A First Course in Stochastic Network Calculus

Dipl.-Math. Michael Alexander Beck

October 30, 2013

Contents

Foreword

This *First Course in Stochastic Network Calculus* is a "First" course from two different perspectives. Obviously, one can see it as an introductory course to Stochastic Network Calculus. On the other side it was my first course I have given about SNC to a few students in June 2012. It builds on the lecture *Performance Modelling in Distributed Systems [5]* at the University of Kaiserslautern. The concepts of stochastic network calclulus parallel those of deterministic network calculus. This is why I refer to *[5]* several times to stress these connections. This document however is thought of as a stand-alone course and hence a deep study of [5] is not necessary (but recommendable).

This course contains a rather large probability primer to ensure that students can really grasp the expressions, which appear in stochastic network calculus. A student familiar with probability theory might skip this first chapter and delve directly into the stochastic network calculus. For each topic exercises are provided, which can (and should) be used to strengthen the understanding of the presented definitions and theory.

This document is still in progress and hopefully will evolve at some day into a fully grown course about Stochastic Network Calculus, providing a good overview over this exciting theory. Hence, please provide feedback to `beck@cs.uni-kl.de`.

- Michael Beck

1 Basics of Probability

In this chapter, we give a refresher about the basics of probability theory. We focus on material needed for constructing a stochastic network calculus. The following definitions and results serve as a foundation to build the calculus upon. The material presented here (and much more) is standard in probability theory and can be found - in much more detail - in many textbooks [3, 1, 4, 2].

1.1 Probability Spaces and Random Variables

We start our refresher with the very basic definitions needed to construct a probability space.

Definition 1.1. We call a non-empty set $\Omega \neq \emptyset$ *sample space*. A *σ-algebra* $\mathcal{A}$ on Ω is called *event space*. If $\omega \in A$ for some *event* $A \in \mathcal{A}$, we say that the outcome ω lies in the event A.

A σ-algebra (also called σ-field) of Ω is a collection of subsets of Ω, with the properties:

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \quad \Rightarrow \quad A^c \in \mathcal{A}$
- $A_1, A_2, A_3, \ldots \in \mathcal{A} \quad \Rightarrow \quad \{\bigcup_{i=1}^{\infty} A_i\} \in \mathcal{A}$

Example 1.2. As first example we consider a fair die. The possible outcomes of a die are the following:

$$\begin{aligned}\Omega = \{&\text{The die shows one pip, The die shows two pips,}\\ &\text{The die shows three pips , The die shows four pips,}\\ &\text{The die shows five pips, The die shows six pips}\}\end{aligned}$$

A simple σ-algebra would be just the *power set* of Ω, defined by:

$$\mathcal{A} = \text{Pot}(\Omega) = 2^{\Omega} = \mathcal{P}(\Omega) := \{A : A \subset \Omega\}$$

Applied to our example this would be:

$$\begin{aligned}\mathcal{A} = \{&\emptyset, \{\text{The die shows one pip}\}, \{\text{The die shows two pips}\}, \ldots,\\ &\{\text{The die shows one pip, The die shows two pips}\},\\ &\{\text{The die shows one pip, The die shows three pips}\}, \ldots\}\end{aligned}$$

Example 1.3. Given a subset $A \in \Omega$ we can always construct a σ-algebra by $\mathcal{A} := \{\emptyset, A, A^c, \Omega\}$. This is called the σ-algebra over Ω induced by A.

Definition 1.4. The pair $(\Omega, \mathcal{A})$ is called *measurable space.* A function $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$ with

- $\mu(\emptyset) = 0$
- $\mu(A) \geq 0 \quad \forall A \in \mathcal{A}$
- $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad \forall A_i \in \mathcal{A} \quad \text{with } A_i \cap A_j = \emptyset \text{for all } i \neq j$

is a *measure* on Ω. We name a triple $(\Omega, \mathcal{A}, \mu)$ *measure space.*

Definition 1.5. A measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ is a *probability measure* if $\mathbb{P}(\Omega) = 1$. The triple $(\Omega, \mathcal{A}, \mathbb{P})$ is a *probability space.*

Example 1.6. We continue our previous example. If the die is fair we assume all outcomes (i.e. how many pips the die shows) to be equally likely. Hence we give each outcome some "chunk" of probability:

$$\mathbb{P}(\{\text{The die shows } i \text{pips}\}) = \frac{1}{6} \quad \forall i = 1, 2, \ldots, 6 \tag{1.1}$$

