Friedrich-Alexander-Universität
Erlangen-Nürnberg

Decision Theory

Lecture 9

Michael Hartisch

Friedrich-Alexander Universität Erlangen-Nürnberg, Department Data Science June 17, 2024

Decision Theory

Recap: What did we do?

- decision under uncertainty
- ordered weighted averaging (OWA)
- weighted ordered weighted averaging (WOWA)

Decision Theory

OWA

- Generalisation of Minimax/Maximin, Average, Hurwicz
- Give importance to best, second-best, third-best, ..., worst
- Maximin: all importance on worst
- Average: equal importance to all
- Hurwicz: importance to best and worst

Decision Theory

OWA

	s_1	$\dots$	s_n
a_1	e_{11}	$\dots$	e_{1n}
a_2	e_{21}	$\dots$	e_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_m	e_{m1}	$\dots$	e_{mn}

- Given weight vector $w = (w_1, \dots, w_n)$
- $w_i \in [0, 1]$ and $\sum_{i \in [n]} w_i = 1$
- For fixed alternative a_i, consider outcomes $(e_{i1}, \dots, e_{in})$
- Sort them from largest to smallest: $(e_{i\pi(1)}, \dots, e_{i\pi(n)})$
- OWA is $\sum_{j \in [n]} w_j e_{i\pi(j)}$

Decision Theory

Example

- Outcomes of alternative a are $(5, 2, 6, 3)$
- Weight vector is $(1/2,\ 0,\ 1/4,\ 1/4)$
- Sort outcomes from largest to smallest: $(6, 5, 3, 2)$
- Apply weights:
$$OWA(a) = 1/2 \cdot 6 + 0 \cdot 5 + 1/4 \cdot 3 + 1/4 \cdot 2 = 4.25$$
- To calculate OWA, it does not matter whether we minimise or maximise
- But this will be important to decide which of the alternatives we choose

Decision Theory

OWA: Example

Maximize, $w = (0, 0, 1/2, 1/2)$

	s_1	s_2	s_3	s_4
a_1	4	6	1	2
a_2	6	2	2	4
a_3	8	2	5	0
a_4	0	9	3	9

- $OWA(a_1) = 0 \cdot 6 + 0 \cdot 4 + 1/2 \cdot 2 + 1/2 \cdot 1 = 1.5$
- $OWA(a_2) = 0 \cdot 6 + 0 \cdot 4 + 1/2 \cdot 2 + 1/2 \cdot 2 = 2.0$
- $OWA(a_3) = 0 \cdot 8 + 0 \cdot 5 + 1/2 \cdot 2 + 1/2 \cdot 0 = 1$
- $OWA(a_4) = 0 \cdot 9 + 0 \cdot 9 + 1/2 \cdot 3 + 1/2 \cdot 0 = 1.5$
- We choose alternative a_2

OWA

- Can be considered as mix between ”and” (andness) and ”or” (orness)
- Worst-case: be good in the first scenario and in the second scenario and ...
- Best-case: be good in the first scenario or in the second scenario or ...
- Degree of orness depends on weight vector w:

$$orness(w) = \frac{1}{n-1} \sum_{j \in [n]} w_j(n-j)$$

- Orness of worst-case is 0, orness of best-case is 1
- Orness of average and median is $1/2$

OWA

- Recall the axioms table:

	Minimax	Hurwicz	Regret	Average
Complete ranking	✓	✓	✓	✓
Labelling	✓	✓	✓	✓
Value scale	✓	✓	✓	✓
Strong domination	✓	✓	✓	✓
Irrelevant alternatives	✓	✓	✗	✓
Constant to column	✗	✗	✓	✓
Row permutation	✓	✓	✗	✓
Column duplication	✓	✓	✓	✗

- Whenever something does not hold for Minimax, Hurwicz, Average, it also does not hold for OWA
- We inherit all counter-examples

Decision Theory

OWA: Combinatorial

- Can apply OWA also to combinatorial problems for minimisation
- Heuristic for K scenarios:
  - For each item, sort the cost vector
  - Multiply with weights and add up
  - I.e., calculate OWA for a single item
  - Get item costs $\hat{c}$
  - Solve with respect to $\hat{c}$
- If item weights are non-increasing (conservative), gives $w_1 K$ approximation
- Special case: $w_1 = 1$, take worst-case of each item, gives K approximation

Decision Theory

WOWA

- OWA cannot assign weights to specific scenarios
- Extension: weighted ordered weighted averaging (WOWA)
- Two sets of weights:
  - w_j for each position in the ordered vector (as before)
  - p_j for each scenario s_j
- Can do the same as OWA and more
- High effort to compute

Decision Theory

Today

- Decision under Risk
  - Basic Concepts
  - μ-Principle
  - μ-σ-Principle

Decision Theory

Situation

$$\begin{array}{c|ccc} & s_1 & \dots & s_n \\ \hline a_1 & e_{11} & \dots & e_{1n} \\ a_2 & e_{21} & \dots & e_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_m & e_{m1} & \dots & e_{mn} \end{array}$$

- For each state s_j, there is a probability p_j
- Also possible: continuous states with continuous probability distribution
- Also possible: combinatorial problem with exponentially many alternatives

Decision Theory

Reminder: Probability

Let $\Omega = \{e_1, \ldots, e_n\}$ be the sample space of a random experiment. A function P that assigns a number $P(A)$ to each event $A \subseteq \Omega$ of a random experiment, such that

1. $P(A) \geq 0$
2. $P(\Omega) = 1$ and
3. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

is called a probability distribution. The number $P(A)$ is called the probability of event A.

Decision Theory

New Term: Certainty Equivalent

- Already familiar with decision based on expected value as a criterion under risk
- Will examine it more closely
- One relevant term here: Certainty Equivalent

Decision Theory

Certainty Equivalent

- Consider alternative a_i
- Which certain amount $S(a_i)$ is as good as the uncertain outcome of a_i?

Example

- You have a lottery ticket:
  - 90% chance of winning 0 Euros
  - 10% chance of winning 1000 Euros
- I offer you instead
  - 1 Euro?
  - ...or 900 Euros?
  - When do you sell?

