Probability and Statistics

Probability

HCMC University of Technology
Dung Nguyen

Sample space

- An experiment is a procedure that is
  - carried out under controlled conditions, and
  - executed to discover an unknown result.
- A trial is a single performance of an experiment.
- An outcome is the result of a trial.
- Deterministic or predictable experiments: only one possible result or outcome.
- Random experiment: different outcomes even when repeated in the same manner every time.
  - Flip a coin
  - Roll a die
  - Roll two dice
  - Measure current

Sample space

- The sample space Ω of an experiment is the set of all possible outcomes
  - $\Omega = \{H, T\}$
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $\Omega = \{(i, j), i, j = 1, 2, ..., 6\}$
- Ω is discrete if it consists of a finite or countable infinite set of outcomes.
- Ω is continuous if it contains an interval of real numbers.
- All subsets of Ω are called events.

Example 1 - Recycle time of a flash

- Randomly select a camera and record the recycle time of a flash. $\Omega = \mathbb{R}^+ = \{x | x > 0\}$, the positive real numbers.
- Suppose it is known that all recycle times are between 1.5 and 5 seconds. Then $\Omega = \{x | 1.5 < x < 5\}$ is continuous.
- It is known that the recycle time has only three values(low, medium or high). Then $\Omega = \{low, medium, high\}$ is discrete.
- Does the camera conform to minimum recycle time specifications? $\Omega = \{yes, no\}$ is discrete.

Events

An event A is a subset of the sample space of a random experiment. (= statement, sentence, clause)

Example

Recycle time of a flash

1. $\Omega = \mathbb{R}^+ = \{x | x > 0\}$: $A_1 = (1, \infty)$, $A_2 = [1, 2]$
2. $\Omega = \{x | 1.5 < x < 5\}$: $A_1 = (1.5, 3), A_2 = (2, 4)$
3. $\Omega = \{\text{low}, \text{medium}, \text{high}\}$: $A_1 = \{\text{low}, \text{medium}\}, A_2 = \{\text{medium}\}, A_3 = \{\text{low}, \text{high}\} = \overline{A}_2$
4. $\Omega = \{\text{yes}, \text{no}\}$ (Does the camera conform to minimum recycle time specifications?).

Operations

Conjunctions	NOT	AND	OR	BUT
Operators	$\overline{A}$	AB	$A+B$	$A-B$

Definition

1. *Complement:*
$$\overline{A} = \Omega \backslash A.$$
Also denoted by A^c *or* A'.
2. *Product = Intersection:* $AB = \{x : x \in A \text{ and } x \in B\}$. *Also denoted by* $A \cap B$.
3. *Sum = Union:* $A + B = \{x : x \in A \text{ or } x \in B\}$. *Also denoted by* $A \cup B$.
4. *Difference (A without B):* $A - B = \{x : x \in A \text{ and } x \notin B\}$.

A = "You pass this course"

B = "You pass the course of Linear Algebra"

Algebra of events' operations

Proposition (De Morgan's Rules)

1. $\overline{A+B} = \overline{A}\ \overline{B}$.
2. $\overline{AB} = \overline{A} + \overline{B}$.

Proposition (De Morgan's Rules)

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$.
2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Proposition (Distributive law)

1. $A(B + C) = AB + AC$.
2. $A + (BC) = (A + B)(A + C)$.

Proposition (Distributive law)

1. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Proposition (Difference laws)

1. $A - (B + C) = (A - B)(A - C)$.
2. $A - (BC) = (A - B) + (A - C)$.

Proposition (Difference laws)

1. $A \backslash (B \cup C) = (A \backslash B) \cup (A \backslash C)$.
2. $A \backslash (B \cap C) = (A \backslash B) \cap (A \backslash C)$.

What is Probability?

- The likelihood or chance that a particular outcome or event from a random experiment will occur.
- A number in the [0,1] interval.