We easily check that $\mathbb{P}(\Omega) = 1$. A sceptic person may ask, how to construct a probability measure from the above equation, since we do not know what the measure of an event like "the die shows one or five pips" is. The answer lies in the physical fact that the die can only show one side at a time. It cannot show one pip and five pips at the same time. Or in other words the outcomes in 1.1 are all disjoint. Hence we can construct by the properties of a measure all events by just adding the probabilities of all outcomes, which lie in that event. In this way the probability for the event "the die shows one or five pips" is - as expected - $\frac{2}{6}$.

This intuitive construction is important enough to be generalised in the next example.

Example 1.7. In the discrete case, i.e., if $\Omega = \{\omega_1, \omega_2, \ldots, \}$ is a countable set, we can define probability weights $p_i = \mathbb{P}(\{\omega_i\}) \geq 0$ with $\sum_{i \in \Omega} p_i = 1$. By defining these weights we construct a uniquely defined probability measure on $(\Omega, \mathcal{P}(\Omega))$ by:

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} p_i \delta_A(\omega_i)$$

with

$$\delta_A(\omega_i) = \begin{cases} 1 & \text{if } \omega_i \in A \\ 0 & \text{if } \omega_i \notin A \end{cases}$$

Example 1.8. If $\Omega = \mathbb{R}$ one usually uses the *Borel σ-algebra* denoted by $\mathcal{B}(\mathbb{R})$. This special σ-algebra is the smallest σ-algebra on $\mathbb{R}$, which contains all open intervals (i.e. we intersect all σ-algebras, which contain all open intervals. Since the power set of $\mathbb{R}$ contains all open intervals and is a σ-algebra, we know that this intersection is not empty.). On this measurable space $(\mathbb{R}, \mathcal{B})$ we can define probability measures by the usage of *probability density functions* (pds). A density $f : \mathbb{R} \to \mathbb{R}_0^+$ fulfills

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

and its corresponding probability measure is given by:

$$\mathbb{P}(A) = \int_A f(x)dx$$

There is a good reason why we use the non-intuitive Borel σ-algebra, instead of $\mathcal{P}(\mathbb{R})$. One can show, that the power set is "too large" to define a measure on it. This means, if one tries to define a non-trivial measure (a measure is trivial if $\mu(A) = 0$ for all $A \in \mathcal{A}$) on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ this leads inevitably to a contradiction (so called Vitali sets).

We have seen in our die example, that the space of outcomes can be something physical. However, we would like to translate this physical descriptions into something more tractable, like real numbers. The next very important definition allows us to do so.

Definition 1.9. A function $X : \Omega \to \Omega'$ between two measurable spaces $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ is *measurable* if

$$\{X \in A'\} \in \mathcal{A} \quad \forall A' \in \mathcal{A}'$$

holds ($\{X \in A'\} := \{\omega \in \Omega : X(\omega) \in A'\}$).

A *random variable* is a measurable function from a probability space into a measurable space, i.e. $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\Omega', \mathcal{A}')$. A random variable X induces a probability measure on $(\Omega', \mathcal{A}')$ by:

$$\mathbb{P}_X(A') := \mathbb{P}(\{X \in A'\}) =: \mathbb{P}(X \in A')$$

The probability measure $\mathbb{P}_X$ is called *distribution* of X.

Often one just speaks of the distribution of X and assumes a corresponding probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to exist. In most cases this creates no problems.

Example 1.10. If we return again to the die example, an intuitive real random variable $X : \Omega \to \mathbb{R}$ would be the following:

$$\begin{aligned}
X(\text{The die shows one pip}) &= 1 \\
X(\text{The die shows two pips}) &= 2 \\
X(\text{The die shows three pips}) &= 3 \\
X(\text{The die shows four pips}) &= 4 \\
X(\text{The die shows five pips}) &= 5 \\
X(\text{The die shows six pips}) &= 6
\end{aligned}$$

Definition 1.11. For a real random variable X we define the *(cumulative) distribution function (cdf)* by:

$$\begin{aligned} F_X :&\mathbb{R} \to [0,1] \\ &x \mapsto F(x) := \mathbb{P}(X \leq x) \end{aligned}$$

By its distribution function F_X the random variable X is uniquely determined.

