviation as random. If this is not the case we get a systematic difference in level, and our measurements becomes biased.

The *validity* concept is also sometimes used for single measurements (i.e. not just for key figures as in Section 5.2.4), i.e. if $\text{Var}(\varepsilon_x)$ is low the validity of the measurement is high.

6.4.4 Bias

The concept of bias may be represented as ι in

$$Y = X + \iota + \varepsilon$$

Bias should always be treated with caution. Usually we tend to think that if we measure more individuals we get a more precise estimate. This is only the case if the measurements are unbiased. If the bias is known, we can of course correct for it, and then we have by definition removed it.

A good example of bias is the use of perching equipment with weighing sensors. In the start of the growing period every animal uses the perch with equal probability, but when the animals become older, the heaviest animals have a lower probability of resting on the perch. The resulting weight measurement is thus systematically underestimated, i.e. they are biased.

6.4.5 Rounding errors

When results from measurement devices are read, the figures are usually rounded to some extent. A systematic rounding method implies that instead of using the measurement, X_O , directly we can assign an interval to the measurement i.e., $X \in]X_O - 0.5, X_O + 0.5]$. Very often, however, the rounding is not systematic. It is much more tempting to round off from 9.41 to 9.4 than from 9.4491 to 9.4. Very often the last figure would be registered as 9.45, although most of the other registered values are with only one digit after the decimal point. However, in animal production this is not likely to cause any major problems, because of the much more serious sources of errors.

6.4.6 Interval censoring

The rounding errors can be generalized to the concept of interval censoring, i.e. where only some threshold values $\{\tau_0, \tau_1, \tau_2, \dots, \tau_n\}$ are registered. For example mating and farrowing, are usually registered on a date level, so that we only know that the measurements are between two of the threshold values.

When the underlying distribution is continuous we can still model the data, based on the probability that the observed value is between the two thresholds.

Example 21 Pregnancy period in sows

The interval from mating to farrowing in sows is approximately distributed as $\mathcal{N}(115, 2.5^2)$. If a sow farrows on day 116 we can calculate the likelihood of this observation as $\Phi(117, 115, 2.5) - \Phi(116, 115, 2.5)$, where $\Phi(x, \mu, \sigma)$ is the probability of value less than x from a normal distribution with mean μ and standard deviation σ .

■

Note that the censoring can also occur with the threshold values of $\pm\infty$, i.e. if an event occurs, it is before or after the specified registration time. Typical examples are when registrations are started in a herd and the previous history is unknown or when data has not yet been registered. An example of the latter case, is when farrowing/calving rates are to be calculated, based on observed matings before the birth takes place. For some of the animals a new event has been registered (remating), while for others no remating has taken place.

6.4.7 Categorical measurements

The concepts used for interval censoring can be further developed to cover the categorical or qualitative measurements, such as heat detection $\{in\ oestrus, not\ in\ oestrus\}$, pregnancy diagnosis $\{pregnant, not\ pregnant\}$, disease diagnosis $\{diseased, not\ diseased\}$, leg weakness score etc. In contrast to the interval censoring case, the registrations are not the threshold values, but rather the number of the interval. i.e. the state of the animal can be $\{d_1, d_2, \dots, d_n\}$.

The measurement w can be thought as a two-step procedure.

Table 6.1: The outcomes from oestrus detection of sows. Sensitivity and specificity of the measurement.

		Actual State
Diagnosis	Not in oestrus	In oestrus
Negative	0.99	0.05
Positive	0.01	0.95

Table 6.2: The outcomes from pregnancy diagnosis of sows. Sensitivity and specificity of the measurement.

		Actual State
Diagnosis	Not pregnant	pregnant
Negative	0.70	0.05
Positive	0.30	0.95

1. Measurement of a continuous variable y_w , where y_w is distributed as $\mathcal{N}(\mu_i, \sigma_w^2)$ if the animal is in state i , and the measurement precision is $1/\sigma_w^2$.
2. Assignment of a state d_j to the measurement depending on a series of threshold values $\{\tau_0, \tau_1, \tau_2, \dots, \tau_m\}$, where m normally (but not necessarily) is equal to $n-1$. The threshold values are specific for the measurement method, such that the animal is said to be in state d_i if $\tau_{i-1} < y_w \leq \tau_i$ (τ_0 is defined as $-\infty$).

The concepts are further illustrated through Example 22.

Example 22 Heat detection and pregnancy test in sows.

Heat detection (oestrus detection) in sows typically has a low error rate and the values shown in Table 6.1 are fairly representative.

This can be modeled as an expected value of 0 for sows not in heat and 1 for sows in heat, σ_w is 0.252 and the threshold value 0.5858. The distributions are shown in Figure 6.1.

Pregnancy testing can be done with an automatic testing device with a relatively low detection precision. A typical example is shown in Table 6.2.

This can be modeled as an expected value of 0 for sows not pregnant and 1 for pregnant sows, σ_w is 0.4610 and the threshold value 0.2417. As shown in Figure 6.2 the two distributions have a considerable overlap. Note that the standard deviation on the underlying scale is almost twice as big for pregnancy diagnosis, as for heat detection.

The elements of Table 6.2 are often called *sensitivity* and *specificity* with terms borrowed from laboratory measurements. In this context the *sensitivity* is the probability of detecting a phenomenon, when it exists, i.e. in the case of pregnancy diagnosis 0.95. The *specificity* is one minus the probability of observing a phenomenon, when it is not there, i.e., 0.7.

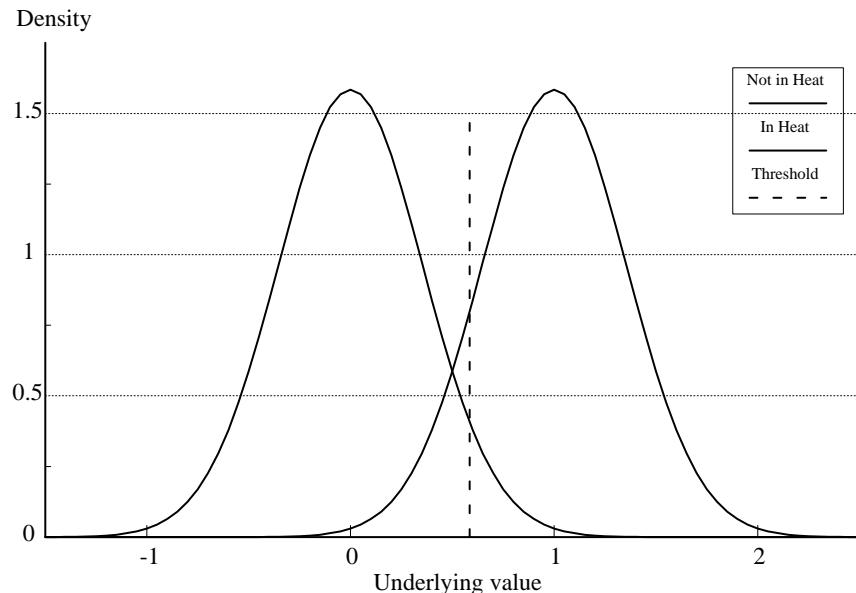


Figure 6.1: Threshold model for heat detection

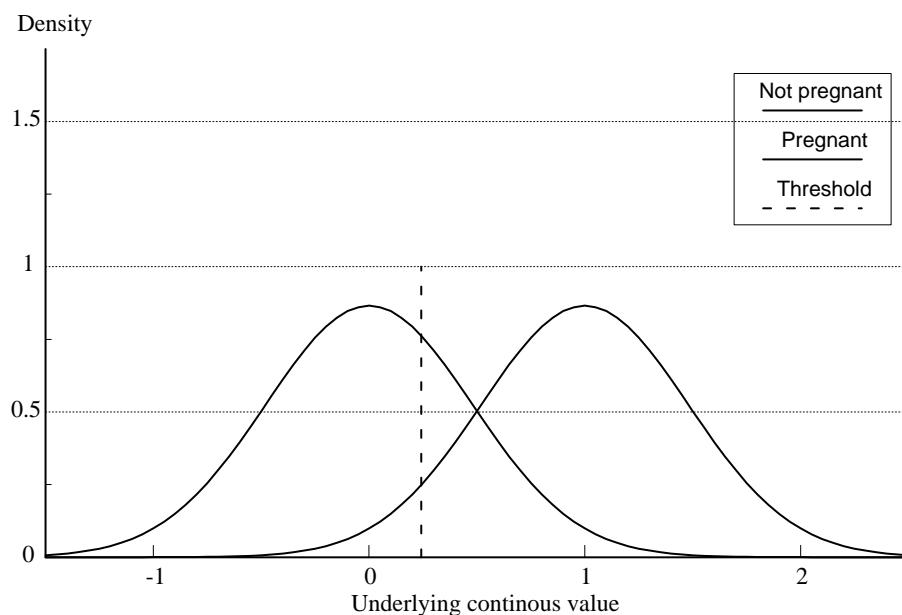


Figure 6.2: Threshold model for pregnancy testing

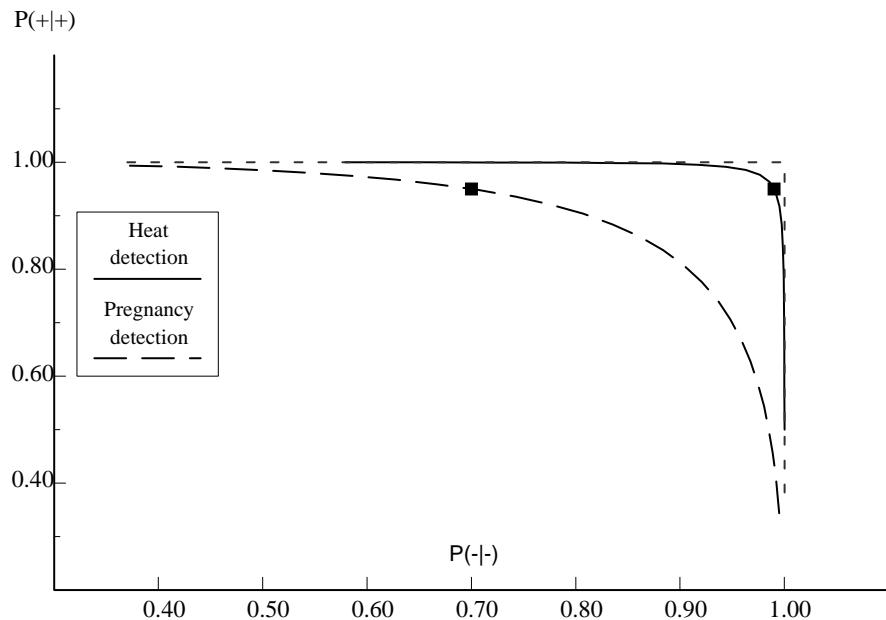


Figure 6.3: Possibilities for calibrating heat and pregnancy detection

Another concept is the predictive value, i.e., the probability that an animal that is diagnosed as pregnant, really is pregnant, and vice versa. This is the value that is most import in a decision context, but it depends on herd level of pregnancy rate as well as the measurement device. Using Bayes rule (Theorem 1), the different values can be calculated.

In a typical measurement apparatus the *threshold value* can be altered, and the variable measured will be continuous. This very closely resembles our model. Correspondingly, the threshold can be altered to give a large spectrum of detection rates. In Figure 6.3 this is shown for the two measurements discussed. The optimal calibration will depend on the proportion of positive and negative animals in the population as well as the decisions taken after the measurements have been made. It is questionable if a general calibration is optimal. The calibration should be tuned to the herd that is using the device.

■

6.4.8 Enhanced models of categorical measurements

The models above are too simple to be of value in the evaluation of measurement method. It is assumed that repeated observations are independent. Observations on the same animal within a short time period will usually tend to give the same

result. The trick is to model the repeated measurement on the continuous scale as discussed above and then to proceed with the classification later on based on the threshold values.

Let Y_{it} be the value measured for animal i at time t . A more detailed model that can account for repeated measurements is

$$Y_{i,t+\Delta} = f(t + \Delta) + A_i + M_{i,t+\Delta} + \varepsilon_{it}, \quad (6.7)$$

where Δ is a small time interval, $f(t)$ is a function representing the effect of the true state of the animal at time t , e.g. *pregnant* versus *not pregnant*, $A_i \sim \mathcal{N}(0, \sigma_A^2)$ is a random effect of animal, M_{it} is an observation from an autocorrelated time series, and ε is the random measurement error. The autocorrelated time series may for instance be modeled as

$$M_{i,t+\Delta} = \exp(-\alpha\Delta)M_{it} + \varepsilon_M, \quad (6.8)$$

where α is a parameter measuring the size of the autocorrelation, and $\varepsilon_M \sim \mathcal{N}(0, \sigma_M^2)$. With this model, measurements with short intervals will be close in values (high repeatability) and measurement far apart will be less correlated

The A and M parameters may have to be modeled separately for each of the underlying states (e.g. one set of models for *pregnant* and an other set for *not pregnant*).

6.4.9 Pre selection

Many registrations are performed on a selected material. An example is weighing of slaughter pigs. Before delivery to the slaughterhouse a visual inspection is performed, and the candidates for delivery are weighed.

In other words, the measurement is a two-step procedure. First a measurement Z_i (visual inspection) is made on every individual in the pen. The measurements are ordered according to size, and the first in the list is weighed on a scale, i.e. the observed value is:

$$Y_i = (X_i | \max(Z_1, \dots, Z_n) = Z_i)$$

where X_i is the true live weight and Y_i is the observed live weight. Clearly we should take care, if we use Y_i as a representative measurement of the live weight of the pigs in the herd.

6.4.10 Registration of interventions

Intervention occurs when the manager believes there is a problem that he can remove by some kind of action. Treatment of disease is an example. Note that this registration is no indication of the problem. It is a combination of the farmer's sensitivity as well as his trust in the action. Symptoms can be registered objectively,

not the diagnosis. Culling reasons is an example of how the registration system almost forces the farmer to diagnose a disease when he culls rather than objectively report symptoms.

6.4.11 Missing registrations

The intervention registration has a parallel in the concept of missing registrations. In fact if we know the behavior of the manager, there is information in the registered value as well as the values not registered.

For example, we would expect a mated sow with no further registration the last 80 days, to have a high probability of farrowing. If it had shown oestrus we would have expected it to have been registered, either as a remating or as a culling. Later on, if there is no farrowing past 118 days after mating, we would suspect that she has not been pregnant, even though there is still a small probability that she will farrow.

6.4.12 Selection and censoring

Some of the most difficult measurement errors are the errors that occur because of what we can call selection within the data, or censoring of observations.

Again the problem is that the farm is under strict management control, and that the traditional assumptions of independency and randomness of observations therefore cannot be justified.

Cows with a high milk production will tend to have longer lactations as well as a higher longevity. Sows with high litter size will have higher longevity. Fast growing slaughter pigs, will be delivered before the slow growing.

If the manager would maintain the same control of the herd, this was not a problem, because e.g. the expected lactation yield in late lactation in old cows would be the same as that previously recorded. But if he changes strategy, and reduce e.g. the culling rate the old value will not be correct anymore, because some of the low producing cows will reach the high age.

Therefore the data should be processed using methods that can take the selection into account.

Example 23 Daily gain in slaughter pigs

To illustrate the problem of censoring in practice, the problem of the calculating average daily gain will be discussed. Most slaughter pig units use a management information system reporting the so-called average daily gain in the herd. It may for instance be calculated straightforward as

$$DG_H = \frac{\sum_{i=1}^k \frac{1}{k_s} SLW_i - \sum_{i=1}^k W_{0i}}{\sum_{i=1}^k t_{si} - t_{0i}}$$

where k_s is a constant for the slaughter waste, SLW_i is the slaughter (carcass) weight of the i th individual, W_{0i} is the weight at the start of the fattening period, t_{si}

is the time at slaughter of pig i and t_{0i} is the time at insertion of the pig. The reason for this formula is that a linear growth is assumed, and under that assumption we obtain the following formula:

$$DG_H = \frac{\sum_{i=1}^k (W_{0i} + DG_i(t_{si} - t_{0i}) - \sum_{i=1}^k W_{0i})}{\sum_{i=1}^k t_{si} - t_{0i}} = \frac{\sum_{i=1}^k DG_i (t_{si} - t_{0i})}{\sum_{i=1}^k t_{si} - t_{0i}} \quad (6.9)$$

This definition would have been correct if the animals were delivered at the same age. Assuming fixed age at insertion, delivery at the same age implies that, $\forall i : t_{si} - t_{0i} = c$. By insertion in (6.9) we get,

$$DG_H = \frac{\sum_{i=1}^k DG_i c}{\sum_{i=1}^k c} = \frac{c \sum_{i=1}^k DG_i}{kc} = \frac{\sum_{i=1}^k DG_i}{k}.$$

Then, average daily gain for the herd is the average of the individual daily gains.

Unfortunately, pigs are not delivered at the same age. Instead, they are delivered at a constant weight, i.e. we have that $DG_i(t_{si} - t_{0i}) = K$. This is a standard example of selection or informative censoring. If we substitute this into (6.9), we obtain

$$DG_H = \frac{\sum_{i=1}^k K}{\sum_{i=1}^k \frac{K}{DG_i}} = \frac{k}{\sum_{i=1}^k \frac{1}{DG_i}}$$

i.e., the daily gain observed in the herd is the harmonic mean of the daily gains of the individual pigs. Compared to the usual arithmetic mean the harmonic mean places more emphasis on the low value. The slowest growing pigs influence the mean more than the fastest growing. They spent more time in the herd, so in that respect, it is reasonable enough.

In fact not every pig is delivered at a constant weight. Before a pen becomes available to new pigs it has to be emptied. Therefore economic considerations mean that some of the pigs are delivered at a lower weight. If the slaughter pig manager uses sectioned production this has even higher effect, because the slow growing pigs occupy a whole section, i.e. more pigs should be delivered at a low weight. An indirect result of this is that sectioned production will have a higher daily gain, because the formula for calculating herd daily gain is more like an arithmetic average, than in normal continuous production, where it closely corresponds to the harmonic mean.

Usually the selection bias in sows or dairy cows will be in the opposite direction of what is seen in Example 23, i.e. the best animals would tend to be longer in the herd, and thus have a higher impact on the herd level.

We would usually think that this would lead to an overestimate of the production curve, but it may in fact draw in the opposite direction if statistical procedures are not used correctly. A typical statistical model for repeated measurements is

$$Y_{it} = \mu_t + A_i + \epsilon_{it}$$

Using standard methods such as the GLM procedure (SAS Institute Inc., 1989) we would declare the animal effect as random with mean 0, and variance σ_A^2 . If only the best animals remain in the herd, the average at high parities of the A -values would be higher than 0. In order to compensate for this, the statistical method would reduce μ_t correspondingly. Toft and Jørgensen (2002) describe a maximum likelihood based method for unbiased estimation of the litter size curve in sows.

6.4.13 Discussion

The list of possible measurement errors has become rather long, in fact so long that it seems like a specification of Murphy's law within measurement in animal production herds (If anything can go wrong, it will). The important message is that if we can identify and describe the error source and type, there will often be statistical methods available that can handle that error, and make correct deductions from the registration possible, despite the errors. If the errors are not detected, faulty conclusions will be the result.

6.5 Data processing

Methods for data processing is outside the scope of this note. Instead the data processing will be classified according to the herd management cycle, and general attributes of the data processing discussed.

If we compare Figure 1.1 with the examples in Section 6.2, we find that almost all of the examples have to do with either the implementation or the monitoring phase of the management cycle, and that the use in the implementation phase has to do with operational decisions that very closely resemble the monitoring process. The data are collected in order to detect deviation from a level previously defined. When this deviation is detected, an action is taken. The heat detection problem is the implementation phase, a detection of oestrus signs above a certain threshold results in a mating. The weighing problem is either operational, i.e. if the weight is above a threshold deliver the animal, or monitoring, i.e. if the average growth rate is below some limit, start an analysis. The climate regulation based on activity is clearly in the implementation phase where in fact the actions based on the measurements is effected wholly automatic, without the intervention of a human decision maker.

For monitoring and operational purposes the data processing leads to very simple information, because the information is only used for selecting between two actions, i.e *Ignore evidence, React to evidence*.

Strangely, the use of data for planning purposes is largely overlooked, when the possibilities for new registrations are discussed. Of those examples mentioned in section 6.2, only the detection of position of free ranging animals is actually a planning parameter. If we know where the animals are, we can plan inspection visits etc. But this can easily be implemented on the smaller scale in a herd. Monitoring

the number of animals in the herd and their growth rate gives a precise prognosis of future productivity, and the best possibility for management of feed inventory, manure, new animals etc. The different approaches to decision support presented in Chapter 9 all rely on a knowledge of the state of nature in the herd, i.e. a complicated processing of data from the herd. It is of interest to note that the information from the data processing is used for much more complex decisions, i.e. the future production plan. Clearly the impact of the correct information becomes larger.

The optimal statistical methods for data processing are to be found within the analysis of time series and longitudinal data, even though the papers presenting the application of the measuring methods in general rely on a more ad hoc solution to the data processing. This adaptation of statistical methods to the specific problems in herd management will be an important research area in the future.

6.5.1 Threshold level

The direct detection of a measurement above a threshold level is in fact directly comparable with the model of quantitative measurement presented in the previous section. With a given measurement precision we can calibrate the threshold to obtain different levels of sensitivity and specificity of the test. If we measure the weight of an animal in order to decide whether to deliver it to the slaughter house, we can set the weight threshold high and then secure that a very low percentage of the animals are delivered too early (with too low a carcass weight), i.e. a high specificity. In doing this, many animals that are actual in the correct weight group will not be delivered, i.e. the sensitivity of the test is too low. The choice of threshold weight will of course depend on the distribution of carcass weight with different threshold weights and thus the economic income.

The measurements are very often repeated, and apart from sensitivity and specificity the detection time becomes an important attribute of measurements in dynamic systems. When the repeated measurements are independent, the sensitivity of the measurement will increase, because the probability of a value above the threshold will increase with the number of repeated measurements. For exactly the same reason the specificity will decrease. In addition, we have to consider the time of detection as well. For example in oestrus detection, it is not very interesting to register the oestrus if the registration is too late to result in a fertile mating.

Usually sensitivity and specificity of threshold level detections should be specified per time unit. For example in oestrus detection, percentage detected the first 12 hour of oestrus, or with threshold detection for control purposes, number of false alarms per year.

Measurements that can rely on a threshold level do not need a very high detail in registration identification. I.e. we do not need the identification of the individual animal. This can result in large differences in the cost of an automatic system.

6.5.2 Change point estimation

Very often the purpose of a measurement is not to detect a value above a certain level but to investigate a change in level. In the statistical literature this is called change point estimation. This comprises detection of sudden changes in the level of a variable as well as changes in trend.

The method of change point detection is often suggested, when the threshold level mentioned above fails, i.e. in heat detection based on temperature measurements, it is suggested to correct for the level of the individual sow. From a conceptual point of view a change-point detection is simply the detection of a change with magnitude above a certain threshold level, and the method is thus very comparable to the method above.

There are important differences, however.

- The change-point estimation needs data-processing, i.e. data should be input and processed in some kind of computing facility before it can be reacted upon.
- The change-point estimation needs identification. When comparing the levels at different measurements, it is necessary to identify each registration item (e.g. animal) and connect the data items. Often this requires a completely different setup, i.e. for automatic registration use of an automatic identification equipment. In addition it can result in a new set of measurement errors, i.e. we need to consider the possibility for mis-specification of identification and the errors this can cause.
- The change-point estimation is less precise. The variance of the difference between two independent and identically distributed variables is twice the variance of each of the variables.
- The change-point detection needs more registrations and the detection will be later than with the threshold value.

A thorough investigation is needed, before it can be concluded that the change-point method will be able to solve problems that the simple threshold technique cannot solve. This consideration should also include the timeliness of the detection. For control purposes it is relevant to ask, how a detection of a change point half a year ago can improve the future production in the herd.

At the herd-level change-point methods are important for planning purposes. Even if there is only a weak evidence for a change in the level, it is important to take this evidence into account when making plans for the future.

6.5.3 Significance test in experiments

This final method is only to illustrate the similarity between the decision support for experimental purposes and for herd management purposes.

The traditional statistical method for experimental design and analysis is of course a processing of data, and it corresponds very closely to the qualitative measurement problem, with sensitivity and specificity. Even though, many data are registered in the experiment, the information from the experiment after processing shall be used to select between two actions (*Reject Null-hypothesis, Do not reject Null-Hypothesis*). Because it is generally impossible to assign utilities to these alternatives, the selection of threshold value is based on a fixed and widely accepted specificity of the experiment, i.e. the so-called type I error, or the significance level (e.g. 5%). If the null hypothesis is true, there is only a low risk of rejecting the null-hypothesis. As the null hypothesis corresponds to the old scientific theory, it is natural for the research community to be conservative in rejection the hypothesis. The statistical procedures modify the threshold value to keep the specificity constant.

The sensitivity of the experiment is controlled by the individual researcher, the so-called type II error, mainly by making the experiment as precise as possible under the economic constraints. An accepted choice is a type II error of 80%. Clearly, the cost of failing to reject the null hypothesis (i.e. document a new theory) is mainly with the individual experimenter, and he/she must evaluate the costs himself.

In conclusion, the experimental procedure is a method for measurement in order to make decisions concerning scientific theories. The value of a correct scientific theory is impossible to estimate. Therefore, the threshold value in experiments cannot be set according to utility, but sensitivity and specificity have to be used instead.

6.6 Utility evaluation examples

6.6.1 Production control

Production Control in sow units is carried out in order to detect deviations from a planned level. The decision to be made based on the production control is whether to continue production according to the original plan or to change production plan. As an example, the Danish Efficiency control has been studied (Jørgensen, 1985). In this control system several traits are registered and presented to the farmers as quarterly averages. These averages fluctuate either due to random influences or due to systematic deviations from the planned level. By observing the production traits during a production period a better knowledge of the expected future levels is obtained. The deviation of the individual trait from the expected level can be weighed with the marginal value of the trait, and an estimate of the economic value of the deviation can be obtained. By comparing these deviations to the expected value of an other plan, the decision whether to alter the production in the future can be made. A similar approach has been used by Huirne (1990).

The improvement in income from these changes in production plan is equal to the value of the production control. The possible decisions are to continue with

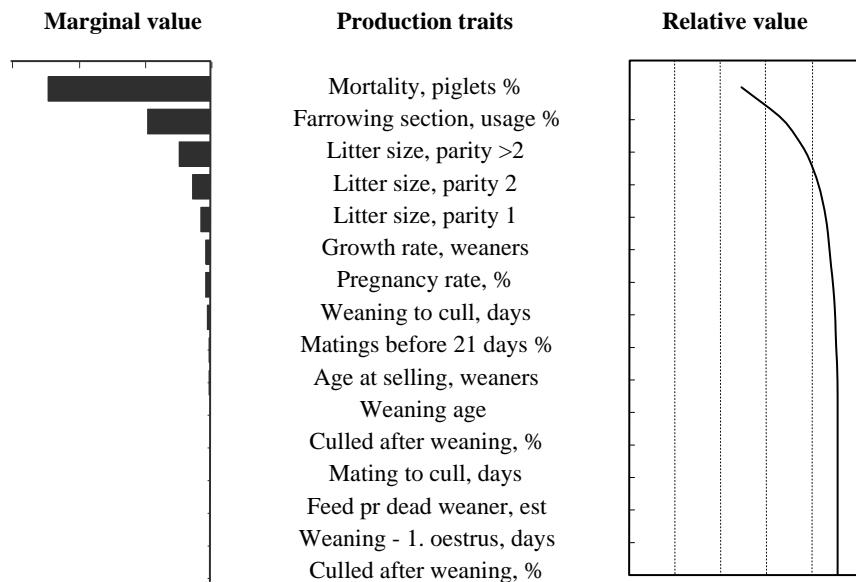


Figure 6.4: Value of different traits in production control in sow units

same production plan or to alter the plan. The first two moments of the prior distribution are the expectation and variance for production traits and expected income under the other production plans. The first two moments of the posterior distribution are the conditional expectations and variance given the observed level of production trait in the control. The loss functions (that are minimized) are minus the expected income from the current plan and minus the expected income from the other plan with the cost of changing plan included.

In the study by Jørgensen (1985) it was shown that the registrations in the Danish efficiency control could be expressed from 16 major traits. The distribution of quarterly averages of these traits were approximated with a 16-dimensional normal distribution. Based on registrations from 100 sow herds in the Danish Applied Pig Research Scheme, 'Den rullende Afprøvning' (Pedersen et al., 1995), the variance in these traits was divided into variance between herds and variance within herds. From these variance-covariance matrices it was possible to specify the prior and posterior distribution of the traits. The loss functions representing expected income in the herd were calculated using standard prices. From these parameters it was possible to calculate the value of the production control. The study showed an overall improvement of 6-7 % in total income from using the control. Furthermore, the marginal values of different production traits from a control point of view were calculated in the study. The sequence of the traits to be included was chosen to make the highest marginal improvement in value for each new trait.

As shown in Figure 6.4, the most important traits to monitor are *piglet mortality*; *usage of farrowing department*; *litter size in parity 1, 2, and 3 or higher*; *growth rate of piglets*; and *pregnancy rate*. It is important to note that the litter size in the different parities treated separately has a value on their own. The recommendations of Sundgren et al. (1979) of treating these traits separately was thus confirmed in this study. These 7 traits accounted for almost all of the value of the control.

The conclusion on the study was that the use of the efficiency control did indeed improve the expected income, but this improvement could be obtained from fewer registrations than currently used. The emphasis should be on controlling the quantitative aspects of the whole herd's production, instead of a detailed control of individual performance.

6.6.2 Recording of sow specific litter size

As mentioned in the previous section, the registration of parity specific litter size in the herd is one of the most important registrations from a production control point of view. This might indicate that litter size of the individual sow is an important trait. Several authors have investigated the possibility for culling sow with respect to litter size, e.g. Strang and King (1970); Treacy (1987); Huirne et al. (1991).

The decision is relatively straight forward; if the sow obtains a litter size lower than a specified norm at a given parity, it should be replaced by a replacement gilt. The improvement in expected income by using information concerning litter size could be used, or, as in the following, the improvement in expected average litter size. Jørgensen (1992) considered a modification of the method used by Huirne et al. (1991). From this paper results concerning detail of information will be presented. Three levels of information were considered. *No information*, i.e., only involuntary culling; *Parity information*, i.e., only the parity of the sow is known; and *Litter size*, i.e., the litter size in each previous parity of the sow is known. The relationship between parity and litter size and involuntary culling was assumed to be known as well as the relationship between litter size in subsequent parities.

In Figure 6.5 the relationship between involuntary culling and average litter size is shown using the three levels of information. The level of involuntary culling is measured by the average age in the herd, if only involuntary culling was used. The level of voluntary culling is assumed to be slightly increasing with parity. As can be seen the culling strategy improves the average litter size, at least for the low level of involuntary culling (i.e., high average age). The difference between the strategy using sow specific information and the strategy using only parity specific information is very low (less than 0.1 pigs per litter).

Furthermore, in Figure 6.6 a more realistic situation is presented, where slightly wrong estimates of the influence of parity on involuntary culling is used when calculating the culling strategy. The use of these erroneous estimates results in a reduction in expected litter size, compared to the situation where no 'Optimal'

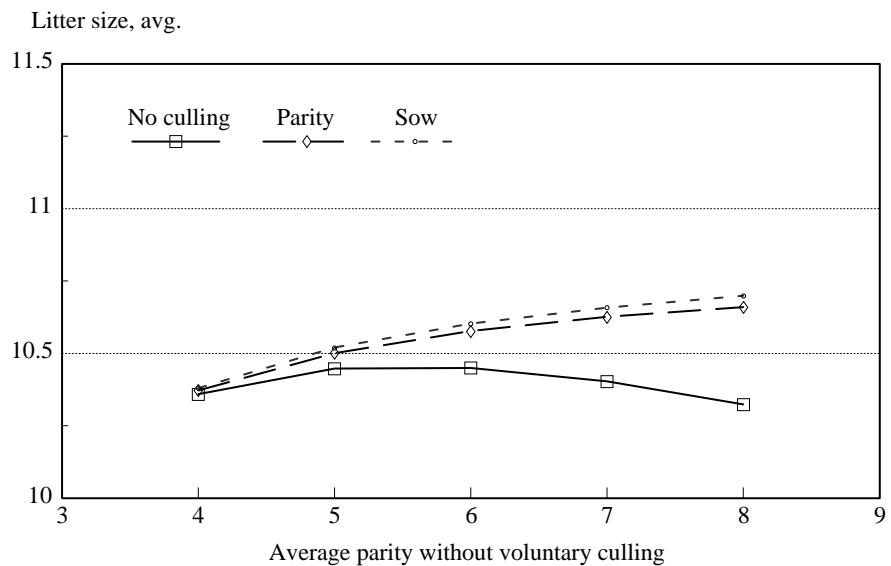


Figure 6.5: Value of different traits in production control in sow units

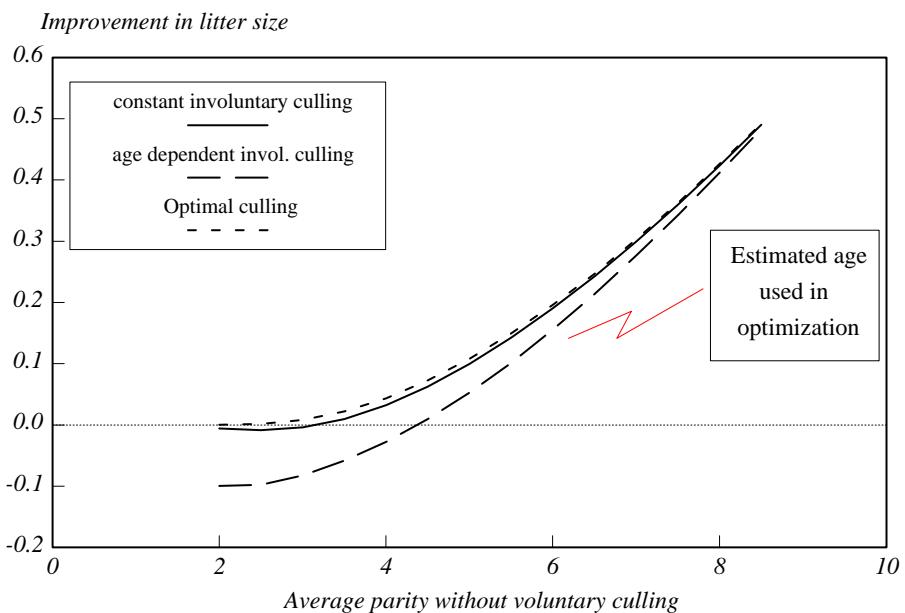


Figure 6.6: Effect of using wrong parameters when calculating 'optimal' strategies

culling strategy is used. The magnitude, 0.1 pigs per litter, is fully comparable to the maximum possible benefit of using the culling strategies.

As a conclusion, due to the low value of sow specific information, decision support models for sow culling do not necessarily need to include the variance between sows with respect to litter size. This gives a considerable reduction in necessary calculations and complexity of the model used. However, the problem is to estimate the influence of parity on the relevant traits in each herd. As this would have to be done on selected data, it is no trivial task. A method for this has been described by Toft and Jørgensen (2002), and later Kristensen and Søllested (2004a,b) have developed a sow replacement model based on parameters estimated from herd specific data using the method of Toft and Jørgensen (2002).

6.6.3 Precision in weighing of slaughter pigs

Weighing of slaughter pigs is often considered necessary in order to obtain efficient production. From a decision point of view, weighing can be used as a control of growth, and for deciding, when to deliver the slaughter pig to the slaughter house. Jørgensen (1993) studied the problem in detail and the following is an extract of that paper, analyzing the use of weighing when selecting pigs for slaughter. Several methods for weighing or estimation of live weight have been suggested: the traditional individual weighing; electronic weighing equipment with electronic identification of the individual pig; weight assessment using the dimensions of the pigs, e.g., through image analysis; and, finally, simple visual assessment of weight. These methods will be expected to differ in precision, or to put it in another term to have different variance. As the cost of weighing differs markedly, an estimate of the value of an increase in this precision would be of interest.

In Denmark slaughter pigs are priced according to their slaughtered weight, and furthermore graded according to meat percentage. Pigs with a slaughter weight in the interval between 70-85 kg will obtain the highest price per kg. Thus, there is an incentive to deliver pigs with the right slaughter weight. A pig producer can only measure the live weight of the pig, and has to decide, whether to deliver or not, based on this criteria. He will necessarily have to cope with a variation around the desired slaughter weight. Furthermore, he has to report how many pigs he will deliver to the slaughter house approximately 3 days in advance. Usually he can only deliver pigs once or twice a week. The pig producer uses the decision rule that if observed live weight of the pig is larger than a threshold weight, the pig is delivered three days afterwards¹. If not, the pig is kept in the herd until the next weighing a week afterwards. However, if the expected return in the next week is lower than the expected return of a new pig, the pig is also delivered, regardless of weight. After delivery the pig is replaced with a new pig after a week for cleaning of the pen. A probabilistic simulation model (Jørgensen, 1991) was used in

¹It is realized that this decision rule is not optimal, a kind of regression equation would improve the decisions. However, this decision rule, is the rule generally recommended by advisers

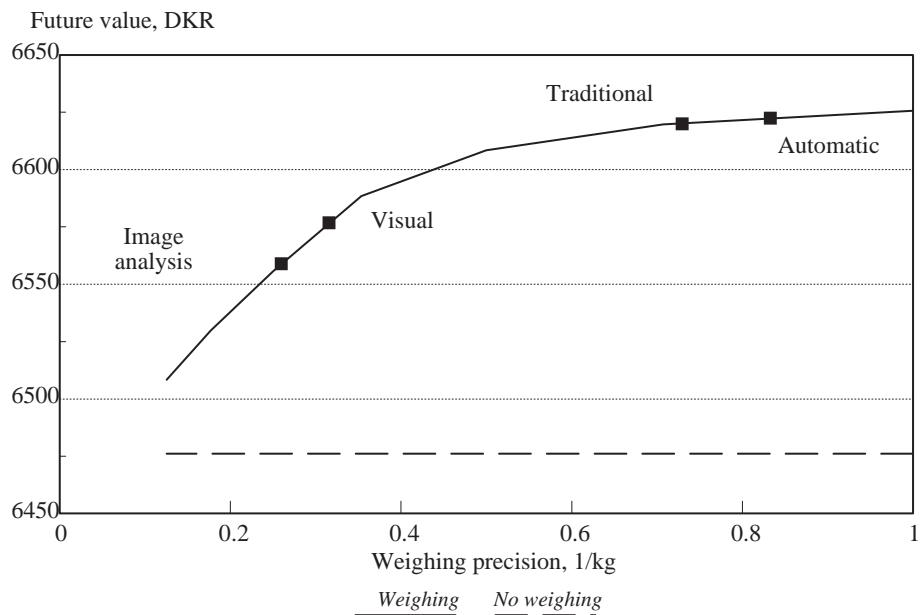


Figure 6.7: Influence of weighing precision

order to calculate proportion delivered on each day after insertion, and corresponding expectation and variance-covariance matrices for total feed consumption and slaughter weight. Assumptions of multivariate normal distribution is used. The expected future value (FV) of the production using an interest rate of 0.1 is calculated for each threshold live weight and the optimal threshold live weight is found. Prices and costs correspond to the level in Denmark in the middle of December 1991. In Figure 6.7 the future value is shown, for different values of the weighing precision, (1 divided with the standard deviation).

The markers in the figure correspond to preliminary estimates from a study at our institute. With a weighing precision of 1 the value of the weighing is 150 DKK in FV compared to a fixed delivery date, or approximately 4 DKK per pig produced. Compared to simple visual estimation of weight the value of weighing is approx. 2 DKK.

It must therefore be concluded that with the current pricing system there is not much economic value of weighing. Automatic weighing equipment in combination with electronic identification cannot even earn the cost for the electronic identification tag. The value of weighing might, however, be found in other production phases, e.g., in growth and feed control and estimation of growth and feed consumption curves for the individual herd.

6.7 Outlook

As shown it is possible to get indications of the value of registrations from a detailed analysis of the decision process, and the related stochastic variation in *state of nature*. The registrations used in their traditional context have shown only a slight improvement in expected utility, or as in the case with the pregnancy tester, even a negative effect. On the other hand, if the registrations are used in a different decision context, they can have positive influence. The estimation of utility value often indicates these other uses of registration.

The notion of negative value of information is important. Researchers trying to develop decision support systems should have this in mind. Also the value of the system should be compared to systems with a lower need for information. Usually it is assumed, that the prior distribution of the traits is known. We usually assume that we know how litter size depends on parity, we know the relationship between growth rate, feed consumption and age, etc. This knowledge is, however, based on experimental results or from other production herds, and might not be relevant for every herd. In practice, the most important problem seems to be how to obtain herd specific estimates of these relationships. If we are not to overtax the animal producers with information from the profusion of possible registrations, we need to identify the decisions made in the herd. Then we should consider, which information is used in the decision.

Finally, we should investigate, whether some relevant data is missing, and whether the data can be obtained in a cost effective way. Then we might suggest to the animal producers, that they begin to use these data.

Chapter 7

Dynamic production monitoring by classical methods

Automatic monitoring of animal production is not very common compared to what is usual in industrial production, where key variables are registered and analyzed statistically during the entire production process in order to quickly alarm about any deviations from the norms.

In animal production current standards for monitoring are as described in Chapter 5. Thus, quarterly or monthly reports with selected key figures is the most common form of monitoring.

In order to improve the production process, a more dynamic monitoring method is desirable. Collecting data (and in what way) has been covered in Chapter 6. It is now time to look at the use of data for monitoring the production process in a dynamic way. Statistics provides us with an analytical framework to handle the uncertainties inherent in especially animal production. This also means that the topics covered here will become more mathematical in nature.

The main problem with the current chapter is, however, that it has not yet been written. Instead, readers are encouraged to consult an excellent book by Montgomery (2005), where in particular Sections 4.1 - 4.3.5, 5.2.4, 8.2, 8.3, 9.4-9.4.2 are recommended. Those sections cover

- Shewhart control charts.
- Moving average control charts.
- Exponentially Weighted Moving Average (EWMA) control charts.

These straight forward monitoring methods have only been applied very little to animal production. A few examples in dairy cows, nevertheless, include Thysen et al. (1988); Thysen (1993b) (in particular the EWMA method is applied). More recent examples are de Vries (2001); de Vries and Conlin (2003, 2005).

Chapter 8

Dynamic production monitoring by state space models

This chapter continues the efforts for creating a consistent framework for dynamic production monitoring. From the herd manager's point of view, animal production is a sequence of decisions based on a continual stream of information. Hence, there is a need for a framework which allows knowledge to accumulate, so that it may, in principle lead to improved understanding of the production and, in consequence, better decisions. Dynamic linear models (DLM) are a possible approach to provide such a framework in which relevant prior knowledge and current information is combined into a coherent tool for decision making. Those models will be the cornerstone of this chapter.

Again, however, the main problem with the current chapter is, that it has not yet been written. Instead, readers are encouraged to consult a paper by Thysen (1993a) presenting the simple dynamic linear model and the multiprocess model. It is a much more accessible text than the textbooks mentioned below, but it is not a proper substitution.

The best textbook on Dynamic Linear Models is in our opinion still West and Harrison (1997). A more accessible textbook is Poole et al. (1994).

Three papers by Madsen and coauthors (Madsen and Ruby, 2005; Madsen et al., 2005; Madsen and Kristensen, 2005) are examples of applications with pig production.

Chapter 9

Decisions and strategies: A survey of methods

9.1 Framework of the survey

In Chapter 1 we defined (Definition 2) the information needs for decision making to include

1. The current state of the unit.
2. A production function describing the immediate production given stage, state and decision and the distribution of the possible random term(s).
3. The distribution of the future state given stage, state and decision.
4. All attribute functions relevant to the farmer's preferences.
5. The utility function representing farmer's preferences.
6. All constraints of legal, economic, physical or personal kind.

The purpose of this chapter is to provide a general overview of techniques available in the determination of optimal decisions and strategies.

The techniques will be described according to their ability to represent the various kinds of necessary information listed above. Furthermore, their potentials for integration of decisions at different levels and time horizons defined in Chapter 1 are discussed. It is not the purpose to describe the various methods in details, but only to provide a general survey relating to the issues of the previous chapters. Details about the methods are left for separate chapters.

9.2 Rule based expert systems

Research concerning expert systems is a development within the area called Artificial Intelligence (AI). The British Computer Society has defined expert systems as

follows (cited from Dindorff, 1992): *An expert system is regarded as the embodiment within a computer of a knowledge-based component from an expert skill in such a form that the system can offer intelligent advice or take intelligent decision about a processing function. A desirable additional characteristic, which many would consider fundamental, is the capability of the system, on demand to justify its own line of reasoning in a manner directly intelligible to the enquirer. The style adopted to attain these characteristics is rule-based programming.* This is just one of many proposed definitions of expert systems.

The fundamental difference between the rule based systems, and the approach that we have presented until now, is that instead of trying to model a system, the rule based expert systems tries to model the expert, or rather the expert's approach to problem solving. Originally, the ambition within AI-research was to make general problem solvers that could be used for any problem area, but this was realized by most researchers within the area to be overambitious. The research efforts had, however, led to new approaches towards problem solving, and within narrow problem (expert) domains, the approach showed some promising results.

Rule based expert systems have three components: a knowledge base containing the expert's knowledge of the domain, an inference engine that decides how and when to use the knowledge, and a user interface.

The knowledge base contains knowledge of a problem domain, as it is described in text books, as well as expert knowledge, e.g. exceptions to general rules, experiences from previous problems, and time-efficient approaches on how to solve problems within the area. The knowledge base is an enhanced data base that apart from data also contains logic rules for the connection between the items in the knowledge base, e.g if x or y then z .

The inference machine contains the mechanism for deduction based on the logical rules in the knowledge base. The deduction can use different inference principles, such as backward chaining and forward chaining. In the rule showed above, the inference machine would start out by finding the value of z , and given the knowledge of z establish the value of x and y . Forward chaining would start out by establishing the values of x and y and deduce the value of z subsequently. In both cases the unknown values are found either by a question to the user or by combining other rules in the knowledge-base. The optimal choice of inference principle depends on the type of problem the expert system is supposed to solve. In very complex expert systems, neither forward nor backward chaining is fast enough, and the so-called heuristic search strategies are needed. These strategies work primarily by searching the knowledge base in an efficient order, focusing on areas, where a solution to the problem is most likely to be found. Both general and problem specific heuristic strategies exist.

The questioning mechanism is a standard part of the user interface. Besides posing questions, the user interface is usually able to explain, why it asks the question, i.e. *I am trying to establish the value of z because I want to know if either x or y is true.* Another facility is the explanation facility, i.e. *I know that neither x*

is true nor y is true because z is not true. Usually the phrases are formulated more user friendly.

Programs for maintaining knowledge bases in connection with inference machine and user interface as an integral part is sold as the so-called expert system shells. This concept originates from the medical diagnosis system, MYCIN. The knowledge base in MYCIN was emptied and the program sold as E-MYCIN (or empty mycin), and was thought to be applicable to any problem domain. Very often these shells are programmed in programming languages where logic deduction can easily be represented, such as LISP or PROLOG, but standard programming languages such as C and Pascal can of course be used.

Rule based expert system can be categorized into several areas (Hayes-Roth et al., 1996). Referring to Figure 1.1 they comprise the planning, check and analysis phase of the management cycle.

In developing rule based expert systems two “players” are essential, of course the expert, but in addition the so-called knowledge engineer. The role of the knowledge engineer is to “extract” the knowledge from the expert and to formulate the knowledge as rules that can serve as input to the knowledge base. Knowledge engineering has in fact become a research area in its own right.

To illustrate the problem of knowledge engineering, the first rule based expert systems were based on very simple “expert” rules very much inspired by the diagnostic systems, e.g. if indications a and b are observed then problem is probably c . Later on it was realized that the expert relies on many information sources and part of being an expert is to know when to draw on which knowledge sources. If we look at the information necessary to make optimal decisions as mentioned in Section 9.1 this can be seen as the result of an expert’s problem solving. An expert system would therefore guide the user through obtaining the necessary information. If it is not possible to obtain the necessary information it would use other information and based on the expert’s experience try to make a sufficiently good plan.

The current trend is that the rule-based system does not function as stand-alone systems, but rather as an integral part of other systems, the so-called knowledge based systems. A typical example could be that the expert system helps in establishing the user’s utility function by asking questions and then uses this utility function when calling an optimizing program. The concept is now incorporated into the wizards and experts known from standard computer program, e.g. spreadsheets and word processors.

The method will not be discussed further in this book.

9.3 Linear programming with extensions

The general linear programming problem may in matrix notation be written as follows:

$$\begin{aligned}
 px &= \text{Min!} \\
 &\quad \text{subject to} \\
 Ax &\leq b \\
 x &\geq 0
 \end{aligned} \tag{9.1}$$

where p is a constant row vector with m elements, A is a constant matrix of size $n \times m$, b is a column vector with n elements, and x is a vector of variables. The problem is to select a vector x that minimizes the linear objective function px and simultaneously meets the linear restraints $Ax \leq b$ and $x \geq 0$.

Eq. (9.1) represents the standard formulation of a linear programming problem. In applied models it is often convenient to define a maximization problem instead, and some of the restraints may be of the kind $a_i x \geq b_i$ or the kind $a_i x = b_i$ (where a_i is a row in A), but any linear programming problem may be rearranged in accordance with the standard formulation of Eq. (9.1).

If we interpret the linear programming problem in relation to a herd management decision problem, then x is a vector of factor levels and $Ax \leq b$ is a set of constraints of legal, economic, physical or personal kind. It should particularly be noticed, that personal restraints may also include constraints on levels of attribute functions (for instance leisure time or monetary gain). This direct representation of constraints is probably the main force of the method. The objective function px has to represent the aggregate utility function.

If we compare the linear programming problem with the information needs of a decision problem (cf. Section 9.1) we observe that all random elements are missing. At least in the standard formulation, the method is deterministic. Also the dynamic linking to the future state of the production unit is missing. A consequence of the latter shortcoming is that only effects at the current stage are represented. In other words, the method is static of nature. Furthermore, we observe, that since the aggregate utility function has to be linear in the factor levels, x , also the production function, all attribute functions and the utility function have to be linear. Examples of linear attribute functions are shown as Eqs. (3.1), (3.4) and (3.7), and a linear utility function is shown as Eq. (3.17). This demand for linear functions and linear restraints is a serious weakness of the method.

Several of the shortcomings mentioned may be redressed or at least adapted by extensions to the method: The linear objective function may be replaced by a quadratic one (quadratic programming); the static nature may be modified by introduction of stages and additional constraints ensuring dynamic links (dynamic linear programming); random terms may be added to the elements of A and b , and the corresponding restraints may be expressed as probabilities (stochastic programming); and often, non-linear functions may be approximated by pieces of linear relations over short intervals. In particular, dynamic linear programming, may be used to link decisions at different levels with different time horizons.

Herd management applications of linear programming are numerous. The most frequent application is no doubt for ration formulation, where least-cost rations meeting the nutritional requirements of the animals in question are met. Most often such programs ignore the effect of feeding on production.

Also examples of application of linear programming for whole-farm planning may be found in literature. Refer for instance to Hansen (1992) and Hardie (1996). Such models are often very large containing thousands of variables and restraints, but since very efficient standard software is available this is hardly a problem.