Certainty Equivalent, Example

- What is it for
  - 75 Euros
  - 90 Euros
  - 100 Euros
  - 125 Euros
  - 150 Euros
- Certainty Equivalent = the amount at which the decision maker would sell
- Subjective

Decision Theory

Certainty Equivalent

- Assume we know the preference function Φ
- It holds (with abuse of notation)
$$\Phi(S(a_i)) = \Phi(a_i)$$
- a_i has the same value as $S(a_i)$

Decision Theory

First Approach

- Natural choice: choose with respect to expected value
- Has shown good properties in decision under uncertainty:
  - Satisfies many axioms
  - Approximation for Minimax
  - Approximation for Regret
- Expected value of a distribution often denoted as μ
- Therefore: μ-Principle

Decision Theory

μ-Principle

- Let $e_{i1}, \ldots, e_{in}$ be outcomes of alternative a_i
- μ-principle formally:

$$\Phi(a_i) = \sum_{j \in [n]} p_j e_{ij} = \mu_i$$

- What about certainty equivalent?

$$\Phi(\mu_i) = \sum_{j \in [n]} p_j \mu_i = \mu_i$$

Hence, certainty equivalent is the expected value.

Decision Theory

Example

- You play a lottery:
  - 90% chance of winning 0 Euros
  - 10% chance of winning 1000 Euros
- Would sell ticket for 100 Euros
- Also: would pay 100 Euros to play

Decision Theory

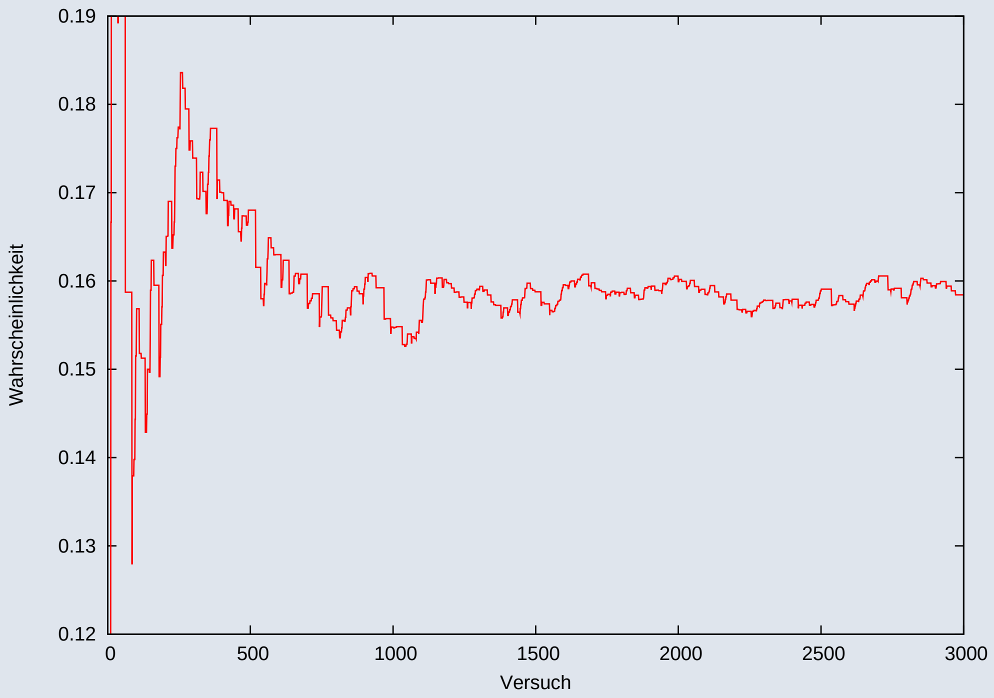
Critique

- Is the μ-Principle reasonable?
- Possible decision situations:
  - Decision about a frequently repeating action (repetition case)
  - Decision about a unique, non-repeating action (single case)

Decision Theory

Repetition Case

- Law of Large Numbers / Central Limit Theorem:
  - Relative frequency of an experiment approaches its expected value



Decision Theory

Example

- We flip a coin, outcome is heads (s_1) or tails (s_2)

	s_1	s_2
a_1	0	10
a_2	5	5

- Both alternatives are equally good in terms of expected value
- Variance can be very different

Decision Theory

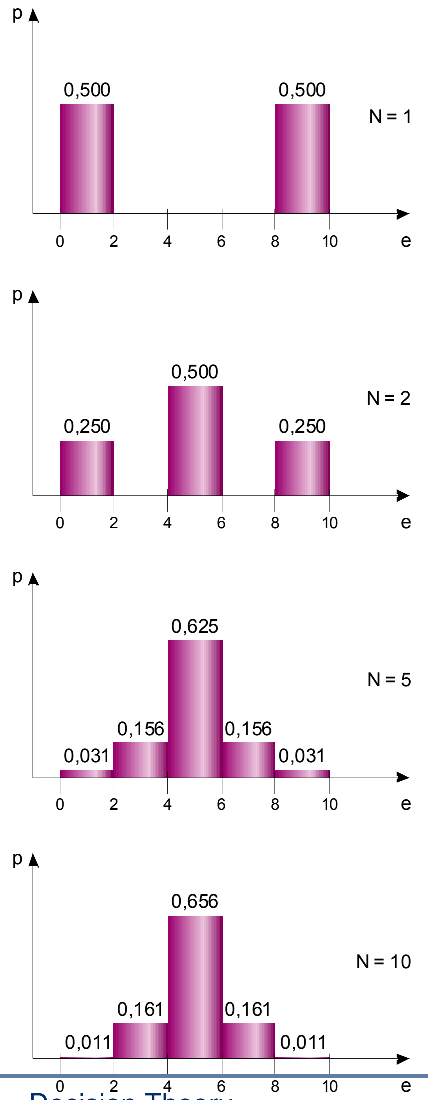
Example

	s_1	s_2
a_1	0	10
a_2	5	5

- Play N rounds in a row
- Expected value per game remains 5
- a_1's winnings range from 0 to $10N$
- a_2's winnings are always $5N$
- Variance decreases with more N
- It generally holds for $Var(X_i) = Var(X_0)$

$$Var(\frac{1}{N}\sum_{i\in[N]} X_i) = \frac{1}{N^2}\sum_{i\in[N]} Var(X_i) = \frac{Var(X_0)}{N}$$