We can see now, why in most cases we are not interested in the original probability space. If we have a real-valued random variable with some distribution function F we can construct the original probability space as follows: Use $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{B}(\mathbb{R})$ and define $\mathbb{P}$ by:

$$\mathbb{P}((-\infty, x]) = F(x)$$

Example 1.12. We say a random variable is exponentially distributed with parameter λ, if it has the following distribution function:

$$F(x, \lambda) = 1 - e^{-\lambda x}$$

The corresponding density $f(x, \lambda)$ is given by differentiating F:

$$F'(x) = \lambda e^{-\lambda x}$$

Example 1.13. A random variable is normally distributed with parameters μ and σ^2, if it has the following distribution function:

$$F(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

Example 1.14. A random variable is Pareto distributed with parameters x_m and α, if it has the following distribution function:

$$F(x, \alpha, x_m) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m \\ 0 & x < x_m \end{cases}$$

Its density is given by:

$$f(x, \alpha, x_m) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \qquad x \geq x_m$$

Exercises

Exercise 1.15. Show that the function $\mathbb{P}$ defined in example 1.7 is a probability measure, i.e. check all the conditions for $\mathbb{P}$ to be a measure, as given in definitions 1.4 and 1.5. Show further that the probability measure is uniquely defined, i.e. assume another function $\mathbb{P}'$ with $\mathbb{P}(\{\omega_i\}) = \mathbb{P}'(\{\omega_i\})$ for all $i \in \mathbb{N}$ and show:

$$\mathbb{P}'(A) = \mathbb{P}(A) \qquad \forall\, A \in \mathcal{A}$$

Exercise 1.16. Let $\mathcal{A}$ and $\mathcal{B}$ be two different σ-algebras on the same space Ω. Show that the intersection of the two σ-algebras $\mathcal{A} \cap \mathcal{B}$ is again a σ-algebra on Ω.

Exercise 1.17. Let X be a real random variable with differentiable density f. Show that the distribution of X is an antiderivative (german: "Stammfunktion") of f, i.e.:

$$\frac{d}{dx}F(x) = f(x)$$

(Hint: Use the Fundamental Theorem of Calculus)

Exercise 1.18. Let $(\Omega, \mathcal{A})$ be a measurable space and $A_1, A_2, A_3, \ldots \in \mathcal{A}$. Show

$$\{\omega : \omega \in A_i \text{ for finite many } i \in \mathbb{N}\} \in \mathcal{A}$$

and

$$\{\omega : \omega \in A_i \text{ for infinite many } i \in \mathbb{N}\} \in \mathcal{A}.$$

Exercise 1.19. Let $(\Omega, \mathcal{A})$ be a measurable space and $A, B \subset \mathcal{A}$. Prove in this sequence:

- $\mathbb{P}(A) \geq \mathbb{P}(B)$ if $A \supseteq B$
- $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$

Exercise 1.20. Many programming languages offer methods to create random numbers in $[0, 1]$. How can you use this to create realisations of an exponentially distributed random variable with parameter λ? Use this method to simulate n distinct exponentially distributed random variables X_i with the same parameter λ and compute the arithmetic mean of these realisations $E = \frac{1}{n}\sum_{i=1}^{n} X_i$. How does $E - \frac{1}{\lambda}$ evolve for large n?

Exercise 1.21. Use the procedure of the previous exercise to generate exponentially distributed random numbers $x_1, x_2, x_3, \ldots$ with parameter λ. Now count the occurences of these random numbers in certain intervals. Define these intervals of length $N \in \mathbb{R}$ by:

$$I_n = [n \cdot N, (n+1) \cdot N) \subset \mathbb{R} \qquad n \in \mathbb{N}_0$$

The number of realisations in one interval is then given by:

$$f^*(n) = \sum_{i=1} \delta_{I_n}(x_i)$$

Plot the sequence $(f^*(n))_{n\in\mathbb{N}}$. Compare this sequence with the density of an exponentially distributed random variable X with the same parameter λ.

Exercise 1.22. A geometrically distributed random variable X is a discrete random variable with probability weights:

$$\mathbb{P}(X = n) = (1-p)^n p \qquad n \in \mathbb{N}_0$$

for some parameter $p \in (0, 1)$. Think about a method to generate geometrically distributed random numbers.

