



LUND
UNIVERSITY

Review of probability theory and statistics

MARIA KIHLM



Literature

Wikipedia or other sites with introductions and reviews of probability theory.

I have used:

https://en.wikipedia.org/wiki/Probability_theory

https://en.wikipedia.org/wiki/Random_variable

https://en.wikipedia.org/wiki/Mutual_exclusivity

https://en.wikipedia.org/wiki/Collectively_exhaustive_events

[https://en.wikipedia.org/wiki/Independence_\(probability_theory\)](https://en.wikipedia.org/wiki/Independence_(probability_theory))

https://en.wikipedia.org/wiki/Conditional_probability

https://en.wikipedia.org/wiki/Law_of_total_probability

https://en.wikipedia.org/wiki/Bernoulli_trial

https://en.wikipedia.org/wiki/Poisson_distribution

https://en.wikipedia.org/wiki/Exponential_distribution

Terminology

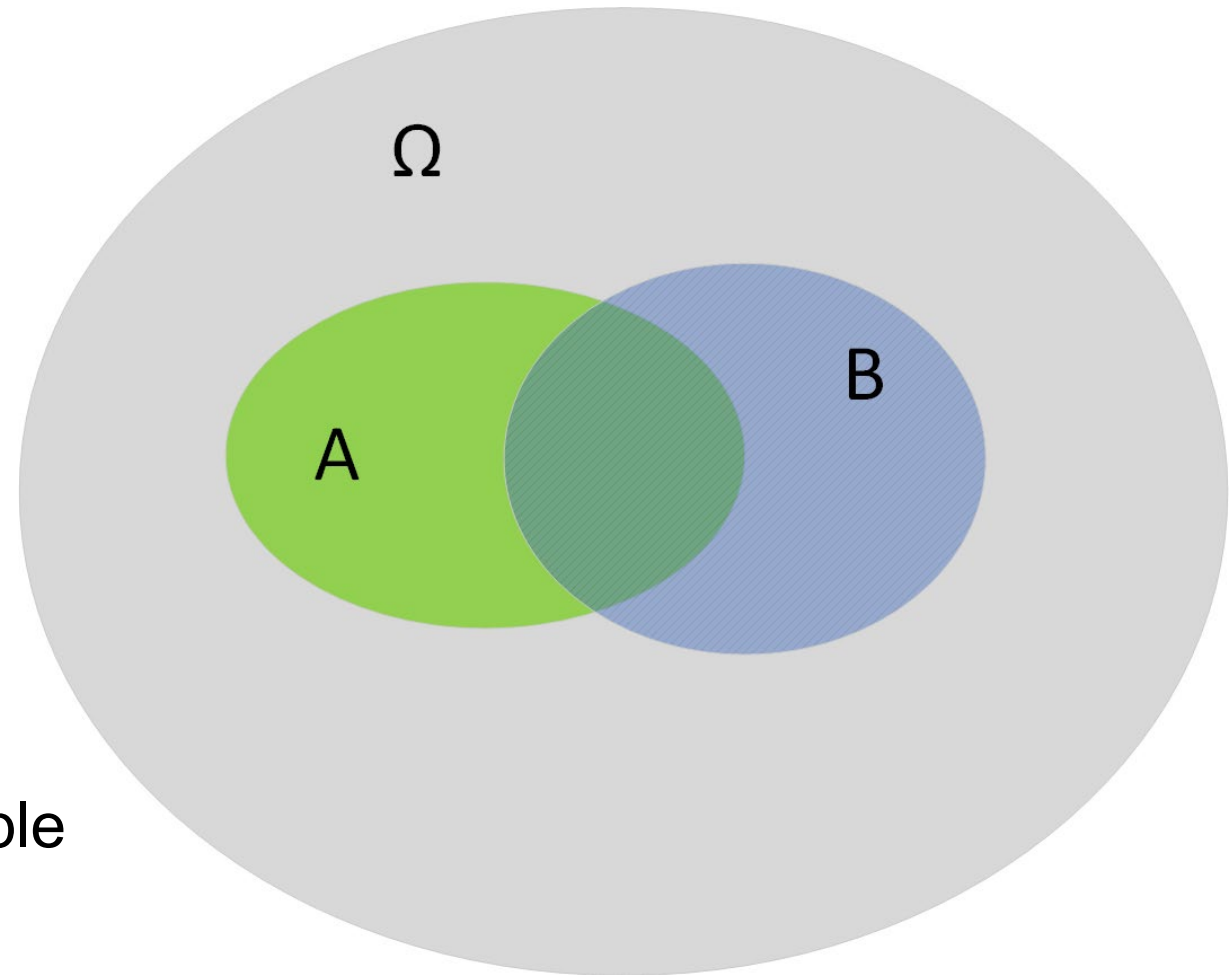
- Probability theory is based on an **experiment**, which can produce a number of **outcomes** (results).
- The **sample space**, Ω , is the set of all outcomes of an experiment.
- An **event** is a set of points in the sample space (i.e a set of outcomes of an experiment).
- **Probability** is a way of assigning every event a value between zero and one, with the requirement that the event made up of all possible results be assigned a value of one.
- A **probability distribution** is the mathematical function that gives the probabilities of occurrence of possible outcomes for an experiment.