Due to the shortcomings of Linear Programming in dealing with important aspects (like dynamics and uncertainty) of herd management, the method is not given high priority in this book. Nevertheless, a short intuition based description is given in Chapter 10, together with practical aspects of modeling by use of linear programming.

9.4 Dynamic programming and Markov decision processes

Consider a production unit which is observed over a number of stages $n = 1, \dots, N$. At the beginning of each stage, we observe the state, $i \in \omega_n$, of the unit. Having observed the state, we have to take an action, $d \in D_n$, concerning the production unit. Usually, the state space, ω_n , and the action space, D_n , are assumed to be finite sets. Depending on the stage, state and action, a reward is gained. The reward may very well be a random variable, but the expected value, $r_i^d(n)$, has to be known. Also the state to be observed at the next stage is a random variable. We shall denote as $p_{ij}^d(n)$ the conditional probability of observing state i at stage $n + 1$ given that state j has been observed and action d taken at stage n . Finally, a strategy, s , is defined as a map assigning to each combination of stage and state an action $s(n, i) \in D_n$. We have now defined the elements of a Markov decision process (or a dynamic programming problem).

The purpose of dynamic programming is to determine a strategy which (in some sense) is optimal. Several optimization techniques are available. The most commonly applied method is called *value iteration* where a value function representing the expected total rewards from the present stage until the end of the planning horizon (i.e. stage N) is maximized. Optimal decisions depending on stage and state are determined backwards step by step as those maximizing the value function. This way of determining an optimal policy is based on the Bellman principle of optimality which says: *An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision* (Bellman, 1957, p. 83). Value iteration is often just denoted as dynamic programming.

If N is large, an infinite planning horizon is often assumed. A relevant optimization technique for infinite stage problems is *policy iteration*. This method was introduced by Howard (1960), who combined the dynamic programming technique with the mathematically well established notion of a Markov chain. A natural con-

sequence of the combination was to use the term Markov decision process to describe the notion. The policy iteration method was a result of the application of the Markov chain environment and it was an important contribution to the development of optimization techniques.

The objective function being maximized during optimization depends on the circumstances. It may represent the total expected rewards, the total expected discounted rewards, the average rewards per stage or the average rewards over some kind of physical output.

If we compare the dynamic programming problem with the information needs of a decision problem (cf. Section 9.1) we observe that most aspects are covered. The current state is simply i , and the rewards directly correspond to production functions; the conditional probabilities $p_{ij}^d(n)$ represent the dynamic random links to future stages; and the objective function represents the farmer's utility function. There are, however, some restrictions on the kind of utility function which can be represented in a dynamic programming model. The restrictions concern the way in which individual stage attributes are aggregated into the over-all utility function as described by Eq. (1.5). In order to be able to use dynamic programming, we implicitly assume that the aggregation may be performed in such a way that we first aggregate attributes at the same stage into a stage specific utility v_n of the kind

$$v_n = g_n(u_{1,t_n}, \dots, u_{k,t_n}),$$

where g_n is a stage specific utility function of arbitrary kind. In the dynamic programming context, v_n is identical to the reward $r_i^d(n)$. The over-all utility (i.e. aggregation over stages) in turn must be calculated as a simple sum of the stage specific utilities v_n , as a discounted sum (cf. Eq. (3.2)), as the average value over stages or as the average value over some kind of physical output or input from production.

The most difficult kind of information to represent in dynamic programming models is the information on constraints. There is no general solution to that problem, but sometimes it may be solved by using an objective function maximizing average rewards relative to the most limiting restriction. An example is maximization of average net returns per unit of milk produced under a milk quota (Kristensen, 1989). In other cases combination of the method with methods like simulation (Ben-Ari and Gal, 1986; Kristensen, 1992) or genetic algorithms (Houben et al., 1995) may circumvent the constraint problem.

A major problem in relation to dynamic programming models is the so-called curse of dimensionality. Since the state space is represented by discrete levels of a set of traits (state variables), models tend to become very big. Thus a model presented by Houben et al. (1994) contained 6.8 million states. Despite the size of the model, optimization was still possible due to a new notion of a hierarchical Markov process described by Kristensen (1987, 1991).

Later the concept was further developed by Kristensen and Jørgensen (2000) into multi-level hierarchical Markov processes in order to allow for simultaneous

optimization of decisions on multiple time scales (at the founder level as well as at the child levels). A general Java software system, MLHMP, for representation and solution of multi-level hierarchical Markov processes has been developed by Kristensen (2003).

Numerous applications of dynamic programming are described in literature. A relevant textbook concerning application in agriculture has been written by Kennedy (1986). The book also contains a survey of agricultural applications. In herd management, the technique has most often been applied for operational decisions like replacement, insemination and medical treatment of animals.

Despite the numerous successful applications, the very concept of a Markov decision process has some built-in limitations:

State space representation: Most often the state of the system (e.g. an animal) being modeled is defined by the values of a number of state variables each representing a trait of the system. The state space is then defined as the cartesian product of all value sets of individual state variables. This leads to very large transition matrices which - even though they are often sparse - are inefficient from a numerical point of view.

Observability: It is generally assumed that the parameters of a Markov decision process are known and that the state space is fully observable. In particular when we are dealing with production assets, it would be more logical to distinguish between directly observable state variables like for instance number of items produced and underlying unobservable asset dependent potential for production capacity. This kind of modeling is referred to as POMDP *Partially Observable Markov Decision Processes*. Refer for instance to Lovejoy (1991) for a survey or Kaelbling et al. (1998) for an introduction.

Markov property: The Markov property implies that the state space at any stage must contain sufficient information for determination of the transition probabilities. In a straight forward formulation of a decision problem this is rarely the case. The trick used most often in order to make the process Markovian is to include *memory variables* in the state space. Even though this solves the problem from a theoretical point of view it contributes significantly to the *curse of dimensionality* of such models.

Even though the multi-level hierarchical Markov decision processes to some extent compensate for these problems by partitioning the state space according to temporal considerations and also an attempt to combine hierachic Markov processes with Bayesian updating in special cases has been done by Kristensen (1993), the state space representation, the lacking observability and the Markov property remain important limitations for the use of Markov decision processes.

Further details of the technique are given in Chapter 13. A very good general textbook has been published by Puterman (1994).

9.5 Probabilistic Expert systems: Bayesian networks

Another part of the research area named Artificial Intelligence are the so called probabilistic expert systems that rely on the Bayesian network. The following description is based on Lauritzen (1995).

In some areas where expert systems are appropriate, the task involves performing a sequence of steps according to specified logical rules. However, other expert systems work in domains that are characterized by inherent uncertainty. This uncertainty is either due to imperfect understanding of the domain, incomplete knowledge of the state of the domain at the time where the task has to be performed, randomness in the mechanisms governing the behavior of the domain or a combination of these. Within these domains probability and statistics can serve to represent and manipulate the uncertain aspect of domains having these characteristics. Probabilistic methods were for some time discarded in this context as requiring too complex specification and computation. However, the work of Pearl (1988) and Lauritzen and Spiegelhalter (1988) demonstrated that these difficulties could be overcome, based on causal networks or as it is now usually termed Bayesian networks. There exist other formalisms for handling uncertainty in expert system, such as the fuzzy sets, but these will not be discussed in the present context.

The rule based systems were mainly constructed through modeling of the behavior of the expert and the encoding of this behavior in terms of rules of various kind. In contrast, probabilistic expert systems are constructed by modeling the domain rather than the expert. The method is thus in close correspondence with the approach used in this book, where the domain is modeled using production functions etc. The probabilistic expert systems specify a graphical model for the variables. The reasoning is then performed by updating the probabilities of the domain in the light of the specific knowledge according to the laws of conditional probability.

The graphical model captures associations between entities in the domain, or rather lack thereof, in terms of conditional independence that in a systematic fashion are encoded in a graph or network with nodes representing the entities themselves and edges representing associations between them. The nodes are represented as dots or circles. The edges are usually directed corresponding to influences of a causal nature and represented as arrows, or, sometimes, undirected corresponding to symmetric associations (e.g. correlations) and represented as lines.

The use of the graphic specification in the probabilistic expert systems plays several roles. For example, it gives a visual picture of the domain information; it gives a concise presentation of domain information in terms of conditional independence relations, and it enables rapid computation and revision of interesting probabilities.

The graphic method can also be used for several important tasks in the specification process. It can be used to learn quantitative and structural aspects, or as it known within general statistics, estimation and model selection.

If we compare the probabilistic expert systems method with the information needs of the decision problem (Section 9.1), it is important to recognize that the method is inherently a static method, even though attempts have been made to model dynamic systems as well. For monitoring and analysis purposes it is ideal, i.e. the method can assign probabilities to observed deviations, whether they are random or not. It can also make a diagnosis in the analysis, that is, indicate probabilities for different causes of the deviation. This can in turn serve as the necessary basis for decisions concerning changes in production plan.

The expert systems can be build as recurrent time slices and can in this manner represent dynamic production functions, predicting the future state for given decisions. The constraints concerning the production function can be modeled, but constraints may cause the same problems as described under Dynamic Programming and Markov decision processes.

To represent decisions in Bayesian networks the decision can be included as a random variable in the model, with the different decisions as level of the variables. When the decision is made, the corresponding level is assigned a probability of 1. This approach does not make any search for optimal decisions.

If decisions have to be included, Decision Graphs should be used instead. They can in fact be fitted into the general framework of Bayesian Networks.

We shall not in this book discuss Bayesian networks further. Instead, readers are referred to an excellent textbook by Jensen (2001) where this important subject is convincingly introduced in Chapters 1 and 2. Readers interested in the algorithmic aspects of the method should consult Cowell et al. (1999).

9.6 Decision graphs

Decision graphs (or influence diagrams as they are also called) were introduced by Howard and Matheson (1981) as a formalism to model decision problems with uncertainty for a single decision maker. The influence diagrams gave a more compact graphical representation of a decision problem than the more traditional decision tree approach as illustrated in Figure 9.1.

A decision graph is very similar to a Bayesian network consisting of chance node and directed edges (“arrows”) denoting causal effect. In addition two more node types are introduced, the decision node shown as a square, and the value node shown as a diamond (cf. Figure 9.1).

Originally, the decision graph was translated to a decision tree within the computer and the standard “average-out and fold-back” algorithm was applied on that tree. In Shachter (1986) a method was suggested for solving the decision problem represented by the decision graph directly, without the translation to a decision tree. This method transformed the decision graph by successively removing nodes in the graph, until at last only one final utility node remained, holding the utility of the optimal policy. In order to solve many similar problems one therefore had to start from scratch every time. The transactions performed on the graph consisted of

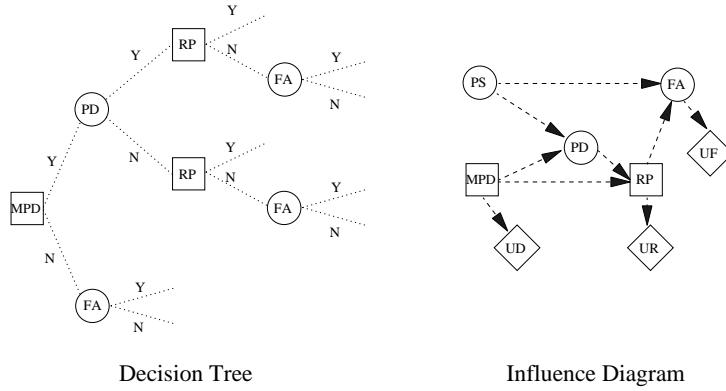


Figure 9.1: Decision tree and decision graph (influence diagram) representation of the pregnancy diagnosis and replacement problem.

four simple transactions, the arc reversal (application of Bayes Theorem), node removal by summing out, expectation of a value node with respect to a change node, and finally removal of a decision node into a value node by maximization. Initially problems were formulated with only one value node. By introducing the concept of separability of the utility function or value function, Tatman and Shachter (1990) showed that the Dynamic Programming Problems could be solved within the decision graph framework, by introducing the separability of the utility function. The requirement for separability is the same as the requirement that the overall utility is calculated as a simple sum of stage specific responses as mentioned in Section 9.4.

Shenoy (1992) proposed another algorithm that gave the solution to the decision graph without disrupting the structure of the graph. Then Jensen et al. (1994) showed how a similar approach could be incorporated within the general framework of Bayesian Networks. This approach has been implemented in the Hugin expert system shell. The similarity between decision graphs and Bayesian networks means, that several important improvements is to be expected. This comprises, e.g approximate solutions by techniques such as Monte Carlo methods (Bielza et al., 1999; Charnes and Shenoy, 1996), easy representation of dynamic models as in dHugin (Kjærulff, 1995) and object oriented design (Bangsø, 2004).

As decision graphs closely corresponds to Dynamic Programming the same comments concerning the information needs of the decision problems can be made. In addition, the current version of decision graphs are inherently static, and no algorithm corresponding to policy iteration has been found (even though R.A. Howard's

research has been central for both developments). If stages of varying time length have to be modeled, time has to be included in the model and the discounting factor has to be incorporated directly in the utility. Furthermore decisions choosing between qualitatively different child processes, such as in Hierarchic Markov processes is currently not possible. It should, however, be noted that it is a very active research area, and continuous progress is to be expected.

Decision graphs at first glance seem to overcome all three problems listed for Markov decision processes in Section 9.4:

State space representation: Instead of a state space formed as a cartesian product, the state variables are represented one by one in decision graphs.

Observability: In decision graphs, variables may be observable or unobservable.

Markov property: The classical algorithms applied to decision graphs as presented by for instance Jensen et al. (1994), implicitly assume *no forgetting* implying that all previous observations done and decisions made are remembered and taken into account. This “no forgetting” assumption is today associated with the concept of “Influence diagrams” as a subclass of the more general concept of “Decision Graphs”.

In practice, however, the performance of influence diagrams has been disappointing. Even though algorithms for optimization are available (refer for instance to Cowell et al., 1999, Chapter 8), the numerical calculations become prohibitive for real world decision problems because of the “no forgetting” assumption. A recent extension by Lauritzen and Nilsson (2001) to influence diagrams called LIMIDs or *Limited Memory Influence Diagrams* relaxes the “no forgetting” assumption thus providing a computationally tractable decision problem without assuming a Markov process. The algorithm presented is exact in some cases and only approximate in other cases.

Thus, the method has obvious possibilities for application within animal production, but so far the only example known to the author is a model for determination of optimal slaughter policies in slaughter pigs (Hansen, 2006). Within crop protection a system has been made for decision making concerning mildew management in winter wheat (Jensen, 1995).

We shall not in this book give details about the algorithms of decision graphs. Instead, we again direct the readers to good textbooks like Cowell et al. (1999); Jensen (2001). Chapter 12 deals with some practical aspects of modeling by use of decision graphs.

9.7 Simulation

As the name implies, a simulation model is simply a model of a system. The model is used for the study of the real system’s behavior under different condi-

tions. Within animal production the term usually refers to computer based dynamic calculation models.

Formally, the simulation model is a computer representation of the production function, the attribute function, and/or the utility function. The degree of detail differs between the different models.

The input to the model consists of two elements, a set of parameters, Φ , and a set of decision rules, Θ . The decision rules specify the setting of input factors as well as other decisions in the system. The term “decision rule” is used rather than decision strategy, because usually no direct mapping between the rule and the state of the whole system exists. A decision rule can e.g. be to use a dynamic programming model to specify a decision strategy every (simulated) year. Another example is to use some simple rule-of-thumb (heuristic) to make culling decisions. The set of parameters can be split in two, $\Phi = (\Phi_0, \Phi_{s\bullet})$, where Φ_0 are the initial values of the parameters at the start of the calculation (State of Nature) and $\Phi_{s\bullet}$ represents parameter values that change during simulations. The elements of $\Phi_{s\bullet}$ are often called state variables, and can be further split into time stages of the model i.e. $\Phi_{s\bullet} = (\Phi_{s1}, \Phi_{s2}, \dots, \Phi_{st}, \dots, \Phi_{sT})$, where T is the number of stages in the planning horizon. It is often convenient to refer to the set of output variables Ω_\bullet that contains calculated values of input factors, production functions, attributes etc. The distinction between the elements of $\Phi_{s\bullet}$ and Ω_\bullet is not clear, but usually, Ω_\bullet is a subset of $\Phi_{s\bullet}$. The elements in Ω_\bullet will usually be traits that at least in principle can be observed in the real system. The term *model input* usually refers to (Φ_0, Θ) .

The purpose of the models is to calculate the expected utility, $\bar{U}(\Theta)$, under a given decision rule, Θ , i.e.,

$$\begin{aligned}\bar{U}(\Theta) &= \int_{-\infty}^{\infty} U(\Theta, \phi) f_{\Phi}(\phi) d\phi \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\Theta, \Phi_{s\bullet} | \Phi_0 = \phi_0) f_{(\Phi_{s\bullet} | \Phi_0)}(\phi_{s\bullet} | \phi_0) f_{\Phi_0}(\phi_0) d\phi_{s\bullet} d\phi_0,\end{aligned}\tag{9.2}$$

where U is the utility function, which in general can refer to any function of the output variables. The functions f_{Φ} , $f_{(\Phi_{s\bullet} | \Phi_0)}$ and f_{Φ_0} are the density functions of Φ , $(\Phi_{s\bullet} | \Phi_0)$ and Φ_0 , respectively. *Simulation models are numerical methods for solving this integral.*

Two different categories of simulation models have been implemented within animal production. Stochastic models and deterministic models, where the stochastic nature of the system is ignored, i.e. the underlying assumption is that $P(\Phi = \phi_c) = 1$ for some ϕ_c , e.g. estimated from various experiments. It is important to realize that such a simplification is only valid, if $U(\Theta, \Phi)$ is linear in the parameters. Since this is almost never the case, we will more or less ignore deterministic simulation models in this book. They may have some justification for system comprehension in animal physiology and nutrition (examples can be found in Whittemore and Fawcett, 1976; Black, 1995; Arnold and Bennet, 1991; Danfær, 1990), but hardly in herd management.

Stochastic models can be further subclassified into Probabilistic Models and Monte Carlo models. Probabilistic models are models such as Markov Chain models (see references under Dynamic Programming and in addition e.g. Jalvingh et al., 1992a,b) and Bayesian Networks. Within the probabilistic models the distribution of the output variables can be directly found within a single run of the model. Reasonable complex models can be specified within this context, at least if the parameters and the traits follow the Gaussian (normal) distribution. Capacity restrictions, interactions between system elements and the inclusion of decision variables will, however, make it impossible to specify the distribution in closed form.

Therefore, the Monte Carlo simulation technique is preferred. It relies on the drawing of random numbers. Every time the model encounters a stochastic variable, a (pseudo)-random variable is drawn from the appropriate distribution and this value is used in the subsequent calculations. Each completed calculation (simulation run) with the model represents a random drawing from the simultaneous distribution of input and output variables. By increasing the number of calculations the distribution of the output variables can be specified to any degree of precision. The expected utility is found from:

$$\bar{U}(\Theta) \approx \frac{1}{k} \sum_{i=1}^k U(\Theta, \phi_i). \quad (9.3)$$

where ϕ_i is a random drawing from the multidimensional distribution of the parameters, and k is the number of random drawings (simulation runs). In addition the standard error on the estimated utility can be found by calculating the variance of $U(\Theta, \phi_i)$. Thus, we can obtain a measure of the precision of the estimated utility, and an indication of how many iterations that are necessary. If, for instance, the standard error of the expected utility is 10% after 100 iterations, it will take 10,000 iterations to obtain a standard error of 1%.

Examples of Monte Carlo simulation models are (in pigs) Singh (1986); de Roo (1987), (in dairy cows) the *SimHerd* model (Sørensen et al., 1992; Østergaard et al., 2000, 2004, 2005) and the *Florida Dairy Computer Program* (de Vries et al., 2004), and (in hens) the *SimFlock* model (McAinsh and Kristensen, 2004; Kristensen and Pedersen, 2003).

Simulation models can also be divided between physiological models of single animals, physiological models of whole herds, and models of whole herds with emphasis on managements strategies. The physiological model of whole herds is e.g. Tess et al. (1983); Pettigrew et al. (1986); Finlayson et al. (1995), while examples of current physiological models is found under the deterministic models, and the whole herd approach under the stochastic models mentioned above. Obviously, it is within the last category that the likely candidates for decision support systems should be found. However, models from the first category have been adapted to serve as decision support systems. The first two approaches are often based on a description of the system with differential (or difference) equations, while the third

approach relies more on the theory behind stochastic processes such as queuing models.

Compared to the other techniques, simulation models are much more flexible, and there is no constraint on the degree of detail in the model. Especially when the so-called object oriented programming method is used, it is possible to achieve a very close correspondence between the elements of the real system and the model (see e.g. Chang et al., 1994; Skidmore and Beverly, 1995; Jørgensen and Kristensen, 1995; McAinsh and Kristensen, 2004). Any model variable can be used as output variable and it is easy to represent capacity restrictions.

Very often the purpose of simulation models is to improve the understanding of a system, i.e. to combine research results from different areas to obtain a comprehensive description of the system, the so-called holistic approach. This purpose should be seen as something different from decision support. When simulation models are used to improve the understanding of the complex system, a fixed and known set of parameters ϕ_0 are used for the initial state of nature, Φ_0 , and the expected value of the utility function or any other output variable is calculated as:

$$\bar{U}(\Theta \mid \Phi_0 = \phi_0) = \frac{1}{k} \sum_{i=1}^k U(\Theta, \phi_{si} \mid \Phi_0 = \phi_0), \quad (9.4)$$

i.e. only the inner part of the integrand in (9.2) is solved.

The knowledge of the systems sensitivity to changes in the parameters is part of, what we call understanding of a system.

In contrast, when simulation models are used to determine “optimal” strategies we want to find the optimal set of decision rules given the precision in our current knowledge of the parameters (state of nature). The parameters used in each simulation run should therefore be a sample from the prior distribution of Φ_0 reflecting the precision in our current knowledge, and not fixed values.

The search for optimal strategies is included in linear programming, dynamic programming and decision graphs, (i.e. simplex algorithm, policy and value iteration). No such search facility is included in connection with simulation models. This is a major drawback of the method.

Within simulation models the search for the optimal set of decision rules is treated as a general problem of multidimensional optimization. Several numerical methods exist that can handle this (see e.g. Press et al., 1989). The choice of method should be made carefully. The flexible form of the simulation models means that the behavior of the expected utility function is unknown, for example if there exist discontinuities and local optima. Such phenomena can make some of the methods go wrong. Another complication is that the expected utility is only estimated with a precision depending on the number of simulation runs within each treatment. The difference between two sets of decision rules may therefore be just a matter of sampling error, rather than a difference in expected utility. The solution to this problem is to do more simulation runs. But there is a trade off between time spent improving the precision in the estimate of one set of decisions rules and

the time spent searching for a better. Guide lines to handle this problem are not available.

The search procedures are most easily demonstrated by borrowing terms from experimental world. A set of decision rule is termed a treatment. Expected values from a given treatment are found by a number of replicates (N) or simulation runs. When searching for optimal decision rules, we have to repeatedly specify new treatments and calculate expected utility for the treatment. If we want to combine a set of treatments simultaneously, we design an experiment with the different treatments included.

A well-established technique for well-behaved expected utility functions, especially with continuous variables in the decision rules, is the gradient search technique. First an experiment is designed to initially explore the expected utility function, e.g. a response surface design. The result from this experiment is analyzed and the response surface estimated. If the optimum is outside the current design, the path of steepest ascent of the response surface is estimated. Then an experiment is made with treatments on the steepest ascent path, until the optimal treatment on this path is found. A new response surface design is made centered around this optimum point. This procedure is repeated until the optimum is found with sufficient precision. Using this procedure, an (at least local) optimum will be found.

Other promising techniques are stochastic search techniques, such as simulated annealing and genetic algorithms. These algorithms start with the selecting of a random initial set of decision rules (treatment) as the current. The expected utility of this is calculated. Then the following steps are carried out iteratively. Select a new treatment candidate based on the current treatment by random permutation. Calculate expected utility for the treatment candidate. Decide randomly according to a specific rule (depending on the technique), whether to use the treatment candidate as current candidate by drawing a random variable. Continue the iterations.

A third possibility is the group of so-called heuristic search strategies. Examples can be found in Reeves (1995).

If we compare the simulation method with the information needs of the decision problem, all the aspects can be covered, and the utility function and capacity restrictions can easily be handled. The curse of dimensionality is not felt immediately. The computation time of a single run of the model grows more or less linearly with the complexity of the model. The problem is the search for optimal solutions. The techniques mentioned are not as efficient as either the simplex, value iteration or policy iteration methods. With the same complexity in the decision rules as in e.g. dynamic programming the curse of dimensionality will be felt, e.g. if the decision to cull an animal should include the states of all other animals in the herd. If the rules are specified more heuristically, such as cull the worst animal, the problems become tractable, but no overall optimal solution is guaranteed. Other decision rules might exist with higher utility.

Probably because of this problem, published results from simulation model research usually have only very few options in the decision rule, and the decision

rules are often of a very general nature. The use of simulation models for decision support is usually suggested to be of the what-if nature, i.e. the user of the model specifies some decision rules and the model calculates the expected output from these decision rules. This approach has advantages because there is no need to attempt to formulate the farmer's utility function. The user of the model can simply look at the list of output variables for different set of decision-rules and decide which set to prefer. Anyhow, it seems that some kind of optimality search within the simulation models would be the best.

The future developments within simulation modeling, will probably be in the area of estimating model parameters, perhaps by directly using the model calculations. More efficient strategies for sampling than the purely random approach, and improved search strategies are of interest, too. Finally, developments within the area of multi criteria optimization, to obtain a better reflection of the farmers utility in the object function should not be overlooked.

Chapter 10

Modeling and sensitivity analysis by Linear Programming

10.1 Introduction

10.1.1 Phases of modeling

This chapter briefly addresses the issue of modeling using Linear Programming techniques. “Linear Programming” is just one of many methods known as *Mathematical Programming*. Other examples are *Integer Programming*, *Convex Programming*, *Quadratic Programming*, *Dynamic Programming* etc. What method to use should always be determined after careful examination of the problem. Regardless of the solution method for a given operations research problem, the way to study the problem is almost always the same. The phases of the an operations research study can be summarized as in Hillier and Lieberman (1996):

1. Formulating the problem.
2. Constructing a mathematical model to represent the system under study.
3. Deriving a solution from the model.
4. Testing the model and the solution derived from it.
5. Establishing controls over the solution.
6. Testing the model and the solution derived from it.
7. Putting the solution to work: implementation.

Each of these phases will be discussed briefly in relation to linear programming in the following sections.

10.1.2 Formulating the problem

Any model regardless of complexity will never be more than an approximation of the real system. When formulating the problem, one must therefore ensure that the relevant information is determined and included in the model. This involves determining such things as the appropriate objectives, the constraints modeling the technical and biological possibilities, interrelationships between the area to be studied and other areas of the organization (farm), the possible alternative courses of action, the time window for making a decision, etc. This process is crucial to the conclusions of the study. The formulation of the problem is therefore often revisited and reconsidered as new insight on the problem is gained.

It is important to recognize that the model is a decision *support* tool. It is just as relevant to list a series of alternative almost optimal solutions as providing one certain optimal. The path from the current solution to the suggested optimal solution(s) is another interesting result that the model should provide. It is not interesting to determine an optimal situation if it is impossible (financially or technically) to get there from the current state of the system.

A key aspect of the problem formulation is the identification of appropriate objectives. These should reflect the decision makers idea of important attributes of the problem and their relative importance is reflected in the utility function of the decision maker.

10.1.3 Constructing a Mathematical Model

When the problem formulation is accomplished, the next step is to translate the problem into a form that is convenient for analysis. Usually this is done by constructing a mathematical model that represent the essence of the problem. The typical mathematical model is,

$$\begin{aligned} & \max f(x_1, x_2, \dots, x_n) \\ & \text{subject to} \\ & G(x_1, x_2, \dots, x_n) \leq b \end{aligned}$$

where x_1, x_2, \dots, x_n represent a set of *decision variables*, the function $f(\cdot)$ is the *objective function* to be maximized subject to the constraints $G(\cdot)$ being less than or equal to the *right-hand side*, b (the right-hand side can be a vector of arbitrary size). The structure of the objective function and the constraints determined the classification of the mathematical program. If both objective function and constraint functions are linear, then the model is referred to as a *linear programming model*.

Mathematical models in general have a lot of advantages over verbal descriptions of the same problem. They describe the problem more concisely and tend to make the overall structure of the problem more comprehensive. The mathematical model often reveals important cause-and-effect relationships, and most impor-

tantly: it makes an abstraction of the problem, that can be solved using efficient general mathematical programming techniques.

Since the model is an abstract idealization of the real world problem, considerable simplifications and approximations are usually needed to keep the model *tractable* (capable of being solved). This implies, that model validation is an essential part of designing the model. The model must ensure that all important courses of an action are identified in order to make a sound decision, but traits that are general for all actions should be excluded from the model, since they don't influence the decision process.

10.1.4 Deriving a solution

Once the model is specified the fun part remains. Obtaining a solution using an appropriate mathematical programming technique. The structure of the mathematical model more or less determines what kind of solution that is obtainable. A common approach is to search for an *optimal* solution, but that is not always possible, so often one will be content to find a solution that is *satisficing*, the term "satisficing" is a combination of satisfactory and optimizing, indicating the idea to search for a solution that is "good enough".

10.1.5 Testing the model and the solution

Most of the model validation is performed during the design of the model, but a lot of work remains in the *post-optimality analysis*. One aspect of the post-optimality analysis is the *sensitivity analysis*. The main purpose of this analysis is to determine which of the parameters of the model, that are the most critical (sensitive) in determining the solution. *Parametric analysis* is conducted to determine which values of the various parameters, that cause the solution to change. For linear programming a lot of standardized methods have been developed, whereas other programming techniques require a lot more research. Besides the sensitivity analysis, post-optimality analysis includes checking if the model behaves as expected, i.e. does an increase in a parameter cause the expected rise (or fall) in the value of a certain decision variable etc.

10.1.6 Establishing control over the solution

Once the model is working at a satisfactory level, procedures for working with and developing the model must be established. It is not likely that the parameters initially chosen in the model will remain constant during the period where the model works as a decision support tool. In fact is it more likely that changes occur constantly due to technological, biological and financial evolution. This requires that the model is constantly re-evaluated and alternations reflecting changes in the assumptions of the model implemented.

Table 10.1: The parameters of the linear programming model. Z is the total value of the objective function.

Resource	Resource usage per unit of activity				Amount of resource available
	1	2	...	n	
1	a_{11}	a_{12}	...	a_{1n}	b_1
2	a_{21}	a_{22}	...	a_{2n}	b_2
:	:	:	:	:	:
m	a_{m1}	a_{m2}	...	a_{mn}	b_m
ΔZ per unit of activity	c_1	c_2	...	c_n	
Level of activity	x_1	x_2	...	x_n	

10.1.7 Implementation

This part involves getting a robust and stable working model derived from the prototype used so far. For the tool to be successful in application certain requirements are needed. The manager must be able to use the tool without too much additional work and the results presented by the tool must be easy to interpret for the manager.

10.2 Using linear programming

10.2.1 The model setup

Let us turn away from the general aspects of modeling and return to the problem at hand: modeling using linear programming. The linear programming approach is especially well suited to problems that can be described as *allocating limited resources among competing activities* in the best possible (i.e. *optimal*) way. The data for the linear programming model is described in Table 10.1. If the LP-problem is written using matrix notation, the following short version can be adapted,

$$\begin{aligned} \max Z &= c'x \\ \text{s.t. } Ax &\leq b \end{aligned}$$

There is a number of assumptions that are required to hold in order to adapt linear programming to a specific problem. These assumptions will be discussed and motivated in the following sections.

10.2.2 Proportionality

The assumption of *proportionality* is about individual activities considered independently of others. Consider the situation where only one activity is chosen at a

level higher than zero. The assumption is then that the objective function and the usage of the resources are proportional to the level of the activity. This implies, that there is no setup cost associated with raising the level of the activity from zero, and that proportionality holds through-out the range of the levels of the activity. Furthermore, the assumption implies that there is a constant marginal return in the objective function.

10.2.3 Additivity

The *additivity* assumption means for each function, that the total function value can be obtained by adding the individual contributions from the respective activities. The additivity assumption is violated if the activities are complementary in a way that increase their joint profit, or if the activities are competitive in a way that decrease their joint profit. The same scenario can arise in the allocation of resources, where a setup cost in switching between products on a machine would violate the additivity assumption by including a cross-product term.

10.2.4 Divisibility

Decision variables often have physical significance only if they have integer values. Examples are number of boars, number of cows in milking and so on, whereas liters of milk produced, tons of crop harvested etc. all are easily divisible.

10.2.5 Certainty

The *certainty* assumption is that all the parameters of the model (the a_{ij} , b_i and c_j values) are known constant. In the real world these parameters would have to be estimated using available data and theory. The parameters would always be affected by some degree of *uncertainty*, reflecting the unknown future (and present) state of the system.

10.2.6 Circumventing the assumptions

If all of the four assumptions should be fulfilled in order to enable modeling by linear programming models, hardly any problem would ever be handled using this technique. Since the literature shows an extensive use of LP-models to handle almost any decision problem, some sort of relaxation of the assumptions must be possible. Please note, that the assumptions are requirements to the nature of the problem. The LP-model will fulfill these assumptions by the definition of an LP-model, so the trick is to ignore the fact that the problem doesn't. When model validation reveals that the model cannot model the essential interactions, the linearity is dropped, and a more complex model is developed. Very often the linear approximation of the real interactions prove sufficient for modeling the problem. When moving from an LP-model to more sophisticated models, the problem of additional data estimation arises. When assuming linearity and independence among

the parameters the guessing of correlations can be avoided. The worst assumption to neglect is the assumption of certainty. This assumption is *never* true, any observable parameter is stochastic. Extensive sensitivity analysis on the estimated parameters is therefore always required before accepting the results of the model.

10.2.7 Construction of constraints

The construction of the constraints to model the resource usage by the activities is a task which requires a lot of experience to be successful. In Rae (1994) the following four steps are recommended as a foolproof way of converting the verbal description into a modeling description:

1. Write down, in words, an accurate description of the relationship that is to be modeled.
2. Convert the written description to an arithmetic statement by introducing numbers and activities.
3. Re-arrange the arithmetic statement so that all activities are gathered on the left-hand side and all constants are gathered on the right-hand side.
4. Tidy up the constraints by gathering similar terms together (e.g. adding up several constraints) and ensure (if possible) that the constraint is specified in linear form.

To illustrate the concept consider the following example, which originally was given as an exercise in Hillier and Lieberman (1996).

Example 24 Identifying constraints of an LP-problem

Problem description in words: A farm family owns 125 acres of land and has \$40,000 in funds available for investment. Its members can produce a total of 3,500 person-hours of labor during the winter months (mid-September to mid-May) and 4,000 person-hours during the summer. If any of these person-hours are not needed, younger members of the family will use them to work on a neighboring farm for \$5 per hour during the winter months and \$6 per hour during the summer.

Cash income may be obtained from three crops (soybean, corn and oats) and two types of livestock: dairy cows and laying hens. No investment funds are needed for the crops. However, each cow will require an investment outlay of \$1,200, and each hen will cost \$9.

Each cow will require 1.5 acres of land, 100 person-hours of work during the winter months, and another 50 hours during the summer. Each cow will produce a net income of \$1,000 for the family. The corresponding figures for

Table 10.2: Person-hours and income associated with three crops per acre.

	Soybeans	Corn	Oats
Winter person-hours	20	35	10
Summer person hours	50	75	40
Net annual cash-income (\$)	600	900	450

each hen are: no acreage, 0.6 person-hour during the winter, and 0.3 more during the summer, and an annual net cash income of \$5. The chicken house can accommodate a maximum of 3,000 hens, and the size of the barn limits the herd to a maximum of 32 cows. Estimated person-hours and income for the three crops are as shown in Table 10.2.

The family wishes to determine how much acreage should be planted in each of the crops and how many cows and hens should be kept to maximize its net cash-income. Formulate an LP-model for this problem.

Modeling the problem: The first thing to do is to determine the activities and the resources that they share. The resources available for the activities on the farm are: funds, land, winter person-hours, summer person-hours, a barn and the chicken house. The activities that compete for these resources are: soybeans, corn, oats, cows, hens, neighboring farms in the winter, and neighboring farms in the summer.

Once the resources and activities are identified, a good approach for the final steps are formulating verbal descriptions of the allocation of resources. Examples are: *the funds invested in cows and hens cannot exceed the funds available or the number of cows cannot exceed the maximum capacity of the barn*. If NC is the number of cows and NH the number of hens, then the above examples translates into following constraints:

$$CC \times NC + CH \times NH \leq TF$$

where CC and CH are the capital cost of cows and hens respectively, and TF the total funds available. And,

$$NC \leq TC$$

where TC is the total capacity of the barn.

By continuing this process, the tableau in Table 10.3 in the form presented in Table 10.1 is derived. The symbols W_W and W_S are the hours spent working on neighboring farms winter and summer respectively, “W-Hours” and “S-Hours” are the person-hours available winter and summer. For additional examples of formulating constraints regarding activities on a farm, the reader is referred to Rae (1994, Chapter 6) ■

Table 10.3: Example 24 formulated as an LP-problem. Empty cells are zeros.

Resource	Resource usage per unit of activity						Amount of resource available
	Activity						
	Soybeans	Corn	Oats	Cows	Hens	W_W	W_S
Funds				1200	9		40,000
Land	1	1	1	1.5			125
W-Hours	20	35	10	100	0.6	1	3,500
S-Hours	50	75	40	50	0.3		4,000
Barn					1		32
C-House						1	3,000
Net income	600	900	450	1,000	5	5	6

10.2.8 Natural extensions of the LP-model

The discussion regarding the assumptions of problems solvable by LP-techniques revealed that certain natural extensions to LP-models should be available. A method to ensure integer values of certain variables must be one of the first things that comes to mind. In the example in the previous section the fact that the number of cows and hens must be integer was completely ignored. When modeling such problems, the modeler must make the decision whether or not to include constraints restricting certain values to integer. In the example, cows would logically be integer because of their substantial usage of resources. The resources used by the last half cow equals the resources used by approximately 75 hens. When ever there is a significant difference between the integer solution and the LP solution, the integer formulation should be used. By the reverse of the above argument, the hens could easily be represented by a continuous variable.

The possibility of restricting certain variables to integer values enables the modeler to include a setup cost by introducing a binary variable (a variable that only has zero and one as possible values), which in fact is a very simple integer variable.

Assume in Example 24 that the production of soybeans in the previous example requires a setup cost in terms of a special machine, that has to be rented. One way of modeling this could be to introduce a binary variable x_B . Then, let c_B the coefficient of the objective function for x_B represent the setup-cost. If NS is the acreage of soybeans, then the following constraint will ensure the inclusion of a setup-cost if production of soybeans is to occur,

$$NS \leq 125x_B \Leftrightarrow NS - 125x_B \leq 0.$$

It is obvious what happens: if $x_B = 0$, then the constraint imposes that $NS = 0$, but if $x_B = 1$, then NS can be as big as the maximum available land (125 acres). If $x_B = 1$, then the setup-cost is deducted in the objective function and otherwise not.

There are numerous algorithms available for solving these mixed-integer or integer programming problems, but the problem of finding a solution with integer values is a lot harder than the linear case, so the time it takes to find an optimal solution increase dramatically. The techniques for modeling special conditions using binary variables is described in details in Williams (1985) to which the interested reader is referred.

One of the other natural extensions is the inclusion of one or more non-linear terms. Generally it is very hard to solve these kinds of problems unless some special conditions regarding the problem are fulfilled. One of the most common demands is that the area formed by the constraints is *convex* (a set is said to be convex, if a straight line between any two points in the set lies within the set itself), but there exist other special cases for which efficient solution procedures are developed. In general, however, it is not likely that even reasonable sized models can be solved in an acceptable time.

10.3 Sensitivity analysis

The discussion in this section is best done using an example and illustrating the concepts in the context of the example. Consider the LP-model:

$$\begin{aligned}
 & \max \quad x_1 + x_2 \\
 \text{s.t.} \quad & 2x_1 + 3x_2 \leq 37 \\
 & 2x_1 + x_2 \leq 23 \\
 & x_1 \leq 10 \\
 & x_2 \leq 9 \\
 & x_1 \geq 0 \\
 & x_2 \geq 0
 \end{aligned} \tag{10.1}$$

This problem is easily solved graphically as in Figure 10.1 by shifting the lines for equal values of the objective function until it only touches the possible region in one point (8,7).

Using this example, shadow prices, price adjustments, capacity adjustments and introduction of additional constraints will be discussed.

10.3.1 Shadow prices

Hillier and Lieberman (1996) define the term *shadow price* as follows:

Definition 8 The *shadow price* (or *dual price*) for a resource measures the *marginal value* of the resource, that is, the rate at which the objective function Z could be increased by (slightly) increasing the amount of the resource being made available.

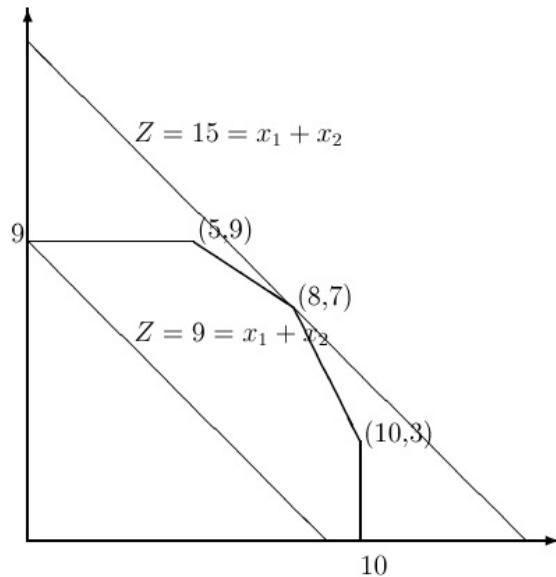


Figure 10.1: Solving the linear program graphically by drawing the constraints defining the possible area and shifting the objective function upwards until the line only touches the area of possibilities in one point (8,7).

■

Please note, that following only is true for sufficiently small increments of the resource. The optimal point must still be defined by the same set of constraints. A popular interpretation of shadow prices, is that it is the price that the producer is willing to pay for one additional amount of a given resource, if the producer were to pay more, then the increase in profit would be smaller than the price paid for the extra capacity.

This gives some information about the shadow prices in problem (10.1). Since the two first constraints define the optimal solution point, then the shadow prices for these must be positive, because an increase in either resource would increase the profit. On other hand the shadow prices for the four last constraints (including the non-negativity constraints) must by the same argument be equal to zero, because an increase in any of the resources would have no effect on the optimal solution.

If the LP-problem is solved using the simplex method, the shadow prices would be easy to read in the final tableau, but when the problem is solved graphically the easiest way to get the shadow price is to increase the right-hand side of the constraint by one and re-optimize. Doing this reveals that the shadow price for the second constraint is $1/4$. The shadow prices are important in the post-analysis of the problem. Assume that the resource land was a limiting factor for a livestock production. If the shadow price of land in the livestock problem is higher than the

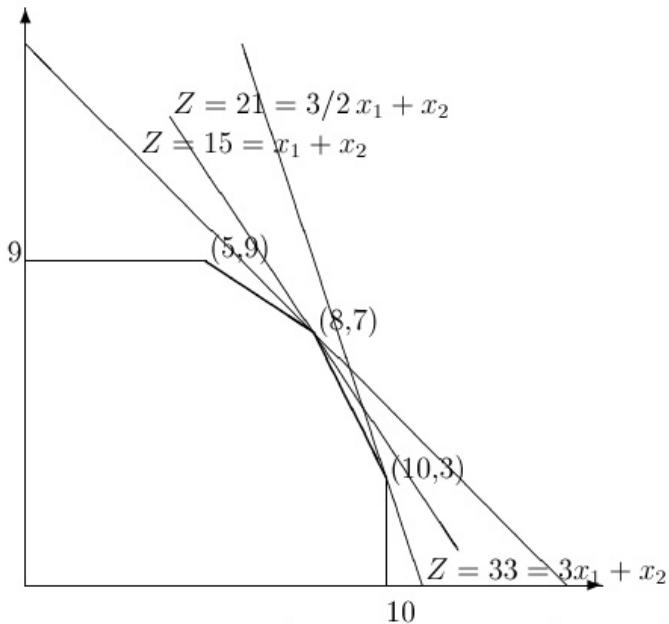


Figure 10.2: The change in price of x_1 causes the slope of the profit line to change, but only the change from $p_1 = 1$ to $p_1 = 3$ causes the optimal production mix to change from the old optimum $(8,7)$ to $(10,3)$.

expected return for the crop currently occupying some of the farmers land, then the production plan should be altered to allow for land utilized by livestock.

10.3.2 Price adjustments

Assume that the price of x_1 suddenly increases due to a change in the demand. The question is: does this call for modifications of the optimal combination of x_1 and x_2 ? The answer is maybe! It depends on how much the price increases. This is illustrated in Figure 10.2. It is seen that an increase from $p_1 = 1$ to $p_1 = 1\frac{1}{2}$ cause no changes, but an increase to $p_1 = 3$ makes the optimal solution change from $(8,7)$ to $(10,3)$.

Figure 10.2 shows that the slope of the profit line becomes steeper and steeper as the price of x_1 increases. The value of p_1 that cause the optimum to shift from $(8,7)$ to $(10,3)$ is the value that makes the slope of the profit line equal the slope of the constraint $2x_1 + x_2 \leq 23$. The same result makes the optimum change from $(8,7)$ to $(5,9)$ when the price of x_1 is decreased enough to make the slope of the profit line equal that of the constraint $2x_1 + 3x_2 \leq 37$. This means, that as long as the price of x_1 is within the interval $2/3$ to 2 (and the price of x_2 is fixed) the optimum will remain unchanged. The corresponding number for changes in the price of x_2 is $1/1$ to $3/2$ for p_1 fixed.

Analysis like this is important to estimate the stability of the solution. If minor changes in certain parameters would cause the optimum to shift from one point to another, then fluctuations in these parameters should be monitored more intensively in order to make the necessary adjustments.

10.3.3 Capacity changes

When changing the capacity of a given resource the same questions as in the previous section arise. Will this affect the current optimal solution? Again, the answer depends on how much the capacity is changed. From our discussion of shadow prices it is clear, that an increase in the capacity of a resource that is currently not a limiting factor will have no effect on the current solution. This situation is hardly of any interest in practical applications, but the reverse situation is very interesting: how much can I reduce this factor, and still keep the current solution? The answer is: as long as the resource is not a limiting factor, i.e. when the current solution gets infeasible if the resource is reduced further.

If (x_1^*, x_2^*) is the optimal solution, then the surplus capacity of resource i is calculated as:

$$b_i - (a_{i1}x_1^* + a_{i2}x_2^*)$$

i.e. the right-hand side minus the allocated activities.

If the surplus capacity is zero, then the resource is a limiting factor, i.e. the constraint helps define the solution. Any changes in such a factor would cause the optimal production to change, but the question is then how much can a limiting resource be increased (or decreased) without changing the set of limiting resources.

In this graphic approach this is done much like finding the optimum solution by shifting the profit line. The constraint is simply shifted parallel upwards and downwards until other constraints become limiting factors. This is done in Figure 10.3 where the result is that the capacity of the constraint $2x_1 + 3x_2 \leq 37$ can be between 29 and 39 without changing the set of limiting factors.

10.3.4 Additional constraints

When adding extra constraints to an LP-problem, two different scenarios occur: either the constraint is fulfilled by the current optimal solution, hence this is still the optimal solution, or the current optimal solution is no longer a feasible solution (i.e. the point where the solution was obtained is violating the new constraint). The way to deal with additional constraints in the graphical solution procedure is simply to add the constraint by drawing the corresponding line and then see whether or not the solution is still optimal. The two different situations are shown in Figure 10.4.

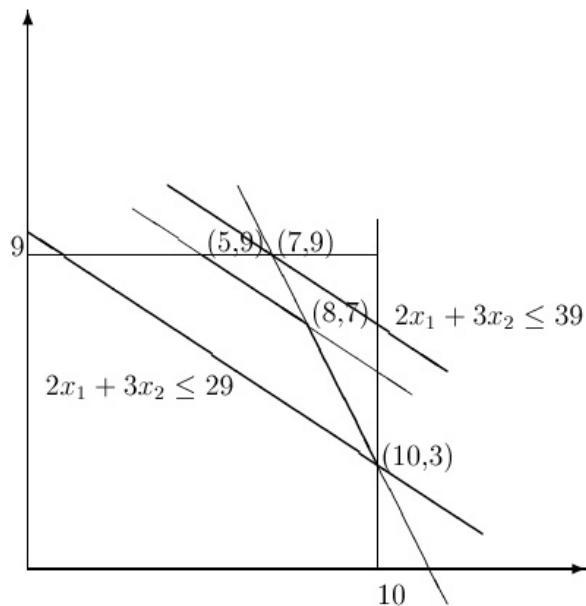


Figure 10.3: The change in capacity cause the shifts from the optimal solution (8, 7) to (7, 9) for capacity above 39 or (10,3) for capacities below 29.

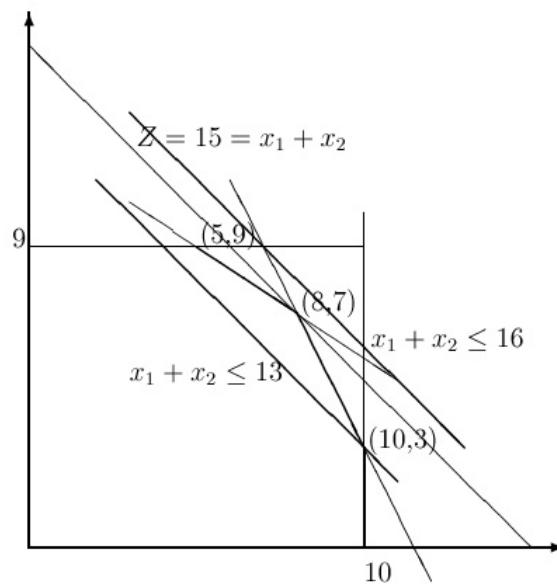


Figure 10.4: Adding the constraint $x_1 + x_2 \leq 16$ has no effect on the optimal solution, but adding $x_1 + x_2 \leq 13$ makes the current optimum infeasible and the new optimum is (10,3).

10.4 Concluding remarks

The art of designing LP-models as decision support tools has been applied to a lot of different problems in the field of herd management. Classic examples include nutrient management (Henry et al., 1995), ration formulation (Gonzales-Alcorta et al., 1994) and selection of sire portfolios (Nash and Rogers, 1995). These are only a very few of the examples found in literature.

The most important feature of any decision support system is a solid model validation. The extensive theory regarding sensitivity analysis enables a methodic validation of LP-models. An implication of this result is that building an initial model using LP-techniques will help the modeler to understand the basic interactions of the involved resources and decisions.

Because of the wide range of applications several possibilities for modeling and solving LP-models exist. Any modern spreadsheet will provide a method for formulating and solving LP-models of reasonable size. The LINDO software package is a tool for solving LP-problems, it is especially well suited for use from within a user interface. The probably most powerful mathematical modeling software available is the GAMS package, which is a combination of modeling language and a series of solvers capable of linear programming, mixed-integer programming, non-linear programming and much more. It is recommended to use GAMS if the model will be large because of the high quality solvers available to GAMS. For initial trials a spreadsheet will prove sufficient.