Types of Probability

- **Subjective probability** is a "degree of belief", derived from an individual's personal judgment or own experience about whether a specific outcome is likely to occur $\implies$ Bias
- **Relative frequency probability** is based on how often an event occurs over a very large sample space.
- **Probability with equally likely outcomes**: Whenever a sample space consists of N possible outcomes that are equally likely, the probability of each outcome is $1/N$.

Definition of Modern Probability – Three Axioms

Denote by $\mathcal{A}$ the collection of all events. Probability is a function

$$P : \mathcal{A} \to [0, 1]$$

such that

1. $P(A) \geq 0$
2. $P(\Omega) = 1$.
3. For two disjoint events A, B

$$P(A + B) = P(A) + P(B)$$

For any event $A \in \mathcal{A}$, $P(A)$ called is the probability of A.

Properties of Probability

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$ and $P(\emptyset) = 0$.
   - Probability of Certainty is 1
   - Probability of Imposibility is 0
   - Probability Zero Does Not Mean Impossible.
3. $P(\overline{A}) = 1 - P(A)$
4. For two disjoint events A, B

$$P(A + B) = P(A) + P(B)$$

In general, for a sequence of disjoint events $A_1, \cdots, A_n$,

$$P(A_1 + \cdots + A_n) = P(A_1) + \cdots + P(A_n)$$

Example

2. If three fair coins are tossed, what is the probability that three faces will not be the same?

3. **Birthday Problem.** Find the probability that, in a set of 10 randomly chosen people, at least one pair of them will have the same birthday.

Addition Rule

1. For two events
$$P(A+B) = P(A) + P(B) - P(AB)$$
2. For three events
$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$
Note the alternating signs.
3. In general,
$$P\left(\sum_i A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i,j,k} P(A_iA_jA_k) - \cdots$$

Example 4 – Semiconductor Wafers

Consider the semiconductor wafer data in the table below.

Contamination level of wafers	Location of Sputtering Tool: Center	Location of Sputtering Tool: Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	940

A wafer is randomly selected from the batch.

- H = high concentrations of contaminants. Then $P(H) = 358/940$.
- C = the wafer being located at the center of a sputtering tool. Then $P(C) = 626/940$.
- $P(HC) = 112/940$
- $P(H+C) = \frac{246+112+514}{940} =$

$$= 0.9277$$
940

- $P(H) + P(C) - P(HC) = \frac{358+626-112}{} = 0.9277$

Example

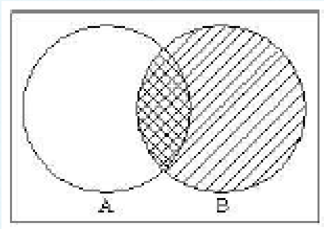
5. A patient arrives at a doctor's office with a sore throat and low-grade fever. After an exam, the doctor decides that the patient has either a bacterial infection or a viral infection or both. The doctor decides that there is a probability of 0.7 that the patient has a bacterial infection and a probability of 0.4 that the person has a viral infection. What is the probability that the patient has both infections?

6. Suppose that both Saturday and Sunday each have probability 0.5 to get rain and that the probability is p that it rains both days. How does the probability of rain during the weekend depend on p?

Conditional Probability

- $P(B|A)$ is the probability of event B occurring, given that event A has already occurred.
- A communications channel has an error rate of 1 per 1000 bits transmitted. Errors are rare, but do tend to occur in bursts. If a bit is in error, the probability that the next bit is also in error is greater that

Conditional probability of A given B
$$P(A|B) = \frac{P(AB)}{P(B)}$$
Visually, conditional probability is the shaded area



Example 7 – Flaws

There are 4 possibilities conditioned on flaws in the below table.

Defective	Surface Flaws		Total
	Yes (F)	No (F′)	
Yes (D)	10	18	28
No (D′)	30	342	372
Total	40	360	400

Compute $P(F), P(D), P(F|D), P(D|F)$.