Decision Theory



Example

- Probabilities for $N = 1, 2, 5, 10$
- On the horizontal axis: Probability for average gain intervals $[0, 2], [2, 4], \ldots, [8, 10]$
- Probability that the result falls near the expected value increases
- 100 repetitions: 85% chance in the interval $[4, 6]$
- 250 repetitions: 99.9% chance in the interval $[4, 6]$

Repetition Case

- With more repetitions, the expected value becomes a more reliable estimate of the outcome
- For alternatives a_1 and a_2 with $\mu_1 > \mu_2$, a_1 becomes significantly better (when maximizing)
- With fewer repetitions, probability distributions overlap more

Decision Theory

Example

	s_1	s_2
a_1	0	10
a_2	4	5

- $\mu_1 = 5 > 4.5 = \mu_2$
- In a single attempt: 50% probability that a_1 is worse than a_2
- With enough repetitions: Probability that a_1 is worse becomes arbitrarily small

Decision Theory

Example

N	$p(e_1 > 4.55)$	$p(e_2 < 4.55)$	$p(e_1 \geq e_2)$
100	0.816	0.841	0.840
250	0.923	0.943	0.942
500	0.978	0.987	0.942
1000	0.998	0.999	0.999

- $\mu_1 > 4.55 > \mu_2$

Decision Theory

Conclusion

- μ-Principle is reasonable when the experiment is repeated frequently
- Examples:
  - Quality assessment of workpieces
  - Tolerance estimates for cutting machines
  - Planning repairs for 1000 machines
  - Tires in the fleet
  - Light bulbs on the company premises
  - ...

Decision Theory

Single Case Decisions

- We have already seen: Expected value is not necessarily the best choice
- Two groups of arguments against it:
  - St. Petersburg paradox
  - Games of chance and insurance

Decision Theory

St. Petersburg Paradox

- We toss a coin (head or tail) until the first head appears
  - On the first toss: Player receives 2 Euros
  - On the second toss: Player receives 4 Euros
  - ...
  - On the n-th toss: Player receives 2^n Euros
- Infinite possible outcomes

Decision Theory

St. Petersburg Paradox

- Let s_i: Head appears on the i-th toss

	s_1	s_2	s_3	$\ldots$	s_n
$p(s_i)$	$1/2$	$1/4$	$1/8$	$\ldots$	2^{-n}
Gain	2	4	8	$\ldots$	2^n

- How much would you pay to play once?
- How much would you demand to play the bank once?
- Expected value:

$$\mu = 2 \cdot \frac{1}{2} + 4\frac{1}{4} + 8 \cdot \frac{1}{8} + \ldots = \infty$$

Decision Theory

St. Petersburg Paradox

- Expected value is infinitely large
- Certainty equivalent for μ-Principle is also infinite
- You would give everything you have, and more, to participate
- For no amount of money would you be willing to play the bank
- But how much would you really offer?
- μ-Criterion not satisfactory here

Decision Theory

Games of Chance

- A company offers a game of chance
- For example:
  - Lottery
  - Roulette
  - Class lottery
- It sells many tickets
- The company wants to make money with it
- Goal of the company: Expected gain of a ticket is less than the price
- According to μ-principle: never participate!

Decision Theory

Insurance

- A company offers insurances
- Many people participate
- The company wants to make money with it
- Goal of the company: Expected payout of an insurance is less than the price
- According to μ-principle: never insure!

Example

- For an individual, the following health care costs per year (in Euros) may occur:

p	50%	19%	0.5%	0.39%	0.1%	0.01%
Costs	100	500	1000	5000	10000	50000

- Expected value: 184.50 Euro
- According to the μ-criterion: only insure if monthly premium is cheaper
- Would we act that way?
- Outliers could ruin us

Decision Theory

Discussion

- Apparently, we do not always act according to the μ-criterion
- Can be interpreted differently:
  - We do not act rationally
  - μ-criterion is not always sensible
- How is it with
  - Lottery
  - Liability insurance

Decision Theory

Caution

- So far, we have only considered absolute gains/losses
- Caution with relative (percentage) values
- Consider the following game:
  - Start with 1 Euro
  - With probability $1/2$: increase by 50%
  - With probability $1/2$: decrease by 40%
- Is it worthwhile in terms of expected value?