Two events A and B

For two events A and B

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $P(A \cap B) = 0$, then A and B are **disjoint** events (also called **mutually exclusive** events).
- If $P(A \cup B) = 1$, then A and B are **collectively exhaustive** events.

(Can of course be generalized to multiple events $E_1, E_2, E_3, \dots$)



Random variables

- A **random variable (stochastic variable)** is a mathematical formalization of a quantity or object which depends on random events.
- When the range of X is finitely or infinitely countable, X is a **discrete random variable** and its distribution is a **discrete probability distribution**, $P(X=k)=p_k$.
- When the range of X is uncountably infinite (usually an interval), then X is a **continuous random variable**, and its distribution can be described by a **probability density function (pdf)**, $f_X(x)$.
- Any random variable can be described by its **cumulative distribution function (cdf)**, $F_X(x)$, which describes the probability that the random variable will be less than or equal to a certain value.

Moments of a random variable

- The **expected value of X , $E[X]$** , of a random variable X is the arithmetic mean of the possible values a random variable can take, weighted by the probability of those outcomes. This is also called the **first moment of X** .
 - Discrete random variable: $E[X] = \sum k \cdot p_k$
 - Continuous random variable: $E[X] = \int x f_X(x) dx$
- The **second moment of X , $E[X^2]$** :
 - Discrete random variable: $E[X^2] = \sum k^2 \cdot p_k$
 - Continuous random variable: $E[X^2] = \int x^2 f_X(x) dx$
- The **Variance for X , $Var(X) = E[X^2] - (E[X])^2$**
- The **standard deviation for X , $\sigma = \sqrt{Var(X)}$**

Some properties for multiple random variables

- Consider two random variables X and Y .
- $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$
- $E[X + Y] = E[X] + E[Y]$
- $Var(X + Y) = Var(X) + Var(Y) + 2 \cdot Cov(X, Y)$
- $Cov(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y]$

Independence in probability theory

- Two events A and B are **independent**, **statistically independent**, or **stochastically independent**, if the occurrence of one does not affect the probability of occurrence of the other, i.e $P(A \cap B) = P(A) \cdot P(B)$.
- Two random variables X and Y are independent if, for every x and y , the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events.
- For **two independent random variables** X and Y :
 - $F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$ for all x, y
 - $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ (continuous) and $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ (discrete)
 - $E[XY] = E[X] \cdot E[Y]$
 - $Var(X + Y) = Var(X) + Var(Y)$

Exercise: A pair of fair dice

A pair of fair dice (the probability of each outcome is $1/6$) is thrown. Let X be the **minimum** of the two numbers that come up.

- (a) Find the probability distribution for X .
- (b) Find $E[X]$ and $Var(X)$.

Solution: A pair of fair dice

Let A be a random variable for the result of the first dice, and B be a random variable for the result of the second dice. Let X be a random variable for the minimum of A and B .

There are 36 possible outcomes of each trial:

A	B	X	A	B	X	A	B	X	A	B	X	A	B	X	A	B	X
1	1	1	2	1	1	3	1	1	4	1	1	5	1	1	6	1	1
1	2	1	2	2	2	3	2	2	4	2	2	5	2	2	6	2	2
1	3	1	2	3	2	3	3	3	4	3	3	5	3	3	6	3	3
1	4	1	2	4	2	3	4	3	4	4	4	5	4	4	6	4	4
1	5	1	2	5	2	3	5	3	4	5	4	5	5	5	6	5	5
1	6	1	2	6	2	3	6	3	4	6	4	5	6	5	6	6	6