(Hint: $\mathbb{P}(X = n) = (1-p)^n p = (1-p) \cdot (1-p)^{n-1} p = (1-p) \cdot \mathbb{P}(X = n-1)$)

Generate some geometrically distributed random variables and plot the result.

Take the sequence $(f^*(n))_{n\in\mathbb{N}}$ of the previous exercise, with parameter $\lambda = -\log(1-p)$ and intervals $I_n = [n, n+1)$ and compare it with the plot of geometrically distributed random variables.

Exercise 1.23. Prove that for X being an exponentially distributed random variable with parameter $\lambda = -\log(1-p)$ it holds that

$$\mathbb{P}(X \in [n, n+1)) = \mathbb{P}(Y = n)$$

where Y is a geometrically distributed random variable with parameter p.

1.2 Moments and Stochastic Independence

Definition 1.24. Let X, Y be real random variables with densities. The *expectaction* of X is given by:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

where f is the density of X. Further is

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx$$

For the special case of $h(x) = |x|^n$ we talk about the *n-th moment* of X:

$$\mathbb{E}(|X|^n) = \int_{-\infty}^{\infty} |x|^n f(x) dx$$

The *variance* of X is given by:

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

The *covariance* of X and Y is given by:

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

Example 1.25. The expectation of an exponential random variable with parameter λ can be calculated by:

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty x \cdot \lambda e^{-\lambda x} dx = \lim_{a\to\infty} (-ae^{-\lambda a} - (-0 \cdot e^{-\lambda \cdot 0})) - \int_0^\infty -e^{-\lambda x} dx \\ &= 0 - \frac{1}{\lambda} \lim_{a\to\infty} (e^{-\lambda a} - e^{-\lambda \cdot 0}) = \frac{1}{\lambda}\end{aligned}$$

Similarly one has for the variance $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

Example 1.26. The expectation and variance of a normal distributed random variable with parameters μ and σ^2 are $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

Example 1.27. The expectation of the Pareto distribution with parameters x_m and α is given by:

$$\begin{aligned}\mathbb{E}(X) &= \int_{x_m}^\infty x \cdot \frac{\alpha x_m^\alpha}{x^{\alpha+1}} dx = \alpha x_m^\alpha \int_{x_m}^\infty x^{-\alpha} dx \\ &= \alpha x_m^\alpha \lim_{a\to\infty} \frac{1}{1-\alpha} a^{-\alpha+1} - \frac{1}{1-\alpha} x_m^{-\alpha+1} \\ &= -\alpha x_m^\alpha \cdot \frac{1}{1-\alpha} \cdot \frac{1}{x_m^{\alpha-1}} = \frac{\alpha x_m}{\alpha - 1}\end{aligned}$$

if $\alpha > 1$. For the cases that $\alpha \leq 1$ the integral $\int_{x_m}^\infty x^{-\alpha} dx = \infty$ does not converge. In this case, we say that the expectation of X does not exist. Analogous one obtains for $\alpha \leq 2$ that the variance of X (or otherwise stated the second moment) does not exist.

From the definition of the expectation as an integral, we know immediately that the expectation is linear and monotone:

$$\mathbb{E}(aX + Y) = a\mathbb{E}(X) + \mathbb{E}(Y) \qquad \forall\, a \in \mathbb{R}$$

$$\mathbb{E}(X) \geq \mathbb{E}(Y) \qquad \text{if } X \geq Y \text{ a.s.}$$

(The expression "a.s." is the abbreviation for almost surely and means that some event occurs with probability one, i.e. $\mathbb{P}(X \geq Y) = 1$). With a bit more effort (which we do not invest here) one can proof for a sequence of non-negative random variables X_n that even

$$\mathbb{E}\left(\sum_{n=1}^\infty X_n\right) = \sum_{n=1}^\infty \mathbb{E}(X_n) \tag{1.1}$$

holds.

The variance is also called the centered second moment, it can be seen as the expected squared deviation from the expectation. Also it can be reformulated by:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(\mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}$$

Further we know that the variance fulfills:

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

To calculate the covariance we need the two-dimensional density of the random vector (X, Y):

$$f(x, y) \geq 0 \qquad \forall\, (x, y) \in \mathbb{R}^2$$
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx\, dy = 1$$

We can define the marginal densities f_X and f_Y of the single random variables X and Y. They can be computed by:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

The covariance can then be expressed by:

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)(x - \mathbb{E}(X))(y - \mathbb{E}(Y)) dx\, dy$$

Further we have:

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X) \tag{1.2}$$
$$\mathrm{Cov}(X, X) = \mathrm{Var}(X) \tag{1.3}$$
$$\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y) \tag{1.4}$$
$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \tag{1.5}$$
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y) \tag{1.6}$$

If $\mathrm{Cov}(X, Y) = 0$ we say the two random variables X and Y are uncorrelated.