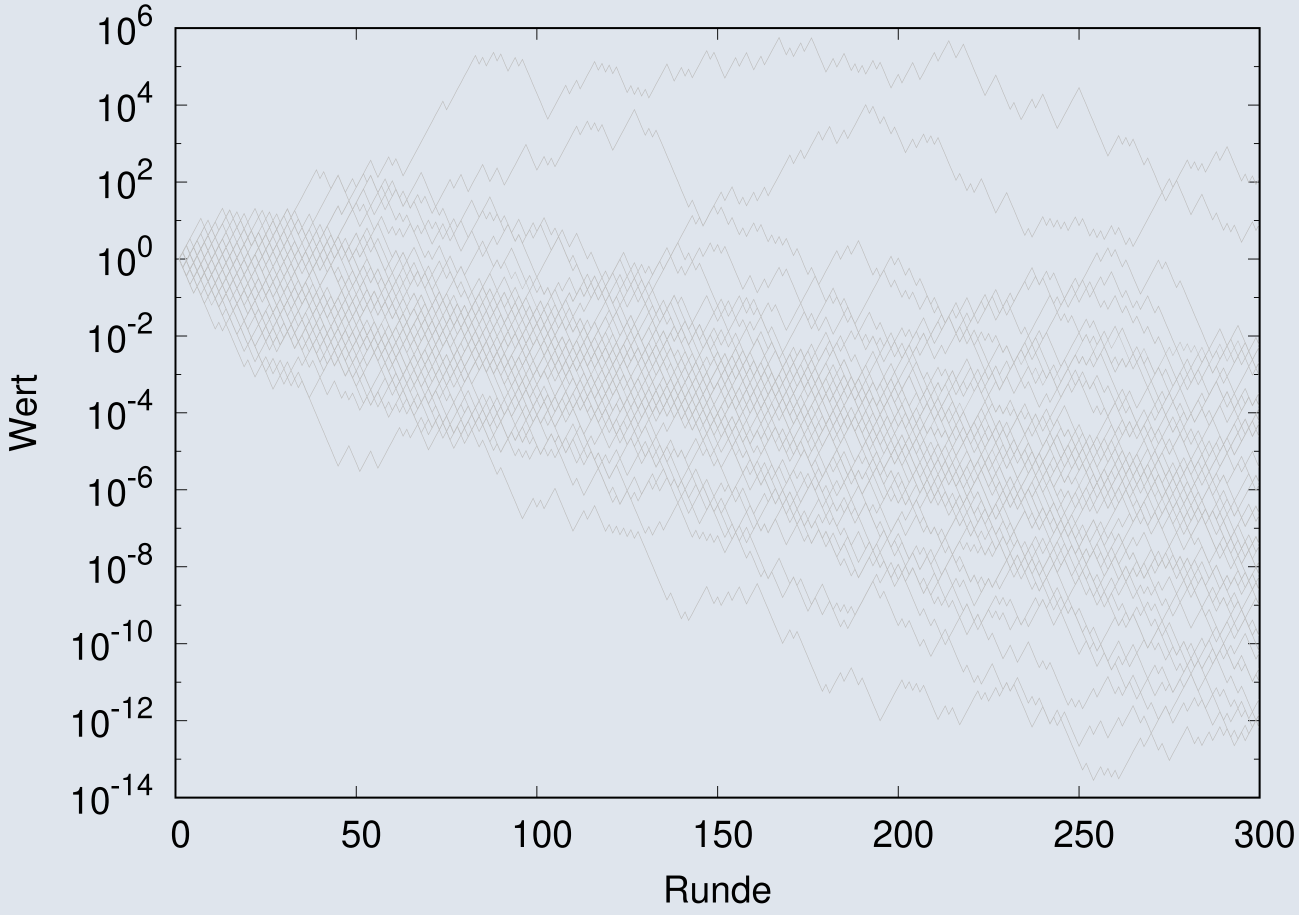
Decision Theory

Caution

- Game:
  - With probability $1/2$: increase by 50%
  - With probability $1/2$: decrease by 40%
- Expected value: $1/2 \cdot 1.5 + 1/2 \cdot 0.6 = 1.05$
- On average, we should make a profit
- Let's test it...

Decision Theory

Simulation



Decision Theory

Why?

- If we win once and lose once:

$$1 \cdot 1.5 \cdot 0.6 = 0.9$$

- Value is not 1.05
- Arithmetic mean: $(x_1 + x_2 + \ldots + x_n)/n$
- Geometric mean: $\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$
- We need to use geometric mean here

Decision Theory

Alternative Criterion

- μ-Principle: consider only expected value
- μ-σ-Principle: also consider standard deviation/variance (”sigma”)
- Reminder:
  - σ is standard deviation

$$\sigma = \sqrt{\sum_{j=1}^{n} p_j (x_j - \mu)^2}$$

  - σ^2 is variance

$$\sigma^2 = \sum_{j=1}^{n} p_j (x_j - \mu)^2$$

Decision Theory

μ-σ-Principle

- We decide depending on μ and σ
- Can also explain lottery/insurance
- Let's assume we maximize $\Phi(\mu, \sigma)$
- The larger μ, the larger Φ should be
- What happens with σ?
  - Φ increases as σ grows: risk-seeking
  - Φ decreases as σ grows: risk-averse

Decision Theory

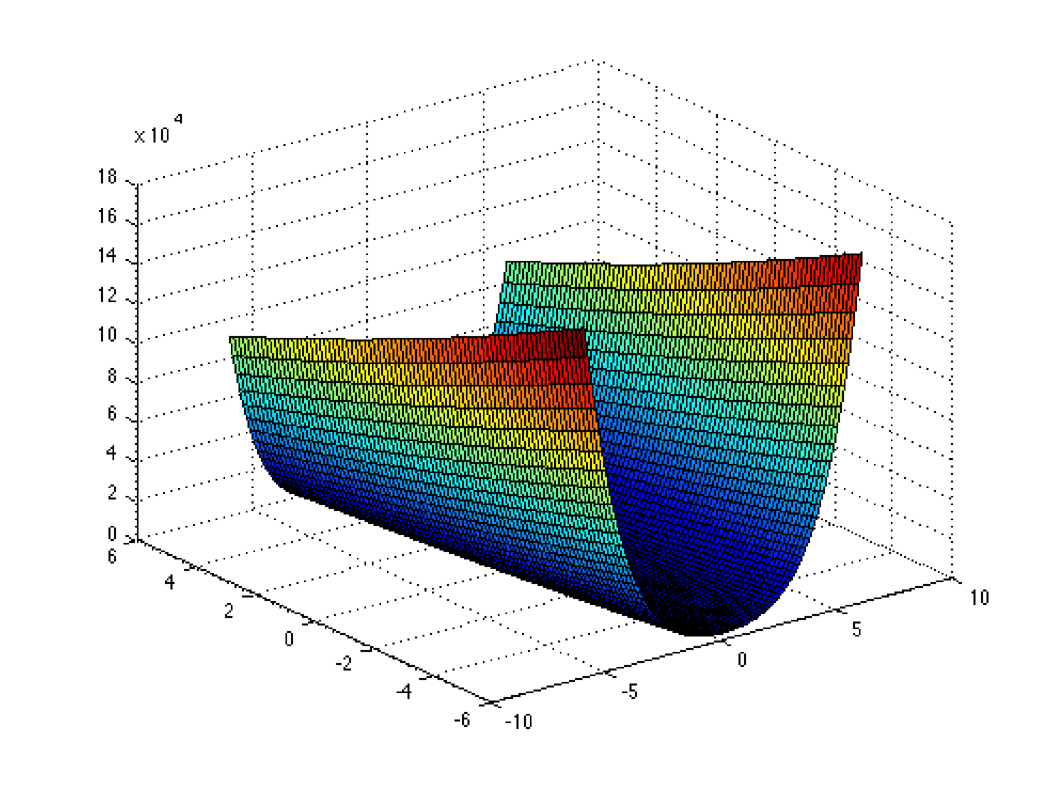
μ-σ-Principle

- Let's assume we can derive Φ
- It has two variables $\rightarrow$ partial derivatives
- $\partial\Phi(\mu,\sigma)/\partial\mu$ is partial derivative with respect to μ
- $\partial\Phi(\mu,\sigma)/\partial\sigma$ is partial derivative with respect to σ
- Formalization of cases:
  - $\partial\Phi(\mu,\sigma)/\partial\mu > 0$
  - $\partial\Phi(\mu,\sigma)/\partial\sigma > 0$: risk-seeking
  - $\partial\Phi(\mu,\sigma)/\partial\sigma < 0$: risk-averse
- Risk-neutral: σ doesn't matter (μ-Principle)