10.5 Exercise

To check the understanding of the discussed topics consider the following small exercise, which essentially originates from Dent et al. (1986): Robert McCull owns 200 ha of irrigated pasture, for which carrying capacities have been estimated to be as follows:

	Breeding ewe flock (ewes per ha)	Spring/summer fattening of cattle (steers per ha)
Spring	15	5
Summer	24	4
Autumn	15	0
Winter	8	0

The expected gross margins are \$24 per ewe and \$60 per steer. Mr McCull wishes to maximize the total farm gross margin.

1. Determine graphically the optimal production program for Mr McCull.
2. Assuming the sheep gross margin remains unchanged at \$24 per ewe, by how much must the gross margin for steers decline before steer numbers should be reduced? Explain your reasoning.

3. Assuming the original gross margin for ewe remains unchanged at \$24 per ewe, to what must the steer gross margin increase before an increased cattle number becomes profitable?
4. Assuming the original gross margin for sheep and cattle of \$24 and \$60, respectively, calculate the shadow price of spring feed.
5. Mr McCull is considering increasing his winter production by the use of artificial nitrogen fertilizer. He expects that the winter carrying capacity will increase by 0.1 ewes per hectare for each kilogram of nitrogen applied. Assuming that nitrogen costs \$90 per kg, would you recommend that he apply nitrogen, and if so, how much? What would your recommendations be if nitrogen were to cost only \$0.2 per kg.

Chapter 11

Bayesian networks

Using Bayesian Networks one can easily build a probabilistic expert system incorporating knowledge of the domain. It consists of two parts: a Directed Acyclic Graph (DAG) specifying the probabilistic interdependencies between the random variables and a set of conditional probability tables specifying the strength of the dependencies. Decision graphs discussed in Chapter 12 add decisions- and utility nodes to the concept.

BNs and DGs give a consistent framework for handling reasoning and decision making under uncertainty. An example could be pregnancy testing of sows. Here fusion of different measurement methods as well as allowing for specification of sensitivity and specificity in connection with the methods is necessary in order to get an overall result.

Since, a chapter on Bayesian networks has not been written yet, the readers are encouraged to study an excellent textbook by Jensen (2001), where in particular Chapters 1 and 2 provide a good introduction to the concept.

For good examples, reference is made to McKendrick et al. (2000b) and Otto and Kristensen (2004).

Chapter 12

Decision graphs: Potential use and current limitations

12.1 Introduction

The purpose of this chapter is to illustrate the potential use of the decision graph technique within herd management. Decision graphs were (under the name of *Influence Diagrams* introduced by Howard and Matheson (1981) as a formalism to model decision problems with uncertainty for a single decision maker. The influence diagrams gave a more compact graphical representation of a decision problem than the more traditional decision tree approach. In Shachter (1986) a method was suggested for solving the decision problem represented by the influence diagram directly, without the translation to a decision tree. This method transformed the influence diagram by successively removing nodes in the graph, until at last only one final utility node, holding the utility of the optimal policy. In order to solve many similar problems one therefore had to start from scratch every time. Shenoy (1992) proposed another algorithm that gave the solution to the influence diagram without disrupting the structure of the diagram. Then Jensen et al. (1994) showed how a similar approach could be incorporated within the general framework of Bayesian Networks. This approach has been implemented in the Hugin expert system shell. We refer to Jensen et al. (1994) for a presentation of the method. In addition, we refer to the short introduction given in Section 9.6.

As concerns software for decision graphs, reference is made to the Hugin¹ system implementing classical influence diagrams under the “no forgetting” assumption by use of the strong junction tree approach (Jensen et al., 1994) and the Esthauge LIMID Software System² implementing Limited Memory Influence Diagrams as described by Lauritzen and Nilsson (2001).

¹<http://www.hugin.com>

²<http://www.esthauge.dk>

12.2 From decision tree to influence diagram

The decision tree is a method for decision analysis. An excellent general description can be found in Hillier and Lieberman (1996). The concept is also illustrated in Example 5 of Chapter 4. In order to further illustrate the method we will elaborate on the pregnancy diagnoses decision tree in Figure 9.1 of Chapter 9.

The sequential decisions start from the left of the diagram. A decision node (a square) represents the decision to Make Pregnancy Diagnosis (MPD). From the decision node two branches originate, “Yes” and “No”. If we follow the “yes branch”, we encounter a chance node (a circle), the outcome of the pregnancy diagnosis, i.e. either “yes” (positive diagnosis) or “no” (negative). After the chance node, another decision node is encountered, “Replace” (RP), with “yes” and “no” as actions. If we decide to keep the animal, the final chance node is encountered, “Farrowing” (FA) with possible outcomes, “yes” or “no”. If we initially decide not to make pregnancy diagnosis, we have to wait and see what the outcome is at farrowing time.

At each branching in the nodes, we can assign a utility and a probability. If the pregnancy diagnosis is made, the cost is the additional work. If the animal is replaced, the income from the slaughter value of the animal plus the expected value of a new animal is received. If we decide to let the animals farrow, we add the feeding cost until farrowing, and the future value of the present animal. If it farrows, we obtain in addition the income from the litter produced. The probability of each branching in the chance nodes can be found similarly. If we make a pregnancy diagnosis the probability of the two outcomes depends on the Pregnancy State (PS) of the animal and the precision of the measurement method, but by applying Bayes formula, the probabilities can be found. (Note that the state of nature, PS, is not represented in the diagram). Similarly the probability of farrowing given PS and PD can be found and assigned to each branch leading from FA. To solve the decision tree diagram, one starts from the outmost branches (the leaves). In a chance node we calculate the expected utility. I.e. the expected utility in the FA node is the probability of farrowing times the income from a litter plus the probability of not farrowing times the income from not producing a litter (0). At a decision node we use maximizing instead of taking the expectation. I.e. we make the decision that have the highest expected utility. Gradually we move from the leaves to the stem/root of the decision tree.

As a real tree, the decision tree very soon becomes confusing to look at, i.e. the idea behind the graphic representation is soon lost. Furthermore, we had to do some calculations outside the tree, because some of the chance nodes were not represented (i.e. the pregnancy state). In addition some chance node are represented at several places in the diagram (e.g. the FA node)

The Influence diagram solved these problems to a large degree, when they were introduced by Howard and Matheson (1981). The use of the circle as a symbol for a chance node, and the square for a decision node was maintained, but in addition the diamond was added to represent the utility in the diagram. The branches

that represented the outcomes of decisions/chances, were removed and replaced by arrows that showed what influenced the decisions and chance nodes. The utility representation of the pregnancy diagnosis problem is also shown in Figure 9.1. The chance node PS (Pregnancy State) is added. The MPD node influences the node UD (Utility Diagnosis), and the outcome of the pregnancy diagnosis, and the next decision RP. The outcome of the pregnancy diagnosis is in addition influenced by the PS node. Note that PS influences FA directly disregarding the outcome of PD. The utility from replacement (UR) and from farrowing (UF) finishes the picture. (Strictly speaking, in the original approach only one mutual utility node was present).

Originally the influence diagrams were automatically translated to decision trees and solved using the same algorithm as mentioned above. Shachter (1986) showed how the influence diagram could be solved directly by removing nodes subsequently, based on 4 different transactions, the arc reversal (application of Bayes Theorem), node removal by summing out, expectation of a value node with respect to a chance node, and finally removal of a decision node into a value node by maximization. The first attempts were based on discrete valued state variables, but Shachter and Kenley (1989) implemented influence diagrams for Gaussian (normal) distributed variables.

12.3 From Bayesian network to influence diagram

If we look at the graphic representation of the decision graphs and compare with the Bayesian network, we find that they have the same nodes, i.e. the chance nodes, represented as circles. The two additional elements in decision graphs, the decision variables (squares) and utility functions (diamonds), can in fact also be handled within the framework of Bayesian networks by putting some restrictions on the solution method. This was first shown in Jensen et al. (1994). In fact, if the subsequent decisions are not sequential (i.e. the result of a previous decision does not influence other decisions) they can be directly represented in the Bayesian network by using the calculation trick, shown in Jensen (1996, Chapter 6). In realistic decision problems this is seldom the case.

The exact procedure for solving influence diagrams as it has been implemented in HUGIN, can be found in Jensen et al. (1994). Here we will summarize the method based on the diagram in Figure 9.1. An important difference between Bayesian networks and influence diagrams is the dynamic aspect, i.e. the sequential decisions impose an ordering on the elements. First of all, the decisions are ordered according to the sequence in which they are performed, ie. $U_D = \{MPD, RP\}$ in this case. The ordering is secured by putting directed edges between the decision nodes on the graph. Second, the chance nodes are put into sequential groups as well, i.e $U_R = \{I_0, I_1, I_2\}$, where I_0 is the set of observations before the first decision (in this case it is empty), I_1 is the set of observations between the first and the second decision, i.e. $I_1 = \{PD\}$, and finally the nodes that

are observed after the second decision or never observed, i.e. $I_2 = \{PS, FA\}$. We thus obtain a partial ordering,

$$\emptyset \prec \{MPD\} \prec \{PD\} \prec \{RP\} \prec \{PS, FA\}.$$

In Bayesian networks, the approach, as described in Jensen (1996, Section 4.5), would be to make a junction tree from the graph, by moralizing and triangulating the graph. In this case, however, a so-called strong junction tree is made. This means that the ordering, we have made, should be preserved in the junction tree, implying that the variables at the right should be eliminated before the variables on the left.

This ordering, which is a consequence of the “no forgetting” assumption discussed in Section 9.6, means that the computational complexity of influence diagrams with decision nodes in general is much larger than in Bayesian networks without.

As described in Jensen et al. (1994) we can then use the junction tree for propagation of evidence by simply interchanging maximization for calculation of expectation.

One note of caution is that the ordering should be preserved in the subsequent calculations made by in the network. In standard Hugin, evidence can be entered anywhere in the tree, and the evidence is distributed properly. This is not possible with influence diagrams. For example, it does not make sense to study the effect of a known pregnancy state, because evidence of pregnancy state is defined only to be available after all decisions have been made. Similarly the effect of known outcome of pregnancy diagnosis can only be studied for given level of the decision MPD. Unfortunately, the program does not have a facility that checks that evidence is input in the correct sequence. However, if the decisions are assigned fixed values the evidence can be entered.

The rather new concept of *LImited Memory Influence Diagrams* described by Lauritzen and Nilsson (2001) relaxes the “no forgetting” assumption as mentioned in Section 9.6. The solution algorithm for this class of models is called *Single Policy Updating*, and all calculations are done in an ordinary junction tree just like those used with Bayesian networks (i.e. a *strong* junction tree is not needed). From a computational point of view, LIMIDs are therefore very appealing, but on the cost of exactness. In many cases, Single Policy Updating only leads to near-optimal solutions, even though they in most cases are very close to the optimum.

12.4 From Dynamic Programming to influence diagrams

The central paper in this respect is Tatman and Shachter (1990). Until the publication of their paper, influence diagrams were constructed with only one utility node. They showed that if the utility function was separable, i.e. either additive or multiplicative, the influence diagram technique corresponded to the Dynamic Pro-

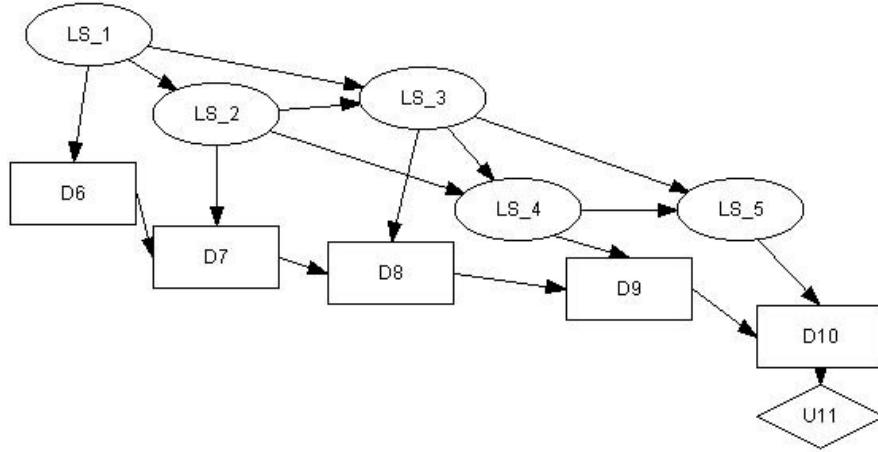


Figure 12.1: Influence Diagram representing part of the replacement problem for sows. LS_i is litter size in parity i, Dx represents decisions and Ux utility.

gramming method, i.e. Dynamic Programming is a special case of the influence diagram method.

Because of the more flexible approach towards representing the causal relations, the use of influence diagrams could often result in computational savings, by reducing the complexity of the problems.

To illustrate the potential of their techniques, Tatman and Shachter (1990) showed some examples. Their example (B) has close correspondence to the techniques used in solving the replacement problem in animal production, as exemplified by Huirne et al. (1991). Therefore, their example (B) is adapted to the replacement problem. The core in the replacement model by Huirne et al. (1991) is a model for the relation between litter size in subsequent parities. This can be illustrated as in Figure 12.1. (Note that the decision and utility part is not adequately represented in the figure).

The litter size is observed at subsequent parities and the decision based on the observation. The litter size is influenced by the litter size in the two previous parities as indicated by the arrows. Each litter size node has some 20 different levels (0, . . . , 19).

To solve this in standard dynamic programming, Huirne et al. (1991) used state augmentation, i.e. they made a state variable consisting of (LS_i, LS_i + 1), as illustrated in the next figure, i.e. the state variable had $20^2 = 400$ states. This corresponds to the influence diagram in Figure 12.2.

The resulting transition matrix consisted of $400 \times 400 = 160,000$ elements. If the influence diagram shown in Figure 12.1 was used directly, the problem would only consist of matrices of $20 \times 400 = 8000$ elements. If information from further previous litters were included, the difference would be even more pronounced. It

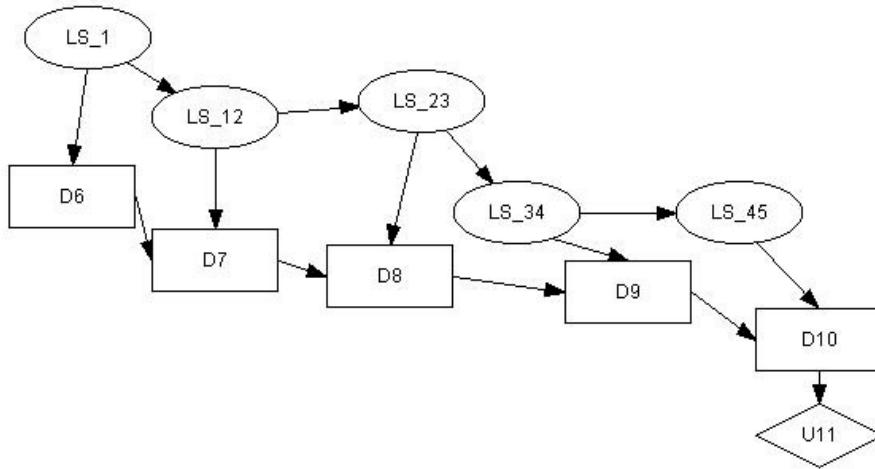


Figure 12.2: Representation of the state augmentation, LS_{ij} is the combined state variable of litter size in parity i and j .

is worth noting that this reduction in dimensionality is unrelated to the reductions due to hierachic Markov processes (cf. Section 13.3.2). Similar savings in state space is expected in representing other independent components of the state space, e.g. state for pregnancy ability (number of re-matings).

However, in the current framework of decision graphs, only problems with finite planning horizon can be solved, as every time step has to be represented in the diagram. Furthermore, the use of semi-Markov processes rather than Markov processes, with the resulting flexibility in using different time scales has not been implemented either. Finally, the advantages of the hierachic Markov process, where in fact a decision can select a process with a different causal structure is not yet possible.

12.5 Examples

The following is a presentation of examples of the possibility for using influence diagrams. The examples have not been constructed in detail but the overall structure of the decision problem should be clear, and quantification of the elements in the example should be possible. The intention is to stimulate ideas for application of the technique within animal production.

12.5.1 The two-sow problem

We shall once again return to the “two-sow problem” introduced in Example 20 of Chapter 6. As the reader will probably remember, a pig producer has bought two sows with the same expected farrowing date. One of the sows has to be culled because there is only room for one sow in the farrowing department. The litter size

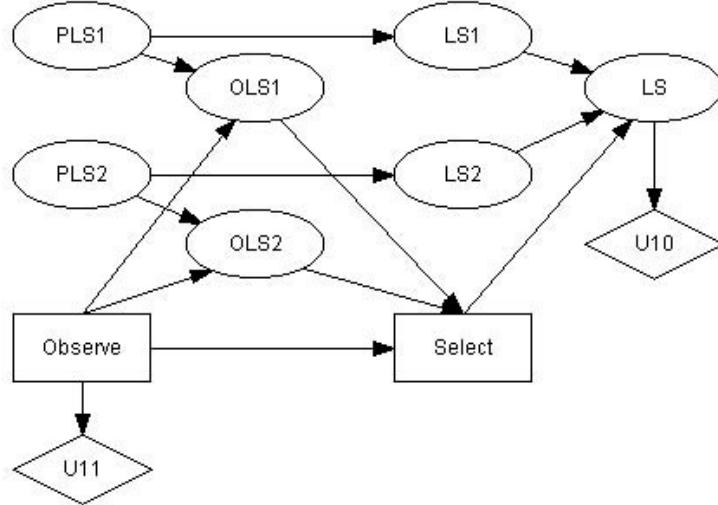


Figure 12.3: Influence diagram representation of the two sow problem.

of a sow is correlated with the litter size in her previous litter. Information about the litter size in the previous litter can be bought from the farmer, who sold the two sows. How much should he be willing to pay for the information, and which sow should he keep?

The example can be formulated as an influence diagram as shown in Figure 12.3. The litter size LS_i of sow i depends on the previous litter size $PLSi$. We can decide to observe i.e. buy the information concerning the previous litter size, and obtain Observed Litter Size (OLS_i). Based on the observed value we can decide which of the sows we should keep, and the resulting litter size is kept in the node LS .

12.5.2 The registration problem for the whole cycle

In fact the registration problem does not only concern the previous litter size, but also the use of heat-detection, pregnancy diagnosis etc. We can easily add nodes for Heat detection, Pregnancy diagnosis and the corresponding decision nodes and utility nodes to the diagram as shown in Figure 12.4 for the one sow case.

The decision nodes are UHD (use heat detection), UPD (use pregnancy diagnosis), and ULS (use litter size information). To each decision a utility node has been assigned. An additional chance node has been assigned, PS, representing pregnancy state. The network in 12.4 can be seen as a combination of Figure 9.1 and 12.3.

In Figure 12.5 the decision example is specified for the two sow problem. In addition nodes for herd level of pregnancy rate (HPR), heat detection rate (HDR) and efficiency of pregnancy diagnosis (Herdetect) has been added. These nodes will represent the uncertainty in the parameters used in the calculation. In addition

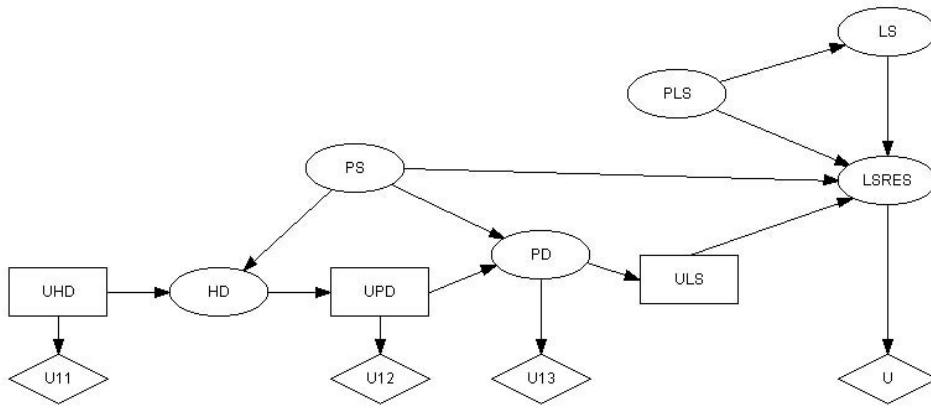


Figure 12.4: Influence diagram representing the decisions in a whole gestation period for a single sow.

the nodes will make the system “learn” from the observations made. A trick is used to ease the specification of the conditional probabilities. The nodes HD_i , PDi and $PLSi$ represent the outcome of the observations. Additional nodes XXO_i represent the information that decisions are based upon, e.g. if a pregnancy diagnosis is performed the states of PDO will be identical to the state of PD . If observations are not made, the state “Not observed” of PDO will have probability 1, i.e. the variable contains no information about the true state.

The network in Figure 12.5 is an example of a network that should be used cautiously when entering evidence. The herd pregnancy rate is one of the nodes, where it would be tempting to ask: What if the pregnancy rate was 0.75, what would the optimal decision be. This is not allowed, because herd pregnancy rate is defined as an unobservable node, i.e. it does not influence any decision.

Note also that the complexity of Figure 12.5. Even with as few as two animals in the network, the decision graph is not readily understood. Of course the problem can be generalized to three sows, but now the diagram becomes almost impossible to understand. A method exists for easy representation of replicated processes, the so-called frames where a rectangular box around part of the network signify that this part of the network is to be replicated. This approach is eg. used in the specification in BUGS³ (Gilks et al., 1993). This to some extent solves the problem of too complicated networks, and is especially suited for automatic learning within the network (Buntine, 1994). Such a representation for the network is shown in 12.6. Unfortunately this technique cannot yet be used within Hugin.

³The BUGS programme and manual can be found by visiting the web-adress with URL: <http://www.mrc-bsu.cam.ac.uk/bugs/>

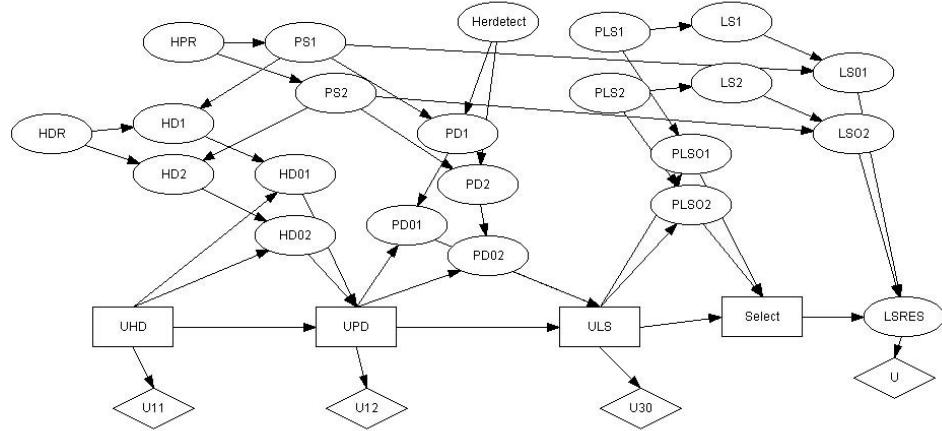


Figure 12.5: Influence diagram representing the decisions within gestation for two sows.

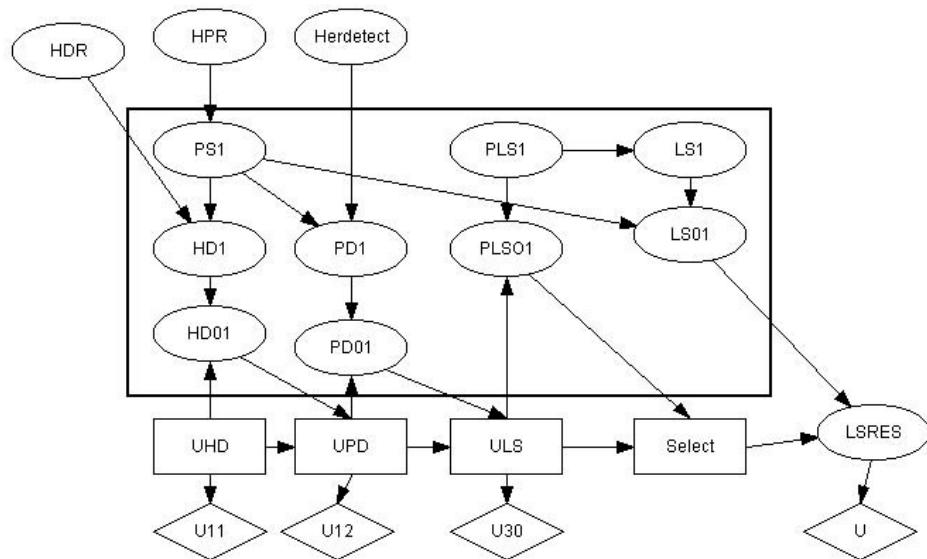


Figure 12.6: Frame representation of the general N-sow problem. Nodes within the frame are replicated automatically. (not yet possible within Hugin).

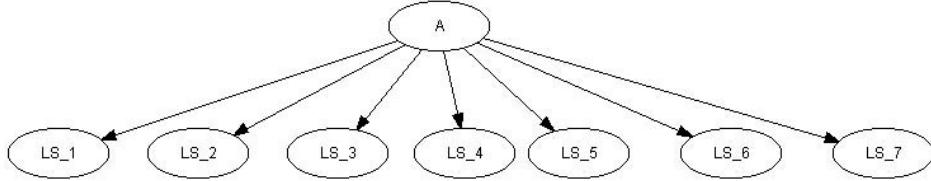


Figure 12.7: Representing model for litter size with additive effect of sow, A .

Another aspect is the computational complexity. Already with a few sows the problem becomes difficult to handle on a standard PC. As the complexity increases exponentially with the number of animals, clearly any computer will be overburdened if realistic problems are to be solved. Therefore, approximate methods need to be implemented within the diagram. In a similar framework, considering optimal number of sows mated, Greve (1995) implemented a heuristic algorithm and was able to solve problems of realistic size, i.e. herd sizes in excess of 500 sows.

12.5.3 The repeated measurement problem

The dynamic programming model by Huirne et al. (1991) assumed only effect of the two previous litters as described in Section 12.4. A more natural approach would be to assume an additive effect specific for each sow, i.e. a network like Figure 12.7.

The network corresponds to a statistical model represented as:

$$LS_{ij} = \mu_i + A_j + \varepsilon_{ij}$$

with μ_i as the mean of parity i , $A_j \sim \mathcal{N}(0, \sigma_A^2)$ as the effect of sow j and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ as an independent and random residual.

Using ordinary Bayesian networks such a system can be handled very effectively, but not when influence diagrams are used. This is because of the strong junction tree mentioned earlier. The junction tree without decisions consists of cliques with $\{A, LS_p\}$, and is thus of a moderate size. However as A is an unobservable effect, it has to be included in the last set in the elimination ordering, when the influence diagram is compiled. That is, the triangulation should be made so that A can be eliminated before any of the other nodes. This can only be attained if links are added between litter size in every parity, and then we are back where we started. The problem shown in Figure 12.7 has a clique in the strong junction tree of size $O(N^7)$, where N is the number of state levels in each litter size node. The

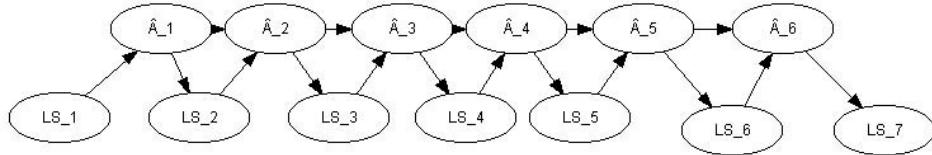


Figure 12.8: Efficient representation of the additive litter size model. A replaced by the observable estimate \hat{A}_p . Decisions are based on \hat{A}_p .

same problem arises if any other of the current algorithms for solution in influence diagrams is used.

Fortunately, this problem can be circumvented. The trick is to use the underlying model to determine observable, \hat{A}_{ij} values, i.e. estimate of the sow effect based on results from her previous litters, so that $LS_{i+1,j} = \mu_{i+1} + \beta(\hat{A}_{ij} - \mu_i) + \varepsilon$. In the case of the model specified, this can easily be found from the multivariate normal distribution. Then we obtain a network like Figure 12.8 instead. Note that the LS nodes are influenced by the estimate. From a causal point of view this of course nonsense, but it provides the correct answer and reduces the complexity drastically.

In 12.8 we have explicitly specified the learning or estimation process based on the observations, and based the decision on the learning. As described in Jensen (1996), Bayesian networks are an efficient method for a decision maker to learn posterior distributions from observed data. Therefore it seems a bit surprising that at the current state of development, when we specify the decision makers subsequent decisions within the network, we have to presume that he does not use this efficient learning method, but only bases his decisions on observed values. The alternative is to specify the learning algorithm directly, as in 12.8.

The figure corresponds to a decision maker that uses a Bayesian network like 12.7, as a DSS-tool and base his decisions on the mode of the probability distribution of A after evidence from previous litters has been added to the network.

Further illustration of the possibility for using this updating technique in connection with decision support can be found in Kristensen (1993) and Jørgensen (1992). For examples of applications, reference is made to Kristensen and Søllested (2004a,b); Lien et al. (2003).

12.5.4 Optimal timing of matings

After weaning most sows show oestrus/heat within 4 to 6 days. Based on the observation, the manager decides to mate the sow. In order to maximize the resulting conception rate and litter size the general advice is to repeat matings with 12 hours interval up to a maximum of three times, if the sow continues to show oestrus. The

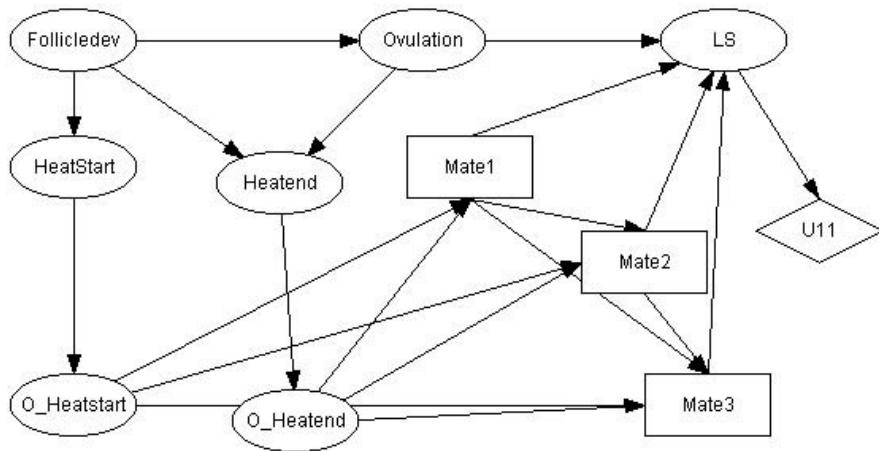


Figure 12.9: Influence diagram representing the problem of optimal timing of matings.

reason for this is that fresh semen (< 24h) should be available around the newly released ovum to increase fertilization success.

Ovulation can, in general, not be observed (see Soede, 1993, for methods for experimental methods for observing ovulation) and the pig producer has to rely on the observable signs of heat. Time of ovulation and start of heat are both depending on the hormonal development that can be summarized as follicular development (it is a complex interaction between the hormones LH, FSH and Oestrogen). In other words they are conditionally independent given follicle development. The end of oestrus is determined by follicular development and possibly a feedback mechanism from ovulation. An influence diagram with the relevant mechanism is shown in Figure 12.9.

Start of heat (Heatstart) and end of heat (Heatend) is checked at regular intervals, e.g. three times a day. The strength of the oestrous signs develops almost following a gaussian curve. Experienced observers will detect weaker oestrous signs than inexperienced, that is, the heat will be observed ($O_{heatstart}$) earlier on average. Based on the observed oestrous sign, the pig producer shall decide the timing and frequency of matings (Mate1, Mate2 and Mate3). Depending on the timing of matings and the unobserved ovulation time (ovulation) the litter size (LS) will vary. In figure 12.9 the utility nodes are omitted from the mating decisions. Further refinements of the diagram would be to include decisions concerning oestrus detection schedule and connected costs.

12.5.5 The feed analysis problem

This example is modified from Pedersen (1996). In dairy production a major part of the feed mix consists of locally produced roughage, such as beets, silage and

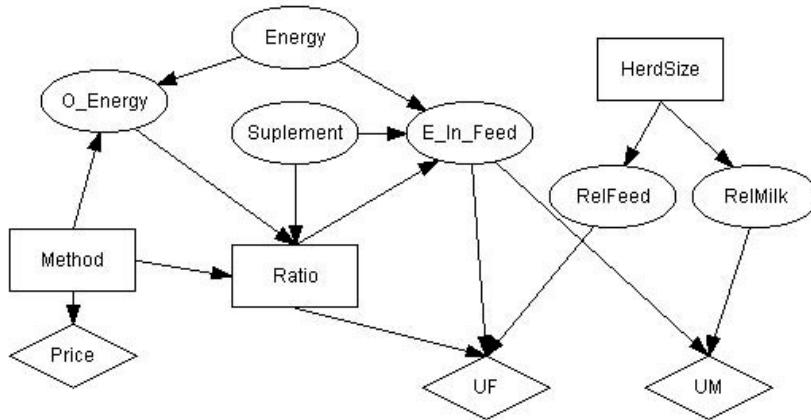


Figure 12.10: Influence diagram representing the problem of optimal timing of matings.

straw. Knowledge of the energy contents of the feed is used in e.g. feed budgeting and feed planning. The energy content of roughage can be estimated using several methods. The simplest is a table look-up for standard figures for energy content in different categories of roughage. Another simple approach is to estimate dry matter content and digestibility subjectively, based on experience concerning the look and feel of the roughage. A more precise evaluation requires a laboratory assessment of the value. The digestibility can be measured using *in vivo* techniques that are the most precise. Alternatively, different *in vitro* techniques that simulate the *in vivo* techniques can be used to a much reduced price. Other aspects of the feed evaluation can be measured using different techniques, such as NIR.

The dairy farmer thus has to decide, how he should analyze the feed. Before this question can be answered, we have to consider the decisions that the feed analysis influences. In Figure 12.10 this is illustrated.

The roughage has a certain true and unobservable energy content (Energy). The farmer can decide to observe the energy level (*O_Energy*) using different methods (Method). These methods differ in price. Together with the observable energy level in the roughage and the energy level in the supplement feed (Suplement) the farmer decides the mixing ratio between these two feed stuffs. The resulting energy in the feed (*E_In_Feed*) is found based on the mixing ratio and the true energy levels in the two feed stuffs. The energy level in the feed will in turn influence the feed intake and the feed costs (UF). In addition the milk production level will be influenced (UM). As the relative price of the feed analysis increases with decreasing herd size, because the information and the feed mixture are used on fewer animals, the utility should be corrected to take this into account. The two nodes Relfeed and Relmilk serve this function.

12.6 Outlook

When we judge the possible uses of the decision graphs, we should consider, that it is a very recent technique, and thus still have some shortcomings.

When we look upon the technique as a generalization of the Markov Chain process for Dynamic Programming, the more static approach used in influence diagrams is clearly a constraint. Easy specification of a dynamic system, such as used in dHugin (Kjærulff, 1995) would clearly be a step forward and should make it possible to utilize techniques comparable to policy iteration. To the authors' knowledge, this has not been developed yet. The flexible time scales, based on discounting the utilities, would represent an even further improvement.

The frame approach used in BUGS, see footnote on page 192, would ease the specification of multi-component systems. It may also be used to represent decision that chooses between qualitatively different subsystems, similar to Hierachic Markov processes (HMP). The subsystem could be represented as different frames, and the decisions a matter of selecting the frame.

Such developments would make the influence diagram method able to solve similar problems as is currently solved by HMP. In addition, savings in complexity similar to those described in section 12.4 would be obtained.

The spinoff from the statistical research in Bayesian networks is expected to result in numerical methods for handling very large systems, as illustrated by the early attempts by Bielza et al. (1999) and Charnes and Shenoy (1996) to use Monte Carlo techniques to optimize the decision strategies. The savings in complexity by basing decisions on the learning process in Bayesian networks illustrated in section 12.5.3 is another obvious improvement.

Chapter 13

Dynamic programming and Markov decision processes

13.1 Introduction

13.1.1 Historical development

In the late fifties Bellman (1957) published a book entitled *Dynamic Programming*. In the book he presented the theory of a new numerical method for the solution of *sequential decision problems*. The basic elements of the method are the *Bellman principle of optimality* and *functional equations*. The idea may be illustrated as follows.

Consider a system being observed over a finite or infinite time horizon split up into periods or *stages*. At each stage, the *state* of the system is observed, and a *decision* (or an *action*) concerning the system has to be made. The decision influences (deterministically or stochastically) the state to be observed at the next stage, and depending on the state and the decision made, an immediate *reward* is gained. The expected total rewards from the present stage until the end of the planning horizon is expressed by a *value function*. The relation between the value function at the present stage and the one at the following stage is expressed by the *functional equation*. Optimal decisions depending on stage and state are determined backwards step by step as those maximizing the right hand side of the functional equation. This way of determining an optimal *policy* is based on the Bellman principle of optimality which says: *An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision* (Bellman, 1957, p. 83).

During the following years, Bellman published several books on the subject (Bellman, 1961; Bellman and Dreyfus, 1962; Bellman and Kalaba, 1965). The books were very enthusiastic, and the method was expected to be the solution to a very wide range of decision problems of the real world. The expectations were

so great, and they were adduced with such a conviction, that Johnston (1965) ironically compared dynamic programming to a new religion. Others regarded the method as a rather trivial computational device.

Similar stories might be told regarding other new numerical methods, as for instance linear programming. However, after some years, the applicational scopes of the methods are encircled. Most often the conclusion is that the method is neither an all-embracing technique nor a triviality. Between these extremities a rather narrow range of problems remains where it is a powerful tool. Other problems are, in some cases, not suitable to be solved by the method. In other cases alternative methods are better.

This also turned out to be the case in dynamic programming. One of the basic elements of dynamic programming is the sequential approach, which means that it fits sequential decision problems best. Obvious examples of sequential decisions in animal production include replacement of animals (it is relevant to consider at regular time intervals whether the present asset should be replaced or it should be kept for an additional period), insemination and medical treatment. Dynamic programming is a relevant tool, but if the traits of the animal are well defined and their precise behavior over time is known in advance, there are other methods that might be applied to determine the optimal decisions analytically. On the other hand, if the traits of the animal are affected by *random* variation over time and among animals, the decisions will depend on the present observations of the traits. In that case dynamic programming is an obvious technique to be used in the determination of optimal decisions and policies.

Having identified dynamic programming as a relevant method to be used with sequential decision problems in animal production, we shall continue on the historical development. Howard (1960) published a book on *Dynamic Programming and Markov Processes*. As will appear from the title, the idea of the book was to combine the dynamic programming technique with the mathematically well established notion of a *Markov chain*. A natural consequence of the combination was to use the term *Markov decision process* to describe the notion. Howard (1960) also contributed to the solution of infinite stage problems, where the *policy iteration method* was created as an alternative to the stepwise backward contraction method, which Howard called *value iteration*. The policy iteration was a result of the application of the Markov chain environment and it was an important contribution to the development of optimization techniques.

The policy iteration technique was developed for two criteria of optimality, namely maximization of total expected *discounted* rewards and maximization of expected *average* rewards per stage. Later on, Jewell (1963) presented a policy iteration technique for the maximization of average rewards over time for *semi-Markov decision processes*, which are Markov decision processes of which the stage length is a random variable. Howard (1971) presented a value iteration method for semi-Markov decision processes.

For the sake of completeness it should also be mentioned that *linear programming* was early identified as an optimization technique to be applied to Markov

decision processes as described by, for instance, Hadley (1964), but no animal production models known to the authors have applied that technique. This is in accordance with a conclusion of White and White (1989) that policy iteration (except in special cases) is more efficient than linear programming.

Since the publication of the first mentioned book by Howard (1960) an intensive research in Markov decision programming has been carried out. Many results have been achieved concerning the relations between the various optimization techniques and criteria of optimality. Reviews of these developments are given by van der Wal and Wessels (1985) as well as White and White (1989).

13.1.2 Applications in animal production

The dominant area of application in animal production has been for solving the animal replacement problem either alone or in connection with insemination and medical treatment. It is however expected that recent methodological developments will broaden the applicational scope.

Already three years after the book by Howard (1960), an application to the dairy cow replacement problem was published by Jenkins and Halter (1963). Their model only included the trait "lactation number" (at 12 levels), and the permanent value of the study was only to illustrate that Markov decision programming is a possible tool to be applied to the problem. A few years later, however, Giaevers (1966) published a study which represents a turning-point in the application of the method to the animal (dairy cow) replacement problem. He considered all three optimization techniques (value iteration, policy iteration and linear programming), described how to ensure that all mathematical conditions were satisfied, and presented an eminent model to describe the production and feed intake of a dairy cow. The work by Giaevers (1966) has not got the credit in literature that it deserves (maybe because it is only available on microfilm). In a review by van Arendonk (1984) it is not even mentioned.

During the following 20 years, several dairy cow replacement models using Markov decision programming were published, but from a *methodological* point of view none of them have contributed anything new compared to Giaevers (1966). Several studies, however, have contributed in *other* ways. Smith (1971) showed that the rather small model of Giaevers (1966) with 106 states did not represent the upper limit. His state space included more than 15 000 states. Kristensen and Østergaard (1982) as well as van Arendonk (1985, 1986) and van Arendonk and Dijkhuizen (1985) studied the influence of prices and other conditions on the optimal replacement policy. Other studies (Killen and Kearney, 1978; Reenberg, 1979) hardly reached the level of Jenkins and Halter (1963).

Even though the sow replacement problem is almost identical to that of dairy cows, very few *early* studies on sows have been published. The only exceptions known to the authors are Huirne et al. (1988, 1991, 1993) and Jørgensen (1992).

A study by Ben-Ari et al. (1983) deserves special attention. As regards methodology it is not remarkable, but in that study the main difficulties concerning appli-

cation to animal production models were identified and clearly formulated. Three features were mentioned:

Uniformity: The traits of an animal are difficult to define and measure. Furthermore the random variation of each trait is relatively large.

Reproductive cycle: The production of an animal is cyclic. It has to be decided *in which* cycle to replace as well as *when* to replace inside a cycle.

Availability: Only a limited supply of replacements (in that case heifers) is available.

The first feature in fact covers two different aspects, namely *uniformity* because the traits are difficult to define and measure, and *variability* because the traits vary at random among animals and over time. The third feature is an example of a *herd restraint*, i.e. a restriction that applies to the herd as a whole and not to the individual animal. Other examples of herd restraints are a production quota or a limited housing capacity. We shall therefore consider the more general problem of herd restraints.

We may conclude that until the mid-eighties, the methodological level concerning the application of Markov decision programming to animal production models was represented by Giaever (1966). The main difficulties that the method should overcome had been identified by Ben-Ari et al. (1983). If we compare the approach of Giaever (1966) to the difficulties that it ought to solve, we may conclude that the problems related to *variability* are directly solved, and as it has been shown by Kristensen and Østergaard (1982) as well as van Arendonk (1985); van Arendonk and Dijkhuizen (1985); van Arendonk (1986), the problems concerning the *cyclic production* may readily be solved without any methodological considerations. The only problem concerning variability and cyclic production is that in order to cover the variability, the state variables (traits) have to be represented by many levels, and to deal with the cyclic production a state variable representing the stage of the cycle has to be included. Both aspects contributes significantly to an explosive growth of the state space. We therefore face a *dimensionality* problem. Even though all necessary conditions of a Markov decision process are met, the solution in practice is prohibitive even on modern computers. The problems concerning uniformity and herd restraints are *not* solved by the approach of Giaever (1966).

13.2 Variability and cyclic production: Markov decision programming

As mentioned in the introduction, Markov decision programming is directly able to take the variability in traits and the cyclic production into account without any adaptations. In order to have a frame of reference, we shall briefly present the theory of traditional Markov decision programming originally described by Howard (1960).



Figure 13.1: Stages of a Markov decision process.

13.2.1 Notation and terminology

Consider a discrete time Markov decision process with a finite *state* space $U = \{1, 2, \dots, u\}$ and a finite *action* set D . A *policy*, s is a map assigning to each state $i \in U$ an action $s(i) \in D$. Let p_{ij}^d be the *transition probability* from state i to state j if the action $d \in D$ is taken. The *reward* to be gained when the state i is observed, and the action d is taken, is denoted as r_i^d . The time interval between two transitions is called a *stage*.

We have now defined the elements of a traditional Markov decision process, but in some models we further assume that if state i is observed, and action d is taken, a physical quantity of m_i^d is involved (e.g. Kristensen, 1989, 1991). In this study we shall refer to m_i^d as the *physical output*. If $s(i) = d$, the symbols r_i^d , m_i^d and p_{ij}^d are also written as r_i^s , m_i^s and p_{ij}^s , respectively.

An *optimal* policy is defined as a policy that maximizes (or minimizes) some predefined objective function. The optimization technique (i.e. the method to identify an optimal policy) depends on the form of the objective function or - in other words - on the criterion of optimality. The over-all objective to maximize net revenue of the entire herd may (depending on the circumstances) result in different criteria of optimality formulated as alternative objective functions. The choice of criterion depends on whether the planning horizon is finite or infinite.

13.2.2 A simple dairy cow replacement model

For any dairy cow it is relevant to consider at regular time intervals whether it should be kept for an additional period or it should be replaced by a heifer. If the line of Figure 13.1 represents time, the markers indicate where we consider to replace. The time interval between two markers is called a stage and in this example we assume the stage length to be one year which for convenience is assumed always to be equal to a lactation period. At the beginning of each stage, we observe the state of the animal in production. The state space must be defined in such way that all relevant information is given by the state. In this very simple example we assume, that the only relevant information is whether the cow is low, average or high yielding. Thus we have got one state variable (milk yield) and three states.

Table 13.1: Rewards (gross margins), r_i^d , depending on state, i , and action, d .

State	$d = 1$ (Keep)	$d = 2$ (Replace)
$i = 1$ (low yielding)	10,000 DKK	9,000 DKK
$i = 2$ (average yielding)	12,000 DKK	11,000 DKK
$i = 3$ (high yielding)	14,000 DKK	13,000 DKK

 Table 13.2: Physical outputs (expected milk yields), m_i^d , depending on state, i , and action, d .

State	$d = 1$ (Keep)	$d = 2$ (Replace)
$i = 1$ (low yielding)	5,000 kg	5,000 kg
$i = 2$ (average yielding)	6,000 kg	6,000 kg
$i = 3$ (high yielding)	7,000 kg	7,000 kg

Having observed the state, we have to take an action concerning the cow. We assume that the action is either to keep the cow for at least an additional stage or to replace it by a heifer at the end of the stage.

The economic net returns (gross margin) from the cow will of course depend on whether it is low yielding or high yielding **and** whether it is kept or replaced. In the model this is represented by a reward depending on state and action as appearing in Table 13.1. Those amounts are simply the annual net returns from a low, average and high yielding cow respectively. If the action replace is taken, we assume that the replacement takes place at the *end* of the stage at a cost of 1,000 DKK.

In this example, we shall define the milk yield during a stage as the physical output. In Table 13.2, the expected milk yield depending on state and action is shown. Since replacement is assumed to take place at the end of a stage, the milk yield will not depend on the action.

If a cow has been low yielding during a stage, there is large risk, that it will also be low yielding during the following stage if it is kept. This is illustrated by the transition probabilities from state i at a stage to state j at the following stage. We assume that the probability to remain at the same level of milk yield is 0.6. The probability of transition to an other level is assumed to be 0.3 from low (or high) to average and 0.1 from low to high or vice versa, if the cow is kept. On the other hand, if it is replaced, we assume that there are equal probabilities of the new heifer to be low, average or high yielding. All transition probabilities are shown in Table 13.3.

All parameters of a traditional Markov decision process are now defined, and we may consider what policy to follow. A policy is defined as a map (rule) that tells us which action to take if a certain state is observed. An example of a logical policy in this very simple case would be to replace if the cow is low yielding and keep if it is average or high yielding.

Our problem is now to determine an *optimal* policy, which in some sense maximizes the net returns of the dairy farmer. In the following section the phrase

Table 13.3: Transition probabilities from state i at a stage to state j at the following stage.

State at present stage	$d = 1$ (Keep)			$d = 2$ (Replace)		
	$j = 1$ (L)	$j = 2$ (A)	$j = 3$ (H)	$j = 1$ (L)	$j = 2$ (A)	$j = 3$ (H)
	0.6	0.3	0.1	0.333	0.333	0.333
$i = 2$ (A)	0.2	0.6	0.2	0.333	0.333	0.333
$i = 3$ (H)	0.1	0.3	0.6	0.333	0.333	0.333

“in some sense maximizes” is clarified. Afterwards, a survey of optimization techniques is given. Throughout the discussion we shall refer to this numerical example as *the simple dairy model*.

13.2.3 Criteria of optimality

Finite planning horizon

A farmer, who knows that he is going to terminate his production after N stages, may use the maximization of total expected rewards as his criterion of optimality. The corresponding objective function h is

$$h(s^1, \dots, s^N) = E \left(\sum_{n=1}^N r_{I(n)}^{s^n} \right), \quad (13.1)$$

where E denotes the expected value, s^n is the policy at stage n , and $I(n)$ is the (unknown) state at stage n . Applying this criterion to the simple dairy model means that the total expected net returns over a fixed number (N) of years are maximized.

If the farmer has a time preference, so that he prefers an immediate reward to an identical reward later on, a better criterion is the maximization of total expected discounted rewards. If all stages are of equal length, this is equal to applying the objective function

$$h(s^1, \dots, s^N) = E \left(\sum_{n=1}^N \beta^{n-1} r_{I(n)}^{s^n} \right), \quad (13.2)$$

where $0 < \beta < 1$ is the discount factor defined by the interest rate and the stage length.

Infinite planning horizon

A situation where the stage of termination is unknown (but at least far ahead) is usually modeled by an infinite planning horizon (i.e. $N = \infty$). In that case the optimal policy is constant over stages. The function 13.1 cannot be applied in this

situation, but since $\beta < 1$, the function 13.2 will converge towards a fixed value for N becoming very large. Thus the objective function is given by

$$h(s) = E \left(\sum_{n=1}^{\infty} \beta^{n-1} r_{I(n)}^s \right). \quad (13.3)$$

Since, usually, each animal and its future successors are represented by a separate Markov decision process, this criterion together with the criterion 13.2, are equal to the maximization of total discounted net revenues *per animal*. Such a criterion is relevant in a situation where a limiting housing capacity is the only (or at least the *most* limiting) herd restraint.

An alternative criterion under infinite planning horizon is the maximization of expected average reward per unit of time. If all stages are of equal length, the objective function in this situation is

$$h(s) = g^s = \sum_{i=1}^u \pi_i^s r_i^s, \quad (13.4)$$

where π_i^s is the limiting state probability under the policy s (i.e. when the policy is kept constant over an infinite number of stages). This criterion maximizes the average net revenues per stage, i.e. *over time*. In the simple dairy model, application of this criterion indicates that average annual net returns are maximized. It may be relevant under the same conditions as criterion 13.3 if an animal and its future successors are represented by a separate Markov decision process. Practical experience shows that the optimal replacement policies determined under criteria (13.3) and (13.4) are almost identical.