Example 8 – Confirmed cases of COVID-19 in Japan

Age Bracket	M	F	Total
0	9	4	13
10	4	3	7
20	20	46	66
30	49	40	89
40	63	59	122
50	93	84	177
60	93	59	152
70	68	58	126
80	50	29	79
90	7	5	12
Total	456	387	843

Example 9 – Sampling Without Replacement

A batch of 50 parts contains 10 made by Tool 1 and 40 made by Tool 2. Select 2 parts randomly. What is the probability that the 1st part came from Tool 1 AND the 2nd part came from Tool 2?

$$P(E_1) = P(\text{1st part came from Tool 1}) = 10/50$$

$$P(E_2|E_1) = P(\text{2nd part came from Tool 2 given that 1st part came from Tool 1}) = 40/49$$

Therefore

$$P(E_1E_2) = P(\text{1st part came from Tool 1 and 2nd part came from Tool 2}) = (10/50)(40/49) = 8/49$$

The product rule:

$$\boxed{P(AB) = P(A)P(B|A) \qquad = P(BA) = P(B)P(A|B)}$$

Example

10 Roll two dice and observe the numbers coming up.

(a) Define two events by: A="the sum is six," and B="the numbers are not equal." Find and compare $P(B)$ and $P(B|A)$.

(b) Let E="the number showing on the first die is even," and F="the sum of the numbers showing is seven." Find and compare $P(F)$ and $P(F|E)$.

11 Consider the tossing of a pair of dice. What is the probability of a number greater than 4 with the second die if a number less than 4 turned up on the first die?

Independence

- Repeated independent trials
- The outcome of any trial of the experiment does not influence or affect the outcome of any other trial
- The trials are said to be physically independent
- Physical independence is a belief
- It cannot be proved that the trials are independent; we can only believe

Independence of two events

Definition

A is independent of B if

$$P(A|B) = P(A)$$

That is, knowing B occurred doesn't impact whether A occurred.

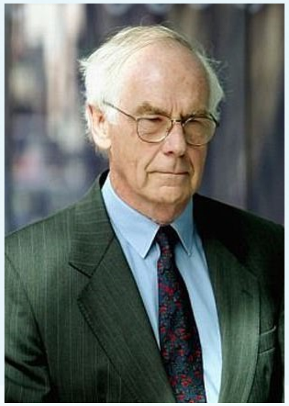
The following statements are equivalent

1. A and B are independent
2. $P(A|B) = P(A)$
3. $P(B|A) = P(B)$
4. $P(AB) = P(A)P(B)$

The case of Sally Clark I

- Solicitor Sally Clark was the victim of a miscarriage of justice when she was found guilty (in 1999) for the murder of two children (Christopher, 11 weeks, 1996)+(Harry, 8 weeks, 1997)
- convicted in November 1999
- 1st appeal (October 2000) dismissed
- 2nd appeal (January 2003), released
- Died in her home in March 2007 from alcohol poisoning

The case of Sally Clark II



Sir Roy Meadow

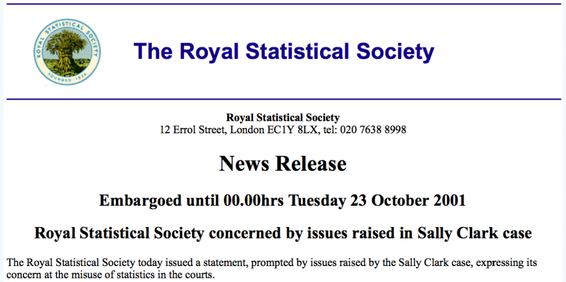
- Studied medicine at Worcester College, Oxford
- British paediatrician at St James's University Hospital.
- Awarded the Donald Paterson prize of the British Paediatric Association in 1968.
- Knighted for services to child health in 1998.
- "Meadow's Law": "In a single family, one sudden infant death is a tragedy, two is suspicious and three is murder, until proved
- An expert witness in cases of

The argument

SIDS = Sudden infant death syndrome = cot death = crib death = the sudden unexplained death of a child of less than one year of age

Sir Roy Meadow testified that

- The frequency of SIDS in an affluent family non-smoking families is about 1 in 8500.
- The chance of two children from an affluent family suffering SIDS was $(1/8500)^2 = 1/73M$.