Decision Theory

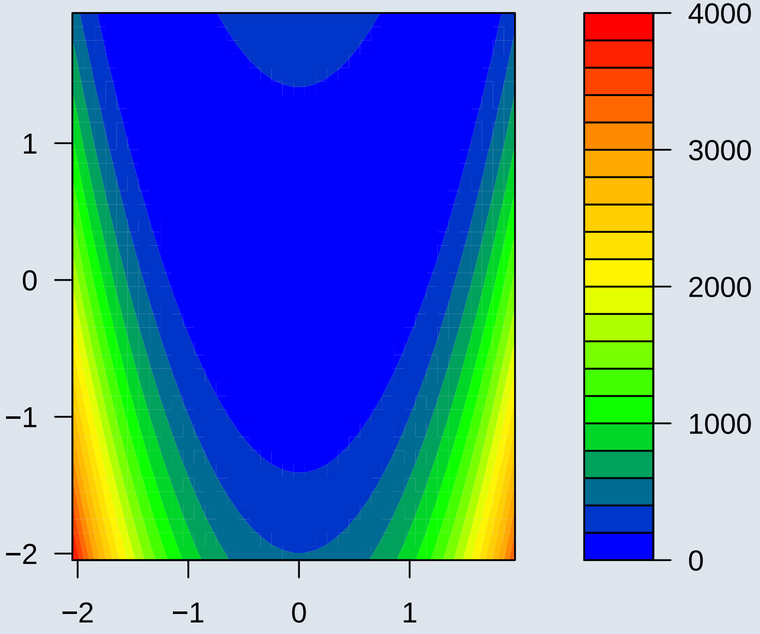
Level Curves

- Φ is a function with two inputs
- How would I draw it?



Level Curves

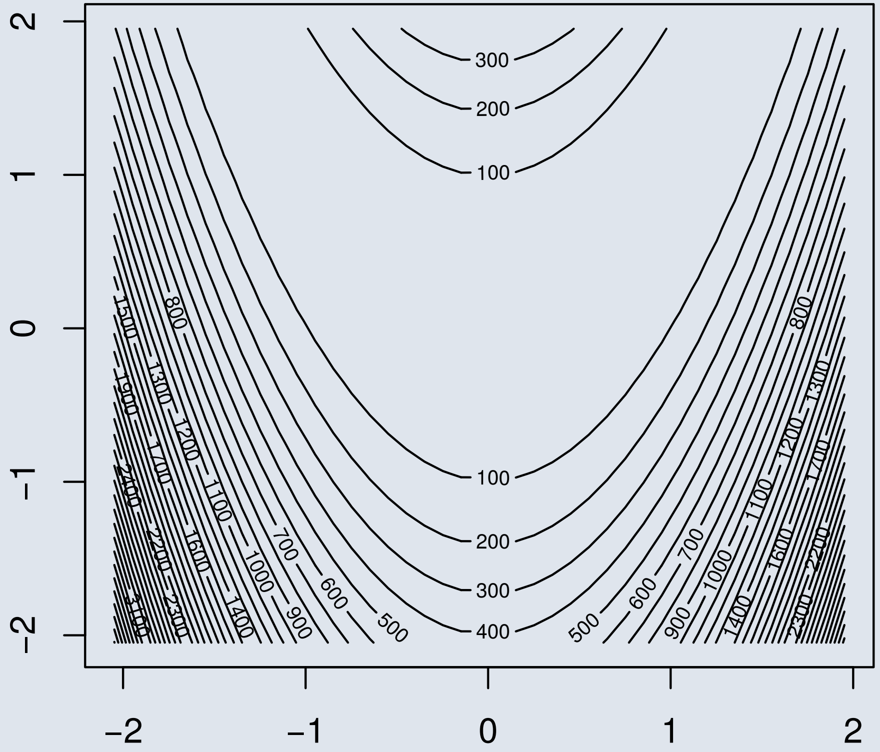
- Φ is a function with two inputs
- How would I draw it?



Decision Theory

Level Curves

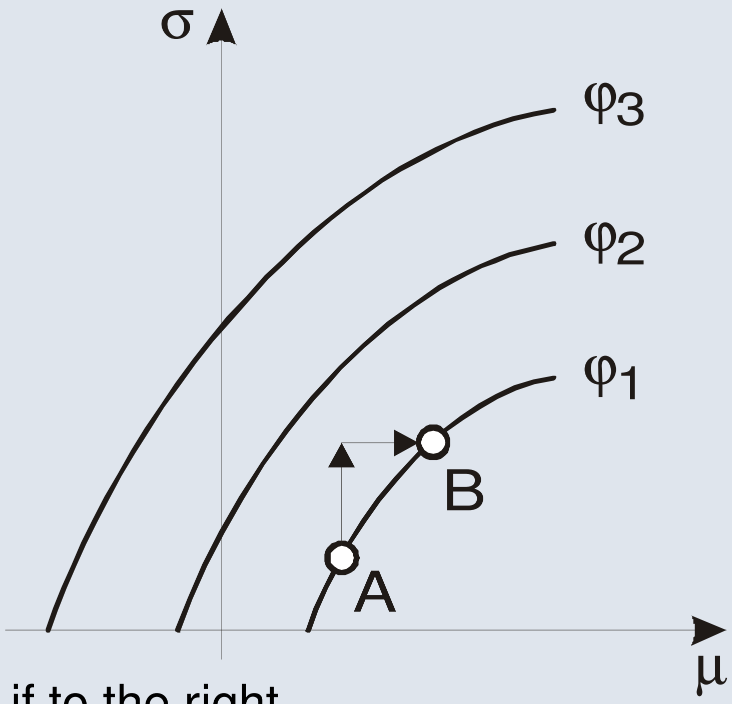
- Φ is a function with two inputs
- How would I draw it?