If a herd restraint (e.g. a milk quota) is imposed on the physical output, a relevant criterion may be the maximization of expected average reward per unit of physical output using the objective function

$$h(s) = g^s = \frac{g_r^s}{g_m^s} = \frac{\sum_{i=1}^u \pi_i^s r_i^s}{\sum_{i=1}^u \pi_i^s m_i^s}. \quad (13.5)$$

In case of a milk quota, the physical output m_i^s is the milk produced by a cow in state i under policy s , and accordingly, average net returns per kg milk is maximized in the simple dairy model. The function 13.5 is also relevant if the criterion is the maximization of the expected average reward over time in a model where the stage length varies. In that case the physical output represents the stage length. It should be noticed that if $m_i^d = 1$ for all i and d , the function 13.5 is identical to 13.4. The symbol g_r^s is the average reward over stages (equal to g^s of Eq. (13.4)) and g_m^s is the average physical output over stages.

13.2.4 Optimization techniques in general Markov decision programming

Value iteration

Under finite planning horizon the *value iteration method* is excellent. The optimal policies are determined sequentially using the functional equations

$$f_i(n) = \max_d \left\{ r_i^d + \beta \sum_{j=1}^u p_{ij}^d f_j(n-1) \right\}, \quad i = 1, \dots, u, \quad (13.6)$$

where the action d maximizing the right hand side is optimal for state i at the stage in question. The function $f_i(n)$ is the total expected discounted rewards from the process when it starts from state i and will operate for n stages before termination. Thus $f_i(0)$ is the salvage value of the system when it is in state i . At each stage an optimal policy is chosen using Eqs. (13.6). If the objective function (13.1) is used, $\beta = 1$ in Eq. (13.6). Otherwise β is the discount factor.

Under infinite planning horizon, the value iteration method may be used to approximate an optimal policy. Under the objective function 13.3 it is possible to show that (Howard, 1960)

$$\lim_{n \rightarrow \infty} f_i(n) = f_i, \quad i = 1, \dots, u, \quad (13.7)$$

where f_i for fixed i is a constant. By using Eqs. (13.6) over a large number of stages, we will sooner or later observe that $f_i(n+1)$ is almost equal to $f_i(n)$ for all i . Further, we will observe that the same policy is chosen during several stages. We can feel rather sure that such a policy is close to be optimal, but there is no guarantee that it is identical to an optimal policy. For practical purposes, however, the approximation usually suffices.

Since the objective function (13.4) is just a special case of function (13.5), where $m_i^s = 1$ for all i and d , we shall only consider the criterion given by (13.5). In this case $f_i(n)$ is the total expected rewards when the process starts from the beginning of a stage in state i and will operate *until n units of physical output have been produced*. Under the criterion given by the objective function (13.4), the production of n units of output is just the operation of the process over n stages. It is assumed that the physical output only takes integer values (for practical purpose this is just a question of selecting an appropriate unit). According to Howard (1971) an optimal policy for producing n units of output (i.e. a policy that maximizes the expected reward of producing n units) is determined recursively by the relations ($i = 1, \dots, u$):

$$f_i(n) = \max_d \left\{ a \left(\frac{nr_i^d}{m_i^d} + f_i(0) \right) + (1-a) \left(r_i^d + \sum_{j=1}^u p_{ij}^d f_j(n-m_i^d) \right) \right\}, \\ n = 1, \dots$$

where

$$a = \begin{cases} 1, & m_i^d \geq n \\ 0, & m_i^d < n \end{cases} \quad (13.8)$$

This is under the assumption that the reward/output rate has the constant value of r_i^d/m_i^d during the entire stage. However, since the physical output is bounded, it is easily seen that for n sufficiently large, $a = 0$. Hence we get for $i = 1, \dots, u$

$$f_i(n) = \max_d \left\{ r_i^d + \sum_{j=1}^u p_{ij}^d f_j(n-m_i^d) \right\}, \quad \text{large } n. \quad (13.9)$$

Thus in the long run, the assumption concerning constant reward/output rate in all states will have no effect. The equivalence of Eq. (13.7) is in this case

$$\lim_{n \rightarrow \infty} (f_i(n) - f_i(n-1)) = g, \quad (13.10)$$

and sooner or later the policy will not differ from step to step of Eqs. (13.9).

Further details on the value iteration method are given by Howard (1960, 1971). It should particularly be noticed that m_i^d , which in this study is interpreted as a physical output (e.g. milk yield), in the study by Howard (1971) is interpreted as the expected *stage length* when state i is observed under the action d . Thus, in his model the criterion (13.5) is the expected average reward over time. Compared to Eq. (13.9), Howard (1971) described a more general case where the stage length is a random variable of which the distribution is given by the action and the present state as well as the state to be observed at the next stage. Furthermore, the reward depends on the state combination, the action and the stage length. The interpretation as physical output has been discussed by Kristensen (1991).

The value iteration method is identical to what is usually referred to as *dynamic programming, successive iteration* or *successive approximation*.

The simple dairy model optimized by value iteration

If we assume the discount factor, β , to be 0.85 and the salvage value of the system to be zero (independently of terminal state and action), we may directly apply the value iteration method as described by Eqs. (13.6). In Table 13.4, the results are shown stage by stage backwards from $n = 1$ to $n = 64$ (some stages omitted).

For any time horizon n , we may read the optimal policy directly from the table as the decisions d_1 for low yielding cows, d_2 for cows of average milk yield and

Table 13.4: The value iteration method of Eqs. (13.6) applied to the simple dairy model.

n	d_1	$f_1(n)$	$\Delta f_1(n)^*$	d_2	$f_2(n)$	$\Delta f_2(n)^*$	d_3	$f_3(n)$	$\Delta f_3(n)^*$
1	1	10000	10000	1	12000	12000	1	14000	14000
2	1	19350	9350	1	22200	10200	1	25050	11050
3	2	27870	8520	1	30870	8670	1	34081	9031
4	2	35299	7429	1	38275	7405	1	41622	7541
5	2	41639	6340	1	44597	6322	1	47988	6366
6	2	47030	5391	1	49981	5384	1	53385	5397
7	2	51612	4582	1	54561	4580	1	57969	4584
8	2	55507	3895	1	58455	3894	1	61864	3895
9	2	58817	3310	1	61765	3310	1	65175	3311
10	2	61631	2814	1	64579	2814	1	67989	2814
20	2	74437	554	1	77385	554	1	80795	554
30	2	76958	109	1	79906	109	1	83316	109
40	2	77455	21	1	80402	21	1	83812	21
50	2	77552	4	1	80500	4	1	83910	4
60	2	77572	1	1	80519	1	1	83929	1
61	2	77572	1	1	80520	1	1	83930	1
62	2	77573	1	1	80520	1	1	83930	1
63	2	77573	1	1	80521	1	1	83931	1
64	2	77574	0	1	80521	0	1	83931	0

$$^* \Delta f_i(n) = f_i(n) - f_i(n-1)$$

210 13.2 Variability and cyclic production: Markov decision programming

Table 13.5: The value iteration method of Eqs. (13.9) applied to the simple dairy model.

n	d_1	$f_1(n)$	$\Delta f_1(n)^*$	d_2	$f_2(n)$	$\Delta f_2(n)^*$	d_3	$f_3(n)$	$\Delta f_3(n)^*$
1	1	10000	10000	1	12000	12000	1	14000	14000
2	1	21000	11000	1	24000	12000	1	27000	13000
3	2	33000	12000	1	36000	12000	1	39500	12500
4	2	45167	12167	1	48100	12100	1	51800	12300
5	2	57355	12189	1	60253	12153	1	64027	12227
6	2	69545	12190	1	72428	12175	1	76228	12201
7	2	81734	12189	1	84612	12183	1	88420	12192
8	2	93922	12188	1	96798	12186	1	100609	12189
9	2	106109	12188	1	108985	12187	1	112797	12188
10	2	118297	12188	1	121172	12187	1	124984	12188
11	2	130484	12187	1	133359	12187	1	137172	12188
12	2	142672	12187	1	145547	12187	1	149359	12188
13	2	154859	12187	1	157734	12187	1	161547	12187
14	2	167047	12187	1	169922	12187	1	173734	12187
15	2	179234	12187	1	182109	12187	1	185922	12187

$$^*\Delta f_i(n) = f_i(n) - f_i(n-1)$$

d_3 for high yielding cows. If, for instance, we assume a time horizon of $n = 10$ stages, we see that low yielding cows should be replaced whereas average and high yielding cows should be kept. We also observe, that for values of n higher than 3, the optimal policy does not vary over stages.

The columns $f_i(n)$ show the expected present value of the chain (i.e. the present cow and its future successors), when the present cow is in state i and n stages remain in the time horizon. From the table we observe, that the present values of the three states converges towards fixed values in accordance with Eq. (13.7). Those values appear to be (close to) 77,574 DKK, 80,521 DKK and 83,931 DKK for low, average and high yielding cows respectively. The differences between these figures represent the relative values of the three states. For instance, $80,521 - 77,574 = 2,947$ is the economic advantage of having a cow of average milk yield instead of a low yielding cow.

In Table 13.5, the corresponding results for the value iteration method under the criterion (13.4) are shown. The optimal policies appear to be exactly identical to those of Table 13.4. Under this criterion, however, the value functions $f_i(n)$ does *not* converge towards fixed values, because no discounting is involved. Instead, we see from Table 13.5 that the *increments*, $\Delta f_i(n) = f_i(n) - f_i(n-1)$, in accordance with Eq. (13.10), converge towards a fixed value of 12,187 DKK independently of state. Thus, 12,187 is the numerical value of g , the annual net returns being maximized under this criterion.

It is also possible to calculate the economic benefit of, for instance, state 2 (average milk yield) over state 1 (low milk yield) from Table 13.5. If we use the

Table 13.6: Equations and expressions to be used in the policy iteration cycle with different objective functions.

Obj. func.	Linear equations of Step 2			Expression Step 3
	Equations, $i = 1, \dots, u$	Unknowns	Add. eq.	
(13.3)	$f_i^s = r_i^s + \beta \sum_j p_{ij}^s f_j^s$	f_1^s, \dots, f_u^s	-	$r_i^d + \beta \sum_j p_{ij}^d f_j^s$
(13.4)	$g^s + f_i^s = r_i^s + \sum_j p_{ij}^s f_j^s$	g^s, f_1^s, \dots, f_u^s	$f_u^s = 0$	$r_i^d + \sum_j p_{ij}^d f_j^s$
(13.5)	$g^s m_i^s + f_i^s = r_i^s + \sum_j p_{ij}^s f_j^s$	g^s, f_1^s, \dots, f_u^s	$f_u^s = 0$	$r_i^d - g^s m_i^d + \sum_j p_{ij}^d f_j^s$

figures relating to $n = 15$, the benefit mentioned is $(182, 109 - 15 \times 12, 187) - (179, 234 - 15 \times 12, 187) = -696 - (-3, 571) = 2,875$. This figure is very close to the corresponding value (2,947) calculated from Table 13.4. This observation confirms the remark in relation to Criterion (13.4), that in practice results from this criterion only slightly differ from those under Criterion (13.3).

Policy iteration

Under infinite planning horizon, the *policy iteration method* may be applied. Unlike the value iteration method it always provides an optimal policy. It covers all three objective functions (13.3), (13.4) and (13.5). The iteration cycle used for optimization has the following steps:

1. Choose an arbitrary policy s . Go to 2.
2. Solve the set of linear simultaneous equations appearing in Table 13.6. Go to 3.
3. For each state, i , find the action d' that maximizes the expression given in Table 13.6, and put $s'(i) = d'$. If $s' = s$, then stop, since an optimal policy is found. Otherwise redefine s according to the new policy (i.e. put $s = s'$ and go back to 2.

From the equations and expressions of Table 13.6, we see that also with the policy iteration method, the objective function (13.4) is just a special case of (13.5), where $m_i^s = 1$ for all i and s . For the objective functions (13.3) and (13.4) the policy iteration method was developed by Howard (1960), and for the function (13.5) a policy iteration method was presented by Jewell (1963). Like Howard (1971), Jewell (1963) interpreted m_i^d as the expected stage length.

Under Criterion (13.3), f_i^s is the total present value of the expected future rewards of a process starting in state i and running over an infinite number of stages following the constant policy s . Under Criterions (13.4) and (13.5), f_i^s is the *relative value* of state i under the policy s . The difference in relative values between two states equals the amount of money a rational person is just willing to pay in order to start in the highest ranking of the two states instead of the lowest ranking.

212 13.2 Variability and cyclic production: Markov decision programming

The absolute value of f_i^s is determined arbitrarily by the additional equation of Table 13.6, where the relative value of state u is defined to be zero. The interpretation of relative values is discussed in details by Kristensen (1991).

Linear programming

Under an infinite planning horizon, linear programming is a possible optimization technique. When the criterion (13.3) is applied, the linear programming problem becomes (Ross, 1970)

$$\begin{aligned} \sum_{i=1}^u x_i &= \text{Max!} \\ &\text{subject to} \\ x_i - \beta \sum_{j=1}^u p_{ij}^d x_j &\geq r_i^d, \quad \text{all } d \in D, \quad i = 1, \dots, u. \end{aligned} \tag{13.11}$$

It appears from 13.11 that each combination of state and action is represented by exactly one restriction. An action d is optimal in state i if, and only if, the corresponding restriction is satisfied as an equation when the values of x_1, \dots, x_u arises from an optimal solution to the linear programming problem. The *optimal* values of x_1, \dots, x_u are equal to the present values f_1^s, \dots, f_u^s under an optimal policy.

If the objective function (13.4) is applied, the linear programming problem becomes

$$\begin{aligned} \sum_{i=1}^u \sum_{d \in D} r_i^d x_i^d &= \text{Max!} \\ &\text{subject to} \\ \sum_{d \in D} x_i^d - \sum_{j=1}^u \sum_{d \in D} p_{ij}^d x_j^d &= 0, \quad i = 1, \dots, u \\ \sum_{i=1}^u \sum_{d \in D} x_i^d &= 1 \\ x_i^d &\geq 0, \quad d \in D, \quad i = 1, \dots, u. \end{aligned} \tag{13.12}$$

In this case an action d is optimal in state i if, and only if, x_i^d from the optimal solution is strictly positive. The optimal value of the objective function is equal to the average rewards per stage under an optimal policy. The optimal value of $\sum_{d \in D} x_i^d$ is equal to the limiting state probability π_i under an optimal policy.

Using Criterion (13.5), we may solve the following linear programming problem (Kennedy, 1986):

$$\begin{aligned}
 x_u &= \text{Max!} \\
 &\text{subject to} \\
 -x_i + \sum_{j=1}^{u-1} p_{ij}^d x_j - m_i^d x_u &\leq -r_i^d - p_{iu}^d a, \quad d \in D, \quad i = 1, \dots, u-1 \\
 \sum_{j=1}^{u-1} p_{uj}^d x_j - m_u^d x_u &\leq -r_u^d - p_{uu}^d a + a, \quad d \in D \\
 x_i &\geq 0, \quad i = 1, \dots, u,
 \end{aligned} \tag{13.13}$$

where a is a pre-determined relative value of state u chosen sufficiently large to ensure that all other relative values are positive. The optimal value of the objective function of the linear programming problem is equal to the expected average reward per unit of output as defined in Eq. (13.5) under an optimal policy. The optimal values of the variables x_1, \dots, x_{u-1} are equal to the relative values of the states $1, \dots, u-1$, provided that the relative value of state u is equal to a . As it appears, each combination of state and action is represented by one, and only one, restriction. An action is optimal in a state if, and only if, the corresponding restriction is satisfied as an equation in the optimal solution.

Since Criterion (13.4) is just a special case of (13.5) with all physical outputs set to the value 1, the linear programming problem (13.13) may also be used in the determination of an optimal policy under Criterion (13.4).

13.2.5 Discussion and applications

Under finite planning horizon, the value iteration method is perfect, but in herd management models the planning horizon is rarely well defined. Most often the process is assumed to operate over an unknown period of time with no pre-determined stage of termination. In such cases the abstraction of an infinite planning horizon seems more relevant. Therefore we shall pay specific attention to the optimization problem under the criteria (13.3), (13.4) and (13.5) where all three techniques described in the previous sections are available.

The value iteration method is not exact, and the convergence is rather slow. On the other hand, the mathematical formulation is very simple, and the method makes it possible to handle very large models with thousands of states. Further it is possible to let the reward and/or the physical output depend on the stage number in some pre-defined way. This has been mentioned by van Arendonk (1984) as an advantage in modeling genetic improvement over time. The method has been used in a lot of dairy cow replacement models as an approximation to the infinite stage optimum. Thus it has been used by Jenkins and Halter (1963); Giaever (1966); Smith (1971); McArthur (1973); Stewart et al. (1977, 1978); Killen and Kearney

(1978); Ben-Ari et al. (1983); van Arendonk (1985, 1986); van Arendonk and Dijkhuizen (1985); DeLorenzo et al. (1992); McCullough and DeLorenzo (1996a,b). Some of the models mentioned have been very large. For instance, the model by van Arendonk and Dijkhuizen (1985) contained 174 000 states (reported by van Arendonk, 1988). In sows, the method has been used by Huirne et al. (1988).

The policy iteration method has almost exactly the opposite characteristics of the value iteration method. Because of the more complicated mathematical formulation involving solution of large systems of simultaneous linear equations, the method can only handle rather small models with, say, a few hundred states. The solution of the linear equations implies the inversion of a matrix of the size $u \times u$, which is rather complicated. On the other hand, the method is exact and very efficient in the sense of fast convergence. The rewards are not allowed to depend on the stage except for a fixed rate of annual increase (e.g. inflation) or decrease. However, a seasonal variation in rewards or physical outputs is easily modeled by including a state variable describing season (each state is usually defined by the value of a number of state variables describing the system).

An advantage of the policy iteration method is that the equations in Table 13.6 are *general*. Under any policy, s , we are able to calculate directly the economic consequences of following the policy by solution of the equations. This makes it possible to compare the economic consequences of various non-optimal policies to those of the optimal. Further we may use the equations belonging to the criterion (13.5) to calculate the long run technical results under a given policy by redefining r_i^s and m_i^s . If for instance $r_i^s = 1$ if a calving takes place and zero otherwise, and m_i^s is the stage length when state i is observed under policy s , then g^s , which is the average number of calvings per cow per year, may be determined from the equations. Further examples are discussed by Kristensen (1991). For an example where the equations are used for calculation of the economic value of culling information, reference is made to Kristensen and Thysen (1991).

The policy iteration method has been used by Reenberg (1979) and Kristensen and Østergaard (1982). The models were very small, containing only 9 and 177 states, respectively.

13.3 The curse of dimensionality: Hierarchical Markov processes

In order to combine the computational advantages of the value iteration method with the exactness and efficiency of the policy iteration method Kristensen (1988, 1991) introduced a new notion of a hierarchical Markov process. It is a contribution to the solution of the problem referred to as the “curse of dimensionality” since it makes it possible to give exact solutions to models with even very large state spaces.

Table 13.7: Rewards (annual net returns) and outputs (annual milk yield) of extended model.

i	Lactation	Yield	Rewards, r_i^d		Outputs, m_i^d	
			$d = 1$ (K)	$d = 2$ (R)	$d = 1$ (K)	$d = 2$ (R)
1	1	Low	8,000	7,000	4,000	4,000
2	1	Ave.	10,000	9,000	5,000	5,000
3	1	High	12,000	11,000	6,000	6,000
4	2	Low	10,000	9,000	5,000	5,000
5	2	Ave.	12,000	11,000	6,000	6,000
6	2	High	14,000	13,000	7,000	7,000
7	3	Low	10,000	9,000	5,000	5,000
8	3	Ave.	12,000	11,000	6,000	6,000
9	3	High	14,000	13,000	7,000	7,000
10	4	Low	8,000	8,000	4,500	4,500
11	4	Ave.	10,000	10,000	5,500	5,500
12	4	High	12,000	12,000	6,500	6,500

13.3.1 The curse of dimensionality

In order to illustrate how the curse of dimensionality arises, we shall reexamine the simple dairy model used in the previous sections. First we shall realize that the *age* of the animal is not represented. Thus, if a cow remains high yielding it will never be replaced according to the optimal policies shown in Tables 13.4 and 13.5. This is certainly not realistic, and furthermore, the milk yield also depends on the lactation number. In order to account for age we shall introduce an additional state variable representing the lactation number of the cow. For convenience, we shall assume that the new variable may take the values 1, 2, 3 or 4 indicating that the maximum age of a cow in this (still very simple) model is assumed to be 4 lactations.

The milk yields (i.e. outputs m_i^d) and economic net returns (i.e. rewards r_i^d) assumed for this slightly extended model appear in Table 13.7. Because a cow is always replaced after 4 lactations, the rewards are identical under both actions for those states representing 4th lactation.

The transition matrices of the extended model are shown in Tables 13.8 and 13.9. It should be emphasized that the state variable concerning milk yield should be interpreted relatively for a cow of the parity in question. As long as a cow is kept, it is assumed to change relative level of milk yield with the same probabilities as in the simple model, but when a replacement takes place, the new heifer is assumed to be low, average or high yielding with equal probabilities.

Inspection of the new matrices clearly illustrate that this very modest extension of the model causes a rather dramatic increase of the dimensions.

Now, suppose that we in addition to lactation and milk yield also want to take the genetic merit into account. We shall assume that the genetic merit of the cow

Table 13.8: Transition matrix of the extended model under action 1 (Keep).

	j	1	2	3	4	5	6	7	8	9	10	11	12	
	l_j	1	1	1	2	2	2	3	3	3	4	4	4	
i	l_i	Y	L	A	H	L	A	H	L	A	H	L	A	H
1	1	L	0	0	0	0.6	0.3	0.1	0	0	0	0	0	0
2	1	A	0	0	0	0.2	0.6	0.2	0	0	0	0	0	0
3	1	H	0	0	0	0.1	0.3	0.6	0	0	0	0	0	0
4	2	L	0	0	0	0	0	0	0.6	0.3	0.1	0	0	0
5	2	A	0	0	0	0	0	0	0.2	0.6	0.2	0	0	0
6	2	H	0	0	0	0	0	0	0.1	0.3	0.6	0	0	0
7	3	L	0	0	0	0	0	0	0	0	0	0.6	0.3	0.1
8	3	A	0	0	0	0	0	0	0	0	0	0.2	0.6	0.2
9	3	H	0	0	0	0	0	0	0	0	0	0.1	0.3	0.6
10	4	L	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
11	4	A	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
12	4	H	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0

Legends: l_i = lactation number of state i . Y = Milk yield indicated as L=Low, A=Average or H=High.

Table 13.9: Transition matrix of the extended model under action 2 (Replace).

	j	1	2	3	4	5	6	7	8	9	10	11	12	
	l_j	1	1	1	2	2	2	3	3	3	4	4	4	
i	l_i	Y	L	A	H	L	A	H	L	A	H	L	A	H
1	1	L	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
2	1	A	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
3	1	H	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
4	2	L	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
5	2	A	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
6	2	H	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
7	3	L	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
8	3	A	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
9	3	H	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
10	4	L	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
11	4	A	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0
12	4	H	1/3	1/3	1/3	0	0	0	0	0	0	0	0	0

Legends: l_i = lactation number of state i . Y = Milk yield indicated as L=Low, A=Average or H=High.

is either "Bad", "Average" or "Good". When a cow is replaced we assume that the probability of the new heifer to be either genetically "Bad", "Average" or "Good" is 1/3 each. The total size of the state space then becomes $3 \times 4 \times 3 = 36$. The milk yields m_i^d and rewards r_i^d appear from Table 13.10. The transition matrices of this 36-state model are now very large. They are shown in Appendix D as Tables D.1 - D.6 for actions "Keep" and "Replace", respectively.

This stepwise extension of the model clearly illustrates that each time a new state variable at n levels is added to the model, the size of the state space is increased by a factor of n . When, in a real model, several traits are represented by state variables at a realistic number of levels, the size of the state space very soon reaches prohibitive dimensions (millions of states). As an example, consider the dairy cow replacement model presented by Houben et al. (1994). The traits considered, when a decision was made, were

- The age of the cow (204 levels).
- Milk yield in present lactation (15 levels).
- Milk yield in previous lactation (15 levels).
- Time interval between 2 successive calvings (8 levels)
- Clinical mastitis - an infectious disease in the udder (2 levels - yes/no).
- Accumulated number of mastitis cases in present lactation (4 levels).

In principle the size of the state space is formed as the product of the number of levels of all traits, i.e. $204 \times 15 \times \dots \times 4 = 11,750,400$ states. In practise it is smaller because some combinations are impossible and because the traits related to previous lactation are not considered during first lactation. Exclusion of such not feasible states resulted in a model with 6,821,724 different states. The model described only considers traits that vary over time for the same animal. If furthermore, we wish to include permanent traits of the present animal (like the genetic merit of the 36-state model) being considered for replacement, the state space would become even larger. In order to circumvent this curse of dimensionality, Kristensen (1988, 1991) introduced a new notion of a hierarchical Markov process, which is described in the following sections.

13.3.2 Notation and terminology

A hierarchical Markov process is only relevant under infinite planning horizon, and there is no relevance of the criterion (13.4) because the special situation where the physical output equals 1 in all stages has no computational advantages over other values. Therefore we shall only consider the criteria (13.3) and (13.5).

A hierarchical Markov process is a series of Markov decision processes called *child processes* built together in one Markov decision process called the *funder*

Table 13.10: Rewards and outputs of the 36 state model.

i	Gen. merit	Lact.	Yield	Rewards, r_i^d		Outputs, m_i^d	
				$d = 1$ (K)	$d = 2$ (R)	$d = 1$ (K)	$d = 2$ (R)
1	Bad	1	Low	6,000	5,000	3,000	3,000
2	Bad	1	Ave.	8,000	7,000	4,000	4,000
3	Bad	1	High	10,000	9,000	5,000	5,000
4	Bad	2	Low	8,000	7,000	4,000	4,000
5	Bad	2	Ave.	10,000	9,000	5,000	5,000
6	Bad	2	High	12,000	11,000	6,000	6,000
7	Bad	3	Low	8,000	7,000	4,000	4,000
8	Bad	3	Ave.	10,000	9,000	5,000	5,000
9	Bad	3	High	12,000	11,000	6,000	6,000
10	Bad	4	Low	6,000	6,000	3,500	3,500
11	Bad	4	Ave.	8,000	8,000	4,500	4,500
12	Bad	4	High	10,000	10,000	5,500	5,500
13	Ave.	1	Low	8,000	7,000	4,000	4,000
14	Ave.	1	Ave.	10,000	9,000	5,000	5,000
15	Ave.	1	High	12,000	11,000	6,000	6,000
16	Ave.	2	Low	10,000	9,000	5,000	5,000
17	Ave.	2	Ave.	12,000	11,000	6,000	6,000
18	Ave.	2	High	14,000	13,000	7,000	7,000
19	Ave.	3	Low	10,000	9,000	5,000	5,000
20	Ave.	3	Ave.	12,000	11,000	6,000	6,000
21	Ave.	3	High	14,000	13,000	7,000	7,000
22	Ave.	4	Low	8,000	8,000	4,500	4,500
23	Ave.	4	Ave.	10,000	10,000	5,500	5,500
24	Ave.	4	High	12,000	12,000	6,500	6,500
25	Good	1	Low	10,000	9,000	5,000	5,000
26	Good	1	Ave.	12,000	11,000	6,000	6,000
27	Good	1	High	14,000	13,000	7,000	7,000
28	Good	2	Low	12,000	11,000	6,000	6,000
29	Good	2	Ave.	14,000	13,000	7,000	7,000
30	Good	2	High	16,000	15,000	8,000	8,000
31	Good	3	Low	12,000	11,000	6,000	6,000
32	Good	3	Ave.	14,000	13,000	7,000	7,000
33	Good	3	High	16,000	15,000	8,000	8,000
34	Good	4	Low	10,000	10,000	5,500	5,500
35	Good	4	Ave.	12,000	12,000	6,500	6,500
36	Good	4	High	14,000	14,000	7,500	7,500

process¹. A child process is a finite time Markov decision process with N stages and a finite state space $\Omega_n = \{1, \dots, u_n\}$ for stage n , $1 \leq n \leq N$. The action set D_n of the n th stage is assumed to be finite, too. A policy s of a child process is a map assigning to each stage n and state $i \in \Omega_n$ an action $s(n, i) \in D_n$. The set of all possible policies of a child process is denoted Γ . When the state i is observed and the action d is taken, a reward $r_i^d(n)$ is gained. The corresponding physical output is denoted as $m_i^d(n)$. Let $p_{ij}^s(n)$ be the transition probability from state i to state j where i is the state at the n th stage, j is the state at the following stage and d is the action taken at stage n . Under the Criterion (13.3) we shall denote the discount factor in state i under the action d as $\beta_i^d(n)$ assuming that the stage length is given by stage, state and action.

Assume that we have a set of v possible child processes each having its own individual set of parameters. The founder process is then a Markov decision process running over an infinite number of stages and having the finite state space $\{1, \dots, v\}$. Each stage in this process represents a particular child process. The action sets of the main process are the sets Γ_ι , $\iota = 1, \dots, v$, of all possible policies of the individual child processes (to avoid ambiguity the states of the founder process will be denoted by Greek letters ι, κ etc.). A policy σ is a map assigning to each state, ι , of the founder process an action $\sigma(\iota) \in \Gamma_\iota$. The transition matrix of the founder process has the dimension $v \times v$, and it is denoted $\Phi = \{\phi_{\iota\kappa}\}$. The transition probabilities are assumed to be independent of the action taken. The reward f_ι^σ and the physical output h_ι^σ in state ι of the founder process are determined from the total rewards and output functions of the corresponding subprocess as

$$\begin{aligned} f_i^s(n) &= r_i^s(n), \quad n = N \\ f_i^s(n) &= r_i^s(n) + \beta_i^s(n) \sum_{j=1}^{u_{n+1}} p_{ij}^s(n) f_j^s(n+1), \quad n = 1, \dots, N-1, \\ \text{and,} \\ f_\iota^\sigma &= \sum_{i=1}^{u_1} p_i(0) f_i^s(1), \quad s = \sigma(\iota), \end{aligned} \tag{13.14}$$

and analogously for h_ι^σ (except for the discount factor). The symbol $p_i(0)$ is the probability of observing state i at the first stage of the child process. Finally, the expected discount factor in state ι under the action s is denoted as B_ι^s and calculated as follows

¹In some texts, a child process is called a *sub process*, and the founder process is called the *main process*.

$$\begin{aligned}
b_i^s(n) &= \beta_i^s(n), \quad n = N, \\
b_i^s(n) &= \beta_i^s(n) \sum_{j=1}^{u_{n+1}} p_{ij}^s(n) b_j^s(n+1), \quad n = 1, \dots, N-1, \\
&\text{and,} \\
B_i^s &= \sum_{i=1}^{u_1} p_i(0) b_i^s(1).
\end{aligned} \tag{13.15}$$

13.3.3 Optimization

Since the founder process is just an ordinary Markov decision process, the policy iteration cycle described previously might be used directly for optimization. In practice Steps 1 and 2 are easily carried out, but Step 3 is prohibitive because of the extremely large number of alternative actions $s \in \Gamma_\iota$ (as mentioned above s is an entire policy of the ι th child process). To circumvent this problem Kristensen (1988, 1991) constructed an iterative method, where a value iteration method is applied in the child processes and the results are used in Step 3 of the policy iteration method of the founder process. The different versions of the method cover the criteria of optimality under infinite planning horizon defined as (13.3) and (13.5). Since criterion (13.4) is a special case of (13.5) it is also indirectly covered.

The general form of the iteration cycle of a hierarchical Markov process has the following steps:

1. Choose an arbitrary policy, σ . Go to 2.
2. Solve the following set of linear simultaneous equations for $F_1^\sigma, \dots, F_v^\sigma$ and in case of Criterion (13.5) for g^σ :

$$g^\sigma h_\iota^\sigma + F_\iota^\sigma = f_\iota^\sigma + B_\iota^\sigma \sum_{\kappa=1}^v \phi_{\iota\kappa} F_\kappa^\sigma, \quad \iota = 1, \dots, v. \tag{13.16}$$

In case of Criterion (13.5) the additional equation $F_v^\sigma = 0$ is necessary in order to determine a unique solution. Go to 3.

3. Define

$$T_\iota = \sum_{\kappa=1}^v \phi_{\iota\kappa} F_\kappa^\sigma \tag{13.17}$$

under Criterion (13.3) and $T_\iota = 0$ under Criterion (13.5). For each child process, ι , find by means of the recurrence equations

$$\begin{aligned}\tau_{\iota,i}(n) &= \max_d \left\{ r_i^d(n) - m_i^d(n)g^\sigma + \beta_i^d(N)T_\iota \right\}, \quad n = N \\ \tau_{\iota,i}(n) &= \max_d \left\{ r_i^d(n) - m_i^d(n)g^\sigma + \beta_i^d(n) \sum_{j=1}^{u_{n+1}} p_{ij}^d(n)\tau_{\iota,j}(n+1) \right\}, \\ &\quad n = 1, \dots, N-1\end{aligned}\tag{13.18}$$

a policy s' of the child process. The action $s'(n)$ is equal to the d' that maximizes the right hand side of the recurrence equation of state i at stage n . Put $\sigma'(\iota) = s'$ for $\iota = 1, \dots, v$. If $\sigma' = \sigma$, then stop since an optimal policy is found. Otherwise, redefine σ according to the new policy (i.e. put $\sigma = \sigma'$ and go back to 2).

When the iteration cycle is used under Criterion (13.3) all physical outputs ($m_i^d(n)$) and accordingly also h_ι^σ are put equal to zero. The iteration cycle covering this situation was developed by Kristensen (1988).

Under Criterion (13.4) all physical outputs, $m_i^d(n)$, and all discount factors $\beta_i^d(n)$ and B_ι^σ are put equal to 1, but under Criterion (13.5) only the discount factors are put equal to 1. The iteration cycle covering these situations was described by Kristensen (1991).

13.3.4 Discussion and applications

The hierarchical Markov process is specially designed to fit the structure of animal decision problems where the successive stages of the child processes correspond to the age of the animal in question. By appropriate selection of state spaces in the child processes and the founder process it is possible to find optimal solutions to even very large models. The idea is to let the number of states in the child processes (where a value iteration technique is applied) be very large and only include very few states in the main process (where the technique is directly based on the policy iteration method). Thus we have got a method which is at the same time fast, exact and able to handle very large models.

Kristensen (1987) used the technique in a dairy cow replacement model which in a traditional formulation as an ordinary Markov decision process would have contained approximately 60,000 states, and later Kristensen (1989) in a model with approximately 180 000 states. In both cases the number of states in the founder process was only 5, reducing Step 2 to the solution of only 5 simultaneous linear equations (versus 180,000 in a traditional formulation). Also Houben et al. (1994) used the method in a dairy cow replacement model. The reduction of the system of equations was in their case from 6,821,724 to just 1, because no permanent traits were considered in the model. An interesting aspect of their model is that for

the first time a disease is taken into account. Haran (1997) also built a dairy cow replacement model based on a hierarchical Markov process.

In a different application area, Mourits et al. (1999) used a hierarchical Markov process in the determination of optimal feeding level and insemination age for dairy heifers. Similar ideas for use in fattening of bull calves have been presented by Makulska and Kristensen (1999).

In sows, Huirne et al. (1993) seem to have applied a technique which in many aspects is similar to a hierarchical Markov process, but they have not explained their method in all details. Also Jørgensen (1992) has applied a technique which is inspired of a hierarchical Markov process in a sow replacement model, and later (Jørgensen, 1993), he used the hierarchical method in the determination of optimal delivery policies in slaughter pigs. Also Broekmans (1992) used the method in the determination of optimal delivery policies in slaughter pigs taking random variation in prices into account. More recent applications to slaughter pigs include Kure (1997) and, in particular, Toft et al. (2005) who studied the influence of epidemic diseases on the slaughter policies. Verstegen et al. (1998) used the technique in an experimental economics study investigating the utility value of management information systems. They used a formulation involving Bayesian updating of traits as described by Kristensen (1993).

Naturally the hierarchical model just described may also be formulated as an ordinary Markov decision process. In that case each combination of child process (founder state), stage and state should be interpreted as a state. We shall denote a state in the transformed process as $(\iota n i)$, and the parameters are

$$\begin{aligned} r_{\iota n i}^d &= r_i^d(n), \\ m_{\iota n i}^d &= m_i^d(n), \\ \beta_{\iota n i}^d &= \beta_i^d(n), \\ p_{(\iota n i)(\kappa m j)}^d &= \begin{cases} p_{ij}^d(n), & \iota = \kappa \wedge m = n - 1 \\ \phi_{\iota \kappa} p_i(0), & n = N \wedge m = 1 \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \quad (13.19)$$

where the parameters mentioned on the right hand side of the equations are those belonging to the ι th child process except for $p_i(0)$ which belongs to child process κ . This formulation of course has the same optimal policies as the hierarchical formulation, so it is only computational advantages that make the hierarchical model relevant. A comparison to traditional methods may therefore be relevant.

Since the policy iteration method involves the solution of a set of u equations (where u is the number of states) it is only relevant for small models. The value iteration method, however, has been used with even very large models and may handle problems of the same size as the hierarchical formulation, but the time spent on optimization is much lower under the hierarchical formulation. To recognize this, we shall compare the calculations involved.

Step 3 of the hierarchical optimization involves exactly the same number of operations as one iteration of the value iteration method (Eq. 13.6). The further needs of the hierarchical method are the calculation of the rewards and *either* the physical output *or* the expected discount factor of a stage in the founder process according to Eqs. (13.14) and (13.15). Since the calculations at each stage is only carried out for one action, the calculation of both founder state parameters involves approximately the same number of operations as one iteration under the value iteration method if the number of alternative actions is 2. If the number of actions is higher, the calculations relatively involves a lower number of operations than an iteration under the value iteration method. These considerations are based on the assumption that the value iteration method is programmed in an efficient way, so that the sum of Eq. (13.6) is not calculated as a sum of all u elements, but only as a sum of those elements where p_{ij}^d is not zero according to Eq. (13.19). Otherwise the hierarchical technique will be even more superior. Finally the system of linear equations of Step 2 of the hierarchical cycle must be solved, but in large models with only a few states in the founder process the time spent on this is negligible.

If we use the considerations above in a practical example, the advantages of the hierarchical technique becomes obvious. As reported by Kristensen (1991) a model with 180,000 state combinations was optimized by the hierarchical technique under 100 different price conditions. The number of iterations needed ranged from 3 to 6 corresponding to between 6 and 12 iterations of the value iteration method. If the latter method was used instead, a planning horizon of 20 years would be realistic (cf. van Arendonk, 1985). Since each state in the model equals 4 weeks, this horizon represents 260 iterations, which should be compared to the equivalence of from 6 to 12 when the hierarchical technique was applied.

13.3.5 The numerical example formulated and solved as a hierarchical Markov process

In order to illustrate the hierarchical technique, we shall formulate the numerical example (the 36-state model) as a hierarchical Markov process. A Java based software tool for hierarchical Markov processes has been developed by Kristensen (2003)². The following example is pre-defined and comes with the program.

The three classes of the genetic merit are defined as states in the founder process of a hierarchical Markov process. Thus the number of child processes is also 3, and each child process represents a dairy cow of a certain genetic merit. When a new heifer is purchased, we assume, like before, that the probability distribution over main states is uniform, so that the probability of entering either one is 1/3. The maximum age of a cow was assumed to be 4 lactations, and the states of the child process are defined from the relative level of milk yield. Further a dummy state

²The program is available on World Wide Web. A Windows installer file may be downloaded from URL: <http://www.prodstyr.ihh.kvl.dk/software/mlhmp.html>

Table 13.11: Parameters of the hierarchical Markov process: Transition probabilities of founder process and initial state probabilities of child processes.

State ι	Transition probabilities, founder			Initial state probabilities, children			
	$\kappa=1$	$\kappa=2$	$\kappa=3$	$p_i(0)$			
				$i = 1$	$i = 2$	$i = 3$	$i = 4$
1	1/3	1/3	1/3	1/3	1/3	1/3	0
2	1/3	1/3	1/3	1/3	1/3	1/3	0
3	1/3	1/3	1/3	1/3	1/3	1/3	0

of length, reward and output equal to 0 is included at each stage of the child processes. If the cow is replaced at the end of a stage, the process enters the dummy state with probability 1 at the next stage, and for the rest of the duration of the subprocess it will stay in the dummy states. Stage numbers in the child processes directly correspond to lactation numbers.

Thus, for all child processes, the probability of staying at the same relative level of milk yield (state in the subprocess) is 0.6, and if the present state is “Average”, the probability of transition to either “Low” or “High” is 0.2 each. The probability of transition (if kept) from “Low” or “High” to “Average” is in both cases 0.3, and from “Low” to “High” and vice versa the probability is 0.1. The initial state probabilities of the child processes represent the probability of a new heifer being either “Low”, “Average” or “High” yielding. Thus, all initial state probabilities are 1/3.

Like before, the physical outputs are interpreted as milk yields and the rewards are defined as the economic net returns. All parameters of the hierarchical model are shown in Tables 13.12 and 13.11.

We shall determine an optimal solution under the following 3 criteria of optimality:

1. Maximization of total expected discounted rewards, i.e., the objective function (13.3). In this case the physical outputs of Table 13.12 are ignored, and a discount factor $\beta_i^d(n) = \exp(-r)$, where r is the interest rate, is applied (for states where the stage length is not zero).
2. Maximization of average rewards over time. In this situation we use the objective function (13.5) letting the output represent stage length. No discounting is performed in this case.
3. Maximization of average rewards over output defined as in Table 13.12. Thus the objective function (13.5) is applied, and no discounting is performed.

In Table 13.13, optimal policies under the three criteria are shown. It appears that the policies under the first two criteria are quite similar, but under the third

Table 13.12: Parameters of the hierarchical Markov process, child processes.

ι	n	i	$p_{ij}^1(n)$				$p_{ij}^2(n)$				$r_i^2(n)$	$m_i^2(n)$	
			$j=1$	$j=2$	$j=3$	$j=4$	$r_i^1(n)$	$m_i^1(n)$	$j=1$	$j=2$	$j=3$	$j=4$	
1	1	1	0.6	0.3	0.1	0.0	6,000	3,000	0.0	0.0	0.0	1.0	5,000
1	1	2	0.2	0.6	0.2	0.0	8,000	4,000	0.0	0.0	0.0	1.0	7,000
1	1	3	0.1	0.3	0.6	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
1	1	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
1	2	1	0.6	0.3	0.1	0.0	8,000	4,000	0.0	0.0	0.0	1.0	7,000
1	2	2	0.2	0.6	0.2	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
1	2	3	0.1	0.3	0.6	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
1	2	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
1	3	1	0.6	0.3	0.1	0.0	8,000	4,000	0.0	0.0	0.0	1.0	7,000
1	3	2	0.2	0.6	0.2	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
1	3	3	0.1	0.3	0.6	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
1	3	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
1	4	1	-	-	-	-	6,000	3,500	-	-	-	-	6,000
1	4	2	-	-	-	-	8,000	4,500	-	-	-	-	8,000
1	4	3	-	-	-	-	10,000	5,500	-	-	-	-	10,000
1	4	4	-	-	-	-	0	0	-	-	-	-	0
2	1	1	0.6	0.3	0.1	0.0	8,000	4,000	0.0	0.0	0.0	1.0	7,000
2	1	2	0.2	0.6	0.2	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
2	1	3	0.1	0.3	0.6	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
2	1	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
2	2	1	0.6	0.3	0.1	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
2	2	2	0.2	0.6	0.2	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
2	2	3	0.1	0.3	0.6	0.0	14,000	7,000	0.0	0.0	0.0	1.0	13,000
2	2	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
2	3	1	0.6	0.3	0.1	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
2	3	2	0.2	0.6	0.2	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
2	3	3	0.1	0.3	0.6	0.0	14,000	7,000	0.0	0.0	0.0	1.0	13,000
2	3	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
2	4	1	-	-	-	-	8,000	4,500	-	-	-	-	8,000
2	4	2	-	-	-	-	10,000	5,500	-	-	-	-	10,000
2	4	3	-	-	-	-	12,000	6,500	-	-	-	-	12,000
2	4	4	-	-	-	-	0	0	-	-	-	-	0
3	1	1	0.6	0.3	0.1	0.0	10,000	5,000	0.0	0.0	0.0	1.0	9,000
3	1	2	0.2	0.6	0.2	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
3	1	3	0.1	0.3	0.6	0.0	14,000	7,000	0.0	0.0	0.0	1.0	13,000
3	1	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
3	2	1	0.6	0.3	0.1	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
3	2	2	0.2	0.6	0.2	0.0	14,000	7,000	0.0	0.0	0.0	1.0	13,000
3	2	3	0.1	0.3	0.6	0.0	16,000	8,000	0.0	0.0	0.0	1.0	15,000
3	2	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
3	3	1	0.6	0.3	0.1	0.0	12,000	6,000	0.0	0.0	0.0	1.0	11,000
3	3	2	0.2	0.6	0.2	0.0	14,000	7,000	0.0	0.0	0.0	1.0	13,000
3	3	3	0.1	0.3	0.6	0.0	16,000	8,000	0.0	0.0	0.0	1.0	15,000
3	3	4	0.0	0.0	0.0	1.0	0	0	0.0	0.0	0.0	1.0	0
3	4	1	-	-	-	-	10,000	5,500	-	-	-	-	10,000
3	4	2	-	-	-	-	12,000	6,500	-	-	-	-	12,000
3	4	3	-	-	-	-	14,000	7,500	-	-	-	-	14,000
3	4	4	-	-	-	-	0	0	-	-	-	-	0

Legends: ι = Child process (Founder State), n = Stage number, i = State at Stage n , j = State at Stage $n + 1$.

Table 13.13: Optimal policies under the three criteria (c_1, c_2, c_3) defined in the text (actions: 1 = “Keep”, 2 = “Replace”).

Subprocess	Stage	State 1			State 2			State 3		
		c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
1	1	2	2	1	2	2	1	2	1	1
1	2	2	2	1	2	2	1	2	1	1
1	3	2	2	1	2	2	1	2	2	1
2	1	1	1	1	1	1	1	1	1	1
2	2	2	1	1	1	1	1	1	1	1
2	3	2	2	1	2	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
3	2	1	1	1	1	1	1	1	1	1
3	3	1	1	1	1	1	1	1	1	1

Table 13.14: The performance of the hierarchical technique compared to the policy and value iteration methods under the three criteria (c_1, c_2, c_3) defined in the text.

	Hierarchical model			Policy iteration			Value iteration		
	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
Number of iterations	4	3	2	4	4	2	100	100	100
Computer time, rel.	1	0.86	0.43	268	267	139	48	46	11

criterion the optimal policy differs significantly. A more detailed example of the effect of criterion of optimality was discussed by Kristensen (1991).

In order to compare the efficiency of the hierarchical technique to the traditional policy and value iteration methods, the problem of the example was transformed to an ordinary Markov decision process and optimized by those methods. The transformed model has $3 \times 4 \times 4 = 48$ states, which is not larger than the policy iteration method may be applied without problems. In Table 13.14 some performance data of the three optimization techniques are compared.

The superiority of the hierarchical technique over the policy iteration method is due mainly to the time spent on solving the linear simultaneous equations of Step 2. In the hierarchical case a system of 3 equations is solved, whereas 48 equations are solved under the ordinary policy iteration method.

In this numerical example the performance of the hierarchical technique is even more superior to the value iteration method than expected from the theoretical considerations of Section 13.3.4. In the present case an iteration of the hierarchical model is performed even faster than one of the value iteration method applied to the same (transformed) model. The reason is that the value iteration algorithm has not been programmed in the most efficient way as defined in Section 13.3.4. On the contrary, the sum of Eq. (13.6) has been calculated over all 48 states of the trans-

formed model. Since only 4 transition probabilities from each state are positive, the sum could be calculated only over these 4 states.

13.4 Uniformity: Bayesian updating

13.4.1 Principles of learning from observations

As discussed earlier, it is obvious that the traits of an animal varies no matter whether we are considering the milk yield of a dairy cow, the litter size of a sow or almost any other trait. On the other hand, it is *not* obvious to what extent the *observed* trait Y_n at stage n is, for instant, the result of a permanent property of the animal X_1 , a permanent damage caused by a previous disease X_2 or a temporary random fluctuation e_n . Most often the observed value is the result of several permanent and random effects. With Y_n , X_1 , X_2 and e_n defined as above the relation might for instance be

$$Y_n = m + X_1 + aX_2 + e_n, \quad (13.20)$$

where m is the expected value for an arbitrarily selected animal under the circumstances in question, and $a = -1$ if the animal has been suffering from the disease, and $a = 0$ otherwise. In this example, X_1 only varies *among* animals, whereas e_n also varies *over time* for the same animal. The effect of the damage caused by the disease X_2 is in this example assumed to be constant over time when it has been “switched on”. The value of X_2 is a property of the individual disease case (the “severity” of the case).

In a replacement decision it is of course important to know whether the observed value is mainly a result of a permanent effect or it is just the result of a temporary fluctuation. The problem, however, is that only the resulting value Y_n is observed, whereas the values of X_1 , X_2 and e_n are unknown. On the other hand, as observations of Y_1, Y_2, \dots are done, we are learning something about the value of the permanent effects. Furthermore, we have got a *prior* distribution of X_1 and X_2 , and each time an observation is done, we are able to calculate the *posterior* distribution of X_1 and X_2 by means of the Kalman-filter theory in connection with Dynamic Linear Models (described for instance by West and Harrison, 1997) if we assume all effects to be normally distributed.

A model as described by Eq. (13.20) fits very well into the structure of a hierarchical Markov process. Thus we may regard Y_n as a state variable in a child process, and the permanent effects X_1 and X_2 as state variables of the founder process. We then face a hierarchical Markov process with *unobservable founder state*. Kristensen (1993) discusses this notion in details, and it is shown that under the assumption of normally distributed effects, we only have to keep the present expected values of X_1 and X_2 , the currently observed value of Y_n and (in this example) the number of stages since the animal was suffering from the disease (if it has been suffering from the disease at all). The expectations of X_1 and X_2 are

sufficient to determine the current posterior distribution of the variables, because the variance is known in advance. Even though the posterior variance decreases as observations are done, the decrease does *not* depend on the *values* of Y_1, Y_2, \dots , but only on the number of observations done.

In the study of Kristensen (1993), a more general case involving several traits each being influenced by several unobservable effects is sketched, and a numerical example involving only a single trait is given. An example concerning replacement of sows has been given by Jørgensen (1992). It was demonstrated in both studies that the Bayesian approach in some cases may result in state space reduction without loss of information. In multi-trait updating models a Kalman filter technique may be relevant as described by Kristensen (1994).

13.4.2 Applications

The described principles for Bayesian updating have in particular been used in combination with Dynamic Linear Models (see West and Harrison, 1997, for a detailed description of the concept). In addition to the model by Jørgensen (1992), such models with built-in learning capabilities have been developed by Kristensen and Søllested (2004a,b) for sow replacement taking all litter size results into account. Lien et al. (2003) developed a model for determination of optimal length of leys in an area with winter damage problems, where the productivity of the ley is learned over time as observations are done. Finally, the principles are used in a slaughter pig marketing model with emphasis on information from online weighing equipment (Kristensen et al., 2006).