The Royal Statistical Society

Royal Statistical Society
12 Errol Street, London EC1Y 8LX, tel: 020 7638 8998

News Release

Embargoed until 00.00hrs Tuesday 23 October 2001

Royal Statistical Society concerned by issues raised in Sally Clark case

The Royal Statistical Society today issued a statement, prompted by issues raised by the Sally Clark case, expressing its concern at the misuse of statistics in the courts.

The Royal Statistical Society in 2001 issued a press release that summed up the two major flaws in Meadow's argument.

The two major flaws

1. $P(A_1A_2) \neq P(A_1)P(A_2)$
   - There is strong evidence that the SIDS does have genetic or environmental factors that may correlate within a family.
   - $P(\text{second death}|\text{first death}) = 1/77$
2. $P(\text{murder}|\text{evidence}) = P(\text{evidence}|\text{murder})$.
   - Suppose that a DNA sequence occurs in 1 in 10000 people. Does this mean that if a suspect 's DNA matches that found at a crime scene, the probability that he is guilty is 10000:1
   - 1 million $\Longrightarrow$ The DNA gives of odds of 100:1 against the suspect being guilty.

Example

12 Consider a random experiment in which a fair coin is flipped and a die is rolled, with A denoting the event that the coin shows a head and B denoting the event that the die shows number 1. Are A and B independent?

13 Suppose a random experiment is to roll a single die. Let A be the event that the outcome is even, and let B be the event that the outcome is a multiple of three. Are A and B independent?

Independence of three events

Definition (Pairwise independence)

Events A, B, and C are pairwise independent if

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C), \quad P(BC) = P(B)P(C)$$

$$P(A|B) = P(A), \quad P(B|C) = P(B), \quad P(C|A) = P(C),$$
$$P(B|A) = P(B), \quad P(C|B) = P(C), \quad P(A|C) = P(A).$$

Example (14)

Suppose two fair coins are flipped. Let

- A be the events "first coin shows heads"
- B be the event "second coin shows heads"
- C be the event "both coins show heads or both coins show tails."

A, B, C are pairwise independent.

Independence of three events

Definition (1st)

Events A, B, and C are (mutally) independent if they are pairwise independent and if $P(ABC) = P(A)P(B)P(C)$.

Example (15)

Suppose two fair coins are flipped. Let

- A be the events "first coin shows heads"
- B be the event "second coin shows heads"
- C be the event "both coins show heads or both coins show tails."

Prove that A, B, C are not independent.

Definition (2nd)

Events A, B, and C are (mutally) independent if

$$P(A|B) = P(A), \quad P(A|C) = P(A), \quad P(B|C) = P(B), \quad P(A|BC) = P(A)$$

Remark I

- Stochastic independence does not necessarily mean that the events are physically independent.
  - Physical independence is, in essence, a property of the events themselves.
  - Stochastic independence is a property of the probability measure.
- In the future, we shall be using the word stochastically in conjunction with the word independent only on rare occasions.

Independence of n events

Definition (1st)

The events $A_1, \cdots, A_n$ *are independent (or mutually independent) if, for every subset* $A_{i_1}, \cdots, A_{i_k}$ *of k of these events* $(k = 2, 3, \cdots, n)$,

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

Remark

This is a recursive definition: the product rule applies to the "big" intersection and also to all smaller ones as well

Definition (2nd)

The events $A_1, \cdots, A_n$ *are independent (or mutually independent) if, for every subset* $A_{i_1}, \cdots, A_{i_k}$ *of k of these events* $(k = 2, 3, \cdots, n)$,

$$P(A_{i_1} | A_{i_2} \cdots A_{i_k}) = P(A_{i_1})$$

Example I

16 Assume $P(H) = p, P(T) = 1 - p$ for an unfair coin. Compute $P(\text{HTHTT})$ and $P(\text{get 2H and 3T, in any order})$

17 A system composed of n separate components is said to be a parallel system if it functions when at least one of the components functions. For such a system, if component i, which is independent of the other components, functions with probability $p_i, i = 1, \ldots, n$, what is the probability that the system functions?