Definition 1.28. Two events $A, B \in \mathcal{A}$ are *stochastically independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n \subset \mathcal{A}$. The sets $\mathcal{A}_i$ are *stochastically independent* if for each selection $\mathcal{A}_{i_1}, \mathcal{A}_{i_2}, \ldots, \mathcal{A}_{i_k}$ of the $\mathcal{A}_i$ and all subsets $A_{i_j} \in \mathcal{A}_{i_j}$ it holds that:

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_k})$$

The second definition is stronger than the *pairwise stochastic independence*, which is given by $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$ and a family of sets $A_i \in \mathcal{A}$ and also stronger than the demand $\mathbb{P}(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i)$. The following simple example makes this clear.

Assume one throws two dice. We define three events:

$$\begin{aligned} A &= \{\text{The first die shows an even number of pips}\} \\ B &= \{\text{The second die shows an even number of pips}\} \\ C &= \{\text{The sum of pips shown by both dice is even}\} \end{aligned}$$

We can easily check that $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ and that $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$. But this pairwise stochastic independence does not infer stochastic independence of the familiy $\mathcal{D} = \{A, B, C\}$ since:

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

Definition 1.29. Let $X_i : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ be a set of random variables $(i = 1, \ldots, n)$. They are *stochastically independent*, if their pre-image-σ-algebras are stochastically independent sets. In expression, if for every selection $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ and all $A'_{i_j} \in \mathcal{A}'$ $(j = 1, \ldots, k)$ holds:

$$\mathbb{P}\left(\bigcap_{j=1}^{k} \{X_{i_j} \in A'_{i_j}\}\right) = \prod_{j=1}^{k} \mathbb{P}(X_{i_j} \in A'_{i_j})$$

Let X_i be a set of stochastically independent random variables $(i = 1, \ldots, n)$. Denote by $X = (X_1, \ldots, X_n)$ the corresponding random vector, by F_X its cdf and by f_X its density (if it exists). Then it holds that:

- $F_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \qquad \forall\, x_i \in \mathbb{R}$
- $f_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \qquad \forall\, x_i \in \mathbb{R}$
- $\mathbb{E}(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} \mathbb{E}(X_i)$ and $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$[1]

In the modelling of stochastic behaviour it is more common to *assume* the independence of events, instead of *proving* that something is stochastically independent. Usually, when one assumes that two outcomes have no effect on each other, this is modelled via stochastic independence. One example would be the (simultaneous) throwing of two fair dice. You can model each die as one random variable, which are stochastically independent of each other. Alternatively you could model both dice by just one random variable with values in $\{1, \ldots, 6\} \times \{1, \ldots, 6\}$ and show that all events concerning only one die (e.g. the first die shows one pip) are stochastically independent from the events, which concern only the other die (see the upcoming example for this).

Definition 1.30. Let A, B be two events in the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathbb{P}(B) > 0$. The *conditional probability* of A given B is defined by:

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

[1] Watch out, that from $\mathrm{Cov}(X, Y) = 0$ does not necessarily imply stochastic independence between X and Y! See exercise 1.41

The conditional probability of some event A given B is the probability that A occurs, if we already know, that B will occur (or has occured).

Example 1.31. We consider the aforementioned two dice experiment. We model the throwing of two dice as a random variable X, mapping from a suitable probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into the measurable space $(\Omega' := \{1, \ldots, 6\}^2, \mathcal{P}(\Omega'))$. Since the dice are fair, we assume each outcome in Ω' to be equally likely, i.e. each pair $(i, j) \in \Omega'$ has probability $\mathbb{P}_X((i,j)) = \frac{1}{36}$.