13.5 Herd restraints: Parameter iteration

One of the major difficulties identified in the introduction was *herd restraints*. All the replacement models mentioned in the previous sections have been single-component models, i.e., only one animal is considered at the same time, assuming an unlimited supply of all resources (heifers or gilts for replacement, feed, labor etc) and no production quota. In a multi-component model all animals of a herd are simultaneously considered for replacement. If all animals (components) compete for the same limited resource or quota, the replacement decision concerning an animal does not only depend on the state of that particular animal, but also on the states of the other animals (components) of the herd.

If the *only* (or at least the *most limiting*) herd restraint is a limited housing capacity, the number of animals in production is the scarce resource, and accordingly the relevant criterion of optimality is the maximization of net revenues per animal as it is expressed in the criteria (13.1), (13.2), (13.3) and (13.4). Thus the optimal replacement policy of the single component model is optimal for the multi-component model too.

If the only (or most limiting) herd restraint is a milk quota, the situation is much more complicated. Since the most limiting restriction is a fixed amount of milk to produce, the relevant criterion of optimality is now the maximization of average net revenues per kg milk yield as expressed in criterion (13.5), because a policy that maximizes net revenues per kg milk will also maximize total net revenues from the herd which was assumed to be the objective of the farmer.

By following a policy which is optimal according to criterion (13.5) we assure at any time that the cows which produce milk in the cheapest way are kept. Thus the problem of selecting which cows to keep in the long run (and the mutual ranking of cows) is solved, but the problem of determining the optimal number of cows in production at any time is *not* solved. If for instance, it is recognized 2 months before the end of the quota year that the quota is expected to be exceeded by 10 percent, we have to choose whether to reduce the herd size or to keep the cows and pay the penalty. The problem is that both decisions will influence the possibilities of meeting the quota of the next year in an optimal way. To solve this short run quota adjustment problem we need a true multi-component model.

An other example of a herd restraint is a limited supply of heifers. If the dairy farmer only uses home-grown heifers for replacement, the actions concerning individual cows become inter-dependent, and again a multi-component model is needed in order to solve the replacement problem. Ben-Ari and Gal (1986) and later Kristensen (1992) demonstrated that the replacement problem in a dairy herd with cows and a limited supply of home grown heifers may be formulated as a Markov decision process involving millions of states. This multi-component model is based on a usual single-component Markov decision process representing one cow and its future successors. Even though the hierarchical technique has made the solution of even very large models possible, such a model is far too large for optimization in practice. Therefore, the need for an approximate method emerged, and a method called *parameter iteration* was introduced by Ben-Ari and Gal (1986).

The basic idea of the method is to approximate either the present value function $f_i(n)$ (objective function (13.3)) or the relative values f_i^s (objective functions (13.4) and (13.5)) by a function G involving a set of parameters a_1, \dots, a_m to be determined in such a way that $G(i, a_1, \dots, a_m) \approx f_i(n)$ or $G(i, a_1, \dots, a_m) \approx f_i^s$.

In the implementation of Ben-Ari and Gal (1986) the parameters were determined by an iterative technique involving the solution of sets of simultaneous linear equations generated by simulation.

In a later implementation Kristensen (1992) determined the parameters by ordinary least squares regression on a simulated data set. The basic idea of the implementation is to take advantage from the fact that we are able to determine an optimal solution to the underlying (unrestricted) single-component model. If no herd restraint was present, the present value of the multi-component model would equal the sum of the present values of the individual animals determined from the underlying single-component model. Then it is argued in what way the restraint will logically reduce the (multi-component) present value, and a functional expres-

sion having the desired properties is chosen. The parameters of the function are estimated from a simulated data set, and the optimal action for a given (multi-component) state is determined as the one that maximizes the corrected present value. (A state in the multi-component model is defined from the states of the individual animals in the single-component model, and an action defines the replacement decision of each individual animal).

Ben-Ari and Gal (1986) compared the economic consequences of the resulting optimal multi-component policy to a policy defined by dairy farmers, and they showed that the policy from the parameter iteration method was better. Kristensen (1992) compared the optimal multi-component policies to policies from usual single-component models in extensive stochastic simulations and showed that the multi-component policies were superior in situations with shortage of heifers.

The parameter iteration method has been applied under a limited supply of heifers. It seems to be realistic to expect, that the method and the basic principles of Kristensen (1992) may be used under other kinds of herd restraints as for instance the short time adjustment to a milk quota as mentioned above.

13.6 General discussion

In Section 13.1, the main difficulties concerning animal production models were identified as *variability* in traits, *cyclic production*, *uniformity* (the traits are difficult to define and measure) and *herd restraints*. We are now able to conclude that the difficulties of variability and the cyclic production are directly solved by the application of Markov decision programming, but when the variability of several traits are included we face a problem of *dimensionality*. The formulation of the notion of a *hierarchical Markov process* contributed to the solution of the dimensionality problem, but did not solve it. The upper limit of number of states to be included has been raised considerably, but not eliminated.

This is for instance clearly illustrated when we formulate *multi-component* herd models in order to deal with herd restraints. In that case we still have to use approximate methods to determine an “optimal” replacement policy. On the other hand it has been demonstrated by Kristensen (1992) that the *parameter iteration* method applied to a multi-component herd model (even though it is only approximate) is able to improve the total net revenue compared to the application of a usual single-component (animal) model in a situation with shortage of heifers. The parameter iteration method is an important contribution to the problem of determining optimal replacement policies under herd restraints.

In other situations with a limiting herd restraint it may be relevant to use an alternative criterion of optimality maximizing average net revenue per unit of the limiting factor. This method has been successfully applied in a situation with milk production under a limiting quota.

Recent results have also contributed to the solution of the uniformity problem. The Bayesian updating technique described in Section 13.4 has turned out to be a promising approach, which has been applied to several problems in farm management. It might also be a solution to the problem of including animal health as a trait to be considered. As concerns other traits such as litter size or milk yield the Bayesian approach may in some cases result in a reduction of the state space without loss of information (Jørgensen, 1992; Kristensen, 1993). Thus it contributes indirectly to the solution of the dimensionality problem.

New developments in the area of Markov decision programming for herd management support includes the notion of multi-level hierarchical models, where actions are defined at different levels with different time horizons. Thus the method is able to optimize decisions with different time horizons simultaneously as described by Kristensen and Jørgensen (2000). A good example of this technique is a model developed by Nielsen et al. (2004) for optimization of decisions concerning winter feeding level, summer grazing strategy, start of fattening and slaughtering of organic steers.

Also the development of a standard tool (the Java based MLHMP software Kristensen, 2003) for (multi-level hierarchical) Markov decision processes is considered to be promising for construction of decision support models in the future.

Chapter 14

Monte Carlo simulation

14.1 Introduction

The use of simulation techniques has a long and established tradition within animal production. Use of simulation models to understand herd management problems was the theme of the influential book *Systems Analysis in Agricultural Management* (Dent and Anderson, 1971), and since then several simulation models have been formulated within the different animal production systems. This chapter is not an attempt to review these previous efforts.

Decision making in animal production has long been seen as a game without a conscious opponent (a game against nature), and use of the Bayesian decision strategy of maximum expected utility has been the accepted criterion for choosing between the available actions. The underlying principle of acting based on the prior (subjective) probability distribution of the state of nature is an important part of the framework. This is the foundation of the so-called Bayesian statistical methods. The use of Bayesian methods within statistics has been rather sporadic until recently. A reason for this is a philosophical debate concerning the existence of subjective probabilities, especially whether it is acceptable to use subjective probabilities in the analysis of experiments. Another reason, is the computational complexity that is inherent in the Bayesian methods. This complexity barrier has to a large degree been broken down, because of numeric techniques, and the use of Bayesian networks to exploit the independencies between elements within the system. Because of this, Bayesian methodology is currently a very active research area within statistics.

The systems approach used in simulation modeling has many things in common with Bayesian modeling, and to the authors' opinion, the reason researchers use the modeling technique is an inherent wish to express their prior knowledge of the quantitative and qualitative structure of the system they investigate. This prior knowledge is based on results from many different experiments, observations in practice, and scientific theory building. Bayesian methodology offers the theoreti-

cal framework for this specification. In addition it offers a more stringent approach to parameter estimation and model testing than the traditional systems approach.

This note is an attempt to describe, how this framework can be used within the systems approach.

14.2 The simulation method

In Section 9.7 we have already given a rather detailed definition of the simulation method. The reader is encouraged to read that section again, before continuing with this chapter, which only covers the Monte Carlo technique. Here, we shall only give a very short summary for reference purpose.

A Monte Carlo simulation model is described by a set of parameters $\Phi = (\Phi_0, \Phi_{s\bullet})$, where Φ_0 represents the *state of nature* specified initially before the simulation, and $\Phi_{s\bullet} = (\Phi_{s1}, \Phi_{s2}, \dots, \Phi_{st}, \dots, \Phi_{sT})$ represents parameters that change values during simulations (often called *state variables*). Further concepts include the decision rule, Θ , and a set of output parameters Ω (typically a subset of $\Phi_{s\bullet}$).

As defined in Section 9.7, the purpose of a Monte Carlo simulation model described by the parameters $\Phi = (\Phi_0, \Phi_{s\bullet})$ is to calculate the expected utility, $\bar{U}(\Theta)$, under a given decision rule, Θ , i.e.,

$$\begin{aligned}\bar{U}(\Theta) &= \int_{-\infty}^{\infty} U(\Theta, \phi) f_{\Phi}(\phi) d\phi \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\Theta, \Phi_{s\bullet} | \Phi_0 = \phi_0) f_{(\Phi_{s\bullet} | \Phi_0)}(\phi_{s\bullet} | \phi_0) f_{\Phi_0}(\phi_0) d\phi_{s\bullet} d\phi_0,\end{aligned}\tag{14.1}$$

where U is the utility function, which in general can refer to any function of the output variables. The functions f_{Φ} , $f_{(\Phi_{s\bullet} | \Phi_0)}$ and f_{Φ_0} are the density functions of Φ , $(\Phi_{s\bullet} | \Phi_0)$ and Φ_0 , respectively.

The Monte Carlo simulation technique is a numerical method for solving this integral. It relies on the drawing of random numbers. Every time the model encounters a stochastic variable, a (pseudo)-random variable is drawn from the appropriate distribution and this value is used in the subsequent calculations. Each completed calculation (simulation run) with the model represents a random drawing from the simultaneous distribution of input and output variables. By increasing the number of calculations the distribution of the output variables can be specified to any degree of precision. The expected utility is found from:

$$\bar{U}(\Theta) \approx \frac{1}{k} \sum_{i=1}^k U(\Theta, \phi_i).\tag{14.2}$$

where ϕ_i is a random drawing from the multidimensional distribution of the parameters, and k is the number of random drawings (simulation runs).

14.2.1 Systems approach

Standard statistical methods for analysis of data, e.g. from experiments help in finding a sufficient model for the data. A sufficient model is such that the deviation from the model is random and independent from any known sources. Such a model will usually have very few parameters. In model comparisons between models with different number of parameters, the model with fewest parameters is preferred, if there is no substantial evidence for the model with most parameters. This approach can be justified for several reasons, from the practical experience that models with few parameters are more robust, when applied to other data sources; from the Taylor expansion that approximates any function to a desired accuracy with a (low order) polynomial; and, from the difficulty of separating effects of different parameters if no prior knowledge of the parameters exists.

Usually each set of data is treated separately from previous experience. In contrast, the simulation research method uses the systems approach, where existing knowledge of a system is build into the model.

The primary consideration within the system approach is the understanding of a whole system. A system can e.g. be defined as in Clausen (1969).

A system is a *part of reality* that in some manner is detached (seems different) from the *rest of reality*. This *rest* is the environment of the system. Evidently the *rest* is also a system. These two systems (the system and its environment) are characterized by their individual structure. But they are also interdependent. There is said to be a communication between them.

Such a system cannot be understood by focusing on single elements of the system, the interaction and structure of the elements is just as important.

The modeling steps can be divided into several steps, e.g. definition of the modeling goals, system analysis, system synthesis, verification, validation and inference. The *goal* of the simulation model can be to develop a decision support system or to integrate current knowledge concerning the system to enhance the understanding of the system etc. In the *system analysis* step the components, interactions, relationship and dynamic behavior of the system is isolated. The behavior of each element is described using information from different sources, e.g. research reports, new experiments. The object oriented approach is one of the techniques that can be used. In the *system synthesis* step the model is developed and programmed according to the specifications in the system analysis step. The Object oriented method makes it possible to make the structure and elements of the program closely correspond to the real system. *Verification* is the check of whether the model behaves according to the system specification. *Validation* is the step where the model output is compared to data from the real world, and finally the *inference* step, where the model is used according to the *goal*. These steps are performed cyclically with subsequent refinement of model until the final model can be used for inference.

The model structure defined in the system analysis is based on current knowledge concerning the causal relations within the system. This is the qualitative element of the model. The quantitative aspects of the model, the model parameters are usual combined estimates from several experiments. In addition, the estimates are very often modified based on experience learnt in the cyclic feed back in the model development.

Though the nomenclature is specific for the system approach of model development, many of the steps closely correspond to other methods.

The system approach can be compared with the point of view of Bayesian statisticians. The system analysis species the prior knowledge of the system. From this prior knowledge the behavior of the system can be predicted. A Bayesian would compare these predictions to observed data, and use standard objective methods to update the knowledge of the system.

Previously, this approach was not possible for models of larger systems. However, improved numerical methods within the Bayesian branch of statistics have become available. Within the statistical area called *Highly Structured Stochastic Systems* (HSSS) techniques are developed that will help improving the techniques used in the system approach towards simulation (refer to Besag et al. (1995) for examples)¹.

Still the complexity of the standard simulation model (number of parameters) is much larger than the systems that has been treated within HSSS.

14.2.2 The purpose of simulation modeling

Simulation models have been formulated with two different purposes. Very often the purpose of simulation models is to improve the understanding of a system, i.e. to combine research results from different areas to obtain a comprehensive description of the system. The other purpose is to use the model as management tool for decision support. These two purposes are related, but the differences put some limitation on the conclusions that may be reached.

The difference is most easily seen if we compare with Eq. (14.1). When simulation models are used to improve the understanding of the complex system, a fixed and known set of parameters are used for the initial state of nature, Φ_0 , and the expected value of the utility function or any other output variable is calculated as:

$$\overline{U}(\Theta \mid \Phi_0 = \phi_0) = \frac{1}{k} \sum_{i=1}^k U(\Theta, \phi_{si} \mid \Phi_0 = \phi_0), \quad (14.3)$$

i.e. only the inner part of the integrand in Eq. (14.1) is solved.

This corresponds closely to experimental design, where “noise” effects are controlled, such as climate, breed of the animals, differences between animals. The control of these effects make the “interesting” effects much more precise, in our

¹The MCMC Preprint service, <http://www.statslab.cam.ac.uk/mcmc/> has several references.

case we need fewer simulation runs or in the experimental case, we need fewer animals.

Typical investigations under this framework, are sensitivity analyses, where the effect of small changes in the parameters in Φ_0 . In this manner important parameters can be isolated and research efforts directed towards a better understanding of these parameters. The knowledge of the systems sensitivity to changes in the parameters is part of, what we call understanding of a system.

In contrast, when simulation models are used to determine “optimal” strategies we want to find the optimal set of decision rules given the precision in our current knowledge of the parameters (state of nature). The parameters used in each simulation run should therefore be a sample from the prior distribution of Φ_0 reflecting the precision in our current knowledge, and not fixed values.

14.2.3 Dimensionality of simulation modeling

The dimensionality of a simulation model is the number of parameters that serves as input to the model.

The dimensionality is best understood when Eq. (14.1) is considered. It can be split into three parts.

The dimensionality of state of nature : This is the number of parameters in Φ_0 specifying the initial conditions. Typically this number is comparatively “small” (less than 100 different variables). Many of the parameters have independent prior distributions, reducing the problem of specifying the covariance between the parameters. These parameters can often be estimated using Bayesian techniques. The dimensionality of the state of nature can be compared to the dimensionality of e.g. Markov chain probabilistic simulation approaches. Other numerical integration methods may be considered for the outer part of the integrand in Eq. (14.1)

The dimensionality of the systems’ state space : The number of parameters in $\Phi_{s\bullet}$. This will typically be *very* large, as each element in the simulation model and the interactions between the elements are represented. That is, the integration of the inner part of Eq. (14.1) is only tractable with the Monte Carlo simulation model.

The dimensionality of the decision space : Nothing general can be said of this dimension. With simple heuristic rules the dimension may be very low, and with complex models the dimensionality may approach the dimensionality of the system state space. If the set of decision rules are too complicated, the search for optimality will usually be intractable.

The major problem with too many parameters in the model is that it is not possible to distinguish effects of the different parameters based on observations. I.e. estimation, or learning, of the values of the parameters will result in very strong dependency between the parameter estimates.

14.3 Why use the Monte Carlo Approach

Within animal production several simulation models exist. These models are either deterministic or probabilistic e.g. Markov Chain or Monte Carlo. Because of the flexibility of the latter approach, the method as such does not lead to any fixed set of assumptions. The assumptions differ from model to model. The two other methods have some assumptions implied. These assumptions are not always stated in the model. Therefore, some of the consequences of the assumptions will be reviewed briefly.

14.3.1 The Deterministic Approach

The assumptions are of course that the real system has no stochastic variation. In order to understand the effect of the assumption, we can approximate the utility function or any output variable by a second order Taylor expansion around the value $\bar{\Phi}$,

$$U(\Phi) \approx U(\bar{\Phi}) + U'(\Phi - \bar{\Phi}) + \frac{1}{2}(\Phi - \bar{\Phi})^\top U''(\Phi - \bar{\Phi}), \quad (14.4)$$

where U' is the vector of first derivatives of $U()$ with respect to the parameters, and U'' is the matrix of second derivatives of $U()$ with respect to the parameters.

The expectation and variance of this expression can in general not be found directly, but if we assume multivariate normal distribution of the parameters, we have:

$$E(U(\Phi)) \approx U(\bar{\Phi}) + \frac{1}{2}tr(U''\text{Var}(\Phi)) \quad (14.5)$$

$$\text{Var}(U(\Phi)) \approx U'\text{Var}(\Phi)U'^\top + \frac{1}{2}tr(U''\text{Var}(\Phi))^2 \quad (14.6)$$

The major problem with deterministic models is the term that depends on the variance in Eq. (14.5). This term only vanishes if $U'' = 0$, i.e. if $U()$ is *linear* in the parameters or if $\text{Var}(\Phi) = 0$. Neither case is likely to happen.

The problem with the deterministic models is not so much, that the variance on the output is ignored, but that the calculated expectation is systematically wrong. Results from a deterministic model will not be correct even if all causal relations and expectations of parameter values are correct. The value of such models is of course dubious, and knowledge of the magnitude of the systematic error is of extreme importance, i.e. the second term in Eq. (14.5).

Probably (hopefully!) the severity of these problems does not interfere with the use of deterministic models to understand causal relations in order to formulate research hypothesis.

14.3.2 The Markov chain approach

The Markov Chain approach is usually used for finding the steady state distribution, i.e. the expected long term probability of being in each of the states. As the states

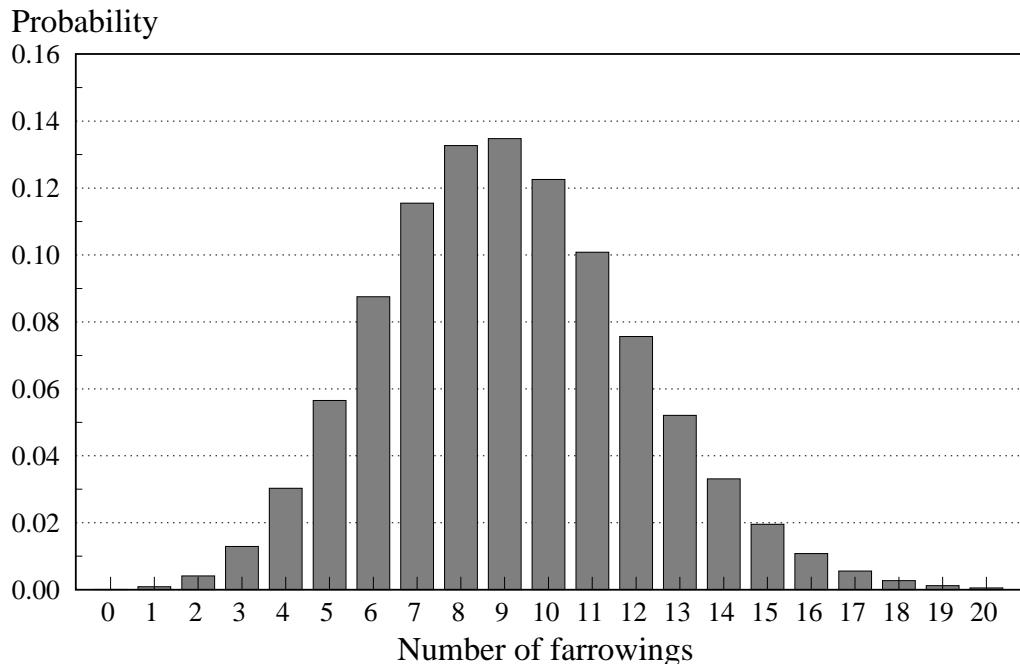


Figure 14.1: Probability distribution of number of animals in the farrowing state when steady state is reached in a Markov Chain simulation model with a herd size of 200 sows

are defined for an individual, the probability refers to a single individual. The whole herd can be studied if independency of the individuals is assumed.

Within the whole herd problem, we are interested in the distribution of how many animals that are in each of the states. This can also be found from the steady state distribution, even though usually only the steady state probability and the expected number are reported. The distribution is multinomial with the total number of animals and the steady state probabilities as parameters.

An example will illustrate this. In a sow herd model based on a step length of 1 week, the steady state distributions shows that approximately 4 % of the sows will be in the farrowing state. In a sow herd of 200 sows the probability distribution of sows in this state is shown in Figure 14.1.

We can assign income (utility) to this curve. The income will increase linearly with number of farrowings per week until the number of available pens is reached. If the number of farrowings is above the number of available pens, the income will most likely have a progressively lower slope, as the number of surplus sows increases, because there are not enough pens. Due to this non-linearity we cannot use the expected probability to calculate the income, but must take the expectation for each of the possible number of farrowings. On the other hand, such a direct application of the income curve to the probability distributions will not suffice either. The problem is that the capacity restrictions will make the farmer react in

case of surplus. This means that the probabilities used in the Markov Chain model are not correct, because they do not only depend on the state of the animal but also on the state of the other animals. The remedy to this problem within the Markov Chain approach is to increase the state space to include the states of all the animals simultaneously. This will make the problem grow prohibitively large, i.e. we are confronted with the curse of dimensionality as discussed in Chapter 13.

Though the Markov Chain approach enables us to obtain the full probability distribution, we still have the problem that this probability distribution is only correct, if the animals are independent of each other. If capacity restrictions etc. exist this is not so.

14.3.3 Conclusion

The two simulation approaches have their limitation because of the implied assumptions of the methods.

Monte Carlo simulation offers a flexible approach for solving these inherent problems. Of course, Monte Carlo simulation has other limitations, such as computation time and lack of a robust search strategy for optimality. Optimally, Monte Carlo simulation should be used to supplement other probabilistic approaches such as Markov Chain or Bayesian Networks.

14.4 An object oriented model of a pig herd

To illustrate the modeling process, a description of a research project with an object oriented simulation model of a pig herd is presented in the following. The description should mainly serve as an illustration of the object oriented concept. It is based on Jørgensen and Kristensen (1995).

As in other simulation models the herd is described consisting of the *animal*, and its biological *states*, as well as the housing system, the *confinement*. The rest of the production system plays a major role, i.e. the observation, the updating and filtering of information, decision support systems, the decisions and the corresponding actions that are carried out. An outline of the production system is shown in Figure 14.2.

The system, a pig herd, is not stable, and rely on the manager's continuous correction of the production flow. The *manager* uses his knowledge or *belief* in the *state* of the herd. This belief is based on *observations* with varying precision of the herd, e.g. observations of weight of the individual animal. The *state* is measured using a *measuring device* with given precision. The resulting *observation* is sent to a *Belief Management System (BMS)* that handles updating and filtering of the observation. The result is a *belief* in the *state*. This *belief* can be used in a *decision support system (DSS)*. The output from the *DSS* is a *suggested decision*. The *Manager* then makes the final *decision* based on the information from the *DSS*. Finally the *decision* is turned into an *action* by the *staff* in the herd. *Actions*

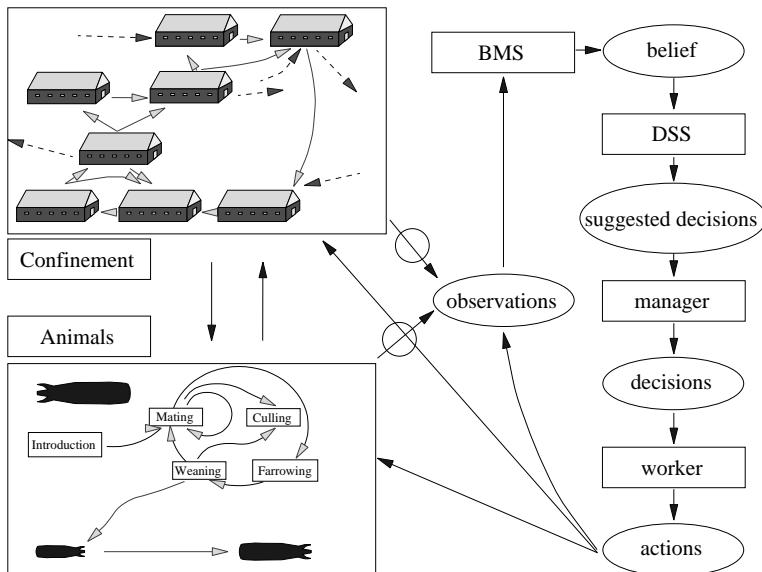


Figure 14.2: The elements of a pig production herd

comprise e.g. matings, weanings, movements, and observations. The objects in the model communicate by a message passing scheme that also handles the dynamic updating of the model.

14.4.1 General layout

The simulation model is programmed in Borland Pascal 7.0 as a program to be run under Windows. The program has two basic classes: *DinaPigObjects* and *Messages* sent between two *DinaPigObjects*. A *message* consists mainly of a message *time*, a *Receiver* and *Sender ID*, a *command* and the *strength* of the command. Similarly the main features of a *DinaPigObject* are an *ID* and a method for handling *Messages*. The main structure in the program is a sorted list or queue of *Messages*, sorted according to message *time*. The simulation is carried out by retrieving the first *Message* in the queue. The *system time* is updated to the message time. The receiver of the message (a *DinaPigObject*) is asked to handle the *message*. The *message* is then deleted and the next *message* in the queue is processed. The handling of the *message* will typically produce additional *messages* to other objects, as well as to the receiver itself. The scheduling of the simulation model is handled by the message processing. The posting of a message from the receiver to itself with a time lag allows for adjustment of the updating frequency of the model. The first step in the handling of messages by a *DinaPigObject* is an updating from the previous to the current system time. The scheduling method is thus a modification

of the event-driven method, where additional events are generated with fixed time intervals, but the updating may occur with varying time steps.

This comparatively simple structure can handle the full complexity of the simulation model.

Seven main descendants of the *DinaPigObject* have been defined, i.e. the *animals*, the physiological *states* of the animals, the *confinement*, the *manager* and the staff, the *measuring device*, the *belief management system* and the *decision support system*. These will be described in the following. Different descendants of the *Message* class are used to handle the *observations*, the *beliefs*, the *suggested decision*, *decisions* and *actions*.

14.4.2 Modeling the confinement

Objects with relation to the confinement are based on a class that inherits the features of *DinaPigObject* but in addition has three lists assigned. A list of points defining the outline of the confinement, a list of resources, e.g. drinking nipples, feeders, and an item list of other items. The main descendants of this class are *Departments*, *Sections* and *Pens*. The item list of a *department* consists of *Sections*. The item list of a *section* consists of *pens* and the item list of a *pen* consists of *animals*.

14.4.3 Modeling the animal and its physiological states

The model of the animal itself is very simple, as the detailed model is handled in the description of the different physiological states of the animal. An *animal* simply has a list of states assigned and a list of internal messages between the states of the animal. The message handling starts with an updating of each state in the state list to the current simulation time.

State description

Each animal has a list of state's assigned. Each state models a physiological process. As mentioned the states are assigned an animal. Each state has a transition function that updates the previous physiological state to the state at the current time. In addition some states are observable and some states can be influenced by outside events. Current states are growth process, death-process, generic disease process, puberty, oestrus, pregnancy and lactation.

14.4.4 Modeling the measuring device

The modeling of the physiological state of the animal relies, as shown above, on detailed models and a number of parameters. This is in order to get as close as possible an approximation to the real, biological system.

This detailed knowledge of the biological systems is in general not available at the herd level. Instead, the pig producers have to rely on observations in the herd.

It is important to make this distinction in the model. Therefore, all information to the pig producer has to be based on observations. To represent this in the model the concept of a measuring device is used. The concept covers both actual devices, e.g. weighing scales, pregnancy testers, as well as virtual devices, e.g. assessment of the health state or weight of an animal by eye.

In the model the manager sends a message to the measuring device in order to obtain an observation on an animal. The measuring device in turn sends a message to the animal and the animal returns a message with the requested value in the strength parameter of the message. The measuring device transforms the strength and returns the value to the manager.

The simplest measuring device is a device for making quantitative measurements. It simply adds a random noise to the value received from the animal. The measurement device is thus defined with a measurement precision (i.e. the reciprocal of the variance of the random noise). Another class of measuring devices is used for categorical measurements, e.g. pregnancy testing, heat detection, disease diagnose. These devices return a measurement such as pregnant/not pregnant, diseased/healthy. These measurements are prone to mistakes, just as the quantitative measurements. The definition of these devices relies on the concept of an underlying continuous distribution with threshold values to define the intervals for the different categories. The strength returned from the animal is added a random noise (with a given precision) and subsequently categorized according to the threshold values. In this manner different precision and sensitivity can be modeled. E.g. the disease state of the animal described above will return a 0 if the animal is healthy and 1 if the animal is infected. This value is added a random value and the animal is classified as diseased if the resulting value is above e.g. 0.75.

14.4.5 Modeling the belief management system

The observations made should be used for adjusting the pig producers belief in the state of his herd. In order to represent this in the model the concept of a Belief Management System is introduced. This is a general framework that should cover simple systems that e.g. calculate monthly averages of the production trait as well as complicated Bayesian Networks that uses Bayesian methods for a correct handling of the observation. (The term BMS is inspired by the terminology used in relation to causal or Bayesian Networks, e.g. Jensen, 1996, 2001). Ideally, even the subjective processing of information in the mind of the pig producer should be covered by this concept. However, our work has not yet carried us that far.

The “ideal” Belief Management System should simply use every measurement in the herd in order to estimate the parameters of the detailed biological models using advanced statistical methods. The contrast between this ideal and the method used in the real world, by pig producers as well as researchers, is striking. Most management information systems rely on rather simple techniques, such as quarterly or monthly averages of key production traits (cf. Chapter 5). A few incorporate methods from production control such as CUSUM-charts. Recently,

more advanced techniques have been applied, such as Kalman filtering techniques. These techniques combined with the techniques of Bayesian nets are promising candidates for the future BMS. Even though these techniques can solve very complex problems, the size of the problems will still set an upper limit to what can be solved. So, a BMS will have to rely on an approximation. It is important to realize that a more complex system is not necessarily better from an economic point of view, if the cost of obtaining the information is included.

It is not the intention to program the BMS directly in the simulation model. Instead, the simulation model communicates through a standard interface with the BMS. The observations from the model are transformed into beliefs. When the manager or DSS needs information concerning beliefs, they will send a message with a request. The BMS returns the belief.

The belief itself can be represented with varying simplicity. If the BMS only calculates monthly averages, it would be natural to let the belief in the production of an animal be the monthly average. More advanced BMS would use the prediction of the future monthly average based in the historical data. The belief may even be based on the individual animals own performance, and may include the variance as well as the expectation.

Clearly, some representation of beliefs would be better than others, but this would be seen as a result (i.e. poorer production due to worse decisions), rather than as a defined standard for the BMS. This ensures that every method of data processing can be described as a BMS.

Some of the aspects of a BMS are currently handled in the DSS, especially in those based on Markov Chain methodology. From a conceptual point of view, the distinction between BMS and DSS has proved to be valuable. In any case many information systems exist that would be categorized as BMS without any DSS element in it. Similarly many DSS exist with no calibration to the actual registrations in the herd.

14.4.6 Modeling the decision support system

The DSS in the simulation model is not intended to be programmed directly into the model, at least not for the more complicated models. Instead interface between DSS and the simulation model is established. The DSS can be of varying complexity including simple 'rules of thumb' as well as complicated systems, such as those based on Markov decision programming. Some of these DSS have included elements that in our terminology would be called BMS. This is, however, only a problem from a more formalistic point of view.

The DSS responds to messages from the Manager requesting suggested choice of actions. A well-established problem solved by the DSS is the selection between the actions cull the animal or keep the animal. Other important operational actions are buying and selling of animals, movement of animals (weaning - regrouping etc), mating (actual decision 'refrain' from mating sows in heat), induction of events, influence nutritional status/growth/meat quality through feeding, observe,

timing of actions. Many of these actions are not currently covered by DSS but are handled by heuristics. The heuristic rules will often differ between herds and the economic basis of the rules is not always adequate.

The DSS will usually base its suggested decision on the belief in the herd. Therefore, the BMS and DSS will often be tightly connected

14.4.7 Modeling the manager and the staff

From the start of project, the role intended for the manager in the simulation model was to transform suggested decisions from the DSS into actual decisions. This transformation should be based on e.g. different objectives, additional information not included in the DSS, constraints unknown to the DSS, and a possible irrational component.

An example of such a transformation is that the DSS suggest that the farmer should cull sow no. 101. The farmer decides to keep the sow. Possible reasons could be that he will wait and see, if there is a replacement gilt coming in heat at the right time, that the sow is easy to manage, that one of the other sows in the group has a sore leg and he wants to have a surplus “just in case”, that the sow is the only one remaining from the foundation of the herd or that the piglets have a “cute” coloring. From the pig producers point of view all these reasons are valid and should be taken into account although they are not included in the DSS.

This original role has still been kept, but some of the roles originally placed in the DSS and BMS are still modeled as part the manager. This corresponds to the role played by the manager in previous simulation models, where the management were mainly described as a set of heuristic rules.

The work with the model has shown it of value to define *experts*, management *tasks* and management *schedules* in order to structure the program. The experts are descendants of the *manager* class and handle a subset of the management function, e.g. buying and selling of animals, weighing of animals, mating of animals. *Tasks* are mainly lists of linked actions, i.e. the movement of the pigs from one pen to another involves several actions, e.g. find a section with empty pens, select the best pen, move the pigs from the old pen etc. Typically the *task* sends a message to other objects and then waits for the responding messages before continuing. A *schedule* is a description of tasks to be performed with regular intervals, e.g. weekly or monthly.

14.5 Representing model dynamics

The dynamic systems that are modeled do of course evolve in continuous time. In order to simulate the system on the digital computer, we have to make a decision, how to represent the continuous time scale on the computer.

14.5.1 Selection and effect of stepping method

First of all, a variable called *system time* is defined.

Within animal simulation the most frequent approach is the *fixed-time* incrementing procedure. Using this procedure, first the system time is advanced with one unit. Then each element in the model is *updated* to this time, by determining the events that has happened, and the changes in the state variables. This procedure is in close accordance with the formulation of the models with difference equations and discrete time Markov Chains.

Another approach *next-event incrementing* that is the most important in simulation of event driven systems (e.g. queuing systems). With this method the system clock is incremented by a variable amount. Each time an element of the system is updated, the time of the next event for that element is calculated (and elements influenced by it). These future events are then stored in a sorted list. The *next-event incrementing* simple sets the system clock to the time of the first element in the list, updates this element (and generates new events for the element) and continues with the next element on the list.

Each approach has its advantages. Continuous growth processes will be difficult to represent by the next event approach, while the handling of almost simultaneous events is difficult in the fixed time approach. If only the first three animals in oestrus should be mated, how do we decide which of the animals, it was, when they in principles show heat simultaneously.

Of course the methods can be combined, a simple approach is to include an artificial event with fixed time steps for each element in the model. The elements will then be updated at least as frequent as the length of these time steps. This is the approach used in the pig simulation model in the previous section.

The choice of the length of each time step in the fixed time incrementing approach is another important decision to make in simulation models. Formally the output of the simulation model should not depend on our choice of time stepping. If the simulation should claim to represent the real system, a choice of calculation method should obviously have no influence. Very often the initial specification of the model does not allow changes in time step length. Therefore before we proceed we will look on the two possible specification methods.

14.5.2 Difference equations

Let us describe a simple growth model.

The animal is given an amount of feed, ΔFU , on a given day. This daily feed intake is split into feed used for maintenance and feed used for growth.

$$\Delta FU_{\text{maintenance}} = k_{\text{maintenance}} W^{0.75} \quad (14.7)$$

$$\Delta FU_{\text{growth}} = \Delta FU - \Delta FU_{\text{maintenance}} + \varepsilon_{fg} \quad (14.8)$$

The random term ε_{fg} is included in order to model the stochastic element in the model.

The growth can then be calculated:

$$\Delta W = k_{\text{efficiency}} \Delta FU_{\text{growth}} + \varepsilon_g$$

And the weight and total feed summarised

$$FU_{t+1} = FU_t + \Delta FU_t \quad (14.9)$$

$$W_{t+1} = W_t + \Delta W \quad (14.10)$$

These so-called state variables are then kept until the next period.

The stochastic elements are often assumed to be distributed independently and $\mathcal{N}(0, \sigma_\varepsilon^2)$ (A different variance for each residual).

This is the difference equation approach and it seems to be the standard formulation of models within animal production.

Note that the length of the time step (1 day) is not formulated explicitly. By selecting this time step in the model description, we are stuck with it. If we want to change the length of the time steps in the model the formulas must be reformulated. If we take the weekly allowance of feed and base the calculation on that, we get a different answer than by applying the formula seven times in succession. As the animal grows, the metabolic weight changes and this results in a lower daily gain. The time step one day seems “natural”, but we may be interested in making shorter time steps, in fact the feeding of animals is seldom an event that only occurs once a day. Again the equations that we would have to use would be different (even though the difference would be smaller).

14.5.3 Differential equations

To avoid this, the model can be quantified on a continuous time scale as differential equations.

The rate of feed energy used for maintenance and the rate of feed energy used for growth.

$$\frac{dFU_{\text{maintenance}}}{dt} = k_{\text{maintenance}} W^{0.75} \quad (14.11)$$

$$\frac{dFU_{\text{growth}}}{dt} = \frac{dFU}{dt} - \frac{dFU_{\text{maintenance}}}{dt} + \varepsilon_{fg} \quad (14.12)$$

The growth rate can then be calculated

$$\frac{dW}{dt} = k_{\text{efficiency}} \frac{dFU_{\text{growth}}}{dt} + \varepsilon_g$$

We also have to specify an equation for the rate of feed available. The simplest case is that the feed is available at a constant rate.

$$\frac{dFU}{dt} = k_{\text{feed}}$$

The weight and total feed consumption can then be found, as a solution to the differential equations.

The random terms in the model are normally assumed to follow the so-called Wiener process.

Solution of the equations

Even though the formulas in the difference equations and the differential equations look alike, it should be emphasized, they are *not* the same. However, in general such differential equations have to be solved by numerical methods, and one of these methods, Euler's method, uses the difference equations as approximations. It is important to recognize that Euler's method is an inefficient method for solving differential equations, and other far more precise and fast methods exist. One of these, the Runge-Kutta method, should be easy to implement in any simulation model. Other methods use a variable time stepping approach, i.e. the length of the time step is varied, until the difference in the function between the two time step is sufficiently small. See any textbook on solving ordinary differential equations for reference (e.g. Press et al., 1989). The Wiener process in continuous time corresponds to the Normal distribution in fixed time interval. Further treatment of this process can be found in textbooks concerning Stochastic processes (e.g. Hoel et al., 1972).

Because of the interaction between model elements, the variable time step method can be difficult to use in a herd simulation model. But a good idea is to compare model outputs with time steps of different length, to choose the method that results in the best trade off between computing time and accuracy. To be honest, very few simulation models within animal production have been through such scrutiny, at least none has been published to our knowledge. The usual time step is either one day or one week.

14.5.4 State transitions in continuous time

A similar discussion, as the one made above, can be made for the modeling of state transitions. Markov chains cover both discrete and continuous time.

The Markov property is that if we know the current state, the future is independent of the past. In Markov chain models this property is ensured by defining a large discrete state space. The transition probability from the state thus only depends on the state, and not on the history. In Monte Carlo simulations models we do not have this property, instead the state space is defined by a mixture of continuous and discrete variables, as well as discrete states. Naturally these variables will interact. The changes in the continuous variables will depend on the discrete variables, and vice versa.

It is often convenient to represent the transition from state to state in continuous time by the intensity of transition. The intensity of transition is defined as

$$h(t) = \frac{f(t)}{1 - F(t)}$$

The function $h(t)$ is also called the hazard rate. $h(t)$ corresponds to the probability that an event will occur within the next infinitesimal small time step, if it has not occurred yet. If we look at a time interval the probability that the event will occur

within the interval is.

$$P(\text{Event}) = \exp\left(-\int_0^t h(t)dt\right)$$

Several important statistical distributions can be easily formulated using the intensity. A process with constant intensity will follow the exponential distribution. Processes with increasing intensity can be used to represent aging processes.

In a simulation model we would usually want to represent the intensity just as we represent other continuous variable in the model, i.e. if the death intensity is linearly related to the weight.

$$\frac{dh_{\text{death}}}{dt} = k_{d0} + k_{d1}W$$

Instead of modeling the intensity directly in this manner, it is often better to model the logarithm of the intensity, because $h(t) > 0$, as can be seen from the definition. This constraint is automatically ensured if the model is on the logarithmic scale.

14.5.5 Competing events

When events are modeled in discrete time intervals, it is important to consider if more than one event can occur in the interval. Typically the events (state transitions) influence the subsequent intensity of transition, and Eq. (14.5.4) cannot be applied directly. The probability of transition from a given state is not simply the sums of the probabilities defined in Eq. (14.5.4).

A method for handling this problem is to calculate the time of transition rather than the probability of transition. If transitions take place within the time interval the first of the transitions are selected. Then the remaining part from the first transition to the end of the time step is treated, in similar manner until no more transitions take place.

14.5.6 Discussion

In the literature concerning simulation models, the difference between the two kinds of models is seldom recognized. Usually one must assume from the context that the equations are difference equations, but on the other hand, a term such as energy used for maintenance, seems only to be natural, if it is understood in terms of differential equations, i.e. in continuous time.

It can strongly be advised to specify the model on a continuous scale, i.e. as differential equations, as the number of artifacts introduced in the modeling process will be minimized in this manner.

14.6 Stochastic specification

The specification of the stochastic part of the model is essential for the validity of the model. As mentioned previously, simulation modeling within animal pro-

duction relies very much on the systems approach and not so much on statistical methods. Therefore, there has been comparatively little effort to describe, how to proceed with the specification of the stochastic part of model. The concept behind relies on piecing together of data from several sources, and experiments and it is not understood formally what is actually done.

In the following guidelines for the specification of animal production herds will be presented. The guidelines are based on our own experience and are very much influenced by the specification of e.g. statistical programs such as BUGS (see Liu, 1996; Spiegelhalter et al., 1996, where the latter is the BUGS manual which is available on the internet). The overall goal is that it should in principle be possible to estimate every parameter in the model using a formal statistical approach. The additive model used within animal breeding and statistical methods for time series analysis has also inspired the approach.

The underlying idea is that the stochastic specification should reflect the uncertainty in the parameters. Some variables we know only a little about, either because they are not easy to observe, or because they vary much, and we have to observe e.g. the specific herd for a while before we know the level in the herd. The prior distribution of such variables should be more vague, than parameters we are sure of. As mentioned this is especially important when considering the use of the simulation model for decision support purposes.

The hierachic system of the production herd should be represented, where it is possible. Animals within a herd at a given time are not statistically independent individuals. But if we model the herd level separately we can catch much of the dependency, and treat the individuals as independent conditioned on the herd level. Within slaughter pig production the same refers to section and pen within section.

14.6.1 Example: use of results from an experiment

When using results from different sources, one very often has to rely on published results, e.g. the mean level of a parameter and a standard error of the parameter. Usually correlations between estimated parameters is not published, though they can be high.

A relevant and very simple model is

$$k_{0_{ij}} = \mu + H_0 + L_{0_i} + A_{0_{ij}}$$

where $k_{0_{ij}}$ is the parameter value for the j th animal within the i th litter with herd 0 (the experimental herd). μ is the true mean of the parameter, H_k is the random effect of herd $H_k \sim \mathcal{N}(0, \sigma_H^2)$, H_0 is one instance of H_k . L_{0_i} is the random effect of litter i , $L_{0_i} \sim \mathcal{N}(0, \sigma_L^2)$ and $A_{0_{ij}}$ is the effect of animal $A_{0_{ij}} \sim \mathcal{N}(0, \sigma_A^2)$

The published results return the information that

$$\mu + H_0 \sim \mathcal{N}(\hat{k}_0, \hat{\sigma}_e^2)$$

where σ_e is the standard error of the estimate.

We want to use information from this experiment in our simulation model, i.e. we want to know the distribution of the parameter value for a new animal in a new herd.

$$k_{kij} = \mu + H_k + L_{ki} + A_{kij} \quad (14.13)$$

$$\text{E}(k_{kij}) = \hat{k}_0 \quad (14.14)$$

$$\text{Var}(k_{kij}) = \sigma_e^2 + \sigma_H^2 + \sigma_L^2 + \sigma_A^2 \quad (14.15)$$

The expectation of the value is the estimated level from the experiment, but the variance is still not known. The experiment does not give any indication of the variance between herds, as it is only carried out within a single herd. The experiment might give us information concerning the two final variance terms, but that depends on the design. If litter is treated as a block in the experiment we can obtain an estimate of σ_L^2 , but the variance between blocks may be larger because the experimental procedures are based on the block design. Anyway, the variance between blocks is seldom reported. In the block design the residual error will give us an estimate of σ_A^2 . If it is not reported, it can usually be found from the standard error of the parameter estimate. If the parameter value is based on litter averages the residual error cannot be used directly as it is a sum of litter and animal effects.

Of course it is not only the estimate of variance but also the precision of the variance estimates that should be considered. Usually the standard error will have a relatively high precision and possibly litter and animal variance as well, while the between herd variance will be very hard to estimate precisely, especially if we consider parameters that need intensive physiological studies to measure.

In order to model uncertainties in the variance estimates the gamma distribution should usually be used

If we combine several experiments, we can pool these experiments, if we are sure that they have measured the same parameter. The difference between the experiments gives us some indication of the variance between *experimental* herds, and this information might be used as an indication of the variance between *emphproduction* herds. Within statistical research this is called *Meta analysis* (e.g. Morris and Normand, 1992).

14.6.2 The principles

The underlying principles in the example above are, that we identify different populations. We consider a population of herds, a population of litters in each herds, and a population of pigs within litters. We assume that the effects of these populations can be described as additive, independent, Normal deviations from the general mean. The correspondence to the model used in quantitative genetics within animal breeding should be clear. Populations such as breed and geographic region may be relevant, as well, and in addition a more abstract one, the population of measurement errors.

The example used above, was static in the description. But if we look at herd results, a general observation that the herd level fluctuates over time. If we consider the effect of the animal, it is also reasonable to assume that the animal's difference from the herd level will change over time, rather than be constant.

A time series representation of this is relevant, i.e. a stationary first order autoregressive process:

$$H_{t+1} = \alpha(H_t - H_0) + (1 - \alpha)\varepsilon_H$$

The variance of ε_H will correspond to the variance between herds, while α determines the dependency of subsequent points. At the herd level the stationary process is the natural assumption, as the variance in the herd level is expected not to increase steadily. With respect to the animal, non-stationary time series may be relevant, e.g. for growing animals. Probably processes of higher order are necessary. A model with changes at different time scale is probably the most appropriate, i.e. the herd level is modeled as the sum of two effects, one changing at a time scale of 1 year and the other at the time scale of 1 week.

Seasonal variation and trends can also readily be implemented in the time series approach. Such effects are most naturally modeled at the herd level.

Sudden changes in level are also part of the real system. The time series approach makes this formulation easy. Random changes can be modeled in the following manner. At each time step there is a probability for a change. If the change occurs then the herd level changes randomly, not with the current time step, but with a much larger, i.e. the correlation between the current herd level becomes much lower. The magnitude of the sudden changes can of course be modeled directly, i.e. the change is simply added to the current herd level.

$$H_{t+1} = S_t + \alpha(H_t - H_0) + (1 - \alpha)\varepsilon_H$$

where S_t is the magnitude of the sudden impact. The resulting change is an abrupt change in level, followed by a gradual decline towards the mean level.

These techniques are primarily used on the herd level, but are also relevant on the animal level, eg. to model increased mortality after moving the animal to a new environment (e.g. weaning) and the effect of boar stimulation on puberty attainment.

The last two effects will have to be modeled on a continuous scale even though they model discrete events. As discussed in the previous section this can be done by modeling transition intensities rather than probabilities.

Probability distributions other than the normal distributions cannot directly handle such aspects as different populations and time series fluctuations. The approach to be used, therefore is to model the parameters of the probability distribution rather than the probability distribution it self. Often the parameters are transformed using a suitable function such as the logarithmic or the logit-transform. This technique corresponds to the generalized linear model approach within statistics (not to be confused with the general linear model).

An example of this approach is given in Chapter 6 when considering categorical measurements.

14.6.3 What happens if the a priori distribution is too vague

If the model gives results that vary too much or don't make sense at all, it is an indication that our prior knowledge of the system is more precise than we have indicated in the a priori assumptions.

This is *not* a problem, in fact it is the advantage of using the stochastic approach to modeling, i.e. we can find the posterior distribution of the parameters. We can document a procedure of specification of vague priors, running the model, finding the a posteriori distribution of the parameters even the latent parameters and then use this knowledge in the final model specification.

The only problems are that it is time consuming, and there will be a very high correlation between some of the parameters. This is a result of the over parametrization of the model, e.g. high growth rate can be a result of either high feed intake or high utilization of the feed. From a herd point of view the cause is not as interesting as the result, and the high correlation is an indication that a simpler model is just as valid for our purpose.

The easiest way of updating the distribution is simply to define a plausible region of outcomes of the model, discard those simulation runs that are outside the plausible region, and investigate the distribution of the input parameters in those simulation runs that have been kept. Either the parameter space can then be redefined or the input values in the remaining data set can be used as basis for re-sampling. See further in section 14.11.3.

14.6.4 Initializing simulation model from herd data bases

In the context of Markov chain decision modeling the concept of steady state distribution was presented, i.e. in the long run, the expected distribution over states would be constant, irrespective of the initial value. The same applies to Monte Carlo Models, except for very special cases. When studying decision rules in general, the effect of initial conditions can therefore be removed with long simulation runs. An arbitrary start configuration of the herd can simply be selected, and after a period depending on the system in question, the effect of the initial configuration will be diminished. This initial period is simply discarded from considerations concerning the decision rules.