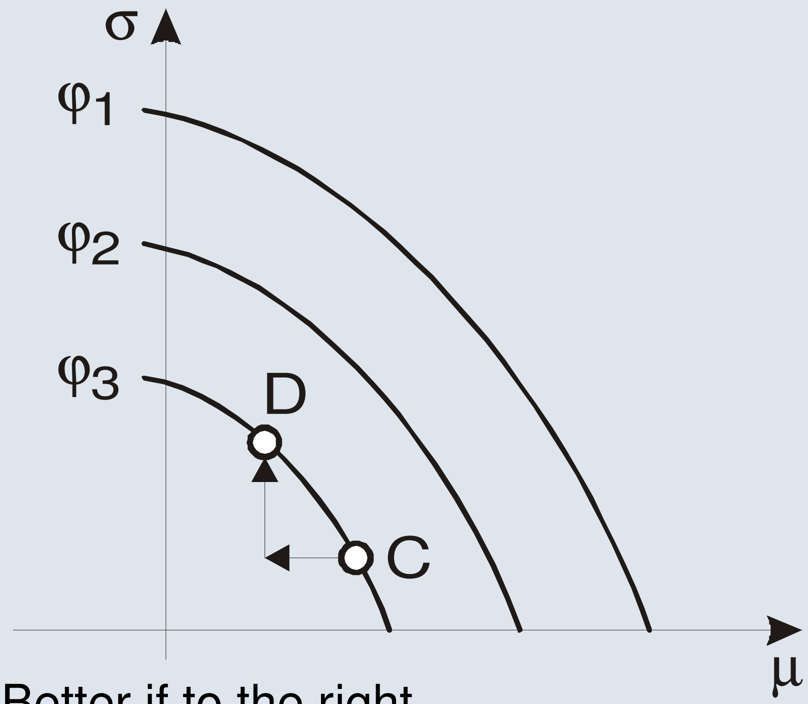
Decision Theory

Level Curves

$\varphi_1 > \varphi_2 > \varphi_3$



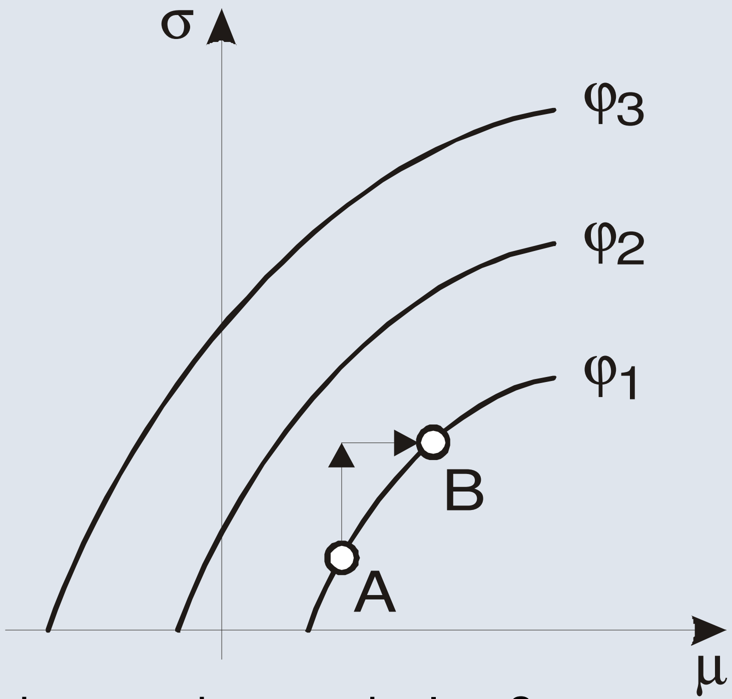
- Better if to the right
- Worse if high
- Risk-averse

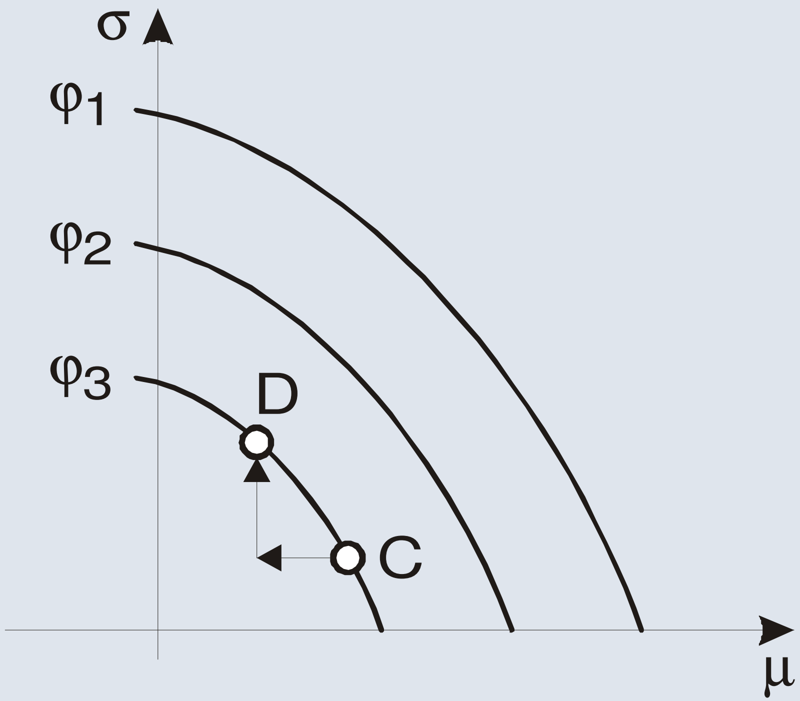


- Better if to the right
- Better if high
- Risk-seeking

Level Curves

$\varphi_1 > \varphi_2 > \varphi_3$





- What is the certainty equivalent?
- Move along the level curve to the intersection with the μ-axis
- Equally good situation with variance equal to zero