Solution: A pair of fair dice, continued

(a)

For the probability distribution for X , count the number of outcomes where $X=1, 2, 3, 4, 5, 6$ and divide with 36 (the total number of outcomes).

$$P(X = 1) = \frac{11}{36}; P(X = 2) = \frac{9}{36}; P(X = 3) = \frac{7}{36};$$
$$P(X = 4) = \frac{5}{36}; P(X = 5) = \frac{3}{36}; P(X = 6) = \frac{1}{36}$$

(b)

$$E[X] = \sum_{k=1}^6 k \cdot P(X = k) = 1 \cdot \frac{11}{36} + 2 \cdot \frac{9}{36} + 3 \cdot \frac{7}{36} + 4 \cdot \frac{5}{36} + 5 \cdot \frac{3}{36} + 6 \cdot \frac{1}{36} \approx 2.53$$

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{k=1}^6 k^2 \cdot P(X = k) \approx 8.36 \Rightarrow \text{Var}(X) \approx 1.96$$

Conditional probability and Baye's theorem

- **Conditional probability** is a measure of the probability of an event occurring, given that another event is already known to have occurred.
- For two events A and B , the probability of A given that B has occurred = $P(A|B)$
 - $P(A | B) = \frac{P(A \cap B)}{P(B)}$
 - If A and B are independent: $P(A|B) = P(A)$
- While conditional probabilities can provide useful information, limited information is often supplied or at hand. Therefore, it can be useful to reverse or convert a conditional probability using **Bayes' theorem**:

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

Law of total probability

- The **law (or formula) of total probability** is a fundamental rule relating marginal probabilities to conditional probabilities. It **expresses the total probability of an outcome** which can be realized via several distinct event.
- If $\{B_n: n = 1, 2, 3, \dots\}$ is a finite or countably infinite set of **mutually exclusive and collectively exhaustive events** (i.e the events are disjoint and encompass the range of all possible outcomes for B)):

$$P(A) = \sum_n P(A \mid B_n) \cdot P(B_n)$$

- The summation can be interpreted as a weighted average, and consequently the marginal probability, $P(A)$, is sometimes called the average probability.

Exercise: Three coins

Consider 3 coins where two are fair, yielding heads with probability 0.50, while the third yields heads with probability 0.75.

**If you randomly select one of the coins and tosses it 3 times, yielding 3 heads
- what is the probability this is the biased coin?**

Solution: Three coins

$P(\text{fair coin}) = 2/3$ and $P(\text{biased coin (or **not** a fair coin)}) = 1 - 2/3 = 1/3$

$P(\text{head} \mid \text{fair coin}) = 0.5$ and $P(\text{head} \mid \text{biased coin}) = 0.75$

Each coin toss is independent of the other tosses, which means that

- $P(3 \text{ heads} \mid \text{fair coin}) = P(\text{head} \mid \text{fair coin})^3 = (0.5)^3$
- $P(3 \text{ heads} \mid \text{biased coin}) = P(\text{head} \mid \text{biased coin})^3 = (0.75)^3$

Also, $P(3 \text{ heads}) = P(3 \text{ heads} \mid \text{fair coin}) \cdot P(\text{fair coin}) +$
 $P(3 \text{ heads} \mid \text{biased coin}) \cdot P(\text{biased coin}) = (0.5)^3 \cdot 2/3 + (0.75)^3 \cdot 1/3$

Use Baye's theorem:

$$P(\text{biased coin} \mid 3 \text{ heads}) = \frac{P(3 \text{ heads} \mid \text{biased coin}) \cdot P(\text{biased coin})}{P(3 \text{ heads})}$$

Bernoulli trials and the geometric distribution

- A **Bernoulli trial** (or **binomial trial**) is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted (the trials are independent of each other).
 - $P(\text{success}) = p$ and $P(\text{failure}) = 1 - p$
- The **geometric distribution** gives the probability that the first occurrence of success requires k independent trials, each with success probability p .
- For a geometric random variable X
 - $p_k = (1 - p)^{k-1} \cdot p$ for $k = 1, 2, 3, \dots$
 - $E[X] = 1/p$
 - $Var(X) = \frac{1-p}{p^2}$

Binomial distribution

- The **binomial distribution** with parameters n and p is the discrete probability distribution of **the number of successes in a sequence of n independent experiments**, each with its own Boolean-valued outcome: *success* (with probability p) or *failure* (with probability $q = 1-p$).
- For a binomial distributed random variable X , with n trials:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

- $E[X] = n \cdot p$ and $Var(X) = n \cdot p \cdot q$

Exercise: Stop-and-wait ARQ

A sender transmits messages to a receiver using Stop-and-wait ARQ.