If the interest is in decision support in an actual herd, it is not only the long run properties that are of interest, but also the short run development of the herd. Therefore it is of interest to base the simulations on the current state of the herd, i.e. the start configuration is based on the recorded values from the herd.

In other words, we need a transformation from the belief in the herd to the actual state of the herd. The state of the herd comprises the state of the animals and the state of the manager's belief in the state of the herd.

The herd average in litter size for 1st parity (H_1) has as prior distribution $\mathcal{N}(9.5, 0.25)$. The effect of the sow (deviation from herd mean, A) is distributed as $\mathcal{N}(0, 1.25)$ and the independent residual ε is distributed as $N(0, 5)$. 20 sows were observed with an average of 10 and variance 6.6 , i.e. the average distributed as $\mathcal{N}(0, 0.33)$, i.e. according to the model:

$$Y_{ij} = H_i + A_j + \varepsilon_{ij}$$

where Y_{ij} is the litter size in parity i for sow j .

The posterior distribution of the herd level is then

$$\mathcal{N}\left(\frac{9.5 \times 4}{3+4} + \frac{10.0 \times 3}{3+4}, \frac{1}{3+4}\right) = \mathcal{N}\left(9.7, \frac{1}{7}\right)$$

The simulated value of H_1 is drawn from this distribution. With known H_1 we can find the value of A_j conditioned on the observed litter size:

$$A_j | (Y_{1j}, H_1) \sim \mathcal{N}\left(\frac{1.25}{5+1.25}(Y_{1j} - H_1), 1.25 - \frac{1.25^2}{5+1.25}\right) = \mathcal{N}(0.2(Y_{1j} - H_1), 1.0)$$

From the specification the individual sow values can be sampled.

Note that for the belief management part of model, the herd distribution is available, but the uncertainty concerning the herd level should be included in the sow values, i.e. the herd levels and the sow values will be correlated, and the conditional variance of the sow value will be $1 + 1/7 \approx 1.15$.

Though, this procedure seems straightforward, it is mainly because we have ignored some of the problems. First of all, how far back in registrations should we base our estimate of the herd level. Clearly, very old registrations do not tell any thing about the current level. Here the methods for time series analysis and monitoring should be used in order to estimate the current level. A second and more difficult problem is that the procedure is only valid for 1st parity. When the managers have observed the litter size in the first parity, management decisions will decide whether the sow will continue in the system or not. This can be remedied by using the correct statistical method, i.e. we should combine all previous observations for the sow in order to estimate A_j , and in addition the herd levels at each of the sow's previous litters.

In many cases the effect of this can be ignored, when simulation models are initialized. But if for instance the effect of a Monte Carlo replacement DSS is to be explored, we cannot ignore the problem. These systems rely on the combined information from the previous history of the animals. If this information is discarded when the model is started, the use of the simulation model for evaluation purposes does not make sense.

To summarize, the initialization from herd data is not a trivial matter, and it should be thoroughly explored. The initialization is an estimation of the parameters in the model, and on the individual level, and not only the parameters in the state of nature. Because of the complexity of this estimation, the initialization is often

based on several hidden assumptions. It is important to specify and investigate these assumptions.

The safe approach is to base model inference on steady state results.

14.6.5 Probability distributions

Comprehensive discussion of these functions can be found in many standard statistical textbooks (DeGroot, 1970, is the classical reference). Therefore, only some general aspects of the choice of distributions shall be discussed.

It is strongly advised to select distributions with well studied statistical properties. The Gaussian (or Normal) distribution has many properties that make it the natural first choice, but other distributions, especially for discrete events, and for between event intervals should be considered. As an example, a sum of exponentially distributed variables is Gamma distributed. The Gamma distribution is thus a natural choice, if the processes modeled are the result of several underlying physiological steps. Because of previous difficulties with random number generation, the triangular distribution has often been the preferred choice within farm management economics. This cannot be recommended. Two justifications for this will suffice. It is not tractable with statistical methods. Therefore, results from data analysis cannot be directly applied to the model. In addition, it specifies a rather narrow range for the observations. In Bayesian statistics a prior probability of zero, is very strict. No matter what evidence becomes available, this prior will never be changed. Zero probability means this can never happen in all eternity!

14.7 Random number generation

The generation of random numbers is the core of the Monte Carlo technique. We will not go into detail with the technique, but only give a review, with some examples included. A detailed account can be found in Press et al. (1989).

Two aspects are important.

- Generation of a sequence of independent uniform deviates within a specified range.
- Transformation of the uniform deviates into different probability distributions.

14.7.1 Generation of uniform random numbers

Uniform deviates are just random numbers, which lie within a specified range typically 0 to 1, with any number in the range just as likely as any other. In most compilers and programs such as SAS and spreadsheets, there is a facility for generating such numbers. Typically, there is a function call to initiate the random sequence (in Borland Pascal (BP) it is *Randomize*), and a function call to obtain each number in the sequence (in BP *Random*). If *Randomize* is used the computer's clock

is read and the value used for starting the random sequence. This sequence will change each time the program is run. Another possibility is to assign a value to a variable (in BP the *Randseed* variable) and then generate a standardised sequence of random numbers. This latter possibility should be preferred, because it will be possible to replicate the calculations. This can be an advantage, when debugging programs and it may be used in the statistical analysis of the simulation runs in special cases.

The standard procedures can be misleading and should be treated carefully. They usually use the so called *linear congruential* method that generates a sequence of integers I_1, I_2, \dots each between 0 and $m - 1$ by the recurrence relation

$$I_{j+1} = aI_j + c \mod m$$

where a and c are positive integer specific for the system routine and m is usually the maximum integer value. If the values of the constants are chosen properly all integers between 0 and $m - 1$ should eventually turn up in the sequence. In this case the choice of random seed I_0 is not important. In BP the choice of m and the other constants is not documented. At least in BP 7.0 m seems to be 32 bit value (*Longint*), but in earlier versions it may be a 16 bit (word) or even an 8 bit-value (*Integer* - maximum value 32767). Too small a m value results in too few different values in the cycle, and a correlation between the subsequent random numbers. In the former case a real number is returned and the full 32 bit representation of the *Longint* type is utilised, while in the latter case only a number of digits corresponding to the type of parameter is used).

As the last digits in the numbers are much less random than the first digits, when the *linear congruential* method is used, a random deviate should never be split into several values.

Press et al. (1989) show several possibilities for improving the system supplied random generator, such as shuffling the random numbers to avoid the serial correlation between the numbers. Such methods should be implemented in every simulation model.

14.7.2 Sampling from probability distributions in general

The uniform deviates can rarely be used directly, but have to be transformed into the relevant probability distributions. Fortunately, efficient algorithms exist that can perform this transformation. The main principle behind the transformation is illustrated, if we consider the discrete case, e.g. the rolling of a dice. Each outcome $\{1, 2, \dots, 6\}$ has a probability of $1/6$ of happening. The accumulated distribution (that is, the probability of obtaining at least 1, at least 2 etc) is $\{1/6, 2/6, \dots, 6/6\}$. Therefore, the algorithm for simulating the outcome of a dice is to draw a uniform random deviate, u , if u is between 0 and $1/6$ then the simulated outcome is 1, if u is between $1/6$ and $2/6$ the outcome is 2 etc. The principle is illustrated in Figure 14.3a.

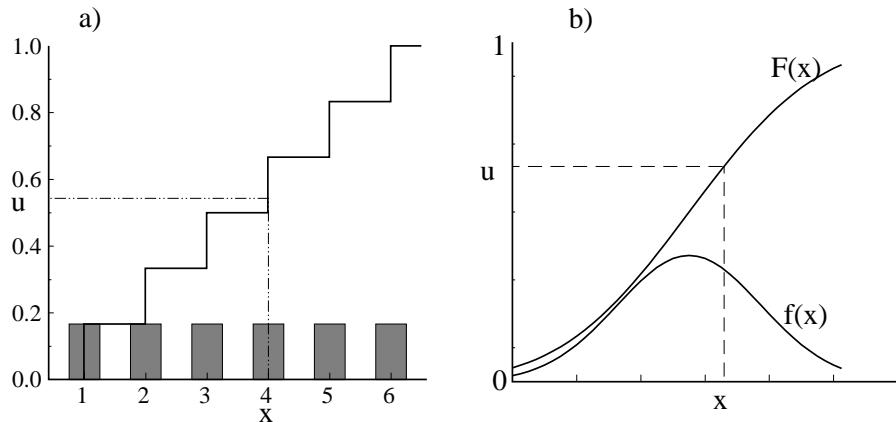


Figure 14.3: Transformation method. (a) The drawing of random values for a discrete value distribution. (b) The drawing from a continuous distribution

Similarly for continuous variable with density $f(x)$. We integrate the function to obtain the (accumulated) probability distribution, $F(x)$. We draw a uniform deviate u , and find the variable x with the density $f()$, by using the expression.

$$u = F(X) \Rightarrow x = F^{-1}(u)$$

The method is illustrated in Figure 14.3b. This general approach can be used for any distribution. However, for many distributions the calculation of F^{-1} is very time consuming. The use of Eq. (14.7.2) is therefore usually restricted to distributions where the inverse can be easily found, such as the Exponential, Weibull and Geometric distribution.

If the calculation of F^{-1} is too time consuming, the so-called acceptance-rejection method can be used. In this method another function $g(x)$ that “covers” or dominates the density function $f(x)$ is selected (“covers” means that $g(x) > f(x)$ for every x). $G(x)$ is defined similarly to $F(x)$, i.e. as the integral of the function over the sample space. Note that $G(x) \rightarrow A$ for $x \rightarrow \infty$, where A is a positive constant $A > 1$, in contrast to $F(x)$ where the limit is 1. The choice of the function $g(x)$ is based on the ease of finding G^{-1} and its proximity to $f(x)$, A should be as close to 1 as possible. An illustration of the method can be seen in Figure 14.4.

The acceptance-rejection algorithm then selects a uniform random deviate on the interval $0, A]$ and finds x by using Eq. (14.7.2) with G instead of F . A new uniform random deviate is then selected on $]0, 1]$ and the value of x is accepted

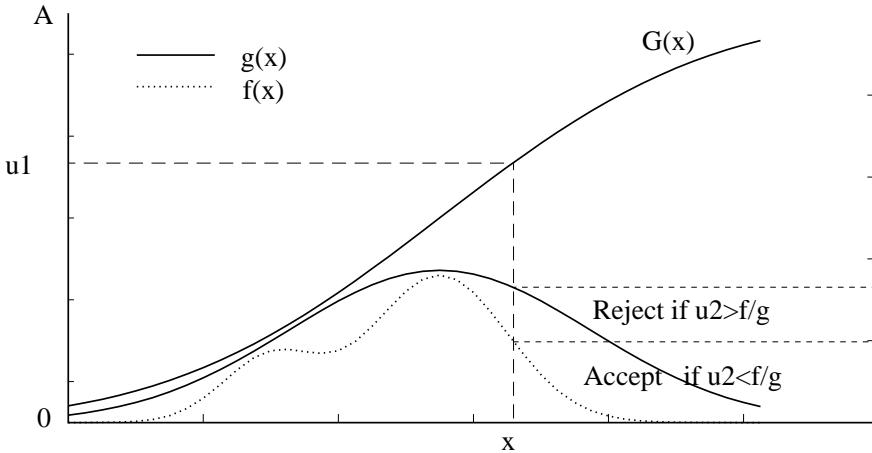


Figure 14.4: The acceptance-rejection method for generating random values

if $u < f(x)/g(x)$, if not it is rejected and the sampling starts a new with a new random deviate on $]0, A]$. The method uses on average A attempts to generate one random deviate. Therefore, A should be as close as 1 as possible.

The principle behind the acceptance-rejection sampling can be used in general. Effective adaptive methods now exist, that makes the specification of $g(x)$ simpler (Gilks and Wild, 1992; Gilks, 1992).

In the case of the normal distribution a third method is used (see Press et al., 1989). This method is based on some of the mathematical properties of the Gaussian distribution.

14.7.3 The multivariate case

Usually, the problem of drawing random numbers from multivariate distribution can be solved by sequentially drawing from the conditional distribution given the instances of the previous variables. In fact, simulation modeling is really a question of representing these conditional distributions (conditioned on the decisions as well).

However, In some cases it is an advantage to be able to draw a set of random variables simultaneously from a multivariate distribution, e.g. the multivariate normal case.

If the variable vector x of dimension $(n \times 1)$ is to be drawn from a $\mathcal{N}(\mu, \Sigma)$ distribution the algorithm is to draw a vector e of n independent normal deviates

with expectation 0 and variance 1. The vector x is then found from

$$x = \mu + Ce$$

where C is a lower diagonal matrix, such that $C^\top C = \Sigma$. Fast algorithms (Cholesky decomposition) for the calculation of C exists.

14.7.4 Metropolis-Hastings method

The Metropolis-Hastings method for drawing random samples from probability distributions has recently become extremely important within statistical research, especially the special case, the Gibbs sampler. For drawing random numbers for Monte Carlo simulation models the use of the method is not the obvious choice, but it may be in the future. In addition the method may be used to estimate the herd specific parameters in the state of nature, Φ_0 , and used as input to the simulation model. Therefore, the principle shall very briefly be discussed. A detailed description of the method can be found in Chib and Greenberg (1995).

The principle is best discussed based on Figure 14.3b. We start out with a point within the area under the density to be sampled. The idea is then to select a new point by moving randomly around within the area. We select the move randomly, but so that we are sure that we follow the “covering” density in our moves. Then a uniform deviate is drawn and the new point is evaluated just as in the acceptance-rejection algorithm. If we accept the new point, we output the point as our random number, if we don’t, we simply output the old point. This procedure is continued. In fact the method specifies a Markov Chain with the desired probability distribution as steady state distribution. Therefore, the methods are also termed Markov Chain Monte Carlo (MCMC) methods.

The benefit from this method is that no integration has to be done, i.e. the function $G(x)$ does not have to be found. Especially for multivariate distributions with high dimension this is an enormous benefit. Furthermore, the random moves can be selected such that first the move is selected in the first dimension, then in the second and so on cyclically through the parameter space.

The draw back, and for our purpose, it is a major one, is that the points are very much serially correlated, i.e. neighbor points are more similar than points far apart.

14.7.5 Sampling from a set of random deviates

If a series of random samplings has been performed previously, it is possible to reuse the samples in re-sampling plan. If a data base of N observations exists, a simple method is to sample each of these observations with equal probability.

One approach of doing this is to shuffle the observation set and then sample sequentially from the data set. A simple algorithm is to add a uniform random deviate to each observation in the data set, and then sort the data set according to this new value.

This approach is particularly well suited to specifying the state of nature distribution ϕ_0 as it becomes easy when comparing treatments in simulation runs, to secure that the same state of nature is used in both cases.

An additional benefit is that the MCMC-based methods for statistical analysis of data automatically deliver such a data set of random samplings from the desired probability distribution. Therefore, there is no need for the additional step of analyzing the output from the MCMC step. As mentioned previously the MCMC method produces a set of random deviates that are serially correlated, and they should therefore not be used directly as they are. The shuffling should remove the problem, but an additional precaution is only to use a fraction of the MCMC output, e.g. every k th observation where k depends on the magnitude of the serial correlation.

14.8 Representing Decision Rules

The primary consideration, when implementing decision rules within the simulation model, is to distinguish between the “real” world as represented by the parameters in the model, and the information that the manager can react upon, i.e. the observable part of the herd. This information is called the belief as mentioned in section 14.4. If e.g. the animal is in the pregnant state, this information should not be available to the manager, the decisions he makes should be based on registered matings, pregnancy diagnosis etc.

One should also be aware that the process of optimization is partly a process of learning about the model parameters. This is one reason that the uncertainty in the state-of-nature should be represented correctly. A very inflexible decision rule can be optimized within a constant state of nature, to even outperform a more flexible decision-rule. But if the state of nature varies as represented by the prior distribution, the flexible approach is the best, because the inflexible rule cannot adapt to the changing circumstances. This is an example of how the model specification and the results from the model are not always independent. To be strict these considerations also refer to the model structure. For example, if we model fluctuations in the herd level as a first order autoregressive process, a decision strategy that uses this model in a BMS will be better off than is actually correct, because we ignore uncertainty in model structure.

The decision rules usually represented within simulation models are rather primitive, and the average farm manager will probably react on a more informed level. It can therefore be advocated to use decision rules that specify how to use DSS-systems within the model to secure that the operational management is close to optimal. As the problem with the current DSS-systems is how to handle constraints, the decision rule should in addition specify how the suggested decisions from the system should be carried out under the constraints.

An additional trick is to let the DSS suggest a decision, and to base the actual decision on a drawing of a random number. E.g., if the DSS suggest the culling of

an animal then the decision is carried out in 0.95 percent of the cases, e.g. we have a penetration ratio of the decision. The penetration ratio can also depend on other variables in the model, i.e. a sow with a very low ranking will have high risk of being culled if she simultaneously has leg weakness. E.g. we model an underlying variable

$$Y = \beta_1 \text{RPO} + \beta_2 \text{LW} + \varepsilon$$

where RPO is the retention Pay Off value and LW is leg weakness score, and ε is random normal deviate. If Y is above a certain threshold the decision is carried out.

It is important to recognize that the random part of the decision rule (the penetration ratio) is part of the state of nature and not a decision variable. The introduction of the penetration ratio corresponds to a mixed strategy rather than the original pure strategy. As it can be shown that there always exist a pure strategy that is at least as good as a mixed strategy, it does not make sense to do any optimization.

14.9 Inference based on the model

The purpose of building the model was to draw inference from the set of decision rules, Θ , the initial state of nature Φ_0 , the output Ω , and response functions based on the values.

The procedures are most easily explained by borrowing terms from experimental world. A set of decision rule is termed a treatment. Expected values from a given treatment are found by a number of replicates (k) or simulation runs. When using the models for inferences new treatments are repeatedly specified and the expected utility calculated for the treatment. If we want to compare a set of treatments simultaneously, we design an experiment with the different treatments included.

As in ordinary statistical design we may choose a design that influences how the state of nature interacts with the desired treatment comparisons, e.g. by blocking different treatments within the same state of nature. The simulation runs are made either to study a relatively short planning horizon, or to study the long run or steady state properties of the different experimental designs. In the latter case, the results from an initial period are discarded before the system is analyzed, to remove the effect of initial conditions.

14.9.1 Definition of sensitivity

As mentioned previously, calculation of sensitivity estimates is one approach towards the understanding of the model's behavior and thus the system. Sensitivity estimates can also be used in other context's, e.g. it plays an important role in some of the search strategies for optimal solutions.

The sensitivity is simply the response function, $U()$, differentiated with respect to a given parameter in the state of nature, Φ_0 (or the set of decision rules, Θ). If we consider every parameter simultaneously, it is the vector of first derivatives, U' ,

as used in Eq. (14.4) on page 238. As shown it makes good sense to calculate U'' , the matrix of second order partial derivatives of $U()$ with respect to the parameters, in addition to U' . In order to estimate these quantities several numerical methods can be chosen. Sometimes the gradient notation $\nabla^k U()$ is used to denote the k th order partial derivatives.

The traditional approach is to calculate the response function for the estimated level of the parameter:

$$\frac{\partial \bar{U}(\Theta, \Phi)}{\partial \Phi_i} \approx \frac{\bar{U}(\Theta, \Phi) - \bar{U}(\Theta, \Phi + \Delta)}{\delta} \quad (14.16)$$

where Φ_i is the i th element of the parameter vector, Δ is a vector with the i th element equal to δ and the rest of the elements zero. δ is chosen small compared to Φ_i . Of course, formally we should let $\delta \rightarrow 0$ in order to estimate the sensitivity. As we have only estimates of the expected utility we run into trouble with this method. The variance of the sensitivity in Eq. (14.16) is $2 \frac{\text{Var}(\bar{U}(\Phi, \Theta))}{\delta^2}$, i.e. if δ is too small the variance of the estimate becomes much too big.

With n_0 parameters we need at least $(n_0 + 1) \times k$ simulation runs to calculate U' and $(2n_0^2 + 1) \times k$ in order to calculate U'' (using the 9 point formula 25.3.25 in Chapter 25 of Abramowitz and Stegun (1964), where k is the number of simulation runs. (Note that in order to calculate U'' for a 10 dimensional multivariate normal distribution with 10 (expectations) + 55 (Variance matrix) parameters we have to do on the order of 8700 k simulation runs). With very complex models with many parameters, the complexity problem is obvious.

Fortunately, research efforts have made the estimation of sensitivity possible within only one simulation run (Rubinstein and Shapiro, 1993). A short review of this technique can be found in Section 14.11.2

14.9.2 Calibration of model and search for optimal strategies

Many of the problems in the initial specification of the model can be formulated as an optimization problem. For example if we know that:

$$f(\Omega) = A \quad (14.17)$$

$$g(\Omega) = B \quad (14.18)$$

where $f()$ and $g()$ are functions of the output variables and A and B are vectors, we can define the response function.

$$R(\Omega) = (A - f(\Omega))^2 + (B - g(\Omega))^2$$

If the (global) minimum of this function is found, the model has been calibrated to our knowledge of the system.

The search for optimal strategies is included in linear programming, dynamic programming and influence diagrams, (i.e. simplex algorithm, policy and value iteration). No such search facility is included in connection with simulation models.

Therefore, the search for the optimal set of decision rules or parameter value is treated as a general problem of multidimensional optimization. Several numerical methods exist that can handle this. The choice of method should be made carefully. The flexible form of the simulation models means that the behavior of the expected utility function is unknown, for example whether discontinuities and local optima exists. Such phenomena can make some methods go wrong. Another complication is, that the expected utility is only estimated with a precision depending on the number of simulation runs within each treatment. The difference between two set of decision rules may therefore be just a matter of sampling error, rather than a difference in expected utility. The solution to this problem is to do more simulation runs. But there is a trade off between time spent improving the precision in the estimation of one set of decisions rules, and the time spent searching for a better.

A well-established technique for well-behaved response functions, especially with continuous variables in the decision rules, is the gradient search. First an experiment is designed to initially explore the response function, e.g. a response surface design. The result from this experiment is analyzed and the response surface estimated. If the optimum is outside the current design, the path of steepest ascent of the response surface is estimated, i.e. the statistical procedure removes the random noise from the individual points. Then an experiment is made with treatments on the steepest ascent path, until the optimal treatment on this path is found. A new response surface design is made centered around this optimum point. This procedure is repeated until the optimum is found with sufficient precision. Using this procedure, an (at least local) optimum will be found.

Refer to a textbook in numerical optimization methods for further techniques, e.g. Press et al. (1989).

14.9.3 Stochastic search techniques

Other promising techniques are stochastic search techniques, such as *simulated annealing* and *genetic algorithms*. These algorithms start with the selection of an arbitrary initial set of decision rules (treatment) as the current. The expected utility of this is calculated by several simulation runs.

Then the following steps are carried out iteratively. A new treatment candidate is selected based on the current treatment. The candidate is selected by some kind of random mechanism, e.g. by randomly changing a decision to the opposite, i.e. if the decision is cull then decide not to cull. Especially within genetic algorithms this random permutation has a whole vocabulary connected with it, such as mutation, cross over, chromosomes etc. Genetic algorithms selects several new candidates at each iteration. Then the treatment candidate(s) is evaluated by calculating the expected utility using this candidate(s).

Based on the expected utility it is decided whether to keep or discard the candidate. This decision is based on the drawing of a random variable. The specific rule for deciding this depends on the technique.

If the treatment candidate is kept the old current treatment is discarded, and the new treatment used instead. Then the iterations are continued.

Simulated annealing follows a so-called annealing schedule, corresponding to a decline in temperature in order to stop the iterations. In the start of the optimization almost any change in the treatment will be accepted, but later on in the schedule only treatment candidates with improved or slightly reduced expected utility will be accepted. This avoids that the procedure get stuck in local optima. Similarly, genetic algorithms has rules concerning the number of so-called generations to be used in the optimization.

There are reasons to expect that the *simulated annealing* will be the best choice of the stochastic techniques. The principles in *simulated annealing* closely correspond to the principles within the Metropolis-Hastings and Gibbs sampling, though the object is to optimize rather than take the expectation. The very large research effort within MCMC-techniques indicates that the progress within the area will be fast. Already the “marriage” between Bayesian Networks and Influence Diagram has exploited the similarity between finding optima and calculating expectation. Both Metropolis Hastings Algorithms and simulated annealing have been used in the framework.

Simulated annealing can then be used in the general framework of Bayesian analysis, while genetic algorithms stand separate, at least at the current state of development.

These stochastic search techniques have not been used to any large extent within herd management. Houben et al. (1995) have used genetic algorithms and Mayer et al. (1996) have studied the use of both genetic algorithms and simulated annealing. For an introduction to the area Ingber (1993) can be recommended.

A third set of possibilities are the so-called heuristic search strategies. Examples can be found in Reeves (1993).

14.9.4 Variance reduction (Experimental planning)

As in the real experimental planning it is an advantage to be able to compare treatments with the same random influences. The block concept of experimental planning ensures that the comparison between treatments can be made within block and the influence of external ‘noise’ can thus be reduced.

The same concept can be used for simulation experiments. As the sequence of random numbers is fully specified by the random seed this is in principle easy. If two treatments are compared using the same random seed in the simulation run the random influences should be the same. However, before the method can be applied it is necessary to follow some guidelines.

The problem with the technique is that the treatment (decision rules) very often changes the number of random drawings to be made. For example, if the decision rule specifies that an animal shall be culled rather than kept, the sequence of random deviates for the other elements in the model changes. Usually this means that the two sequences soon get out of synchronization and thus the effect of the

blocking is lost. However, here it is important to divide the set of parameters in two, i.e. $\Phi = (\Phi_0, \Phi_{s\bullet})$, where Φ_0 are the initial values of the parameters at the start of the calculation (*State of Nature*) and $\Phi_{s\bullet}$ are parameter values that change during simulations as mentioned in Section 14.2. The parameters based on Φ_0 will typically be easy to standardize and blocking can be kept for this purpose. In addition, it can be advised to make the parameters specifying the fluctuations in herd level identical for the different simulation runs. This should be possible without too much difficulty.

For few of the parameters in Φ_0 , the so-called stratified sampling can be of interest. The stratified sampling secures that values are drawn more evenly from the whole distribution than the density of the parameter specifies. E.g. with the exponential distribution, where the tail is very long, it can be of interest to be sure to have a sample of adequate size in the tail end of the distribution. Therefore a fixed proportion of the simulation runs are made from the tail end of the distribution. The expected utility is then calculated as a weighted average of the average in each of the samples. This method becomes too complicated if several parameters are used. Perhaps another example can illustrate the principle as well. In models of disease spread within the herd, it is of interest to study the cases, where the herd is infected, i.e. a sample is made so that the herd is infected. When calculating expected utility the real infection rate is used in order to combine the results from the infected and non-infected simulation runs.

14.10 Verification and model comparison

In literature concerning simulation model the verification and validation of the models are always strongly emphasized, but the application of these project phases is more seldom.

Verification is the control of internal consistency of the model, i.e. if the model behaves as it is supposed to do. This is a natural and essential part of any programming activity, and is usually carried out in simulation models, but by informal methods.

Validation is the control, whether the model is a valid representation of the real system. A review of validation attempts for simulation models can be found in Sørensen (1990). Systematic guidelines for verification and validation are also given by Sargent (1998).

To the authors' opinion the use of the validation concept has not proven very useful. Using the systems approach, where the model building reflects the current scientific theory, the validation of the verified model corresponds closely to testing the theory. See Déry et al. (1993) for a discussion of validation seen in this context. In any case, a controlled and precise experiment would be a more natural choice for tests of scientific hypothesis. The role to be played by the simulation model in this context is to improve the experimental planning.

The validation concept is comparable to the question: *is a map a valid model of a landscape?* Of course there is no general answer to this question. If we ask *Is this map or that map the better model of the landscape*, we can still not obtain a general answer, but now the discussion can be fruitful. If we can define what maps are to be used for, then we can measure the adequacy of each of the models.

If model building is considered from the Bayesian framework, the model is a reflection of the subjective belief in the system. If the assumptions are believed to be true and the model is verified, the model is simply a calculation based on the belief. A subjective belief can not be validated, it can only be updated based on new observations. There may exist different prior belief in the systems structure and parameter values. Decision makers may even have several possible models that may be true, and thus prior probabilities of being true can be assigned to the alternative models. In light of the evidence the most probable model can be found.

Instead of discussing model validation, model comparison, should therefore be preferred. If we look at statistical modeling, a whole set of tools for model comparison and model test is available. These methods offer a stringent method of model evaluation. In line with the Bayesian framework a natural choice would be to use the Bayes factor (see Kass and Raftery, 1995, for a review), but the stochastic approach should make other statistical test's available, as well. For the comparison between two alternative models (hypothesis) the Bayes factor is simple the ratio of the likelihood of data given each of the models. Another definition is that the posterior odds is equal to the Bayes factor multiplied by the prior odds.

14.11 Advanced simulation methods

The area of Monte Carlo methods is rapidly advancing and some of the techniques that have not yet reached the area of farm management, will be illustrated below.

The techniques are mainly applicable with respect to the parameters in Φ_0 , the state of nature, and will, in general, improve the use of existing simulation runs, and make it possible to refine the posterior parameter estimates.

14.11.1 Importance Sampling

Importance sampling is originally a method for increasing the efficiency of integration by Monte Carlo simulation. If we sample more frequently in a region, where the magnitude of the integrand is important, the efforts are used more efficiently. The idea is to sample from a distribution π with density function $f_\pi(x)$ and then to calculate the integral as a weighed average,

$$\int_{-\infty}^{\infty} h(x)dx \approx \frac{1}{N} \sum_{i=1}^N \frac{h(x_i)}{f_\pi(x_i)}$$

where x_1, \dots, x_N is a random sample drawn from the distribution π . Inspired by the simulation integral of Eq. (14.1), we define the function h as $h(\phi) =$

$U(\Theta, \phi)f_\Phi(\phi)$ for a given decision rule, Θ . Assume that we draw a random sample ϕ_1, \dots, ϕ_N from the distribution of Φ with the density function $f_\Phi(\phi)$. We may then calculate the integral of h by importance sampling as,

$$\int_{-\infty}^{\infty} h(\phi)d\phi \approx \frac{1}{N} \sum_{i=1}^N \frac{h(\phi_i)}{f_\Phi(\phi_i)} = \frac{1}{N} \sum_{i=1}^N \frac{U(\Theta, \phi_i)f_\Phi(\phi_i)}{f_\Phi(\phi_i)} = \frac{1}{N} \sum_{i=1}^N U(\Theta, \Phi_i). \quad (14.19)$$

If we compare with Eq. (14.1) and Eq. (14.2) on page 234, it is easily seen that the right-hand side of Eq. (14.19) is just the expected utility determined by simulation, and since,

$$\bar{U}(\Theta) = \int_{-\infty}^{\infty} h(\phi)d\phi = \int_{-\infty}^{\infty} U(\Theta, \phi)f_\Phi(\phi),$$

we see that the left-hand side is just the true expected utility, $\bar{U}(\Theta)$. In other words, our initial approach to the sampling, was in fact a use of importance sampling, where the importance was measured in terms of a prior probability density of the parameters.

In special cases it may be more relevant to sample from a density proportional to the whole integrand as in Eq. (14.19) instead of the density of parameters. A more obvious application of the principle is, if we have performed a simulation experiment with one probability density of the parameters $f_\Phi(\phi)$ and we want to find the expected utility with another density, e.g. $q_\Phi(\phi)$. We can then find the expected utility by use of the original sample ϕ_1, \dots, ϕ_N drawn from the distribution described by the density $f_\Phi(\phi)$ as,

$$\bar{U}_q(\Theta) = \int_{-\infty}^{\infty} U(\Theta, \phi)q_\Phi(\phi) \approx \frac{1}{N} \sum_{i=1}^N \frac{U(\Theta, \phi_i)q_\Phi(\phi_i)}{f_\Phi(\phi_i)} = \frac{1}{N} \sum_{i=1}^N w_i U(\Theta, \phi_i),$$

where $w_i = \frac{q_\Phi(\phi_i)}{f_\Phi(\phi_i)}$.

This approach can be applied with sensitivity analysis and calibration of the model to an observed level. Note that the method requires that both the sampling density and the new density is known.

Usually, the sum of weight factors is close to one, i.e. $\sum_{i=1}^N w_i \approx 1$, but if the distributions differ widely this might not be case. As a precaution the expected utility is therefore divided by the sum of the weights, i.e. the final estimated utility under the new distribution becomes

$$E_q(U(\Theta)) = \frac{\sum_{i=1}^N w_i U(\Theta, \phi_i)}{N \sum_{i=1}^N w_i} \quad (14.20)$$

Note that if the two distributions differ widely, only few of the simulation runs will have a high weight factor, and the variance of the estimated utility will be high.

14.11.2 Sensitivity calculation

An obvious possibility for using importance sampling for calculating sensitivities is to use Eq. (14.16) and calculate by adding δ to the parameters in the density function and use Eq. (14.20) to calculate the expected utility with the new parameters. However, an even better method exists, if certain assumptions are fulfilled.

The gradient (vector of first order partial derivatives with respect to the parameters) is simply calculated using the following formula.

$$\bar{\nabla} \bar{U}_M(\Theta) \approx \frac{1}{k} \sum_{i=1}^k U_M(\Theta, \psi_i) S^{(1)}(\Theta, \phi_i)$$

where

$$S^{(1)}(\Theta, \phi_i) = \nabla \log(p(\Theta, \phi_i))$$

$S^{(1)}$ is called the efficient score function. It has close correspondence to likelihood ratios, etc

For the parameters in Φ_0 , $S^{(1)}$, can usually be found easily. The method generalizes to the matrix of second order partial derivatives as well. Note that no extra simulation runs are necessary. To some extent it is even possible to optimize with respect to the decision parameters. This demands of course that the decision variables have been varied randomly within the simulation run.

For further information concerning this score function method, see Rubinstein and Shapiro (1993).

14.11.3 Calibration with observed values

The following has been adapted from Raftery et al. (1995).

The usual method for stochastic specification in simulation models is to assign prior distribution to the state of nature, Φ_0 , and to treat the output, Ω , as a result of the model. Very often a better approach is to specify a *pre-model joint distribution* of (Φ_0, Ω) , because the prior knowledge is very often based on observations of output values rather than model parameters. The level of production traits etc. is generally known, e.g. the expected milk yield in a herd can be specified even without a simulation model. The model provides us with a sample from the probability distribution of Ω , $g_M(\Omega)$.

The method relies on the importance sampling principle. The sampling has been done with the density based on the pre-model distribution of Φ_0 , $p(\Omega)$. This density has to be estimated from the output of the simulation runs (the estimated density is referred to as $\hat{p}(\Omega)$). Instead we want to sample from the density based on the pre-model distribution of Ω , $q(\Omega)$, i.e. the importance sampling weights are:

$$w_i = \frac{q(\Omega = \omega_i)}{\hat{p}(\Omega = \omega_i)}$$

Based on Eq. (14.20) we can calculate any relevant measure, such as expectation and variance of the parameters in Φ_0 .

14.11.4 Acceptance rejection sampling

The concept behind acceptance-rejection sampling as described in Section 14.7.2 can also be applied to processing of output from simulation runs, by elaborating on the methods above.

The sample from the simulation run corresponds to the “covering” distribution $g(x)$ in Figure 14.4. The weight w_i in the formula above is the ration of the two densities. All that remains is to find a constant such that $aw_i \leq 1$. If we let $a = 1/\max(w_i)$ this is accomplished.

The data set with simulation results is treated sequentially. For each element i we draw a uniform random deviate u_i . If $aw_i \leq u_i$ the element is kept in the data set. If not it is deleted. The resulting database will be a sample from the posterior distribution.

This ‘calibrated’ data set can be used for further calculations.

Note that if the the two densities differ widely only few values will be kept in the database.

14.12 Outlook

Within the last years many improvements in Monte Carlo simulation techniques has been made. The improvements has taken place within optimization, parameter estimation and model comparison and within sensitivity analysis. These new developments has only to a small extent been implemented in the area of herd management. The possibilities for this are obvious. Especially the combination with the Markov Chain decisions support methods, or their generalization Bayesian Networks and Influence Diagrams.

In the future the Monte Carlo simulation will probably be a technique used to supplement the techniques, in order to overcome the inherent complexity problems, due to the herd constraints.

Bibliography

- Abramowitz, M., Stegun, I., 1964. Handbook of Mathematical Functions. Applied Mathematics Series, vol. 55. National Bureau of Standards, Washington.
- Adler, H. C., 1978. Animal welfare in farm animal production. In: Folsch, D. W. (Ed.), Ethology and Ethics of farm Animal Production. Birkhauser, Basel and Stuttgart.
- Anderson, D., Weeks, D., 1989. Cattle liveweight sampled on a continuous versus intermittent basis. *Livestock Production Science* 23, 117–135.
- Anderson, J. R., Dillon, J. L., Hardaker, J. B., 1977. Agricultural Decision Analysis. The Iowa State University Press, Ames, Iowa.
- Anonymous, 1977. Report of committee of experts on animal husbandry and welfare: Policy recommandations. Tech. rep., Dutch National Council for Agricultural Research, The Hague.
- Arnold, R. N., Bennet, G. L., 1991. Evaluation of four simulation models of cattle growth and body composition: Part i - comparison and characterization of the models. *Agricultural Systems* 35, 401–432.
- Bangsø, O., 2004. Object oriented bayesian networks. Ph.D. thesis, Department of Computer Science, Aalborg University, Aalborg.
- Bellman, R. E., 1957. Dynamic Programming. Princeton University Press, Princeton.
- Bellman, R. E., 1961. Adaptive control processes: A guided tour. Princeton University Press, Princeton.
- Bellman, R. E., Dreyfus, S. E., 1962. Applied dynamic programming. Princeton University Press, Princeton.
- Bellman, R. E., Kalaba, R., 1965. Dynamic programming and modern control theory. Academic Press, New York.
- Ben-Ari, Y., Amir, I., Sharar, S., 1983. Operational replacement decision model for dairy herds. *Journal of Dairy Science* 66, 1747–1759.

- Ben-Ari, Y., Gal, S., 1986. Optimal replacement policy for multicomponent systems: An application to a dairy herd. *European Journal of Operational Research* 23, 213–221.
- Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems. *Statistical Science* 10, 3–66.
- Bielza, C., Müller, P., Ríos Insua, D., 1999. Decision analysis by augmented probability simulation. *Management Science* 45, 995–1007.
- Black, J. L., 1995. The evolution of animal growth models modelling energy metabolism in the pig - critical evaluation of a simple reference model. In: Moughan, P. J., Verstegen, M. W. A., Visserreyneveld, M. I. (Eds.), *Modelling Growth in the Pig*. Vol. 78 of Eaap Publication. Wageningen Peers, Wageningen.
- Brambell, F. W. R., 1965. Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry system. Command Paper 2836, HMSO, London.
- Brandl, N., Jørgensen, E., 1996. Determination of live weight in pigs from dimensions by image analysis. *Computers and Electronics in Agriculture* 15, 57–72.
- Brantas, G. C., 1978. Quoted in “introduction to plenary session i. intensive confinement farming: Ethological, veterinary and ethical issues”. In: Proceedings of the 1st World Congress of Ethology and Applied Zootech. Madrid.
- Bressers, H., te Brake, J., Jansen, M., Nijenhuis, P., Noordhuizen, J., 1994. Monitoring individual sows - radiotelemetrically recorded ear base temperature changes around farrowing. *Livestock Production Science* 37, 353–361.
- Bressers, H., te Brake, J., Noordhuizen, J., 1991. Oestrus detection in group-housed sows by analysis of data on visits to the boar. *Applied Animal Behaviour Science* 31, 183–189.
- Broekmans, J. E., 1992. Influence of price fluctuations on delivery strategies for slaughter pigs. *Dina Notat* 7, 1–28.
- Buntine, W. L., 1994. Operations for learning with graphical models. *Journal of Artificial Intelligence Research* 2, 159–225.
- Chang, W., Streeter, D., Jones, L., 1994. An object-oriented model for simulating milking parlor operations. *Journal of Dairy Science* 77, 84–93.
- Charnes, J. M., Shenoy, P. P., 1996. A forward monte carlo method for solving influence diagrams using local computation (abstract). In: Proceedings of the First European Conference on Highly Structured Stochastic Systems. 19.-24. may 1996. Rebild, Denmark.

- Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. *American Statistician* 49, 327–335.
- Clausen, S., 1969. *Kybernetik - Systemer og Modeller*. Institute of Mathematical Statistics and Operations Research. The Technical University of Denmark.
- Cornou, C., 2006. Automated oestrus detection methods in group housed sows: Review of the current methods and perspectives for development. *Livestock Science*, In press.
- Cowell, R. G., Dawid, A. P., Lauritzen, S. L., Spiegelhalter, D. J., 1999. Probabilistic Networks and Expert Systems. *Statistics for Engineering and Information Science*. Springer-verlag, New York, Berlin, Heidelberg.
- Danfær, A., 1990. A dynamic model of nutrient digestion and metabolism in lactating dairy cows. *Beretning fra Statens Husdyrbrugsforsøgs* 671, National Institute of Animal Science.
- Dawkins, M., 1980. *Animal Suffering*. Chapman and Hall, London.
- de Mol, R., Keen, A., Kroeze, G., 1999. Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis with a kalman filter. *Computers and Electronics in Agriculture* 22, 171–185.
- de Roo, G., 1987. A stochastic model to study breeding schemes in a small pig population. *Agricultural Systems* 25, 1–25.
- de Vries, A., 2001. Statistical process control charts applied to dairy herd reproduction. Ph.D. thesis, University of Minnesota.
- de Vries, A., Conlin, B. J., 2003. Design and performance of statistical process control charts applied to estrus detection efficiency. *Journal of Dairy Science* 86, 1970–1984.
- de Vries, A., Conlin, B. J., 2005. A comparison of the performance of statistical quality control charts in a dairy production system through stochastic simulation. *Agricultural Systems* 82, 317–341.
- de Vries, A., van Leeuwen, J., Thatcher, W. W., 2004. Economic importance of improved reproductive performance. In: *Proceedings 2004 Florida Dairy Reproduction Road Show*. Okeechobee, Ocala and Chipley, Florida, pp. 33–43.
- DeGroot, M. H., 1970. *Optimal statistical decisions*. McGraw-Hill, Inc., New York.
- DeLorenzo, M. A., Spreen, T. H., Bryan, G. R., Beede, D. K., van Arendonk, J. A. M., 1992. Optimizing model: insemination, replacement, seasonal variation, and cash flow. *Journal of Dairy Science* 75, 885–896.
- Dent, J., Anderson, J., 1971. *System analysis in agricultural management*. Wiley.

- Dent, J. B., Harrison, S. R., Woodford, K. B., 1986. Farm planning with linear programming: Concept and Practice. Butterworths, Sydney.
- Dijkhuizen, A., Krabbenborg, R., Huirne, R., 1989. Sow replacement: A comparison of farmer's actual decisions and model recommendations. *Livestock Production Science* 23, 207–218.
- Dijkhuizen, A. A., Huirne, R. B. M., Jalvingh, A. W., 1995. Economic analysis of animal diseases and their control. *Preventive Veterinary Medicine* 25, 135–149.
- Dijkhuizen, A. A., Morris, R. S., 1997. Animal Health Economics: Principles and Applications. University of Sidney, Sidney.
- Dijkhuizen, A. A., Morris, R. S., Morrow, M., 1986. Economic optimization of culling strategies in swine breeding herds, using the "porkchop computer program". *Preventive Veterinary Medicine* 4, 341–353.
- Dindorph, U., 1992. Agricultural applications of knowledge based system concepts - exemplified by a prototype on weed control in organic farming. WEEDOF. Ph.D. thesis, Tidsskrift for planteavl Specialserie Beretning nr S 2201.
- Dittmer, S., January 1997. Predicting calving dates with bayesian networks. In: Kristensen, A. R. (Ed.), Proceedings of the Dutch/Danish symposium on animal health and management economics. Vol. 56 of Dina Notat. Copenhagen, pp. 95–101.
- Déry, R., Landry, M., Banville, C., 1993. Revisiting the issue of model validation in or: An epistemological view. *European Journal of Operations Research* 66, 168–183.
- Enevoldsen, C., 1993. Sundhedsstyring i mælkeproduktionen. Ph.D. thesis, Institut for Husdyrbrug og Husdysundhed, Den Kongelige Veterinær- og Landbohøjskole, København.
- Finlayson, J., Cacho, O., Bywater, A. C., 1995. A simulation model of grazing sheep: I. animal growth and intake. *Agricultural Systems* 48, 1–25.
- Firk, R., Stamer, E., Junge, W., Krieter, J., 2002. Automation of oestrus detection in dairy cows: a review. *Livestock Production Science* 75 (219-232).
- Fraser, A. F., Broom, D. M., 1990. Farm Animal Behaviour and Welfare, 3rd Edition. Baillière, London.
- Giaever, H. B., 1966. Optimal dairy cow replacement policies. Ph.D. thesis, University of California, Berkeley, University Microfilms, Ann Arbor, Michigan.
- Gilks, W., 1992. Derivative-free adaptive rejection sampling for gibbs sampling. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), Bayesian Statistics 4. Oxford University Press, pp. 641–649.

- Gilks, W., Thomas, A., Spiegelhalter, D. J., 1993. A language and program for complex bayesian modelling. *The Statistician* 43, 169–178.
- Gilks, W., Wild, P., 1992. Adaptive rejection sampling for gibbs sampling. *Applied Statistics* 41(2), 337–348.
- Gonzales-Alcorta, M. J., Dorfman, J. H., Pesti, G. M., 1994. Maximizing profit in broiler production as price change: A simple approximation with practical value. *Agribusiness* 10 (5), 389–399.
- Goodall, E. A., Sprevak, D., 1985. A bayesian estimation of the lactation curve of a dairy cow. *Animal Production* 40, 189–193.
- Greve, J., 1995. The dynamic ranking approach and its application to piglet production. Dina research report, 35, Faculty of Technology and science. Aalborg University., Aalborg.
- Hadley, G., 1964. Nonlinear and dynamic programming. Vol. Reading, Massachusetts. Addison-Wesley.
- Hansen, A. E., 1992. Styret vækst hos slagtesvin ved ad libitum fodring. Master thesis, Department of Animal Science and Animal Health, The Royal Veterinary and Agricultural University, København.
- Hansen, M. S., 2006. Optimization of delivery strategies for slaughter pigs. Master's thesis, Department of Large Animal Sciences, The Royal Veterinary and Agricultural University.
- Haran, P., 1997. Markov decision processes in the optimisation of culling decisions for irish dairy herds. Master's thesis, School of Computer Applications, Dublin City University.
- Hardie, A. J., 1996. Ph.D. thesis, University of Exeter.
- Hayes-Roth, F., Waterman, D. A., Lenat, D. B., 1996. Building Expert Systems. Addison-wesley.
- Henry, G. M., DeLorenzo, M. A., Beede, D. K., van Horn, H. H., Moss, C. B., Bogges, W. G., 1995. Determining the optimal nutrient management strategies for dairy farms. *Journal of Dairy Science* 78, 693–703.
- Hillier, F. S., Lieberman, G. J., 1996. Introduction to Operations Research, 3rd Edition. Holden-Day Inc., San Francisco.
- Hoel, P., Port, S., Stone, C., 1972. Introduction to Stochastic Processes. Waveland Press, Inc., Illinois.

- Hogeveen, H., Thiesson, B., Thysen, I., Nordhuizen-Stassen, E. N., Maurice, H., Andersen, S. K., Lam, T. J. G. M., van Werven, T., 1994. Mastitis pathogen diagnosis using bayesian belief networks. In: Hogeveen, H. (Ed.), Knowledge-based methods for automated mastitis diagnosis on dairy farms. Department of Herd Helth and Reproduction, Utrecht University, Utrecht, pp. 129–144.
- Houben, E., Huirne, R., Dijkhuizen, A., Kristensen, A. R., 1994. Optimal replacement of mastitis cows determined by a hierachic markov process. *Journal of Dairy Science* 77, 2975–2993.
- Houben, E., Huirne, R. B. M., Dijkhuizen, A. A., Schreinemakers, J. F., 1995. Genetic algorithm to link decisions at animal and herd level on dairy farms. In: Houben, E. H. P. (Ed.), Economic optimization of decisions with respect to dairy cow health management. Department of Farm Management, Wageningen Agricultural University, pp. 101–121.
- Howard, R., 1960. Dynamic Programming and Markov Processes. The M.I.T. Press, Cambridge, Massachusetts.
- Howard, R. A., 1971. Dynamic probabilistic systems. Volume II: Semi-Markov and decision processes. John Wiley & Sons, Inc., New York.
- Howard, R. A., Matheson, J. E., 1981. Influence diagrams. In: Howard, R. A., Matheson, J. E. (Eds.), Readings on the Principles and Applications of Decision Analysis. Strategic Decisions Group, Menlo Park, California, pp. 719–762.
- Hughes, B. O., 1976. Behaviour as an index of welfare. In: Proceedings of the 5th European Conference. Malta.
- Huirne, R., 1990. Computerized Management Support for Swine Breeding Farms. Ph.D. thesis. Department of Farm Management, Agriculturel University , P.O.box 338,6700 AH Wageningen. The Netherlands, Wageningen.
- Huirne, R., Dijkhuizen, A., Renkema, J. A., 1991. Economic optimization of sow replacement decisions on the personal computer by method of stochastic dynamic programming. *Livestock Production Science* 28, 331–347.
- Huirne, R., Hendriks, T., Dijkhuizen, A., Giesen, G., 1988. The economic optimisation of sow replacement decisions by stochastic dynamic programming. *Journal of Agricultural Economics* 39, 426–438.
- Huirne, R., van Beek, P., Hendriks, T., Dijkhuizen, A., 1993. An application of stochastic dynamic programming to support sow replacement decisions. *European Journal of Operational Research* 67, 161–171.
- Ingber, A., 1993. Simulated annealing: Practice versus theory. *J. Mathl. Comput. Modelling.* 18 (11), 29–57.

- Jalvingh, A. W., Dijkhuizen, A. A., van Arendonk, J. A. M., Brascamp, E. W., 1992a. Dynamic probabilistic modelling of reproductive and replacement in sow herds. general aspects and model description. Agricultural Systems 39, 133–152.
- Jalvingh, A. W., Dijkhuizen, A. A., van Arendonk, J. A. M., Brascamp, E. W., 1992b. An economic comparison of management strategies on reproduction and replacement in sow herds using a dynamic probabilistic model. Livestock Production Science 32, 331–350.
- Jenkins, K. B., Halter, A. N., 1963. A multistage stochastic decision model: Application to replacement of dairy cows. Technical Bulletin 67, Agricultural Experimental Station, Oregon State University.
- Jensen, A. L., 1995. A probabilistic model based decision support system for mildew management in winter wheat. Dina research report 39, University of Aalborg.
- Jensen, F., Jensen, F. V., Dittmer, S. L., 1994. From influence diagrams to junction trees. In: de Mantaras, R. L., Poole, D. (Eds.), Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, San Francisco, CA, pp. 367–373.
- Jensen, F. V., 1996. An Introduction to Bayesian Networks. UCL Press, London.
- Jensen, F. V., 2001. Bayesian Networks and Decision Graphs. Statistics for Engineering and Information Science. Springer-Verlag, New York.
- Jensen, K. K., Sørensen, J. T., 1999. The idea of “ethical accounting” for a livestock farm. Journal of Agricultural and Environmental Ethics 11, 85–100.
- Jewell, W., 1963. Markov renewal programming i and ii. Operations Research 11, 938–971.
- Johnston, J., 1965. Dynamic programming and the theory of the firm. Journal of Agricultural Economics 16, 532–543.
- Jørgensen, E., 1992. Sow replacement: Reduction of state space in dynamic programming models and evaluation of benefit from using the model. Dina Research Report 6.
- Jørgensen, E., August 1993. The influence of weighing precision on delivery decisions in slaughter pig production. Acta Agriculturæ Scandinavica, Section A 43 (3), 181–189.
- Jørgensen, E., 1985. Principper bag produktionskontrol i sohold. Phd thesis, Department of Animal Science, The Royal Veterinary and Agricultural University, Copenhagen, in Danish.