Decision Theory

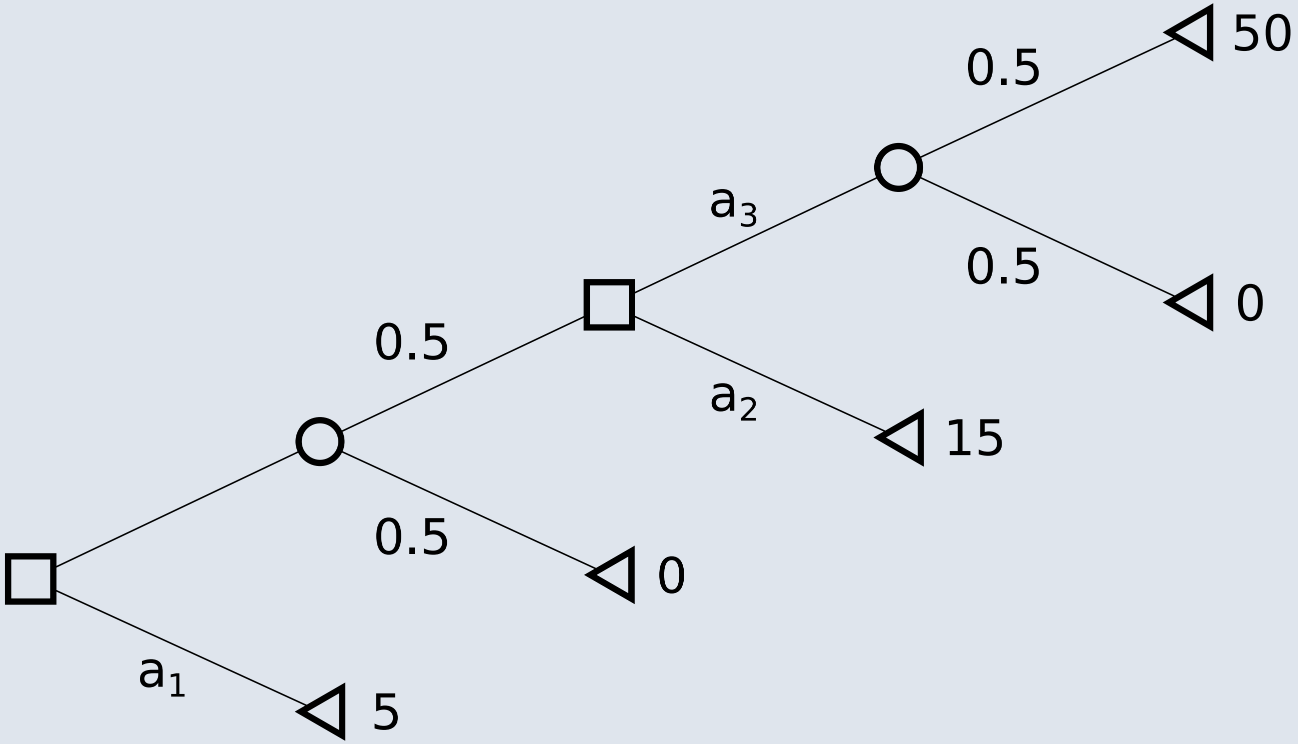
Examples of $\Phi(\mu, \sigma)$

1. $\Phi(\mu, \sigma) = \mu - \alpha\sigma$
2. $\Phi(\mu, \sigma) = \mu - \alpha\sigma^2$

- Difference between the expected value and α times the standard deviation or variance.
- $\alpha > 0$: risk-averse
- $\alpha < 0$: risk-seeking