When a message is sent, the sender waits for an ACK. If the ACK does not come within a certain time period, the message is retransmitted.

Assume that **the probability that the sender receives an ACK within the time period is 0.8.**

In average, how many transmissions are required until the sender receives an ACK?

Solution: Stop-and-Wait ARQ

$P(\text{The sender receives an ACK}) = 0.8 \Rightarrow$

$P(\text{The sender does not receive an ACK}) = 0.2$

If the sender does not receive an ACK, it retransmits the message.

$P(\text{retransmission of a message}) = 0.2$

The outcome of each transmission is independent on other transmissions.

Let X denote the number of transmissions that are required to receive an ACK.

$P(X = 1) = 0.8$ (ACK in the first transmission of the message)

$P(X = 2) = 0.2 \cdot 0.8$ (One failed transmission, then an ACK)

$P(X = k) = 0.2^{k-1} \cdot 0.8$ ($k-1$ failed transmissions, then an ACK)

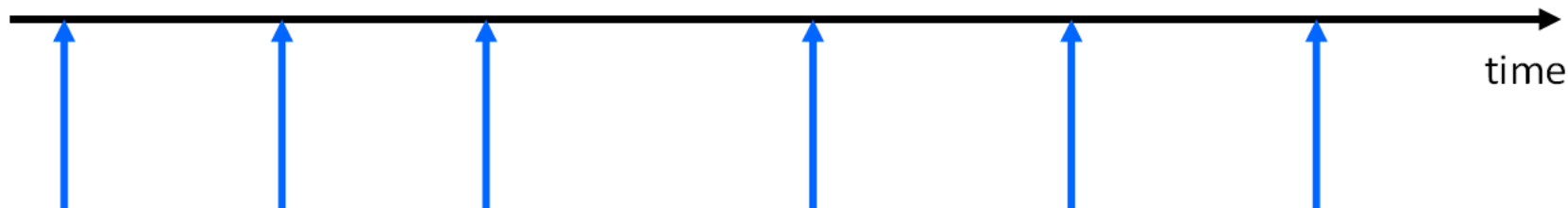
Geometric distribution with $p = 0.8$, which means that $E[X] = 1/0.8 = 1.25$

Poisson distribution

- The **Poisson distribution** is a discrete probability distribution that is very useful in this course.
- It expresses the probability of a given number of events occurring in a time interval, t , if **these events occur with a known constant mean rate, λ , and independently of the time since the last event.**
- For a Poisson distributed random variable X , the number of events k during a time interval t :
 - $P(X = k \text{ during } t) = e^{-\lambda \cdot t} \cdot \frac{(\lambda \cdot t)^k}{k!}$
 - $E[X \text{ during } t] = \lambda \cdot t$, i.e if t is expressed in seconds, λ is the average number of events per second.

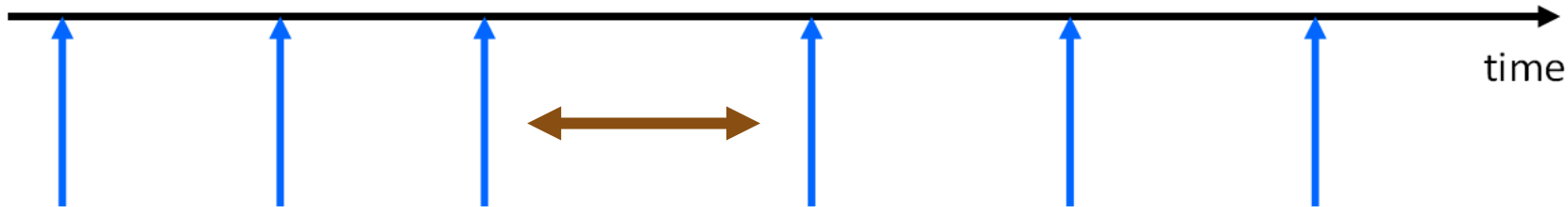
Poisson process

- The Poisson process is a stochastic process that is very useful in this course.
- In a stochastic process used in this course, an event occur randomly. **The time between each event (called the inter-arrival time) can be described with a probability distribution.**
- In a Poisson process, **the number of events in a specific time interval, t , follows a Poisson distribution with mean $\lambda \cdot t$.**
- Two important features for Poisson processes
 - The number of arrivals in each finite interval has a Poisson distribution;
 - The number of arrivals in disjoint intervals are independent random variables.
 - This implies that the events have the **memoryless property**: the existence of one event does not affect the probability (distribution) of other events.