- Jørgensen, E., 1991. Probabilistic growth model for pigs. paper presented at dina-workshop 1991. ebeltoft 3-4 december. dupl.
- Jørgensen, E., 1992. Sow replacement: Reduction of state space in dynamic programming model and evaluation of benefit from using the model. Dina Research Report 6, -21.
- Jørgensen, E., 1993. The influence of weighing precision on delivery decisions in slaughter pig production. *Acta Agriculturae Scandinavica Section A - Animal Science* 43, 181–189.
- Jørgensen, E., Kristensen, A. R., 1995. An object oriented simulation model of a pig herd with emphasis on information flow. In: FACTS 95 March 7, 8, 9. Farm Animal Computer Technologies Conference. Orlando, Florida, pp. 206–215.
- Kaelbling, L. P., Littman, M. L., Cassandra, A. R., 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101 (1–2), 99–134.
- Kass, R., Raftery, A., 1995. Bayes factors. *J. Amer. Statist. Assn.* 90 (430), 773–795.
- Kennedy, J. O. S., 1986. Dynamic Programming Applications to Agriculture and Natural Resources. Elsevier Applied Science Publishers, London and New York.
- Kilgour, R., 1987. The application of animal behaviour and the human care of farm animals. *Journal of Animal Science* 46, 1478–1486.
- Kilgour, R., Dalton, C., 1984. Livestock Behaviour. A practical guide. Granada, London.
- Killen, L., Kearney, B., 1978. Optimal dairy cow replacement policy. *Irish Journal of Agricultural Economics and Rural Sociology* 7, 33–40.
- Kjærulff, U., 1995. dhugin: A computational system for dynamic time-sliced bayesian networks. *International Journal of Forecasting* 11, 89–111.
- Kristensen, A. R., 1986. A model to describe and predict the production and food intake of the dairy cow. *Acta Agriculturae Scandinavica* 36, 305–317.
- Kristensen, A. R., 1987. Optimal replacement and ranking of dairy cows determined by a hierachic markov process. *Livestock Production Science* 16, 131–144.
- Kristensen, A. R., 1988. Hierachic markov processes and their applications in replacement models. *European Journal of Operational Research* 35, 207–215.
- Kristensen, A. R., 1989. Optimal replacement and ranking of dairy cows under milk quotas. *Acta Agriculturae Scandinavica* 39, 311–318.

- Kristensen, A. R., 1991. Maximization of net revenue per unit of physical output in markov decision processes. *European Review of Agricultural Economics* 18, 231–244.
- Kristensen, A. R., 1992. Optimal replacement in the dairy herd: A multi-component system. *Agricultural Systems* 39, 1–24.
- Kristensen, A. R., 1993. Bayesian updating in hierachic markov processes applied to the animal replacement problem. *European Review of Agricultural Economics* 20, 223–239.
- Kristensen, A. R., 1994. A survey of markov decision programming techniques applied to the animal replacement problem. *European Review of Agricultural Economics* 21, 73–93.
- Kristensen, A. R., 2003. A general software system for markov decision processes in herd management applications. *Computers and Electronics in Agriculture* 38, In press.
- Kristensen, A. R., Jørgensen, E., 2000. Multi-level hierachic markov processes as a framework for herd management support. *Annals of Operations Research*. 94, 69–89.
- Kristensen, A. R., Nielsen, L., Hansen, M. S., 2006. Optimal slaughter pig marketing with emphasis on information from on-line live weight assessment. Submitted for journal publication.
- Kristensen, A. R., Pedersen, C. V., 2003. Representation of uncertainty in a monte carlo simulation model of a scanvenging chicken production system. In: EFITA 2003. Fourth European Conference of the European Federation for Information Technology in Agriculture, Food and the Environment. Debrecen, Hungary, pp. 451–459.
- Kristensen, A. R., Søllested, T. A., 2004a. A sow replacement model using bayesian updating in a 3-level hierachic markov process. i. biological model. *Livestock Production Science* 87, 13–24.
- Kristensen, A. R., Søllested, T. A., 2004b. A sow replacement model using bayesian updating in a 3-level hierachic markov process. ii. optimization model. *Livestock Production Science* 87, 25–36.
- Kristensen, A. R., Østergaard, V., 1982. Optimalt udskiftningstidspunkt for malkekøen fastlagt ved en stokastisk model. Beretning fra Statens Husdyrbrugsforsøg 533, National Institute of Animal Science.
- Kristensen, A. R., Thysen, I., 1991. Economic value of culling information in the presence and absence of a milk quota. *Acta Agriculturæ Scandinavica* 41, 129–135.

- Kure, H., 1997. Marketing management support in slaughter pig production. Ph.D. thesis, The Royal Veterinary and Agricultural University.
- Lauritzen, S. L., 1995. Probabilistic expert systems. Tech. Rep. R-95-2013, Aalborg University. Institute for Electronic Systems. Department of Mathematics and Computer Science., Aalborg.
- Lauritzen, S. L., Nilsson, D., 2001. Limits of decision problems. *Management Science* 47, 1235–1251.
- Lauritzen, S. L., Spiegelhalter, D. J., 1988. Local computations with probabilities on graphical structures and their application to expert system (with discussion). *Journal of the Royal Statistical Society B* 50, 157–224.
- Lien, G., Kristensen, A. R., Hegrenes, A., Hardaker, J. B., 2003. Optimal length of leys in an area with winter damage problems. *Grass and Forage Science* 58, 168–177.
- Liu, J., 1996. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing* 6, 113–119.
- Lovejoy, W. S., 1991. A survey of algorithmic methods for partially observed markov decision processes. *Annals of Operations Research* 28 (1-4), 47–66.
- Madsen, T. N., Andersen, S., Kristensen, A. R., 2005. Modelling the drinking patterns of young pigs using a state space model. *Computers and Electronics in Agriculture* 48, 39–62.
- Madsen, T. N., Kristensen, A. R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. *Computers and Electronics in Agriculture* 48, 138–154.
- Madsen, T. N., Ruby, V., 2005. An application for early detection of growth rate changes in the slaughter pig production unit. *Computers and Electronics in Agriculture* 25, 261–270.
- Makulska, J., Kristensen, A. R., 1999. Economic optimization of bull fattening. In: Perspectives of Modern Information and Communication Systems in Agriculture, Food Production and Environmental Control. European Federation for Information Technology in Agriculture, Food and the Environment.
- Marchant, J. A., Schofield, C. P., Jun. 1993. Extending the snake image processing algorithm for outlining pigs in scenes. *Computers and Electronics in Agriculture* 8 (4), 261–275.
URL <http://www.sciencedirect.com/science/article/B6T5M-48TN34R-8F/2/2c7618fa096e82ccb4b955a2f4085bed>
- Mayer, D., Belward, J., Burrage, K., 1996. Use of advanced techniques to optimize a multi-dimensional dairy model. *Agricultural Systems* 50, 239–253.

- McAinsh, C. V., Kristensen, A. R., 2004. Dynamic modelling of a traditional african chicken production system. *Tropical Animal Health and Production* 36, 609–626.
- McArthur, A. T. G., 1973. Application of dynamic programming to the culling decision in dairy cattle. *Proceedings of the New Zealand Society of Animal Production* 33, 141–147.
- McCullough, D. A., DeLorenzo, M. A., 1996a. Effects of price and management level on optimal replacement and insemination decisions. *Journal of Dairy Science* 79, 242–253.
- McCullough, D. A., DeLorenzo, M. A., 1996b. Evaluation of a stochastic dynamic replacement and insemination model for dairy cattle. *Journal of Dairy Science* 79, 50–61.
- McKendrick, I. J., Gettinby, G., Gu, Y., Reid, S. W. J., Reive, C. W., 2000a. Using a bayesian belief network to aid differential diagnosis of tropical bovine diseases. *Preventive Veterinary Medicine* 47, 141–156.
- McKendrick, I. J., Gettinby, G., Gu, Y., Reid, S. W. J., Revie, C. W., 2000b. Using a bayesian belief network to aid differential diagnosis of tropical bovine diseases. *Preventive Veterinary Medicine* 47, 141–156.
- McKinion, J. M., Lemmon, H. E., 1985. Expert systems for agriculture. *Computers and Electronics in Agriculture* 1, 31–40.
- Montgomery, D. C., 2005. *Introduction to statistical quality control*, 5th Edition. John Wiley & Sons, Inc.
- Morris, C., Normand, S., 1992. Hierarchical models for combining information and for meta-analysis. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), *Bayesian Statistics 4*. Oxford University Press, pp. 321–344.
- Mourits, M., Huirne, R., Dijkhuizen, A., Kristensen, A. R., Galligan, D., 1999. Economic optimization of dairy heifer management decisions. *Agricultural Systems* 61, 17–31.
- Nash, D. L., Rogers, G. W., 1995. Herd sire portfolio selection: A comparison of rounded linear and integer programming. *Journal of Dairy Science* 78, 2486–2495.
- Nielsen, B. K., Kristensen, A. R., Thamsborg, S. M., 2004. Optimal decisions in organic steer production - a model including winter feeding level, grazing strategy and slaughtering policy. *Livestock Production Science* 88, 239–250.
- Østergaard, S., Chagunda, M., Friggens, N., Bennedsgaard, T., Klaas, I., 2005. A stochastic model simulating pathogen-specific mastitis control in a dairy herd. *Journal of Dairy Science* 88, 4243–4257.

- Østergaard, S., Sørensen, J. T., Hindhede, J., Houe, H., 2004. Control strategies against milk fever in dairy herds evaluated by stochastic simulation. *Livestock Production Science* 86, 209–223.
- Østergaard, S., Sørensen, J. T., Kristensen, A. R., 2000. A stochastic model simulating the feeding-health-production complex in a dairy herd. *Journal of Dairy Science* 83, 721–733.
- Otto, L., Kristensen, C. S., 2004. A biological network describing infection with mycoplasma hyopneumoniae in swine herds. *Preventive Veterinary Medicine* 66, 141–161.
- Pearl, J., 1988. Probabilistic reasoning in intelligent systems. Morgan Kaufman Publishers, San Mateo.
- Pedersen, B., Ruby, V., Jørgensen, E., 1995. Organisation and application of research and development in commercial pig herds: The Danish approach. In: Hennesy, D., Cranwell, P. (Eds.), *Manipulating Pig Production V. Proceedings of the Fifth Biennial Conference of the Australasian Pig Science Association (APSA)*. Canberra, November 26 to 29, 1995.
- Pedersen, C., 1996. Kan det betale sig at analysere grovfoder. [is there a pay-off from roughage analysis], student report in the course “Advanced Herd Management”.
- Pettigrew, J., Cornelius, S., Eidman, S., Moser, R., 1986. Integration of factors affecting sow efficiency: A modelling approach. *Journal of Animal Science* 63, 1314–1321.
- Poole, A., West, M., Harrison, P. J., 1994. *Applied Bayesian Forecasting and Time Series Analysis*. Chapman-Hall, New York.
- Preinrich, G., 1940. The economic life of industrial equipment. *Econometrica* 8, 12–44.
- Press, W. H., Flannery, B. R., Teukolsky, S. A., Vetterling, W. T., 1989. *Numerical Recipes in Pascal. The art of scientific computing*. Cambridge University Press, Cambridge.
- Puterman, M. L., 1994. *Markov Decision Processes: discrete stochastic dynamic programming*. Wiley - Interscience.
- Rae, A. N., 1994. *Agricultural Management Economics - Activity Analysis and Decision Making*. CAB International.
- Raftery, A., Givens, G., Zeh, J., 1995. Inference from a deterministic population dynamics model for bowhead whales. *Journal of the American Statistical Association*. 90 (430), 402–430.

- Ramaekers, P., Huiskes, J., Verstegen, M., den Hartog, L., Vesseur, P., Swinkels, J., 1995. Estimating individual body weights of group-housed growing-finishing pigs using a forelegs weighing system. Computers and Electronics in Agriculture 13, 1–12.
- Randall, J. M., 1976. Environmental and welfare aspects of intensive livestock housing: a report on an american tour. Silsoe Report 19, National Institute of Agricultural Engineering.
- Rapp, B., 1974. Models for optimal investment and maintenance decisions. Almquist & Wiksell, Stockholm.
- Rasmussen, L. K., 1995. Bayesian network for blood typing and parentage verification of cattle. Vol. 38 of Dina Research Report.
- Rasmussen, S., Nielsen, A. H., 1985. The impact of quotas on the optimal adjustment of milk production at the farm level. European Review of Agricultural Economics 12, 351–364.
- Reenberg, H., 1979. Udkiftning af malkekør. en stokastisk udkiftningsmodel. Memorandum 6, Jordbruksøkonomisk Institut.
- Reeves, C. (Ed.), 1993. Modern Heuristic techniques for combinatorial problems. Halsted Press.
- Reeves, C. (Ed.), 1995. Modern heuristic techniques for combinatorial problems. McGraw-Hill, London.
- Ross, S. M., 1970. Applied probability models with optimization applications. Holden-Day, San Francisco, California.
- Rubinstein, R., Shapiro, A., 1993. Discrete Event Systems. Sensitivity analysis and stochastic optimisation by the score function method. John Wiley & Sons, Chichester.
- Sandøe, P., Hurnik, F. (Eds.), 1996. Welfare of domestic animals. Concepts, theories and methods of measurement. No. 27 in Acta Agriculturæ Scandinavica, section A, Animal Science. Supplementum.
- Sandøe, P., Simonsen, H. B., 1992. Assessing animal welfare: where does science end and philosophy begin? Animal Welfare 1, 257–267.
- Sargent, R. G., 1998. Verification and validation of simulation models. In: Medeiros, D. J., Watson, J. S., Carson, J. S., Manivannan, M. S. (Eds.), Proceedings of the 1998 winter simulation conference. Washington DC, pp. 121–130.
- Scholten, H., Udink ten Cate, A., Eradus, W., Geurts, P., Glas, C., Miedema, R., 1995. Optimization of a fuzzy inference system for estrus detection in cattle. In:

- Udink ten Cate, A., Martin-Clouaire, R., Dijkhuizen, A., Lokhorst, C. (Eds.), 2nd IFAC/IFIP/EurAgEng Workshop on AI in Agriculture, Wageningen, May 29-31 1995. Wageningen, pp. 103–108.
- Shachter, R., 1986. Evaluating influence diagrams. *Operations Research* 34, 871–882.
- Shachter, R. D., Kenley, C., 1989. Gaussian influence diagrams. *Management Science* 35, 527–550.
- Shenoy, P. P., 1992. Valuation-based systems for bayesian decision analysis. *Operations Research* 40, 463–484.
- Singh, D., 1986. Simulation of swine herd population dynamics. *Agricultural Systems* 22, 157–183.
- Skidmore, A. L., Beverly, R. W., 1995. Utilizing object-oriented analysis, design, and programming techniques to develop a dairy herd simulation program. In: FACTS 95 March 7, 8, 9. Farm Animal Computer Technologies Conference. Orlando, Florida, pp. 190–195.
- Slader, R., Gregory, A., 1988. An automatic feeding and weighing system for ad libitum fed pigs. *Computers and Electronics in Agriculture* 3, 157–170.
- Smith, B. J., 1971. The dairy cow replacement problem. an application of dynamic programming. Bulletin 745, Florida Agricultural Experiment Station, Gainesville, Florida.
- Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. BUGS. Bayesian Inference Using Gibbs Sampling. Version 0.5 (version ii). MRC Biostatistics Unit. URL <http://weinberger.mrc-bsu.cam.ac.uk/bugs/Welcome.html>
- Sørensen, J., 1990. Validation of livestock herd simulation models: a review. *Livestock Production Science* 26, 79–90.
- Sørensen, J. T., Enevoldsen, C., 1991. Kvægsundhed, økonomi og mælkekvote. *Dans Veterinærtidsskrift* 74, 298–302.
- Sørensen, J. T., Kristensen, E. S., Thysen, I., 1992. A stochastic model simulating the dairy herd on a pc. *Agricultural Systems* 39, 177–200.
- Stewart, H. M., Burnside, E. B., Pfeiffer, W. C., 1978. Optimal culling strategies for dairy cows of different breeds. *Journal of Dairy Science* 61, 1605–1615.
- Stewart, H. M., Burnside, E. B., Wilton, J. W., Pfeiffer, W. C., 1977. A dynamic programming approach to culling decisions in commercial dairy herds. *Journal of Dairy Science* 60, 602–617.

- Strang, G., King, J., 1970. Litter productivity in large white pigs. *Animal Production* 12, 235–243.
- Sundgren, P.-E., van Male, J., Aumaitre, A., Kalm, E., Nielsen, H., 1979. Sow and litter recording procedures. report of a working party of the eaap commision on pig production. *Livestock Production Science* 7, 393–401.
- Tatman, J. A., Shachter, R. D., 1990. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics* 20, 365–379.
- Tess, M. W., Bennett, G. L., Dickerson, G. E., 1983. Simulation of genetic changes in life cycle efficiency of pork production. i. a bioeconomic model. *Journal of Animal Science* 56, 336–353.
- Thompson, P., Pulvermacher, R., Timms, L., 1995. Pedometer use for estrus detection. In: Proc. Animal Behavior and the Design of Livestock and Poultry Systems. International Conference. Indianapolis, Indiana, april 19-21. 1995. pp. 230–243.
- Thysen, I., 1983. Alternative fremgangsmåder ved optimering af foderrationer. In: Østergaard, V., Neimann-Sørensen, A. (Eds.), Optimum feeding of the dairy cow. Feed composition, feed intake, matabolism and production. Vol. 551 of Report from the National Institute of Animal Science. København, Ch. 17, pp. 17.1–17.28, in Danish.
- Thysen, I., 1993a. Monitoring bulk tank somatic cell counts by a multi-process kalman filter. *Acta Agriculturæ Scandinavica, Section A, Animal Science* 43, 58–64.
- Thysen, I., 1993b. Productivity control and decision support in dairy herds. In: Schiefer, G. (Ed.), Integrated systems in agricultural informatics. Bonn, pp. 149–158.
- Thysen, I., Enevoldsen, C., 1994. Visual monitoring of reproduction in dairy herds. *Preventive Veterinary Medicine* 19, 189–202.
- Thysen, I., Enevoldsen, C., Hindhede, J., Sørensen, J. T., 1988. Et system til styring af mælkeproduktion og reproduktion i kvægbesætningen. In: Østergaard, V., Hindhede, J. (Eds.), Studier i kvægproduktionssystemer. Vol. 649 of Beretning fra Statens Husdyrbrugsforsøg. København, pp. 15–57.
- Toft, N., Jørgensen, E., 2002. Estimation of farm specific parameters in a longitudinal model for litter size with variance components and random dropout. *Livestock Production Science* 77, 175–185.
- Toft, N., Kristensen, A. R., Jørgensen, E., 2005. A framework for decision support related to infectious diseases in slaughter pig fattening units. *Agricultural Systems* 85, 120–137.

- Treacy, D., 1987. In: Proc. of th 6th Conf. Australian Ass. of Anim. Breed. & Gen. University of Western Australia, Perth, WA Australia, Dep. Agric. & Rural Affairs, box 125, Bendigo, Vic. 3550 Australia,, pp. 285–288.
- Turner, M., Gurney, P., Belyavin, C., 1983. Automatic weighing of layer-replacement pullets housed on litter or in cages. *British Poultry Science* 24, 33–45.
- Vaarst, M., Enevoldsen, C., Thysen, I., Thiesson, B., Andersen, S. K., Criél, M., 1996. Mastitis diagnosis in Danish organic herds using Bayesian networks. Vol. 54 of Dina Research Report.
- van Arendonk, J. A. M., 1984. Studies on the replacement policies in dairy cattle.
i. evaluation of techniques to determine the optimum time for replacement and to rank cows on future profitability. *Zeitschrift für Tierzüchtung und Züchtungsbiologie* 101, 330–340.
- van Arendonk, J. A. M., 1985. Studies on the replacement policies in dairy cattle.
ii. optimum policy and influence of changes in production and prices. *Livestock Production Science* 13, 101–121.
- van Arendonk, J. A. M., 1986. Studies on the replacement policies in dairy cattle.
iv. influence of seasonal variation in performance and prices. *Livestock Production Science* 14, 15–28.
- van Arendonk, J. A. M., 1988. Management guides for insemination and replacement decisions. *Journal of Dairy Science* 71, 1050–1057.
- van Arendonk, J. A. M., Dijkhuizen, A. A., 1985. Studies on the replacement policies in dairy cattle. iii. optimum policy and influence of changes in production and prices. *Livestock Production Science* 13, 333–349.
- Van der Stuyft, E., Schofield, C., Randall, J., Wambacq, P., Goedseels, V., 1991. Development and application of computervision systems for uses in livestock production. *Computers and Electronics in Agriculture* 6, 243–265.
- van der Wal, J., Wessels, J., 1985. Markov decision processes. *Statistica Neerlandica* 39, 219–233.
- Versteegen, J. A., Sonnemans, J., Huirne, R. B., Dijkhuizen, A. A., Cox, J. C., 1998. Quantifying the effects of sow-herd management information systems on farmers' decision making using experimental economics. *American Journal of Agricultural Economics* 80, 821.
- Watson, R. T., 2002. Data management. *Databases and organizations*, 3rd Edition. John Wiley & Sons, Inc., New York.
- West, M., Harrison, J., 1997. *Bayesian Forecasting and Dynamic Models*, 2nd Edition. Springer Series in Statistics. Springer, New York.

- White, C. C. I., White, D. J., 1989. Markov decision processes. European Journal of Operational Research 39, 1–16.
- Whittemore, C. T., Fawcett, R. H., 1976. Theoretical aspects of a flexible model to simulate protein and lipid growth in pigs. Animal Production 22, 87.
- Williams, H. P., 1985. Model Building in Mathematical Programming. Wiley.

Appendix A

A short review of basic probability theory and distributions

Probabilities and distributions play an important role in almost all aspects of herd management science. In this appendix a very brief overview of basic concepts is given.

A.1 The probability concept

A.1.1 Interpretation of probabilities

In daily language the probability concept is used intensively. We may for instance claim that

- The probability of the outcome “5” when rolling a dice is $\frac{1}{6}$.
- The probability that cow no. 543 is pregnant is 0.40.
- The probability that a cure for cancer is developed within 5 years is 0.15.

Even though we intuitively have a clear understanding of what is meant by each of the three statements, a closer investigation will reveal that the probability concept is used in three very different senses of the word. We may distinguish among the following interpretations:

A frequency interpretation: For the “dice statement” the probability expresses how frequently we will observe a given outcome if exactly the same experiment is repeated a “large” number of times. The value is rather objective. This interpretation is relevant for the dice example. The statement only makes sense before the dice is rolled.

An objective belief interpretation: For the “cow pregnancy statement”, the probability expresses our belief in a certain (unobservable) state or event. Since the cow *is* either pregnant or open, the probability only refers to our belief - not to the true state of the cow. The belief may be based on an underlying frequency interpretation of similar cases and thus be rather objective.

A subjective belief interpretation: For the “cure for cancer statement” the probability expresses our belief in a certain unobservable (or not yet observed) event. Since no similar data are available, the belief will necessarily be highly subjective.

In herd management, we typically use the frequency interpretation to deal with biological variation (uncertainty about future outcomes) and a belief interpretation for representation of uncertain knowledge. If possible, we prefer the objective belief interpretation, but in some cases (typically when dealing with strategic decisions) subjective beliefs may be the only thing available. For a discussion of the use of subjective probabilities in agricultural decision analysis, reference is made to Anderson et al. (1977).

A.1.2 Experiments and associated concepts

In probability theory, the term “experiment” is used intensively. An experiment can be anything creating an outcome we can observe. For a given experiment we define the *sample space*, S , as the set of all possible outcomes.

An *event*, A , is a subset of S , i.e. $A \subseteq S$, and two events, A_1 and A_2 , are called *disjoint*, if they have no common outcomes, i.e. if $A_1 \cap A_2 = \emptyset$.

Example 25 Experiment: Rolling a dice

Rolling a dice and observing the outcome is an example of an experiment. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Examples of events are:

$$A_1: \{1\}$$

$$A_2: \{1, 5\}$$

$$A_3: \{4, 5, 6\}$$

Since $A_1 \cap A_3 = \emptyset$, we conclude that A_1 and A_3 are disjoint. On the other hand, A_1 and A_2 are *not* disjoint, because $A_1 \cap A_2 = \{1\}$.

■

A.1.3 Probability distributions on a sample space

Let S be the sample space of an experiment. We shall now present a simplified definition of a probability distribution.

Definition 9 A *probability distribution* P on the sample space S is a function, so that

- $P(S) = 1$.
- For any event $A \subseteq S$, $0 \leq P(A) \leq 1$.
- For any two disjoint events A_1 and A_2 , $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

■

Example 26 Probability distribution: Rolling a dice

Just like Example 25, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. A valid probability function on S is, for $A \subseteq S$,

$$P(A) = \frac{|A|}{6}, \quad (\text{A.1})$$

where $|A|$ is the size of A (i.e. the number of elements it contains). Applying Eq. (A.1) for selected events results in the following probabilities:

- $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$.
- $P(\{1, 5\}) = \frac{2}{6} = \frac{1}{3}$.
- $P(\{1, 2, 3\}) = \frac{3}{6} = \frac{1}{2}$.

It should be noticed that many other valid probability functions could be defined, even though the one defined here is the only one that makes sense from a frequency point of view.

■

A.1.4 Independence

If two events, A and B are independent (i.e. if the output of one of the events has no influence at all on the outcome of the other), then

$$P(A \cap B) = P(A)P(B). \quad (\text{A.2})$$

Example 27 Independence: Rolling two dices

If we role two dices (a red and a blue) simultaneously, it is reasonable to assume the outcomes for the two dices to be independent. The sample space, S , for the experiment is

$$S = \{(1, 1), (1, 2), \dots, (1, 6), \dots, (6, 6)\}$$

where the outcome for the red dice is always first in the tuple. We define the probability distribution P on S for any $A \subseteq S$ as

$$P(A) = \frac{|A|}{36}. \quad (\text{A.3})$$

where $|A|$ is the number of elements in A . Now, define two events, A and B :

$$\begin{aligned} A &= \{(6, 1), (6, 2), \dots, (6, 6)\} \Rightarrow P(A) = \frac{6}{36} = \frac{1}{6} \\ B &= \{(1, 6), (2, 6), \dots, (6, 6)\} \Rightarrow P(B) = \frac{6}{36} = \frac{1}{6}. \end{aligned}$$

In words, the event A represents that the outcome of the red dice is “6” and B is the event that the blue dice shows “6”.

Now, define the event C as the outcome that both dices show “6”, i.e. $C = A \cap B = \{(6, 6)\}$. By use of Eq. (A.3) we calculate the probability of C as $P(C) = \frac{1}{36}$, and we notice that Eq. (A.2) also applies in this case, since

$$P(C) = P(A \cap B) = P(A)P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

■

A.1.5 Conditional probabilities and sum rule

Definition 10 Let A and B be two events, where $P(B) > 0$. The *conditional probability* of A given B is written as $P(A | B)$, and it is defined as

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

Example 28 Conditional probabilities: Rolling a dice

Again, let $S = \{1, 2, 3, 4, 5, 6\}$, and $P(A) = \frac{|A|}{6}$. Define the events $B = \{1, 2, 3\}$, and $A = \{2\}$. Then, $A \cap B = \{2\}$, and

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{1, 2, 3\})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

The result is in agreement with the intuitive expectation that if the outcome is known to be “1”, “2” or “3” all three values are equally probable.

■

Let A_1, A_2, \dots, A_n be pairwise disjoint events so that

$$\bigcup_{i=1}^n A_i = S.$$

Furthermore, let B be an event so that $P(B) > 0$. According to the *chain rule* we may calculate $P(B)$ as

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i). \quad (\text{A.4})$$

Example 29 Chain rule: Rolling a dice

Again, let $S = \{1, 2, 3, 4, 5, 6\}$, and $P(A) = \frac{|A|}{6}$. Define the three disjoint events

- $A_1 = \{1, 2\}$,
- $A_2 = \{3, 4\}$,
- $A_3 = \{5, 6\}$.

Then, $A_1 \cup A_2 \cup A_3 = S$. Define the event $B = \{1, 3, 5\}$ with the probability $P(B) = \frac{3}{6} = \frac{1}{2}$. The conditional probabilities $P(B | A_i)$ for $i = 1, 2, 3$ are calculated as

$$P(B | A_1) = \frac{P(B \cap A_1)}{P(A_1)} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2},$$

and analogously for $i = 2$ and $i = 3$. Applying the chain rule enables us to calculate the probability of B indirectly as

$$\begin{aligned} P(B) &= \sum_{i=1}^n P(B | A_i)P(A_i) \\ &= P(B | A_1)P(A_1) + P(B | A_2)P(A_2) + P(B | A_3)P(A_3) \\ &= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{3}{6} = \frac{1}{2}. \end{aligned}$$

■

A.1.6 Bayes' theorem

One of the most important theorems at all in probability calculus is Bayes' theorem which may be stated as follows.

Theorem 1 (Bayes' theorem) *Let A_1, A_2, \dots, A_n be pairwise disjoint events so that $\bigcup_{i=1}^n A_i = S$, where S is the sample space. If B is an event so that $P(B) > 0$, then*

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^n P(B | A_j)P(A_j)}.$$

If we compare with the chain rule (A.4) we notice that Bayes' theorem may also be written as

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{P(B)}, \quad (\text{A.5})$$

implying that the theorem provides us with a formula for “reversing” conditional probabilities. This is extremely important for all kinds of reasoning under uncertainty as the following example will illustrate.

Example 30 Updating of belief

In a dairy herd, the conception rate is known to be 0.40, and the heat detection rate is known to be 0.55.

Define I as the event “insemination” of a dairy cow. Define also the events Π_+ and Π_- as the events “pregnant” and “not pregnant”, respectively. The conditional probability $P(\Pi_+ | I)$ is equal to the conception rate, i.e. $P(\Pi_+ | I) = 0.40$. Accordingly, $P(\Pi_- | I) = 0.60$.

After 3 weeks, the farmer observes the cow for heat. Define H_+ as the event that the farmer actually detects heat, and let H_- be the event that no heat is detected. Thus, $P(H_+ | \Pi_-) = 0.55$, and accordingly, $P(H_- | \Pi_-) = 0.45$. There is a slight risk that the farmer erroneously observes a pregnant cow to be in heat. We therefore assume, that $P(H_+ | \Pi_+) = 0.01$ (the value is a guess).

Now, assume that the cow has been inseminated (implying that the probability of pregnancy is 0.40), and after 3 weeks the farmer observes the cow and concludes that it is *not* in heat. Thus, we have observed the event H_- , and we would like to know the *updated* probability that the cow is pregnant. We apply Bayes' theorem:

$$P(\Pi_+ | H_-) = \frac{P(H_- | \Pi_+)P(\Pi_+)}{P(H_- | \Pi_+)P(\Pi_+) + P(H_- | \Pi_-)P(\Pi_-)}.$$

All conditional probabilities of the above expression are known, and we have

$$P(\Pi_+ | H_-) = \frac{0.99 \cdot 0.40}{0.99 \cdot 0.40 + 0.45 \cdot 0.60} = 0.59.$$

The conclusion is that the negative heat observation result increases our belief in the (unobservable) event “pregnant” from the original 0.40 to 0.59.

A.2 Discrete distributions

For some experiments, the probability is defined by a certain function defined over the sample space. In those cases, we say that the outcome is drawn from a standard distribution. If the sample space is a countable set, we denote the corresponding distribution as discrete.

There exists standard distributions corresponding to many natural phenomena. Three classical examples are presented in the following subsections after the definition of three important concepts describing a distribution.

Definition 11 If X is a random variable representing the (numerical) outcome of an experiment with sample space S , and $P(X)$ is the probability of the outcome, then the *expected value*, $E(X)$, of X is defined as

$$E(X) = \sum_{k \in S} kP(k). \quad (\text{A.6})$$

The *variance*, $\text{Var}(X)$, is defined as

$$\text{Var}(X) = E(X^2) - (E(X))^2. \quad (\text{A.7})$$

The *distribution function* $F(x)$ is defined as the probability of X being less than or equal to x , i.e.

$$F(x) = \sum_{k \leq x} P(k). \quad (\text{A.8})$$

■

A.2.1 The binomial distribution

Consider an experiment with binary outcomes - we either observe a success (s) or a failure (f). Examples of this kind of experiments include for instance:

Mating of a sow: Pregnant (s), not pregnant (f).

Tossing a coin: Heads (s), tails (f).

Testing for a disease: Positive (s), negative (f).

Assume that the probability of success is p , and that the experiment is repeated n times. Let X be the total number of successes observed in the n experiments. The sample space of the compound n experiments is $S = \{0, 1, 2, \dots, n\}$. The random variable X is then said to be *binomially distributed* with parameters n and p . In symbols we write $X \sim \mathcal{B}(n, p)$.

The probability function $P(X = k)$ is (by frequency interpretation) given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad (\text{A.9})$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (\text{A.10})$$

is the *binomial coefficient* which may be calculated directly from the equation or looked up in a table.

The expected value (mean) of a binomial distribution is simply

$$\mathbb{E}(X) = np, \quad (\text{A.11})$$

and the variance is

$$\text{Var}(X) = np(1 - p). \quad (\text{A.12})$$

The binomial distribution is one of the most frequently used distribution for natural phenomena.

A.2.2 The hypergeometric distribution

Consider an experiment where n items are drawn at random from a total population of N items among which D have a special property (we assume that $0 < n \leq N$ and $0 \leq D \leq N$). Let the random variable X denote the number of items having the special property among the items drawn. The sample space is $S = \{n_-, n_- + 1, \dots, n_+\}$, where $n_- = \max\{0, n - N + D\}$ and $n_+ = \min\{n, D\}$. The random variable X is then said to have a *hypergeometric distributed* with parameters n, N and D . In symbols we write $X \sim \mathcal{H}(n, N, D)$.

The probability function $P(X = k)$ is (by frequency interpretation) given by

$$P(X = k) = \frac{\binom{D}{k} \binom{N - D}{n - k}}{\binom{N}{n}}. \quad (\text{A.13})$$

The expected value (mean) of a hypergeometric distribution is

$$\mathbb{E}(X) = \frac{nD}{N}, \quad (\text{A.14})$$

and the variance is

$$\text{Var}(X) = \frac{nD(N - D)}{N^2} \cdot \frac{N - n}{N - 1}. \quad (\text{A.15})$$

A.2.3 The Poisson distribution

If a certain phenomenon occurs at random with a constant intensity (but independently of each others) the total occurrences X in a time interval of a given length (or in a space of a given area) is *Poisson distributed* with parameter λ . In symbols we write $X \sim \mathcal{P}(\lambda)$. Examples of such counts include:

- Number of (non-infectious) disease cases per month.

- Number of feeding system failures per year.
- Number of labor incidents per year.

The sample space for X is $S = \{0, 1, 2, \dots\}$, and the probability function defined on S is (by frequency interpretation):

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}. \quad (\text{A.16})$$

The expected value (mean) of a Poisson distribution is

$$E(X) = \lambda, \quad (\text{A.17})$$

and also the variance is simply

$$\text{Var}(X) = \lambda. \quad (\text{A.18})$$

The Poisson distribution may be used as an approximation for a binomial distribution with “small” p and “large” n .

A.3 Continuous distributions

In some cases, the sample space S of a distribution is *not* countable. If, furthermore, S is an interval on \mathbb{R} , the random variable X taking values in S is said to have a *continuous* distribution.

If the random variable X has a continuous distribution, we have for any $x \in S$:

$$P(X = x) = 0. \quad (\text{A.19})$$

Thus, no probability function exists for a continuous distribution. Instead, the distribution is defined by a *density function*, $f(x)$. By convention we define that $f(x) = 0$ for $x \in \mathbb{R} \setminus S$. The density function has the following properties (for $a, b \in S$ and $a \leq b$):

$$P(X \in S) = \int_S f(x)dx = 1 \quad (\text{A.20})$$

and

$$P(a \leq X < b) = \int_a^b f(x)dx. \quad (\text{A.21})$$

Thus, for a continuous distribution, f can only be interpreted as a probability when integrated over an interval. The *expected value*, $E(X)$, of X is defined as

$$E(X) = \int_S xf(x)dx. \quad (\text{A.22})$$

The *variance*, $\text{Var}(X)$, is calculated as in the discrete case, i.e. $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2$. The distribution function is also for a continuous distribution defined as the probability of X being less than or equal to x , but it is evaluated as an integral,

$$F(x) = \int_{-\infty}^x f(x)dx. \quad (\text{A.23})$$

It follows directly, that

$$F'(x) = f(x). \quad (\text{A.24})$$

In the following subsections, we shall introduce three classical continuous distributions.

A.3.1 The uniform distribution

If $S = [a; b]$, and the random variable X has a *uniform distribution* on S , then the density function is, for $X \in S$,

$$f(x) = \frac{1}{b-a}. \quad (\text{A.25})$$

In symbols, we write $X \sim \mathcal{U}(a, b)$. The expected value and the variance are

$$\text{E}(X) = \frac{a+b}{2}, \quad (\text{A.26})$$

and

$$\text{Var}(X) = \frac{(b-a)^2}{12}. \quad (\text{A.27})$$

A.3.2 The normal distribution

If the sample space S of a random variable, X , is the entire set of real numbers, \mathbf{R} , and X has a *normal distribution* with parameters μ and σ , then the density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (\text{A.28})$$

In symbols, we write $X \sim \mathcal{N}(\mu, \sigma^2)$. The expected value and the variance are simply

$$\text{E}(X) = \mu, \quad (\text{A.29})$$

and

$$\text{Var}(X) = \sigma^2. \quad (\text{A.30})$$

The normal distribution may be used to represent almost all kinds of random outcome on the continuous scale in the real world. Exceptions are phenomena that are bounded in some sense (e.g. the waiting time to be served in a queue cannot be negative). The normal distribution is the cornerstone among statistical distributions.

It can be shown (central limit theorems) that if X_1, X_2, \dots, X_n are random variables of (more or less) any kind, then the sum $Y_n = X_1 + X_2 + \dots + X_n$ is normally distributed for n sufficiently large.

The following theorem describes other important properties of normally distributed random variables.

Theorem 2 Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and assume that X_1 and X_2 are independent. Furthermore, let a and b be arbitrary real numbers, and define $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$ and $Y_3 = aX_1 + b$. Then,

$$Y_1 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2),$$

$$Y_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2),$$

and

$$Y_3 \sim \mathcal{N}(a\mu_1 + b, a^2\sigma_1^2).$$

■

The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *standard normal distribution*. A random variable being standard normally distributed is often denoted as Z . The density function and the distribution function of the standard normal distribution are often denoted as ϕ and Φ , respectively. It follows from Eq. (A.28) that

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \quad (\text{A.31})$$

Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and define

$$Z = \frac{X - \mu}{\sigma}.$$

It then follows from Theorem 2, that $Z \sim \mathcal{N}(0, 1)$. Accordingly the values of the density function f and distribution function F of X may be calculated as

$$f(x) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right), \quad (\text{A.32})$$

and

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (\text{A.33})$$

Thus, if we have a table for (or a computer implementation of) the standard normal distribution, we may use it to calculate the values for any normal distribution.

A.3.3 The truncated normal distribution

If, in a population of individuals (e.g. animals), a numerical trait X has a normal distribution $\mathcal{N}(\mu, \sigma^2)$, and individuals are selected so that only those having $a < X \leq b$ are kept, then the trait X_{ab} among the selected individuals has a *truncated normal distribution*. If the density function of X is denoted as $f(x)$ defined as in Eq. (A.28), and the corresponding distribution function is denoted as $F(x)$, then the density function $f_{ab}(x)$ of X_{ab} is

$$f_{ab}(x) = \frac{f(x)}{F(b) - F(a)} = \frac{\phi(z)}{\sigma (\Phi(\beta) - \Phi(\alpha))}, \quad (\text{A.34})$$

where $z = \frac{x-\mu}{\sigma}$, $\alpha = \frac{a-\mu}{\sigma}$, and $\beta = \frac{b-\mu}{\sigma}$.

The expected value and the variance of X_{ab} are simply

$$\mathbb{E}(X_{ab}) = \mu - \sigma^2 \frac{f(b) - f(a)}{F(b) - F(a)} = \mu - \sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}, \quad (\text{A.35})$$

and

$$\text{Var}(X_{ab}) = \sigma^2 \left(1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right). \quad (\text{A.36})$$

The special cases, where either $a = \infty$ or $b = \infty$ are easily handled by the formulas above. If, for instance $b = \infty$, then

$$\mathbb{E}(X_{a\infty}) = \mu - \sigma^2 \frac{f(\infty) - f(a)}{F(\infty) - F(a)} = \mu + \sigma^2 \frac{f(a)}{1 - F(a)} = \mu - \sigma \frac{\phi(\alpha)}{1 - \Phi(\alpha)}.$$

A.3.4 The exponential distribution

If the sample space is $S = \mathbb{R}_+$, and the random variable X has an *exponential distribution* with parameter λ , then the density function is

$$f(x) = \lambda e^{-\lambda x}. \quad (\text{A.37})$$

In symbols, we write $X \sim \mathcal{E}(\lambda)$. The expected value and the variance are

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad (\text{A.38})$$

and

$$\text{Var}(X) = \frac{1}{\lambda^2}. \quad (\text{A.39})$$

The exponential distribution is in many ways complimentary to the Poisson distribution. If something happens at random at constant intensity, the number of events within a fixed time interval is Poisson distributed, and the waiting time between two events is exponentially distributed.

A.4 Sampling from a distribution

Assume that X_1, X_2, \dots, X_n are sampled independently from the same arbitrary distribution having the known mean μ and the known standard deviation σ . Define the sample mean as

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

Then,

$$\text{E}(\bar{X}) = \mu, \quad (\text{A.40})$$

and,

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}. \quad (\text{A.41})$$

It follows from (A.41) that the standard deviation of \bar{X} is σ/\sqrt{n} .

If, in particular, for all i , $X_i \sim \mathcal{N}(\mu, \sigma^2)$ then $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

A.5 Hypothesis testing

A.5.1 Normal distributions

Assume that X_1, X_2, \dots, X_n are sampled independently from the same normal distribution having the known mean μ and the known standard deviation σ . For some reason we expect (hope) that μ has a certain value, μ_0 , and we would therefore like to test the hypothesis $H_0 : \mu = \mu_0$.

From Section A.4 we know that the sample mean \bar{X} has the distribution $\mathcal{N}(\mu, \sigma^2/n)$. We may use this information to calculate a confidence interval for the mean, μ .

A 95% confidence interval for the sample mean under H_0 is calculated as

$$\mu_0 \pm 1.96 \frac{\sigma}{\sqrt{n}},$$

where we use the facts that $1.96 = \Phi^{-1}(0.975)$, and $-1.96 = \Phi^{-1}(0.025)$. In other words, in the standard normal distribution 2.5% of the observations fall below -1.96, and 97.5% fall below 1.96. Thus 95% fall between the two values.

The idea of calculating a confidence interval is to check whether the sample mean, \bar{X} , falls inside the interval. If the answer is yes, we accept the hypothesis. Otherwise we reject it.

When dealing with basic production monitoring, a good approximation to the confidence interval is

$$\mu_0 \pm 2 \frac{\sigma}{\sqrt{n}}.$$

If *neither* μ nor σ are known, the sample mean becomes student-t distributed (with $n - 1$ degrees of freedom) instead. Then the confidence interval becomes larger because of the uncertainty about σ . However, for large n , the student-t distribution converges towards a standard normal distribution.

A.5.2 Binomial distributions

Assume that we have observed the outcome of X successes out of n binomial trials. We would like to test the hypothesis $H_0 : p = p_0$.

Under the hypothesis, the expected number of successes is $E_0(X) = np_0$, and the variance is $\text{Var}_0(X) = np_0(1 - p_0)$. In order to test the hypothesis, we calculate the likelihood (probability) that the observed value of X is drawn from the distribution $\mathcal{B}(n, p_0)$. This may basically be done in two different ways:

Approximation with the normal distribution $\mathcal{N}(np_0, np_0(1 - p_0))$: This is a reasonable approach if n is big, and p_0 is neither close to 0 nor 1.

Direct use of the binomial distribution function: If n is small or p_0 is close to 0 or 1, this is the only valid option.

Appendix B

Selected topics from linear algebra

In this appendix a very short overview of basic topics from linear algebra is given. The main purpose is to introduce vectors and matrices with their associated operations, but first, a few characteristics of real numbers are described.

B.1 Real numbers

B.1.1 Elements of the real number system

The real number system consists of 4 parts:

- A set, \mathbb{R} , of all real numbers.
- A relation “ $<$ ” on \mathbb{R} . If, $a, b \in \mathbb{R}$, then “ $a < b$ ” is either true or false. It is called the *order relation*.
- A function, “ $+$ ”: $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. The *addition* operation.
- A function, “ \cdot ”: $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. The *multiplication* operation.

B.1.2 Axioms for real numbers

A number of *axioms* apply to real numbers. We shall particularly mention the following:

Associative laws: For the two operations we have:

$$a + (b + c) = (a + b) + c$$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

Commutative laws: For the two operations we have:

$$a + b = b + c$$

$$a \cdot b = b \cdot a$$

Distributive law: $a \cdot (b + c) = a \cdot b + a \cdot c$

Additive identity (“zero element”): There exists an element in R called 0, so that, for all a , $a + 0 = a$

Additive inverse: For all a , there exists a b so that $a + b = 0$, and $b = -a$

Multiplicative identity (“one element”): There exists an element in R called 1, so that, for all a , $1 \cdot a = a$

Multiplicative inverse: For all $a \neq 0$, there exists a b so that $a \cdot b = 1$, and $b = a^{-1}$

B.1.3 Solving equations

Let $a \neq 0$ and b be known real numbers, and let x be an unknown real number. If, for some reason, we know that $a \cdot x = b$, we say that we have a *linear equation*. We can *solve* the equation (determine the value of x) in a couple of stages using the axioms:

$$\begin{aligned} a \cdot x &= b \Leftrightarrow \\ a^{-1} \cdot a \cdot x &= a^{-1} \cdot b \Leftrightarrow \\ 1 \cdot x &= a^{-1} \cdot b \Leftrightarrow \\ x &= a^{-1} \cdot b. \end{aligned}$$

B.2 Vectors and matrices

B.2.1 Definition, notation and examples

A matrix is a rectangular table of real numbers arranged in columns and rows. The *size* of a matrix is written as $n \times m$, where n is the number of rows, and m is the number of columns. We may refer to a matrix using a single symbol, like a, b, x etc. Sometimes we use bold face, $\mathbf{a}, \mathbf{b}, \mathbf{x}$, or underline, $\underline{a}, \underline{b}, \underline{x}$ in order to emphasize that we refer to a matrix and not just a real number.

Examples of matrices:

A 2×3 matrix:

$$\begin{pmatrix} 3 & 6 & 2 \\ -1 & 5 & 23 \end{pmatrix}$$

A 4×3 matrix:

$$\begin{pmatrix} 2 & 5 & 1 \\ -1 & 7 & 3 \\ 3 & 16 & 12 \\ -3 & 4 & 3 \end{pmatrix}$$

Symbol notation for a 2×2 matrix:

$$a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

B.2.2 Special matrices, vectors

A matrix a of size $n \times n$ is called a *quadratic matrix*:

$$a = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

A matrix b of size $1 \times n$ is called a *row vector*:

$$b = (b_{11} \ b_{12} \ \dots \ b_{1n}).$$

A matrix c of size $n \times 1$ is called a *column vector*:

$$c = \begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{n1} \end{pmatrix}.$$

B.2.3 Addition operation

Definition

Just like in real numbers, we have an addition operation for matrices. It is, however, not valid for arbitrary matrices, but only for matrices of same size (say $n \times m$). The operation is defined as an element by element addition of the corresponding real numbers in the two matrices, i.e.

$$a + b = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{pmatrix} =$$

$$\begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \dots & a_{nm} + b_{nm} \end{pmatrix}.$$

Associative and commutative laws

From the axioms for real numbers, it follows directly, that the commutative and associative laws are also valid for matrix addition (when the operation is defined). In other words, if a , b and c are of equal size, then

$$a + (b + c) = (a + b) + c,$$

and

$$a + b = b + a.$$

Additive identity: “Zero element”

We also easily see, that the set of $n \times m$ matrices has a “zero element”, $\underline{0}$, consisting of only zeros, so that for any $n \times m$ matrix, a ,

$$a + \underline{0} = a.$$

Additive inverse

The additive inverse of a matrix a is denoted as $-a$ and it is created just by replacing each of its real number elements with the additive inverse. Thus,

$$a + (-a) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} + \begin{pmatrix} -a_{11} & -a_{12} & \dots & -a_{1m} \\ -a_{21} & -a_{22} & \dots & -a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & -a_{nm} \end{pmatrix} = \underline{0}$$

B.2.4 Multiplication operation

Definition

We also have a multiplication operation for matrices. Like addition, it is not valid for arbitrary matrices, but only for matrices where the number of columns of the first matrix equals the number of rows of the second matrix. Thus, if a is an $n \times m$ matrix, and b is an $m \times k$ matrix, the multiplication operation is defined as

$$a \cdot b = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mk} \end{pmatrix} =$$

$$\begin{pmatrix} \sum_j a_{1j}b_{j1} & \sum_j a_{1j}b_{j2} & \dots & \sum_j a_{1j}b_{jk} \\ \sum_j a_{2j}b_{j1} & \sum_j a_{2j}b_{j2} & \dots & \sum_j a_{2j}b_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_j a_{nj}b_{j1} & \sum_j a_{nj}b_{j2} & \dots & \sum_j a_{nj}b_{jk} \end{pmatrix}.$$

The result of the multiplication is a matrix of size $n \times k$.

Laws for matrix multiplication

It is rather easily proved that the associative law holds for matrix multiplication (when defined), i.e.