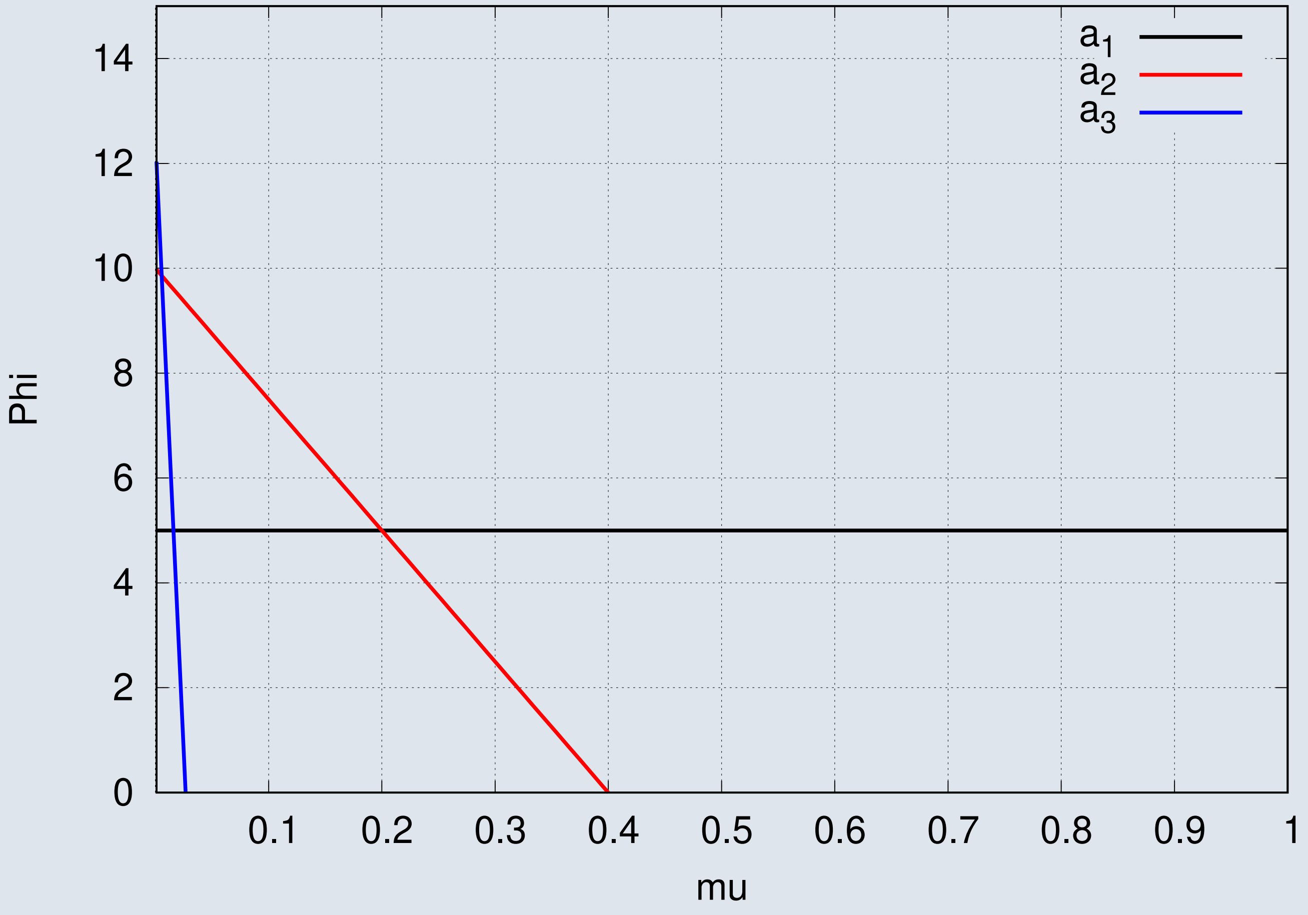
Decision Theory

Example - Maximize



Decision Theory

Example - Maximize



Decision Theory

Example - Maximize

Decision Theory

Example - Maximize

p	0.2	0.1	0.3	0.2	0.2
a_1	10	0	10	10	10
a_2	10	10	10	40	10

- a_1 is dominated by a_2 (absolute dominance)
- $\mu_1 = 9$, $\sigma_1 = 3$
- $\Phi(a_1) = \mu_1 - \alpha\sigma_1^2 = 9 - \alpha \cdot 9$
- $\mu_2 = 16$, $\sigma_2 = 12$
- $\Phi(a_2) = \mu_2 - \alpha\sigma_2^2 = 16 - \alpha \cdot 144$
- It depends on α which alternative is considered better
- How precise is this relationship?

Decision Theory

Example

$$\begin{aligned} 9 - 9\alpha &= 16 - 144\alpha \\ 135\alpha &= 7 \\ \alpha &= 7/135 \end{aligned}$$

- Thus:

$$\Phi(a_1) \begin{cases} < \Phi(a_2) & \text{for } \alpha < 7/135 \\ = \Phi(a_2) & \text{for } \alpha = 7/135 \\ > \Phi(a_2) & \text{for } \alpha > 7/135 \end{cases}$$

- The preference function $\Phi(\mu, \sigma) = \mu - \alpha\sigma^2$ can select an absolutely dominated alternative.

Decision Theory

Additional Metrics

- Besides μ and σ, other metrics can also be used
- For example:
  - Worst-case
  - Quantile values
  - Loss probability
  - Expected loss
  - Range

Decision Theory

Worst-Case

- $\mu -$ min Principle:
$$\Phi(\mu_i, e_i^{\text{min}}) = \lambda\mu + (1 - \lambda)e_i^{\text{min}}$$
- Hodges-Lehmann rule
- Similar to Hurwicz: λ weights two metrics
  - $\lambda = 1$: μ-Principle
  - $\lambda = 0$: Maximin
- Combination is also possible between extremization and satisficing, e.g.,
  - Maximize μ and demand $e^{\text{min}} \geq S$

Decision Theory

Quantile Values

- Given a critical probability p^k
- Quantile value f = the largest outcome for which the probability is not below p^k
- $p^k = 0.5$: Median
- p^k close to 0 or 1: best-case and worst-case

Example

p	0.10	0.06	0.20	0.15	0.17	0.13	0.02	0.17
e_1	-1	1	2	-2	3	0	-1	0

Cumulative probability:

e	-2	-1	0	1	2	3	4
$p(e_1 \geq e)$	1.00	0.85	0.73	0.43	0.37	0.17	0

Decision Theory

Loss Probability

- Probability that a critical value e^k is undershot: $p(e_1 < e^k)$
- Specifies the desired level, minimizing the probability of falling below it

Example

e	-2	-1	0	1	2	3	4
$p(e_1 \geq e)$	1.00	0.85	0.73	0.43	0.37	0.17	0

- For $e^k = -0.5$:

$$p(e_1 < -0.5) = 1 - p(e_1 \geq -0.5) = 1 - 0.73 = 0.27$$

- For $e^k = 1$:

$$p(e_1 < 1) = 1 - p(e_1 \geq 1) = 1 - 0.43 = 0.57$$

Decision Theory

Expected Loss

- Consider all outcomes K_i below the critical value e^k
- $K_i = \{j \in [n] : e_{ij} < e^k\}$
- Expected value of the distance of these outcomes to e^k

$$V_i = \sum_{j \in K_i} p_j(e^k - e_{ij})$$

- Special case $e^k = 0$: Sum only negative outcomes

$$V_i = - \sum_{j \in [n]: e_{ij} < 0} p_j e_{ij}$$

Decision Theory

Range

- Difference between the largest and smallest outcome

$$y(a_i) = \max_{j \in [n]} e_{ij} - \min_{j \in [n]} e_{ij}$$

Decision Theory

Quiz

Question 1

True or false?

The more often I repeat the same experiment, the closer the average result gets to the expected value.

Question 2

True or false?

The more often I repeat the same experiment, the larger the variance becomes.

Question 3

Determine expected value and variance:

p	1/4	1/4	2/4
e	5	7	4

a) $\mu = 4$, $\sigma^2 = 1$

b) $\mu = 5$, $\sigma^2 = 1.5$

c) $\mu = 5$, $\sigma^2 = 2$

Decision Theory

Quiz

Question 1

True or false?

The more often I repeat the same experiment, the closer the average result gets to the expected value.

Solution

Correct: Law of Large Numbers.

Question 2

True or false?

The more often I repeat the same experiment, the larger the variance becomes.

Solution

False: The variance becomes smaller.

Question 3

Determine expected value and variance:

p	1/4	1/4	2/4
e	5	7	4

a) $\mu = 4, \sigma^2 = 1$

b) $\mu = 5, \sigma^2 = 1.5$

c) $\mu = 5, \sigma^2 = 2$

Solution

Correct: b)