Example II

18 Suppose that we toss 2 fair dice. Let E_1 denote the event that the sum of the dice is 6 and F denote the event that the first die equals 4. Prove that E_1 and F cannot be independent.

Example 19 - XOR gate

Let A and B respectively denote the events that inputs #1 and #2 of an Exclusive-OR gate are logical 1. Assume that A and B are physically independent (hence they are stochastically independent) events. Let C denote the event that the output of the Exclusive-OR gate is logical 1

$$C = A \oplus B = A\bar{B} \cup \bar{A}B$$

1. Is the output of the XOR gate really independent of the input? Are A and C independent events?
2. Assume that $P(A) = P(B) = 0.500001$. Is the result still the same?

Probability of a product - Independent

Proposition

For two independent events A, B,

$$P(AB) = P(A)P(B)$$

In general, if $A_1, \ldots, A_n$ are independent events then

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2)P(A_3)\cdots P(A_n)$$

Example 20 – Semiconductor Wafers

Assume the probability that a wafer contains a large particle of contamination is 0.01 and that the wafers are independent; that is, the probability that a wafer contains a large particle does not depend on the characteristics of any of the other wafers. If 15 wafers are analyzed, what is the probability that no large particles are found?

Solution

Let E_i denote the event that the ith wafer contains no large particles, $i = 1, 2, \ldots, 15$.

$$P(E_1 E_2 \cdots E_{15}) = (0.99)^{15} = 0.86.$$

Probability of a product – Dependent

Theorem (Multiplication Rule)

- *The conditional probability can be rewritten to generalized a multiplication rule.*

$$P(AB) = P(A)P(B|A) = P(B)P(A|B)$$

The last expression is obtained by exchanging the roles of A and B.

- *More generally*

$$P(ABC) = P(A)P(B|A)P(C|AB)$$

- *In general*

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 A_2 \cdots A_{n-1})$$

Example 21 – Machining Stages

The probability that a part made in the 1st stage of a machining operation meets specifications is 0.90. The probability that it meets specifications in the 2nd stage, given that met specifications in the first stage is 0.95. What is the probability that both stages meet specifications?

Solution

Let A and B denote the events that the part has met the 1st and the 2nd stage specifications, respectively.

$$P(AB) = P(B|A)P(A) = (0.95)(0.90) = 0.855$$

Example

22 Celine is undecided as to whether to take a French course or a chemistry course. She estimates that her probability of receiving an A grade would be $\frac{1}{2}$ in a French course and $\frac{2}{3}$ in a chemistry course. If Celine decides to base her decision on the flip of a fair coin, what is the probability that she gets an A in chemistry?

23 An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each. Compute the probability that each pile has exactly 1 ace.

Theorem (Total probability - Simple form)

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

Proof.

$$P(A) = P(AB) + P(A\overline{B})$$
$$P(AB) = P(A|B)P(B)$$
$$P(A\overline{B}) = P(A|\overline{B})P(\overline{B})$$

Remark

- *To find $P(A)$, first imagine that B occurred*
- *From $P(A|B)$, we can determine $P(AB)$*
- *Next imagine that $\overline{B}$ occurred*
- *From $P(A|\overline{B})$, we can determine $P(A\overline{B})$*
- *The sum of these two numbers is $P(A)$*

Example 24 – Semiconductor Contaminati

Information about product failure based on chip manufacturing process contamination is given below. Find the probability of failure.

