
Elementary Probability

Exam Number 38119

1. INTRODUCTION

Consider any experiment whose result is unknown, for example throwing a coin, the daily number of customers in a supermarket or the duration of a phone call in a service office. Each of these experiments has a more or less wide variety of possible results. The set of all these results is called **result space** and denoted Ω . In the examples above we have $\Omega_1 = \{\text{head, number}\}$, $\Omega_2 = \mathbb{N}$ and $\Omega_3 = (0, \infty)$. We cannot forecast for certain, which result the experiment will have, but we can tell something about the probability of certain results $\omega \in \Omega$. Often we are not interested in single results but in subsets $A \subseteq \Omega$ containing several results.

We denote the set containing all “interesting” events $A \subseteq \Omega$ as \mathfrak{A} . The A are called **events**.

Definition 1.1. A set $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ is called **σ -algebra** over Ω if

- (1) $\Omega \in \mathfrak{A}$,
- (2) $A^c \in \mathfrak{A}$ for all $A \in \mathfrak{A}$ and
- (3) $\bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{A}$ for all $(A_n)_{n \in \mathbb{N}} \in \mathfrak{A}^{\mathbb{N}}$.

For $\mathfrak{E} \subseteq \mathfrak{P}(\Omega)$ let $\mathfrak{A}(\mathfrak{E})$ the smallest σ -algebra containing \mathfrak{E} , that is

$$\mathfrak{A}(\mathfrak{E}) = \bigcap_{\{\mathfrak{A} | \mathfrak{A} \text{ is } \sigma\text{-algebra and } \mathfrak{A} \supseteq \mathfrak{E}\}} \mathfrak{A}.$$

In Theorem 1.5 we will see that it is not always possible to choose $\mathfrak{A} = \mathfrak{P}(\Omega)$. For countable Ω we will usually choose $\mathfrak{A} = \mathfrak{P}(\Omega)$, but for uncountable $\Omega \subseteq \mathbb{R}^1$ we need the **Borel σ -algebra**

$$\mathfrak{B}^1 = \mathfrak{A}(\{(a, b] \mid a, b \in \mathbb{R}^1\})$$

or some sub- σ -algebra of it. For $\Omega \subseteq \mathbb{R}^n$ we have the n dimensional Borel σ -algebra

$$\mathfrak{B}^n = \mathfrak{A}(\{\times_{i=1}^n B_i \mid B_i \in \mathfrak{B}\}).$$

On \mathfrak{A} we can now define a probability distribution P . We want P to be “realistic”; therefore we require some basic properties for it.

Definition 1.2. A function $P : \mathfrak{A} \mapsto [0, 1]$ is called **probability distribution** or **probability measure** for \mathfrak{A} if

- (1) $P(\Omega) = 1$ and
- (2) $P(\sum_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$ for all pairwise disjoint $(A_n)_{n \in \mathbb{N}} \in \mathfrak{A}^{\mathbb{N}}$.

Definition 1.3. The triplet $(\Omega, \mathfrak{A}, P)$ where \mathfrak{A} is a σ -algebra over Ω and P is a probability distribution for \mathfrak{A} is called **probability space**.

We will show some basic properties of probability distributions in the following

Lemma 1.4. *Let $A, B \in \mathfrak{A}$ and $(A_n)_{n \in \mathbb{N}} \in \mathfrak{A}^{\mathbb{N}}$. Then we have*

- (1) $P(A^c) = 1 - P(A)$,
- (2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

- (3) $P(A) \leq P(B)$ if $A \subseteq B$,
- (4) $P(A \setminus B) = P(B) - P(A)$ if $A \subseteq B$ and
- (5) $P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$ if $(A_n)_{n \in \mathbb{N}}$ is isotonic or antitonic.

Proof. The properties (1)–(4) follow from the definition of a probability distribution. To show (5) we assume first that $(A_n)_{n \in \mathbb{N}}$ is isotonic. Define $A_0 = \emptyset$ and $B_n = A_n \setminus A_{n-1}$ for $n \in \mathbb{N}$. Then the B_n are pairwise disjoint and we have

$$\bigcup_{n \in \mathbb{N}} A_n = \sum_{n \in \mathbb{N}} B_n.$$

Therefore we have

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P\left(\sum_{n \in \mathbb{N}} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \\ &= \lim_{k \rightarrow \infty} \sum_{i=1}^k P(A_i \setminus A_{i-1}) = \lim_{k \rightarrow \infty} \sum_{n=1}^k (P(A_n) - P(A_{n-1})) = \lim_{k \rightarrow \infty} P(A_k). \end{aligned}$$

Now let $(A_n)_{n \in \mathbb{N}}$ be antitonic. Then $(A_n^c)_{n \in \mathbb{N}}$ is isotonic and we have

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = 1 - P\left(\bigcap_{n \in \mathbb{N}} A_n^c\right) = \\ &= 1 - \lim_{n \rightarrow \infty} P(A_n^c) = \lim_{n \rightarrow \infty} P(1 - A_n^c) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

□

Theorem 1.5. *There exists no function $P : \mathfrak{P}([0, 1]) \rightarrow [0, 1]$ such that*

- (1) $P([a, b]) = b - a$ for $0 \leq a \leq b \leq 1$,
- (2) P is translation invariant, that is $P(A + x) = P(A)$ for all $A \in \mathfrak{P}([0, 1])$, $x \in [0, 1]$ such that $A + x \in \mathfrak{P}([0, 1])$ and
- (3) $P(\sum_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$ for all pairwise disjoint $(A_n)_{n \in \mathbb{N}} \in \mathfrak{P}([0, 1])^{\mathbb{N}}$.