We investigate two events. The first event denoted by C, is that the first die shows at least 5 pips. We calculate:

$$\mathbb{P}_X(C) = \mathbb{P}_X(\{5,6\} \times \{1, \ldots, 6\}) = \frac{12}{36} = \frac{1}{3}$$

Next we investigate the event A, which is that the sum of both dice is at least 5 pips, if we already know, that the first die shows an even number of pips. For this, we denote by B the event, that the first die shows an even number of pips. The probability of A is given now, by the conditional probability:

$$\mathbb{P}_X(A|B) = \frac{\mathbb{P}_X(A \cap B)}{\mathbb{P}_X(B)} = \frac{\mathbb{P}_X(\{(2,j)|j \in \{3, \ldots, 6\}\} \cup \{4,6\} \times \{1, \ldots, 6\})}{\mathbb{P}_X(\{2,4,6\} \times \{1, \ldots, 6\})} = \frac{\frac{16}{36}}{\frac{1}{2}} = \frac{8}{9}$$

Example 1.32. Another typical example of dependence is sampling without replacement. Imagine an urn with colored balls in it. Someone draws randomly the balls from the urn, one by one and we might ask, what is the probability that the n-th ball has a certain color. For our example imagine the urn is filled with two blue and one green ball. We ask for the probability that the second ball drawn will be the green one, this event will be denoted by G_2. To model this we need to find a suitable event space Ω first. If we number the blue balls by b_1 and b_2 as well as the green ball by g, we see that the following event space describes all possibilites to draw the balls from the urn:

$$\Omega = \{(b_1, b_2, g),\ (b_2, b_1, g),\ (b_1, g, b_2),\ (b_2, g, b_1),\ (g, b_1, b_2),\ (g, b_2, b_1)\}$$

Since the balls are drawn completely random, we can assume each of the outcomes has the same probability $\frac{1}{6}$, hence we have $\mathbb{P}(G_2) = \frac{1}{3}$. If we now condition on the event, that the first ball is a blue one and want to know again the probability that the second ball is the green one, we have (denoting by B_1 the event of first drawing a blue ball):

$$\mathbb{P}(G_2|B_1) = \frac{\mathbb{P}(G_2 \cap B_1)}{\mathbb{P}(B_1)} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

Definition 1.33. The *moment generating function (MGF)* of a real random variable X is given by:

$$\begin{aligned} \phi_X &: \mathbb{R} \to \mathbb{R} \\ \phi_X &: \theta \mapsto \phi_X(\theta) := \mathbb{E}(e^{\theta X}) \end{aligned}$$

If $\mathbb{E}(e^{\theta X}) = \infty$ we say that the MGF of X does not exist at θ.

Corollary 1.34. *Let X_i be a family of stochastically independent, identically distributed (i.i.d.) variables ($i = 1, \dots, n$) and define $S_n = \sum_{i=1}^n X_i$. Assume $\phi_{X_i}(\theta)$ exists, then it holds that:*

$$\phi_{S_n}(\theta) = (\phi_{X_i}(\theta))^n$$

Proof. We have:

$$\begin{aligned}\mathbb{E}(e^{\theta S_n}) = \mathbb{E}(e^{\theta \sum_{i=1}^n X_i}) &= \mathbb{E}\left(\prod_{i=1}^n e^{\theta X_i}\right) \\ &= \prod_{i=1}^n \mathbb{E}(e^{\theta X_i}) = (\phi_{X_i}(\theta))^n\end{aligned}$$

□

Example 1.35. Let X_i be i.i.d. exponentially distributed random variables with parameter λ. Define the random variable $S_n = \sum_{i=1}^n X_i$. We say that such a random variable is Erlang distributed with parameters n and λ. Its MGF is given by:

$$\mathbb{E}(e^{\theta S_n}) = (\mathbb{E}(e^{\theta X_i}))^n = \left(\lambda \int_0^\infty e^{(\theta-\lambda)x} dx\right)^n$$

Let now $\theta < \lambda$:

$$\begin{aligned}\mathbb{E}(e^{\theta S_n}) &= \left(\lambda \lim_{a\to\infty} \frac{1}{\theta - \lambda}(e^{(\theta-\lambda)a} - e^{(\theta-\lambda)\cdot 0})\right)^n \\ &= \left(\frac{\lambda}{\lambda - \theta}\right)^n\end{aligned}$$

We see, that for $\theta \geq \lambda$ the MGF does not exist.

Example 1.36. Let X be a Pareto distributed random variable with parameters x_m and α. You are asked in the exercises to show that the MGF of X does not exist for any $\theta > 0$, no matter how the parameters x_m and α are chosen.