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not High	0.8

Let F denote the event that the product fails. Let H denote the event that the chip is exposed to high contamination during manufacture. Then

$$P(FH) = 0.02 \quad (P(F|H) = 0.100, P(H) = 0.2)$$
$$P(FH') = 0.004 \quad (P(F|H') = 0.005, P(H') = 0.8)$$
$$P(F) = P(FH) + P(FH') = 0.020 + 0.004 = 0.024$$

Example

25 Box I has 3 green and 2 red balls, while Box II has 4 green and 6 red balls. A ball is drawn at random from Box I and transferred to Box II. Then, a ball is drawn at random from Box II. What is the probability that the ball drawn from Box II is green?

26 In box 1, there are 60 short bolts and 40 long bolts. In box 2, there are 10 short bolts and 20 long bolts. Take a box at random, and pick a bolt. What is the probability that you chose a short bolt?

Remark (A built-in test for checking answers)

$P(A)$ is between $P(A|B)$ and $P(A|\overline{B})$

Theorem (Total probability - Extended form)

$$P(A) = \sum_{i=1}^{k} P(AB_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

where $B_i \cap B_j = \emptyset$, for $i \neq j$ and $S = \bigcup_{i=1}^{k} B_i$ (disjoint).

Remark

- *The numerator is always one of the terms in the denominator*
- *$P(A)$ is between the smallest of $P(A|B_i)$ and the largest of $P(A|B_i)$*

Example 27 - Semiconductor Failures

Continuing the discussion of contamination during chip manufacture, find the probability of failure.

Probability of Failure	Level of Contamination	Probability of Level
0.100	High	0.2
0.010	Medium	0.3
0.001	Low	0.5

Solution

Let F denote the event that a chip fails
Let H, M, L denote the event that a chip is exposed to high, medium, low levels of contamination

$$P(F) = P(F|H)P(H) + P(F|M)P(M) + P(F|L)P(L)$$
$$= (0.1)(0.2) + (0.01)(0.3) + (0.001)(0.5) = 0.0235$$

Example I

28 A bin contains 3 different types of disposable flashlights. The probability that a type 1 flashlight will give over 100 hours of use is 0.7, with the corresponding probabilities for type 2 and type 3 flashlights being 0.4 and 0.3, respectively. Suppose that 20 percent of the flashlights in the bin are type 1, 30 percent are type 2, and 50 percent are type 3. What is the probability that a randomly chosen flashlight will give more than 100 hours of use?

Lemma (Bayes's lemma)

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

Bayes's Formula - Simple form

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})}$$

The numerator is always one of the terms in the denominator

Example I

29 In answering a question on a multiple-choice test, a student either knows the answer or guesses. The probability that the student knows the answer is 0.8. Assume that a student who guesses at the answer will be correct with probability 1/4. What is the conditional probability that a student knew the answer to a question given that he or she answered it correctly?

Example II

30 A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy people tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply that he or she has the disease.) If 0.5 percent of the population actually has the disease, what is the probability that a person has the disease given that the test result is positive?

Example III

31 At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose, however, that a new piece of evidence which shows that the criminal has a certain characteristic (such as left-handedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)}$$

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

Theorem (Bayes's Formula – Extended form)

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$$

Example I

32 Identify the source of a defective item. There are 3 machines: M_1, M_2, M_3. P(defective): 0.01, 0.02, 0.025, respectively. The percent of items made that come from each machine is: 50%, 30%, and 20%, respectively.

33 A bin contains 3 different types of disposable flashlights. The probability that a type 1 flashlight will give over 100 hours of use is 0.7, with the corresponding probabilities for type 2 and type 3 flashlights being 0.4 and 0.3, respectively. Suppose that 20 percent of the flashlights in the bin are type 1, 30 percent are type 2, and 50 percent are type 3. What is the probability that a randomly chosen flashlight will give more than 100 hours of use? Given that a flashlight lasted over 100 hours, what is

Example II

the conditional probability that it was a type 1 flashlight?