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

Due to the restriction on size, it is obvious that the commutative law is *not* valid for matrix multiplication, since most often, the product $b \cdot a$ doesn't exist. Even when $b \cdot a$ exists, it most often differs from $a \cdot b$. Thus, for matrices most often,

$$a \cdot b \neq b \cdot a$$

As concerns, the distributive law, it is easy to show that it holds for matrices. Thus, if a is an $n \times m$ matrix, and b and c are $m \times k$ matrices, then

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Vector multiplication

Since vectors are just special variants of matrices, all rules that apply to matrices are also valid for vectors. In particular, a row vector, a , of size $1 \times n$ may be multiplied with a column vector, b , of size $n \times 1$. The product $a \cdot b$ is a 1×1 matrix (i.e. a real number),

$$a \cdot b = (a_{11} \ a_{12} \ \dots \ a_{1n}) \cdot \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{pmatrix} = \sum_j a_{1j}b_{j1}.$$

The “opposite” product, $b \cdot a$, is a quadratic $n \times n$ matrix:

$$b \cdot a = \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{pmatrix} = \begin{pmatrix} b_{11}a_{11} & b_{11}a_{12} & \dots & b_{11}a_{1n} \\ b_{21}a_{11} & b_{21}a_{12} & \dots & b_{21}a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1}a_{11} & b_{n1}a_{12} & \dots & b_{n1}a_{1n} \end{pmatrix}.$$

Multiplicative identity: “One element”

For the set of $n \times m$ matrices, there exist two quadratic “one elements”, I_n and I_m , having the property that, for an arbitrary $n \times m$ matrix:

$$I_n \cdot a = a \cdot I_m = a.$$

The “one element”, I_n is referred as the *identity matrix* of size n . It is defined as

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

An identity matrix is always a quadratic matrix of which the diagonal elements are 1 and all other are 0.

B.2.5 Other matrix operations

A real number r may be multiplied with a matrix a :

$$ra = ar = r \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} = \begin{pmatrix} ra_{11} & ra_{12} & \dots & ra_{1m} \\ ra_{21} & ra_{22} & \dots & ra_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ ra_{n1} & ra_{n2} & \dots & ra_{nm} \end{pmatrix}.$$

The *transpose*, a' of a matrix a is formed by changing columns to rows and vice versa:

$$a' = \left(\begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{array} \right)' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix}.$$

Thus, the transpose of an $n \times m$ matrix is an $m \times n$ matrix.

The *determinant*, $\det(a)$ of a quadratic $n \times n$ matrix is a real number. Calculation of the determinant is rather complicated for large n . However many software packages support calculation of determinants. Here we shall only show the calculation for $n = 2$ and $n = 3$:

The determinant of a 2×2 matrix is:

$$\det(a) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}.$$

The determinant of a 3×3 matrix is:

$$\det(a) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =$$

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{32}a_{23}a_{11}.$$

A quadratic matrix, a , is called *singular* if, and only if, $\det(a) = 0$. Accordingly, if $\det(a) \neq 0$, then a is said to be *non-singular*.

The *trace* of a quadratic matrix, a , of size $n \times n$ is,

$$tr(a) = \sum_{i=1}^n a_{ii},$$

i.e. the trace is the sum of the diagonal elements.

B.2.6 The multiplicative inverse of a quadratic matrix

If a quadratic matrix, a , is non-singular, it has a unique (multiplicative) inverse matrix, a^{-1} , satisfying

$$a \cdot a^{-1} = a^{-1} \cdot a = I,$$

where I is the identity matrix of same size as a .

Finding the inverse of a matrix is rather complicated, but many software packages support the operation. For very big matrices (millions of rows and columns), inversion is a challenge even to modern computers. Nevertheless, it is crucial in many herd management implementations (and certainly also in animal breeding).

Here we shall only show the calculation for $n = 2$ and $n = 3$:

Inversion of a 2×2 matrix:

$$a^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{\det(a)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

Inversion of a 3×3 matrix:

$$a^{-1} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}^{-1} = \frac{1}{\det(a)} \begin{pmatrix} \left| \begin{array}{cc} a_{22} & a_{23} \\ a_{32} & a_{33} \end{array} \right| & - \left| \begin{array}{cc} a_{21} & a_{23} \\ a_{31} & a_{33} \end{array} \right| & \left| \begin{array}{cc} a_{21} & a_{22} \\ a_{31} & a_{32} \end{array} \right| \\ - \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{32} & a_{33} \end{array} \right| & \left| \begin{array}{cc} a_{11} & a_{13} \\ a_{31} & a_{33} \end{array} \right| & - \left| \begin{array}{cc} a_{11} & a_{12} \\ a_{31} & a_{32} \end{array} \right| \\ \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{22} & a_{23} \end{array} \right| & - \left| \begin{array}{cc} a_{12} & a_{13} \\ a_{21} & a_{23} \end{array} \right| & \left| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right| \end{pmatrix}'$$

B.2.7 Systems of linear equations

Solving systems of linear equation plays an important role in many herd management methods. Often, we need to deal with systems of thousands, or even millions, of equations. To illustrate the use of matrices in this connection we, however, start with a very simple (and naive) example.

Example 31 Old McDonald has a farm ...

McDonald has sheep and geese on his farm, but he has forgotten how many of each species. Let us denote the number of sheep as x_1 and the number of geese as x_2 . The other day, he counted the number of heads of his animals. The number was 25. He knows that sheep and geese have one head each, so he set up the following equation:

$$1x_1 + 1x_2 = 25.$$

He also counted the number of legs, and it was 70. He knows that a sheep has 4 legs and a goose has 2 legs, so he also set up the following equation:

$$4x_1 + 2x_2 = 70.$$

Now, define the matrix a and the vectors b and x as

$$a = \begin{pmatrix} 1 & 1 \\ 4 & 2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad b = \begin{pmatrix} 25 \\ 70 \end{pmatrix}.$$

We may then express the two equations as one matrix equation,

$$\begin{pmatrix} 1 & 1 \\ 4 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 25 \\ 70 \end{pmatrix},$$

or, in short notation:

$$ax = b.$$

Having brought a system of linear equations to the elegant form, $ax = b$, solution for x is just as straightforward as with an equation of real numbers:

$$\begin{aligned} a \cdot x = b &\Leftrightarrow \\ a^{-1} \cdot a \cdot x = a^{-1} \cdot b &\Leftrightarrow \\ I \cdot x = a^{-1} \cdot b &\Leftrightarrow \\ x = a^{-1} \cdot b. \end{aligned}$$

This is true no matter whether we have a system of 2 equations like Example 31, or we have a system of a million equations.

Even though solution by inversion of a matrix as shown above is mathematically correct, it is important to know that from a computational point of view, it is not recommendable, because it is numerically unprecise. If solution of systems of linear equations is needed, the documentation of the software used should be studied carefully in order to find a method which is numerically better. Many software packages have better routines for numerical solution of equations than just inversion. If not, a textbook on numerical methods should be consulted.

B.3 Linear regression

Matrices and vectors also play an important role in linear regression and other kinds of linear models. We shall illustrate that through an example.

Example 32 Linear regression

In a study of children born in Berkeley 1928-29, the height and weight of 10 18-year old girls were measured. The results (in cm and kg, respectively) were as follows:

Girl #, i	Height, x_i	Weight, Y_i
1	169.6	71.2
2	166.8	58.2
3	157.1	56.0
4	181.1	64.5
5	158.4	53.0
6	165.6	52.4
7	166.7	56.8
8	156.5	49.2
9	168.1	55.6
10	165.3	77.8

It seems reasonable to assume, that the weight, Y_i depends on the height x_i according to the following linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, 10. \quad (\text{B.1})$$

Now, define the vectors and matrices

$$Y = \begin{pmatrix} 71.2 \\ 58.2 \\ 56.0 \\ 64.5 \\ 53.0 \\ 52.4 \\ 56.8 \\ 49.2 \\ 55.6 \\ 77.8 \end{pmatrix} \quad x = \begin{pmatrix} 1 & 169.6 \\ 1 & 166.8 \\ 1 & 157.1 \\ 1 & 181.1 \\ 1 & 158.4 \\ 1 & 165.6 \\ 1 & 166.7 \\ 1 & 156.5 \\ 1 & 168.1 \\ 1 & 165.3 \end{pmatrix} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

We may then write the model (B.1) in matrix notation simply as

$$Y = x\beta + \epsilon. \quad (\text{B.2})$$

It should be noticed, that the linear model (B.1) is also able to handle class variables. If, instead of 10 girls, we had observed 5 girls and 5 boys, we could introduce a binary variable, x_{i1} having the value 0 for “girl”, and the value 1 for “boy”. Thus, our data matrix, x and coefficient vector β would just have been,

$$x = \begin{pmatrix} 1 & 1 & 169.6 \\ 1 & 1 & 166.8 \\ 1 & 0 & 157.1 \\ 1 & 1 & 181.1 \\ 1 & 0 & 158.4 \\ 1 & 0 & 165.6 \\ 1 & 1 & 166.7 \\ 1 & 0 & 156.5 \\ 1 & 1 & 168.1 \\ 1 & 0 & 165.3 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix},$$

and the model (B.2) would still hold. ■

The model (B.2) could be extended to include more explanatory variables and more data. This only means that the matrices defined grow bigger. So, in general, if data can be described by a linear model as in (B.2), a *least squares estimate*, $\hat{\beta}$, of the coefficients in the vector β is calculated as

$$\hat{\beta} = (x'x)^{-1}x'Y. \quad (\text{B.3})$$

Furthermore, define the vector of predictions $\hat{Y} = x\hat{\beta}$. An estimate, s^2 , for the residual variance, σ^2 is calculated as

$$s^2 = \frac{1}{n - k}(Y - \hat{Y})'(Y - \hat{Y}), \quad (\text{B.4})$$

where n is the number of observations (number of rows in x), and k is the number of parameters estimated (number of columns in x).

Example 33 Numerical results

For Example 32 (without class variable for gender), we get the following estimates:

$$\hat{\beta} = \left(\begin{pmatrix} 1 & 169.6 \\ 1 & 166.8 \\ 1 & 157.1 \\ 1 & 181.1 \\ 1 & 158.4 \\ 1 & 165.6 \\ 1 & 166.7 \\ 1 & 156.5 \\ 1 & 168.1 \\ 1 & 165.3 \end{pmatrix}' \begin{pmatrix} 1 & 169.6 \\ 1 & 166.8 \\ 1 & 157.1 \\ 1 & 181.1 \\ 1 & 158.4 \\ 1 & 165.6 \\ 1 & 166.7 \\ 1 & 156.5 \\ 1 & 168.1 \\ 1 & 165.3 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 169.6 \\ 1 & 166.8 \\ 1 & 157.1 \\ 1 & 181.1 \\ 1 & 158.4 \\ 1 & 165.6 \\ 1 & 166.7 \\ 1 & 156.5 \\ 1 & 168.1 \\ 1 & 165.3 \end{pmatrix}' \begin{pmatrix} 71.2 \\ 58.2 \\ 56.0 \\ 64.5 \\ 53.0 \\ 52.4 \\ 56.8 \\ 49.2 \\ 55.6 \\ 77.8 \end{pmatrix}.$$

By doing the calculations at the right hand side, we get

$$\beta = \begin{pmatrix} -36.9 \\ 0.582 \end{pmatrix}.$$

For the variance and standard deviation we get ($n = 10, k = 2$),

$$s^2 = 71.5, \quad s = 8.46$$

■

Appendix C

“Advanced” topics from statistics

C.1 Covariance and correlation

Definition 12 Let X and Y be two random variables having expected values μ_X and μ_Y , and let σ_X and σ_Y denote the standard deviations. The *covariance* between X and Y is denoted as $\text{Cov}(X, Y)$, or just σ_{XY} . It is defined as

$$\text{Cov}(X, Y) = \sigma_{XY} = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X\mu_Y.$$

The *correlation* between X and Y is

$$\text{Corr}(X, Y) = \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y}.$$

It follows directly from the definition that

$$\text{Cov}(X, X) = \sigma_X^2,$$

and,

$$\text{Corr}(X, X) = 1.$$

If X and Y are independent, then $E(XY) = \mu_X\mu_Y$, and therefore, $\text{Cov}(X, Y) = 0$, and $\text{Corr}(X, Y) = 0$. Thus, if two variables are independent, the covariance is zero. It must be emphasized that the opposite is *not* necessarily true. The fact that a covariance is zero does not guarantee that the variables are independent.

C.2 Random vectors

Some experiments (see Section A.1.2) produce outcomes that are vectors. Such a vector is called a *random vector*. We write $X = (X_1 X_2 \dots X_n)'$.

Each element, X_i in X , is a random variable having an expected value $E(X_i) = \mu_i$ and a variance $\text{Var}(X_i) = \sigma_i^2$. The covariance between two elements X_i and X_j is denoted as σ_{ij} . For convenience, we also use the notation $\sigma_{ii} = \sigma_i^2$.

A random vector $X = (X_1 X_2 \dots X_n)'$ has an expected value, μ , which is also a vector. It has a “variance”, Σ , which is a matrix:

$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}.$$

The matrix Σ is also called the variance-covariance matrix or just the covariance matrix. Since $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$, we conclude that Σ is symmetric, i.e. that $\sigma_{ij} = \sigma_{ji}$.

Theorem 3 *Let X and Y be independent random vectors of size n . Furthermore, let A be an $m \times n$ matrix and b be a column vector of size m . Define the random vectors Z_1 and Z_2 as $Z_1 = X + Y$ and $Z_2 = AX + b$. Denote the expectations and variance-covariance matrices as μ_X , μ_Y , μ_{Z_1} , μ_{Z_2} , Σ_X , Σ_Y , Σ_{Z_1} and Σ_{Z_2} , respectively. Then*

$$\mu_{Z_1} = \mu_X + \mu_Y, \quad \Sigma_{Z_1} = \Sigma_X + \Sigma_Y,$$

and,

$$\mu_{Z_2} = A\mu_X + b, \quad \Sigma_{Z_2} = A\Sigma_X A'.$$

■

C.3 Multivariate distributions

The distribution of a random vector is called a *multivariate distribution*. Some multivariate distributions may be expressed by a certain function over the sample space. We shall consider two such common multivariate functions: The *multinomial distribution* (discrete) and the *multivariate normal distribution* (continuous).

C.3.1 The multinomial distribution

Consider an experiment with categorical outcomes. Assume that there are k mutually exclusive and exhaustive outcomes. Examples include:

- Rolling a dice $\rightarrow 1, 2, 3, 4, 5, 6$ ($k = 6$)
- Testing for somatic cell count at cow level ($k = 4$) \rightarrow

- $c \leq 200,000$
- $200,000 < c \leq 300,000$
- $300,000 < c \leq 400,000$
- $400,000 < c$

Assume that the probability of category i is p_i with $\sum_i p_i = 1$, and that the experiment is repeated n times. Let $X = (X_1 X_2 \dots X_n)'$ be a random vector defined so that X_i is the total number of outcomes belonging to category i . The sample space of the compound n experiments is $S = \{x \in Z_+ \cup \{0\} \mid \sum_i x_i = n\}$, where $Z_+ \cup \{0\}$ is the set of non-negative integers. Under the conditions mentioned, the random vector X is said to have a multinomial distribution with parameters $p = (p_1, p_2, \dots, p_k)'$ and n .

The probability distribution for $x \in S$ is

$$f(x) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}. \quad (\text{C.1})$$

The expected value, μ , and the variance-covariance matrix, Σ are

$$\mu = np, \quad \Sigma = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \dots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \dots & np_k(1-p_k) \end{pmatrix}. \quad (\text{C.2})$$

It is easy to see that the binomial distribution is a special case of the multinomial distribution with $k = 2$.

C.3.2 The multivariate normal distribution

Definition

A random vector $X = (X_1 X_2 \dots X_n)'$ from the sample space $S = \mathbb{R}^n$ has a multivariate normal distribution if its density function is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{(x-\mu)' \Sigma^{-1} (x-\mu)}{2}}. \quad (\text{C.3})$$

The expected value is μ and the variance-covariance matrix is Σ . We also write $X \sim \mathcal{N}(\mu, \Sigma)$.

Conditional distribution of sub-vectors

Suppose that $X = (X_1 X_2 \dots X_n)'$ is $\mathcal{N}(\mu, \Sigma)$, and we partition X into two sub-vectors $X_a = (X_1 \dots X_j)'$ and $X_b = (X_{j+1} \dots X_n)'$. We partition the mean vector μ and the variance-covariance matrix Σ accordingly and write

$$X = \begin{pmatrix} X_a \\ X_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

Then, $X_a \sim \mathcal{N}(\mu_a, \Sigma_{aa})$, and $X_b \sim \mathcal{N}(\mu_b, \Sigma_{bb})$. The matrix $\Sigma_{ab} = \Sigma'_{ba}$ contains the covariances between elements of the sub-vector X_a and the sub-vector X_b .

For $X_a = x_a$, the *conditional distribution* of X_b is $\mathcal{N}(\nu, C)$, where

$$\nu = \mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a), \quad (\text{C.4})$$

and,

$$C = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}. \quad (\text{C.5})$$

Example 34 Lactations of a dairy cow

Let X_1, X_2, \dots, X_5 denote the 305 days energy corrected milk yield of the same dairy cow. It is then reasonable to assume that $X = (X_1 X_2 \dots X_5)'$ has a 5 dimensional normal distribution. Having observed e.g. X_1, X_2 and X_3 we can predict X_4 and X_5 according to Eqs. (C.4) and (C.5).

■

C.4 Hyper distributions

C.4.1 Motivation

Until now, our assessment to distributions has been under the assumption that a parameter has a fixed (but often unknown) value.

When we, for instance, observe the number, n , of sows conceiving out of N inseminated, we have tested the hypothesis that n is drawn from a binomial distribution with parameter $p = p_0$, where p_0 is some fixed value (e.g. 0.84). For analysis of production results (cf. Chapter 5), this is not really a problem.

However, even though it is interesting to analyze whether an achieved conception rate is “satisfactory”, the *real challenge* is to use the information gained for planning of the future, where *predictions* play an important role. Thus, a central question in production planning is: *If we mate 50 sows, how many can we expect to farrow in 115 days?*

The question can be answered at several levels of insight:

The naive approach returning only a value: Assume that we know the conception rate with certainty, e.g. $p = 0.84$. With $N = 50$ inseminations, the expected number of farrowings is $pN = 0.84 \times 50 = 42$ which we return as the answer.

The semi-naive approach returning a distribution: Here we take the binomial variation into account and return the binomial distribution $\mathcal{B}(50, 0.84)$ by which we can calculate the probability of any $n \leq 50$.

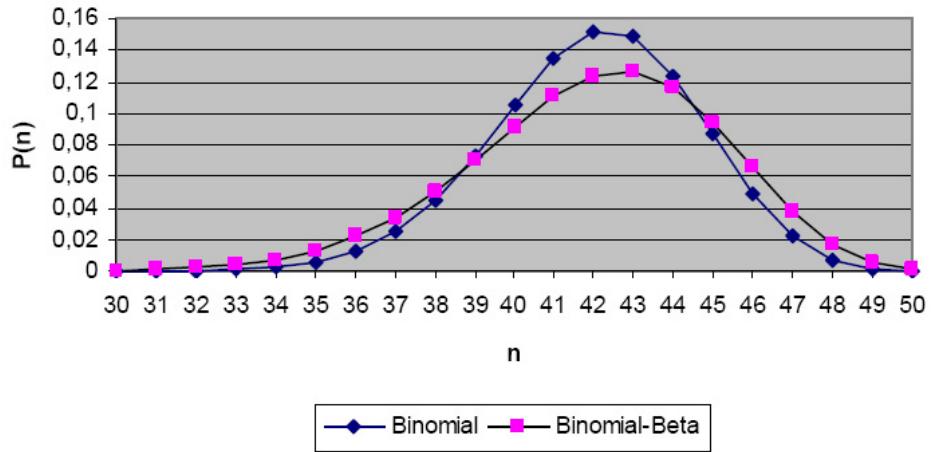


Figure C.1: Distribution of number of farrowings, n , out of 50 inseminated. The distribution under known conception rate, $p = 0.84$ is compared to the distribution with Beta distributed conception rate.

The “correct” approach returning a compound distribution: Here we also take the uncertainty about p into account. We do that by specifying a (hyper) distribution for p . The natural hyper distribution for the probability parameter of a binomial distribution is a so-called beta distribution (see Section C.4.2). The function returned as answer is a binomial distribution with Beta distributed probability.

In Figure C.1, the two distributions are compared. As it is seen from the figure, the uncertainty about the number of farrowings increase if the conception rate is not known with certainty.

In all kinds of planning, representation of uncertainty is important. In general, we use hyper distributions to represent the uncertainty about a true value of a parameter. The hyper distribution reflects our initial belief in the true value.

Hyper distributions learn over time. As new observations are collected, our uncertainty may decrease. This is reflected in the (updated) hyper distribution (the variance of it decreases).

The parameters of a hyper distribution are called *hyper parameters*.

C.4.2 Conjugate families

Definition

Most standard distributions have corresponding families of hyper distributions that are suitable for representation of our belief in parameter values. If the family is closed under sampling, it is called a *conjugate family*. To be closed under sam-

pling means that if our *prior* belief in a parameter is represented by a distribution from a certain family, then the *posterior* distribution after having taken further observations belongs to the same family.

Conjugate family of a binomial distribution

The conjugate family for the probability parameter of a binomial distribution is the family of Beta distributions. The sample space of a Beta distribution is $S =]0; 1[$. The Beta distribution has two parameters α and β . The density function is (where Γ is the gamma function):

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}. \quad (\text{C.6})$$

The expected value and variance are

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \quad (\text{C.7})$$

Theorem 4 Assume that our prior belief in the probability, p , of a binomial distribution is $\text{Beta}(\alpha, \beta)$ and we take a new observation from the binomial distribution in question. If we observe k successes out of n trials the posterior belief in p , given the new observation, is $\text{Beta}(\alpha + k, \beta + n - k)$.

When we specify a Beta prior, we may think of it this way:

- What is our initial guess about the parameter? This guess is put equal to $\alpha/(\alpha + \beta)$.
- How certain are we about our guess? As if we have observed $m = 10, 100, 1000$ trials (cases)? We then put $\alpha + \beta = m$ and from these two equations we can determine the parameters α and β .

Conjugate family of a Poisson distribution

The conjugate family for the parameter λ of a Poisson distribution is the family of Gamma distributions with density function is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}. \quad (\text{C.8})$$

The expected value and variance are

$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \quad \text{Var}(X) = \frac{\alpha}{\beta^2}. \quad (\text{C.9})$$

Theorem 5 Assume that our prior belief in the parameter, λ , of a Poisson distribution is $\text{Gamma}(\alpha, \beta)$ and we take n new observations, $X_1 = x_1, \dots, X_n = x_n$, from the Poisson distribution in question. Our posterior belief in λ given the new observations is then $\text{Gamma}(\alpha + \sum_i^n x_i, \beta + n)$.

When we specify a Gamma prior, we may think of it this way:

- What is our initial guess about the parameter? This guess is put equal to $\alpha/(\beta)$.
- How certain are we about our guess? As if we have observed $m = 10, 100, 1000$ trials (cases)? We then put $\beta = m$ and from these two equations we can determine the parameters α and β .

Other conjugate families

For parameters of some other distributions, the conjugate families are:

Mean of a normal distribution: The family of normal distributions.

Inverse variance (= precision) of a normal distribution: The family of Gamma distributions.

Exponential distribution: The family of Gamma distributions.

Uniform distribution: The family of Pareto distributions.

Multinomial distribution: The family of Dirichlet distributions.

For further information about conjugate families, reference is made to DeGroot (1970).

Appendix D

Transition matrices for Chapter 13

D.1 Probabilities for the actions “Keep” and “Replace”

Table D.1: Transition probabilities from state i to state j of the 36-state model under the action "Keep" ($j = 1, \dots, 12$).

i	g_i	l_i	y_i	Bad genetic merit											
				1st lactation			2nd lactation			3rd lactation			4th lactation		
				L	A	H	L	A	H	L	A	H	L	A	H
1	B	1	L	0	0	0	0.6	0.3	0.1	0	0	0	0	0	0
2	B	1	A	0	0	0	0.2	0.6	0.2	0	0	0	0	0	0
3	B	1	H	0	0	0	0.1	0.3	0.6	0	0	0	0	0	0
4	B	2	L	0	0	0	0	0	0	0.6	0.3	0.1	0	0	0
5	B	2	A	0	0	0	0	0	0	0.2	0.6	0.2	0	0	0
6	B	2	H	0	0	0	0	0	0	0.1	0.3	0.6	0	0	0
7	B	3	L	0	0	0	0	0	0	0	0	0	0.6	0.3	0.1
8	B	3	A	0	0	0	0	0	0	0	0	0	0.2	0.6	0.6
9	B	3	H	0	0	0	0	0	0	0	0	0	0.1	0.3	0.6
10	B	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
11	B	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
12	B	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
13	A	1	L	0	0	0	0	0	0	0	0	0	0	0	0
14	A	1	A	0	0	0	0	0	0	0	0	0	0	0	0
15	A	1	H	0	0	0	0	0	0	0	0	0	0	0	0
16	A	2	L	0	0	0	0	0	0	0	0	0	0	0	0
17	A	2	A	0	0	0	0	0	0	0	0	0	0	0	0
18	A	2	H	0	0	0	0	0	0	0	0	0	0	0	0
19	A	3	L	0	0	0	0	0	0	0	0	0	0	0	0
20	A	3	A	0	0	0	0	0	0	0	0	0	0	0	0
21	A	3	H	0	0	0	0	0	0	0	0	0	0	0	0
22	A	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
23	A	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
24	A	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
25	H	1	L	0	0	0	0	0	0	0	0	0	0	0	0
26	H	1	A	0	0	0	0	0	0	0	0	0	0	0	0
27	H	1	H	0	0	0	0	0	0	0	0	0	0	0	0
28	H	2	L	0	0	0	0	0	0	0	0	0	0	0	0
29	H	2	A	0	0	0	0	0	0	0	0	0	0	0	0
30	H	2	H	0	0	0	0	0	0	0	0	0	0	0	0
31	H	3	L	0	0	0	0	0	0	0	0	0	0	0	0
32	H	3	A	0	0	0	0	0	0	0	0	0	0	0	0
33	H	3	H	0	0	0	0	0	0	0	0	0	0	0	0
34	H	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
35	H	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
36	H	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0

Legends: i = State; g_i = genetic merit of state i indicated as B = Bad, A = Average, H = High; l_i = lactation number of state i ; y_i = milk yield of state i indicated as L = Low, A = Average, H = High.

Table D.2: Transition probabilities from state i to state j of the 36-state model under the action "Keep" ($j = 13, \dots, 24$).

i	g_i	l_i	y_i	Average genetic merit											
				1st lactation			2nd lactation			3rd lactation			4th lactation		
				L	A	H	L	A	H	L	A	H	L	A	H
1	B	1	L	0	0	0	0	0	0	0	0	0	0	0	0
2	B	1	A	0	0	0	0	0	0	0	0	0	0	0	0
3	B	1	H	0	0	0	0	0	0	0	0	0	0	0	0
4	B	2	L	0	0	0	0	0	0	0	0	0	0	0	0
5	B	2	A	0	0	0	0	0	0	0	0	0	0	0	0
6	B	2	H	0	0	0	0	0	0	0	0	0	0	0	0
7	B	3	L	0	0	0	0	0	0	0	0	0	0	0	0
8	B	3	A	0	0	0	0	0	0	0	0	0	0	0	0
9	B	3	H	0	0	0	0	0	0	0	0	0	0	0	0
10	B	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
11	B	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
12	B	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
13	A	1	L	0	0	0	0.6	0.3	0.1	0	0	0	0	0	0
14	A	1	A	0	0	0	0.2	0.6	0.2	0	0	0	0	0	0
15	A	1	H	0	0	0	0.1	0.3	0.6	0	0	0	0	0	0
16	A	2	L	0	0	0	0	0	0	0.6	0.3	0.1	0	0	0
17	A	2	A	0	0	0	0	0	0	0.2	0.6	0.2	0	0	0
18	A	2	H	0	0	0	0	0	0	0.1	0.3	0.6	0	0	0
19	A	3	L	0	0	0	0	0	0	0	0	0	0.6	0.3	0.1
20	A	3	A	0	0	0	0	0	0	0	0	0	0.2	0.6	0.6
21	A	3	H	0	0	0	0	0	0	0	0	0	0.1	0.3	0.6
22	A	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
23	A	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
24	A	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
25	H	1	L	0	0	0	0	0	0	0	0	0	0	0	0
26	H	1	A	0	0	0	0	0	0	0	0	0	0	0	0
27	H	1	H	0	0	0	0	0	0	0	0	0	0	0	0
28	H	2	L	0	0	0	0	0	0	0	0	0	0	0	0
29	H	2	A	0	0	0	0	0	0	0	0	0	0	0	0
30	H	2	H	0	0	0	0	0	0	0	0	0	0	0	0
31	H	3	L	0	0	0	0	0	0	0	0	0	0	0	0
32	H	3	A	0	0	0	0	0	0	0	0	0	0	0	0
33	H	3	H	0	0	0	0	0	0	0	0	0	0	0	0
34	H	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
35	H	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
36	H	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0

Legends: i = State; g_i = genetic merit of state i indicated as B = Bad, A = Average, H = High; l_i = lactation number of state i ; y_i = milk yield of state i indicated as L = Low, A = Average, H = High.

Table D.3: Transition probabilities from state i to state j of the 36-state model under the action "Keep" ($j = 25, \dots, 36$).

i	g_i	l_i	y_i	High genetic merit											
				1st lactation			2nd lactation			3rd lactation			4th lactation		
				L	A	H	L	A	H	L	A	H	L	A	H
1	B	1	L	0	0	0	0	0	0	0	0	0	0	0	0
2	B	1	A	0	0	0	0	0	0	0	0	0	0	0	0
3	B	1	H	0	0	0	0	0	0	0	0	0	0	0	0
4	B	2	L	0	0	0	0	0	0	0	0	0	0	0	0
5	B	2	A	0	0	0	0	0	0	0	0	0	0	0	0
6	B	2	H	0	0	0	0	0	0	0	0	0	0	0	0
7	B	3	L	0	0	0	0	0	0	0	0	0	0	0	0
8	B	3	A	0	0	0	0	0	0	0	0	0	0	0	0
9	B	3	H	0	0	0	0	0	0	0	0	0	0	0	0
10	B	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
11	B	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
12	B	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
13	A	1	L	0	0	0	0	0	0	0	0	0	0	0	0
14	A	1	A	0	0	0	0	0	0	0	0	0	0	0	0
15	A	1	H	0	0	0	0	0	0	0	0	0	0	0	0
16	A	2	L	0	0	0	0	0	0	0	0	0	0	0	0
17	A	2	A	0	0	0	0	0	0	0	0	0	0	0	0
18	A	2	H	0	0	0	0	0	0	0	0	0	0	0	0
19	A	3	L	0	0	0	0	0	0	0	0	0	0	0	0
20	A	3	A	0	0	0	0	0	0	0	0	0	0	0	0
21	A	3	H	0	0	0	0	0	0	0	0	0	0	0	0
22	A	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
23	A	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
24	A	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
25	H	1	L	0	0	0	0.6	0.3	0.1	0	0	0	0	0	0
26	H	1	A	0	0	0	0.2	0.6	0.2	0	0	0	0	0	0
27	H	1	H	0	0	0	0.1	0.3	0.6	0	0	0	0	0	0
28	H	2	L	0	0	0	0	0	0	0.6	0.3	0.1	0	0	0
29	H	2	A	0	0	0	0	0	0	0.2	0.6	0.2	0	0	0
39	H	2	H	0	0	0	0	0	0	0.1	0.3	0.6	0	0	0
31	H	3	L	0	0	0	0	0	0	0	0	0	0.6	0.3	0.1
32	H	3	A	0	0	0	0	0	0	0	0	0	0.2	0.6	0.6
33	H	3	H	0	0	0	0	0	0	0	0	0	0.1	0.3	0.6
34	H	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
35	H	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
36	H	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0

Legends: i = State; g_i = genetic merit of state i indicated as B = Bad, A = Average, H = High; l_i = lactation number of state i ; y_i = milk yield of state i indicated as L = Low, A = Average, H = High.

Table D.4: Transition probabilities from state i to state j of the 36-state model under the action "Replace" ($j = 1, \dots, 12$).

i	g_i	l_i	y_i	Bad genetic merit											
				1st lactation			2nd lactation			3rd lactation			4th lactation		
				L	A	H	L	A	H	L	A	H	L	A	H
1	B	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
2	B	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
3	B	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
4	B	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
5	B	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
6	B	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
7	B	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
8	B	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
9	B	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
10	B	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
11	B	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
12	B	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
13	A	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
14	A	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
15	A	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
16	A	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
17	A	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
18	A	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
19	A	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
20	A	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
21	A	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
22	A	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
23	A	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
24	A	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
25	H	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
26	H	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
27	H	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
28	H	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
29	H	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
30	H	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
31	H	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
32	H	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
33	H	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
34	H	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
35	H	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
36	H	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0

Legends: i = State; g_i = genetic merit of state i indicated as B = Bad, A = Average, H = High; l_i = lactation number of state i ; y_i = milk yield of state i indicated as L = Low, A = Average, H = High.

Table D.5: Transition probabilities from state i to state j of the 36-state model under the action "Replace" ($j = 13, \dots, 24$).

i	g_i	l_i	y_i	Average genetic merit											
				1st lactation			2nd lactation			3rd lactation			4th lactation		
				L	A	H	L	A	H	L	A	H	L	A	H
1	B	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
2	B	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
3	B	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
4	B	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
5	B	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
6	B	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
7	B	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
8	B	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
9	B	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
10	B	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
11	B	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
12	B	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
13	A	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
14	A	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
15	A	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
16	A	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
17	A	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
18	A	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
19	A	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
20	A	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
21	A	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
22	A	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
23	A	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
24	A	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
25	H	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
26	H	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
27	H	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
28	H	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
29	H	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
30	H	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
31	H	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
32	H	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
33	H	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
34	H	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
35	H	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
36	H	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0

Legends: i = State; g_i = genetic merit of state i indicated as B = Bad, A = Average, H = High; l_i = lactation number of state i ; y_i = milk yield of state i indicated as L = Low, A = Average, H = High.

Table D.6: Transition probabilities from state i to state j of the 36-state model under the action "Replace" ($j = 25, \dots, 36$).

i	g_i	l_i	y_i	High genetic merit											
				1st lactation			2nd lactation			3rd lactation			4th lactation		
				L	A	H	L	A	H	L	A	H	L	A	H
1	B	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
2	B	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
3	B	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
4	B	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
5	B	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
6	B	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
7	B	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
8	B	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
9	B	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
10	B	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
11	B	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
12	B	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
13	A	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
14	A	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
15	A	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
16	A	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
17	A	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
18	A	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
19	A	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
20	A	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
21	A	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
22	A	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
23	A	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
24	A	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
25	H	1	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
26	H	1	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
27	H	1	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
28	H	2	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
29	H	2	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
30	H	2	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
31	H	3	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
32	H	3	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
33	H	3	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
34	H	4	L	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
35	H	4	A	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0
36	H	4	H	1/9	1/9	1/9	0	0	0	0	0	0	0	0	0

Legends: i = State; g_i = genetic merit of state i indicated as B = Bad, A = Average, A = High; l_i = lactation number of state i ; y_i = milk yield of state i indicated as L = Low, A = Average, H = High.

Appendix E

Solutions to exercises

E.1 Solutions to exercises of Chapter 4

E.1.1 Gross margin

Revenues:

Milk, 680,000 kg × 2.75	1,870,000
Sold cows	320,000
Sold calves, 110 × 1,250	137,500
<i>Total revenues:</i>	2,327,500

Variable costs:

Penalty, 30,000 kg × 2.70	81,000
Heifers (bought/transferred), 46 × 8,000	368,000
Concentrates, 213,000 SFU × 1.95	415,545
Roughage, 340,000 FE a 1.25	425,000
Various costs	113,400
<i>Total costs:</i>	1,402,945

Gross margin (housing, labor, management) 924,555

Gross margin per cow per year, 9,246

Gross margin per kg milk 1,35

Gross margin per kg quota 1,42

In the present herd, net returns per kg quota is most relevant, since the primary limitation is the size of the quota.

E.1.2 Internal prices

Gross margin, spring barley:

Revenues:

Grain, 64 hkg × 124 DKK	7,936
Straw, 40 hkg × 20 DKK	800
<i>Total:</i>	8.736

Variable costs:

Sowing	427
Fertilizer	889
Chemicals	260
Drying	358
Fuel and machinery	4,060
<i>Total:</i>	5.994

Gross margin per hectare 2,742

Internal price, beets

Variable costs:

Sowing	576
Fertilizer	2,291
Chemicals	1,270
Fuel and machinery	6,520
<i>Total:</i>	10,700

Lost gross margin from barley 2,742

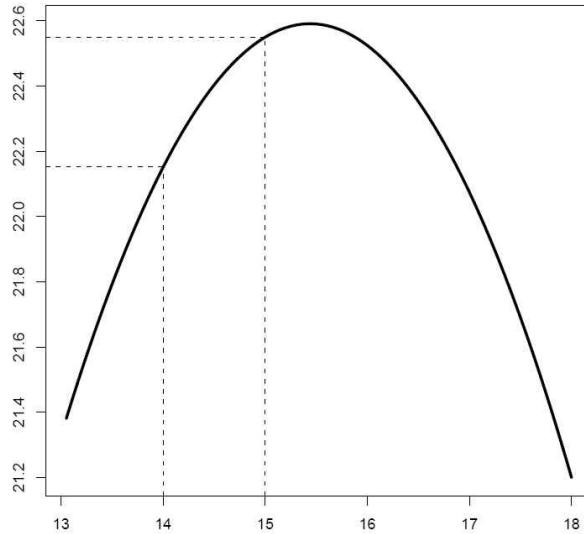
Internal price per hectare 13,442

Internal price per SFU 1.08 DKK

In other words, if industrial by-products are available at a price of 1.00 DKK per SFU, the farmer should *not* grow beets (problem c).

E.1.3 Marginal considerations, I

In the figure below, the marginal increase in milk yield (problem a) is shown as the distance between the two horizontal dashed lines:



The exact marginal value for a given feed intake is calculated from the first derivative of Y with respect to X :

$$\frac{dY}{dX} = 6.546 - 0.424X$$

By insertion we have for $X=14$, $\frac{dY}{dX} = 0.61$.

For the prices given, the marginal revenue for an additional energy unit becomes:

$$\frac{dU}{dX} = \frac{dY}{dX}p_Y = (6.546 - 0.424X)p_Y,$$

where p_Y is the price of milk. The marginal costs is:

$$\frac{dO}{dX} = p_X$$

where p_X is the feed price.

Optimal energy intake is obtained where:

$$\frac{dU}{dX} = \frac{dO}{dX} \Leftrightarrow (6.546 - 0.424X)p_Y = p_X \Leftrightarrow (6.546 - 0.424X) = \frac{p_X}{p_Y}$$

In other words, it is the price ration between feed and milk that determines the optimal feeding level. For the present prices, $X = 14$ is optimal.

E.1.4 Marginal considerations, II

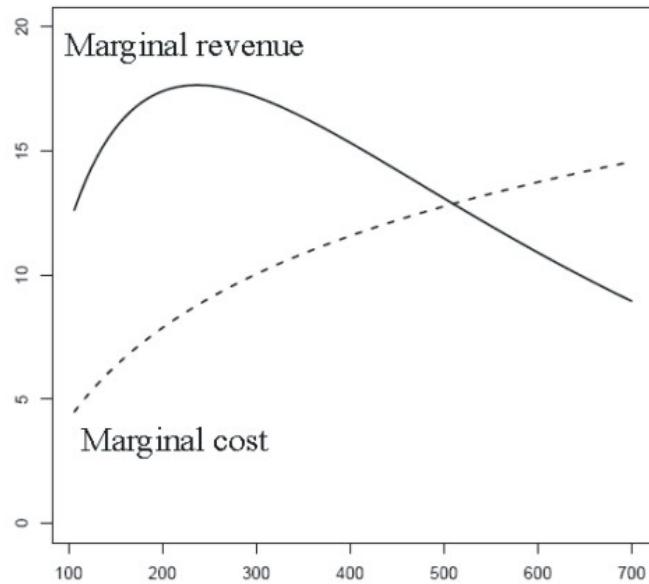
Daily feed intake is:

$$X = 3.56 \ln(V) - 13.61$$

Daily gain is:

$$G = \frac{X}{2.17 \exp(0.00256V)}$$

Marginal revenue is found as daily feed intake multiplied by the slaughter price. Marginal cost is found as daily feed intake multiplied by the feed price. The resulting marginal curves are shown in the figure below.



Optimal slaughter weight is determined as the intersection of the two curves (510 kg).

E.2 Exercises of Chapter 5

E.2.1 Milk yield

- a. Probably one of the “cleanest” expressions for herd milk production capacity. In comparison, 305 days milk yield and lactation yield are influenced by the insemination policy of the herd. As concerns the lactation yield, it is not even standardized.
- b. The precision depends heavily on herd size. It is low in small herds because of the sample uncertainty.
- c. Since revenues from milk sale is the most important source of income in a dairy herd, the significance is huge from an economical point of view.
- d. Standard deviation per day becomes $490/7/24 = 2.92$ kg for first lactation cows and $630/7/24 = 3.75$ kg for second and higher parities. Since the calculated results are average values of 11 and 15 observations, respectively, the standard deviations of the average become $2.92/\sqrt{11} = 0.89$ kg and $3.75/\sqrt{15} = 0.97$ kg. As a rule of thumb for normally distributed data, a deviation is significant if it exceeds the standard deviation multiplied by 2, so in this case, acceptable results would be $26.3 - 2 \times 0.89 = 24.5$ kg and $29.6 - 2 \times 0.97 = 27.7$ kg.

Since the calculated result concerning first lactation cows is clearly below the deduced acceptable value, we conclude that the deviation is significant from a statistical point of view. In contrast, the result concerning other cows is higher than the acceptable value, and accordingly we have to accept the result as a possible consequence of usual random variation. On the other hand, the significance from an economical point of view could call for use of a higher significance level than the usual 0.05.

E.2.2 Daily gain in slaughter pigs

- a. It is necessary to record

- The weight of all pigs inserted
- The weight of all pigs delivered to the slaughter house
- The valuation weight of all pigs at the beginning of the period
- Date of first day of the period
- The valuation weight of all pigs at the end of the period
- Date of last day of the period

It is *not* necessary to observe any weights at animal level.

- b. The following registrations must be recorded for each animal i

- Date of insertion
- Weight at insertion
- Date of delivery
- Weight at delivery

All records must be kept at animal level, so the method implies that the pigs are uniquely identified.

c. It is necessary to record

- The total weight of all pigs inserted
- The date of insertion
- The number of pigs inserted at a given day
- The weight of all pigs delivered to the slaughter house
- The date of delivery
- The number of pigs delivered at a given day

It is *not* necessary to observe any weights at animal level.

d. Using method a, any precision can be achieved, but if a high precision is desired, the labor demands for obtaining the valuation weights are huge. Nevertheless, it is the standard method used in Denmark. Method b is not realistic unless all pigs are uniquely identified (it is used with calves). The labor requirements of method c are small and the precision is high. It is used in the alternative Danish system called “Dynamisk e-kontrol”.

e. The total number of days in feed are $125 \times 10 \times 90 = 112500$. Thus, daily gain has been 796 g.

f. We have for the 4 methods, where σ_c is the standard deviation of the calculated value:

1. In this case $\sigma_o = \sqrt{2(\sigma_v^2 + \sigma_m^2)} = \sqrt{2(0^2 + 0^2)} = 0$. Then, $\sigma_c = \sqrt{\sigma_s^2 + \sigma_o^2} = \sqrt{\sigma_s^2 + 0^2} = \sigma_s = 3.1$.
2. Using the equation provided, we have $\sigma_v = s = \sqrt{RN_i^2 \frac{\sigma_i^2}{n_i} (1 - \frac{n_i}{N_i})} = \sqrt{125 \times 10^2 \frac{5^2}{1} (1 - \frac{1}{10})} = 530$ kg. We further obtain for the total gain $\sigma'_o = \sqrt{2(\sigma_v^2 + \sigma_m^2)} = \sqrt{2(530^2 + 0^2)} = 750$ kg. Re-calculated to standard deviation on daily level we have $\sigma_o = \sigma'_o / 112500 = 750 / 112500 = 0.0067$ kg = 6.7 g. Then, $\sigma_c = \sqrt{\sigma_s^2 + \sigma_o^2} = \sqrt{3.1^2 + 6.7^2} = 7.4$.

3. Using the equation provided, we have $\sigma_v = s = \sqrt{RN_i^2 \frac{\sigma_i^2}{n_i} (1 - \frac{n_i}{N_i})} = \sqrt{1 \times 1250^2 \frac{15^2}{125} (1 - \frac{125}{1250})} = 1591 \text{ kg}$. We further obtain for the total gain $\sigma'_o = \sqrt{2(\sigma_v^2 + \sigma_m^2)} = \sqrt{2(1591^2 + 0^2)} = 2250 \text{ kg}$. Re-calculated to standard deviation on daily level we have $\sigma_o = \sigma'_o / 112500 = 2250 / 112500 = 0.0200 \text{ kg} = 20.0 \text{ g}$. Then, $\sigma_c = \sqrt{\sigma_s^2 + \sigma_o^2} = \sqrt{3.1^2 + 20.0^2} = 20.2$.
4. We have, $\sigma_m \approx 76000 \times 0.03 = 2280 \text{ kg}$. Then, we further obtain for the total gain $\sigma'_o = \sqrt{2(\sigma_v^2 + \sigma_m^2)} = \sqrt{2(0^2 + 2280^2)} = 3224 \text{ kg}$. Re-calculated to standard deviation on daily level we have $\sigma_o = \sigma'_o / 112500 = 3224 / 112500 = 0.0287 \text{ kg} = 28.7 \text{ g}$. Then, $\sigma_c = \sqrt{\sigma_s^2 + \sigma_o^2} = \sqrt{3.1^2 + 28.7^2} = 28.9$.
- g. Using the rule of thumb that deviations below 2 times the standard deviation provides us with the following acceptable results under the 4 methods:
1. $820 - 2 \times 3.1 = 814 \text{ g}$.
 2. $820 - 2 \times 7.4 = 805 \text{ g}$.
 3. $820 - 2 \times 20.2 = 780 \text{ g}$.
 4. $820 - 2 \times 28.9 = 762 \text{ g}$.

Since the calculated result was 796, the farmer should only interfere if method 1 or method 2 has been applied.

- h. The worst consequence is the disturbance in the production system, because the batch is not ready for delivery at the scheduled time. The economic consequences cannot be calculated unless it has been decided how to adjust.

E.2.3 Reproduction in dairy herds

We use the hypothesis $p = 0.5$ and calculate the binomial test variable t_n as

$$t_n = P(n \leq n_o \mid p = 0.5)$$

where n_o is the number of observed pregnancies out of N inseminations. In each of the three cases we have

- a. $t_n = P(n \leq 0.4 \times 10 \mid p = 0.5, N = 10) = P(n \leq 4 \mid p = 0.5, N = 10) = 0.38$.
- b. $t_n = P(n \leq 0.4 \times 20 \mid p = 0.5, N = 20) = P(n \leq 8 \mid p = 0.5, N = 20) = 0.25$
- c. $t_n = P(n \leq 0.4 \times 50 \mid p = 0.5, N = 50) = P(n \leq 20 \mid p = 0.5, N = 50) = 0.10$

- d. The calculated key figures are very imprecise under typical Danish herd sizes (unless a very long period is behind).
- e. The probability of 5 unsuccessful inseminations is $0.5^5 = 0,03125$. Our test statistic under 20 observations become $t_n = P(n \geq 5 | p = 0.03125, N = 20) = 1 - P(n \leq 4 | p = 0.03125, N = 20) = 1 - 0.9999 = 0.0001$. It is *very unlikely* that the observation is caused by random biological variation. The farmer should investigate further.
- f. Our test statistic under 100 observations become $t_n = P(n \geq 5 | p = 0.03125, N = 100) = 1 - P(n \leq 4 | p = 0.03125, N = 100) = 1 - 0.796 = 0.204$. The observation may *very well* be the result of random fluctuations.
- g. Our test statistic under 200 observations become $t_n = P(n \geq 5 | p = 0.03125, N = 200) = 1 - P(n \leq 4 | p = 0.03125, N = 200) = 1 - 0.248 = 0.752$. The observation may *very well* be the result of random fluctuations.

E.2.4 Disease treatments

- a. Sole ulcers are very painful to cows, they cause economic losses and the number seems to be far too high. The summer occurrence indicates a problem with the grazing area.
- b. The number of cases is poisson distributed under the hypothesis that they occur at random. Our test statistic therefore becomes $t_n = P(n \geq n_o | \lambda = 2)$ where n_o is the observed number of cases. Since, for $n_o = 5$, $t_n = 0.05$, we conclude that observation of more than 5 cases is significant.
- c. They are very fluctuating due to random variation.
- d. Treatments are not a good measure of health in a herd. Farmers have different policy concerning call for veterinary assistance.

E.2.5 Validity, reproduction

We have approximately that the number of calvings per cow per year (*CPCPY*) may be calculated as

$$CPCPY = \frac{365}{IT \times RR + CI \times (1 - RR)}$$

where

- IT = Average time of insertion of heifers expressed as days after calving of the cows they replace.
- CI = Average calving interval in days for the cows who are *not* replaced.

- RR = Annual replacement rate.

In this particular case we have

$$1.10 = \frac{365}{IT \times 0.45 + 360 \times 0.55} \Leftrightarrow IT = 297$$

If the farmer wants to increase the number of calvings per cow per year, *CPCPY*, to 1.30, we get

$$1.30 = \frac{365}{297 \times 0.45 + CI \times 0.55} \Leftrightarrow CI = 267$$

Since the pregnancy period is 280 days for cows, this is obviously not possible. The low number of calvings in the herd is due to the fact that culled cows are kept on average to 297 days after calving (which has nothing to do with reproduction). The conclusion is that the validity of *CPCPY* is low as a measure of reproduction efficiency.

E.2.6 Economic significance of days open

In both cases the number of calves per cow per year is increased by $(12/13 - 12/14) \times 0.60 = 0.0396$ calves.

- When there is no milk quota the marginal value is simply

$$\frac{dD}{dY} = b = 2.10$$

The cost of 30 additional days open is

Milk, 255 kg \times 2.10	=	535.50
Calves, 0.0396 \times 1200	=	47.52
Total cost, 30 days	=	583.02

The cost of one additional day open is therefore $583.02/30 = 19.43$ DKK.

- Under a quota, the *total* gross margin T , is the quota multiplied by the average gross margin per kg milk produced:

$$T = 600000 \frac{D}{Y} = 600000 \frac{c + bY}{Y} = 600000 \left(\frac{c}{Y} + b \right)$$

The marginal value of a 1 kg increased milk yield *per cow per year* therefore becomes

$$\frac{dT}{dY} = -600000 \frac{c}{Y^2} = -600000 \frac{-4000}{Y^2} = \frac{24 \times 10^8}{Y^2}$$

Since the present milk yield is $Y = 6500$, the total marginal value becomes

$$\frac{dT}{dY} = \frac{24 \times 10^8}{6500^2} = 56.80$$

The number of cows needed to produce 600000 kg milk is $600000/6500 = 92.3$. Thus, the marginal value of 1 kg milk is $56.80/92.3 = 0.62$ kr.

The cost of 30 additional days open is

Milk, 255 kg \times 0.62	=	158.10
Calves, 0.0396 \times 1200	=	47.52
Total cost, 30 days	=	205.62

The cost of one additional day open is therefore $205.62/30 = 6.85$ DKK.