(Negative) Exponential distribution

- The **exponential distribution** or **negative exponential distribution** is the probability distribution of the **inter-arrival time** between two events in a **Poisson process**.
- The exponential distribution is the continuous analogue of the geometric distribution, and it has the key property of being **memoryless**.
- If X is an exponentially distributed random variable:
 - $P(X \leq t) = F_X(t) = 1 - e^{-\alpha \cdot t}$
 - $E[X] = 1/\alpha$



Exercise: Web server system

We are designing a web server system and we start by analysing the arrivals of HTTP requests. Assume that we find that users send requests according to a Poisson process with mean 10 requests/second.

- (a) What is probability that we receive more than 2 requests during one second?
- (b) Assume that we have received a request. What is the probability that the next request arrives within 0.2 seconds?

Solution: Web server system

Poisson process with mean 10 requests per second.

Let X be the number of requests during time interval t .

$$P(X = k \text{ during } t) = e^{-10 \cdot t} \cdot \frac{(10 \cdot t)^k}{k!}$$

(a) $t = 1 \text{ second}$.

$$P(X = k; t = 1) = e^{-10} \cdot \frac{(10)^k}{k!}$$

$$P(X > 2; t = 1) = 1 - P(X \leq 2; t = 1) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

$$P(X = 0) = e^{-10} \cdot \frac{(10)^0}{0!} \approx 4.54 \cdot 10^{-5}; P(X = 1) = e^{-10} \cdot \frac{(10)^1}{1!} \approx 4.54 \cdot 10^{-4};$$

$$P(X = 2) = e^{-10} \cdot \frac{(10)^2}{2!} \approx 0.00227;$$

$$\Rightarrow \mathbf{P(X > 2; t = 1) \approx 0.997}$$

Solution: Web server system, continued

(b) For a Poisson process with mean λ requests per second, the inter-arrival time between two requests is exponentially distributed with mean $1/\lambda$.

Let Y be the time to the next request.

$$P(Y \leq 0.2 \text{ second}) = 1 - e^{-10 \cdot 0.2} \approx 0.86$$

Exam question 231025: MAC protocol

You are modelling a very simple MAC protocol. A device can send a data packet at any time.

- If a request is received correctly by the receiver, the receiver sends an ACK packet after exactly 1 second.
- If the device does not receive the ACK packet correctly, it retransmits the data packet after another 1 second.
- The cycle repeats until the device has received an ACK.
- The probability that any data or ACK packet are received correctly is 0.8.

What is the average time from when the device starts to send its data packet until it has received an ACK?

In your calculations, you can ignore all processing, transmission, and propagation delays. Also, the devices are perfectly synchronized.

Solution: MAC Protocol

A successful transmission = Data received correctly **AND** ACK received correctly.

We have to assume that a transmission of a Data packet and a transmission of an ACK are independent events (that is, the outcomes of the events do not depend on each other).

$$P(\text{Data received correctly}) = P(\text{ACK received correctly}) = 0.8$$

$$\begin{aligned} P(\text{a successful transmission}) &= P_S \\ &= P(\text{Data received correctly}) \cdot P(\text{ACK received correctly}) \end{aligned}$$

$$P(\text{a failed transmission}) = 1 - P_S$$

Let A be a random variable for how many transmission cycles that are needed for a successful transmission.

Solution: MAC Protocol, continued

If the transmission needs k transmission cycles, then there are $k-1$ failed transmission cycles, and the last cycle is successful. The outcomes of the cycles are independent of each other.

$$P(A = k) = P_S \cdot (1 - P_S)^{k-1} \text{ (geometric distribution)}$$

$$E[A] = 1/P_S \approx 1.56$$

Length of a failed transmission cycle = *2 seconds*.

Length of a successful transmission cycle = *1 second*.

Note that the transmission always requires one successful transmission cycle, that is, the average number of failed cycles is *0.56*.

$$E[\textit{Time for a successful transmission}] = 2 \cdot 0.56 + 1 \cdot 1 = 2.1 \textit{ seconds}.$$



LUND
UNIVERSITY