We state here some useful properties of the moment generating function, without explicitly proving them:

- If two random variables have the same existing MGF, they also have the same distribution.
- For non-negative random variables and varying θ the MGF exists in an interval $(-\infty, a)$ with $a \in \mathbb{R}_0^+$.
- The MGF always exists for $\theta = 0$ and if it exists in the interval $(-a, a)$ it is infinitely often differentiable on that interval.
- The MGF is convex.

Exercises

Exercise 1.37. Prove equations (1.1)-(1.4).

Exercise 1.38. What is the n-th moment of a Pareto distributed random variable and when does it exist?

Exercise 1.39. Prove that the variance of an exponentially distributed random variable is equal to $\frac{1}{\lambda^2}$.

Exercise 1.40. Construct an example, in which for three events $A, B, C \subset \mathcal{A}$ it holds that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, but at least one of the following equalities *does not* hold:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$
$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$$
$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$$

(Hint: Draw three intersecting circles and give each area a certain probability weight, such that the weights sum up to 1.)

Exercise 1.41. Let X and Y be independent and Bernoulli distributed with the same parameter p (this means $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$). Show that the two new random variables $X + Y$ and $X - Y$ are uncorrelated, yet not independent.

Exercise 1.42. Let X be some real random variable, such that the MGF of X exists in an interval around 0. Calculate the first derivative of the MGF at zero $\frac{d}{d\theta}\phi_X(\theta)\big|_{\theta=0}$. Calculate the n-th derivative of the MGF at zero $\left(\frac{d}{d\theta}\right)^n \phi_X(\theta)\big|_{\theta=0}$. (Hint: Use that $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ and use (1.1).)

Exercise 1.43. In this exercise prove first the law of total probability: If $A \in \Omega$ and $B \in \Omega$ are some events in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ it holds:

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

Convince yourself that for some partition $(B_n)_{n \in \mathbb{N}}$ of Ω we similarly have:

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

(A partition fulfills: $\bigcup_{n=1}^{\infty} B_n = \Omega$ and $B_n \cap B_m = \emptyset$ for all $n \neq m$)

Now if we have two real random variables X and Y with densities f_X and f_Y, we can reformulate the law of total probability:

$$\mathbb{P}(X \in A) = \int_{-\infty}^{\infty} \mathbb{P}(X \in A|Y = y) f_Y(y) dy$$

Assume now that X is an exponentially distributed random variable with parameter λ and Y is stochastically independent of X. Show that the probability that Y is smaller than X is equal to the MGF of Y at the point λ, in expression:

$$\mathbb{P}(X > Y) = \mathbb{E}(e^{-\lambda Y})$$

(Hint: Use that $\mathbb{E}(h(Y)) = \int_{-\infty}^{\infty} h(y) f_X(y) dy$)

Exercise 1.44. Show that for the Pareto distribution and every $\theta > 0$ the corresponding MGF does not exist.

Exercise 1.45. The log-normal distribution is defined by the following density:

$$f(x, \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\left(\frac{(\ln x - \mu)^2}{2\sigma^2}\right)} \qquad \forall\, x > 0$$

Its moments are given by $\mathbb{E}(X^n) = e^{n\mu + \frac{1}{2}\sigma^2 n^2}$. Show that for all $\theta > 0$ the corresponding MGF does not exist.

Exercise 1.46. This exercise continues exercises 1.20 and 1.21 in which we have seen, that the arithmetic mean converges to the expectation. Simulate again n exponentially distributed numbers and compute $V = \frac{1}{n}\sum_{i=1}^{n}(x - \frac{1}{\lambda})^2$. How does $V - \frac{1}{\lambda^2}$ evolve for large n? Plot $V - \frac{1}{\lambda^2}$ against n for different values of λ. When converges $V - \frac{1}{\lambda^2}$ fast and when slow?

Bibliography

[1] *H. Bauer.* Probability Theory*. De Gruyter, 1995.*

[2] *W. Feller.* An introduction to probability theory and its applications*, volume Vol I. Wiley and Sons, 1968.*

[3] *J. Jacod and P. Protter.* Probability essentials*. Springer, 2003.*

[4] *U. Krengel.* Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, 2003.*

[5] *J. B. Schmitt. Lecture notes: Performance modeling of distributed systems. http://disco.informatik.uni-kl.de/content/PDS_ WS1112, October 2011.*