Proof. Suppose this function P does exist. Consider the relation \sim on $[1/3, 2/3]$ with

$$x \sim y \iff x - y \in \mathbb{Q}.$$

We have $x \sim x$ for all $x \in [0, 1]$ as $x - x = 0 \in \mathbb{Q}$. Therefore \sim is reflexive. The relation \sim is symmetric as $y - x = -(x - y) \in \mathbb{Q}$ whenever $x - y \in \mathbb{Q}$. It is also transitive as from $x - y \in \mathbb{Q}$ and $y - z \in \mathbb{Q}$ follows $x - z = (x - y) + (y - z) \in \mathbb{Q}$. Hence \sim is an equivalence relation. According to Zorn's Lemma there exists a set $A \in \mathfrak{P}([0, 1])$ containing exactly one element of each equivalence class.

Let the sequence $(q_n)_{n \in \mathbb{N}}$ count through the rational numbers in the interval $[-1/3, 1/3]$ and let

$$A_n = A + q_n$$

for all $n \in \mathbb{N}$. We now show that the A_n are pairwise disjoint. Let

$$a \in A_k \cap A_l$$

for some $k, l \in \mathbb{N}$. This means that both $a \sim q_k$ and $a \sim q_l$. Since \sim is transitive, we also have $q_k \sim q_l$ and therefore $A_k = A_l$.

Let $x \in [2/5, 3/5]$. Then can we find some $a \in A$ such that $x \sim a$. Hence we find some $m \in \mathbb{N}$ such that $x - a = q_m$. It follows that $x \in A + q_m$. Therefore we have

$$[2/5, 3/5] \subseteq \sum_{n \in \mathbb{N}} A_n \subseteq [0, 1].$$

It follows

$$1 \geq P\left(\sum_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n) = \sum_{n \in \mathbb{N}} P(A),$$

which means $P(A) = 0$. This leads to the contradiction

$$1/5 \leq P\left(\sum_{n \in \mathbb{N}} A_n\right) = 0.$$

□

2. EXAMPLES FOR PROBABILITY SPACES

In this section we will examine some examples for probability spaces. First we will look at discrete probability distributions, that is Ω is countable. In this case we can define $p_\omega = P(\omega) = P(\{\omega\})$ for all $\omega \in \Omega$; it is called **density**, in this special case of a countable result space we call it **discrete density function**. The p_ω are sufficient to describe a distribution, since for every $A \in \mathfrak{A}$ we have $P(A) = \sum_{\omega \in A} p_\omega$.

- The most simple distribution for finite Ω and $\mathfrak{A} = \mathfrak{P}(\Omega)$ is given by $P(A) = |A| / |\Omega|$ for all $A \in \mathfrak{A}$. The fact that P is a distribution follows directly from the definition. It is called **Laplace distribution**.
- Let $p \in [0, 1]$, $\Omega = \{1, \dots, n\}$, $\mathfrak{A} = \mathfrak{P}(\Omega)$ and $p_i = \binom{n}{i} p^i (1-p)^{n-i}$ for $1 \leq i \leq n$. As

$$\sum_{i=0}^n p_i = (p + (1-p))^n = 1,$$

we have that $(p_i)_{i \in \Omega}$ is a density. Therefore $(\Omega, \mathfrak{A}, P)$ is a probability space where P is the corresponding distribution. This distribution is called **binomial distribution** and denoted $b(n, p)$.

- For $p \in (0, 1]$, $\Omega = \mathbb{N}$ and $\mathfrak{A} = \mathfrak{P}(\Omega)$ we can define a density by $p_i = p(1-p)^{i-1}$ for $i \in \mathbb{N}$ since

$$\sum_{i=1}^{\infty} p_i = \frac{p}{1-p} \sum_{i=1}^{\infty} (1-p)^i = 1.$$

The corresponding distribution is called **geometrical distribution** and denoted $geo(p)$.

- Let a be a positive real number, $\Omega = \mathbb{N}_0$, $\mathfrak{A} = \mathfrak{P}(\Omega)$ and $p_i = e^{-a} a^i / i!$ for $i \in \mathbb{N}_0$. We have

$$\sum_{i=0}^n p_i = e^{-1} \sum_{i=0}^n \frac{a^i}{i!} = 1.$$

Therefore the p_i form a density. The corresponding distribution is called **Poisson distribution** and denoted $Poi(a)$.

Now we will give some examples for continuous distributions. In this case we have $\Omega = \mathbb{R}^1$ and $\mathfrak{A} = \mathfrak{B}^1$. Given a Riemann-integrable function $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ with $\int_{-\infty}^{\infty} f(x) dx = 1$ we define $P(A) = \int_A f(x) dx$, which is therefore a distribution. The function f is called its **density** or **probability density function**.

A useful tool to describe continuous densities is the **characteristic function**

$$\mathbf{1}_A : \Omega \rightarrow \mathbb{R}^1, \omega \mapsto \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

for some $A \in \mathfrak{A}$.

- The function $f = \mathbf{1}_{[a,b]} / (b - a)$, where $a < b$, defines a continuous density since

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{b - a} \int_a^b 1 dx = 1.$$

Its corresponding distribution is called **uniform distribution** and denoted $U[a, b]$.

- Let the function f be defined as $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty)}(x)$ for some positive real λ . Since

$$\int_{-\infty}^{\infty} f(x) dx = \lambda \int_0^{\infty} e^{-\lambda x} dx = 1,$$

we have a density. Its corresponding distribution is called **exponential distribution** and denoted $exp(\lambda)$.

- The **normal distribution** is denoted $N(\mu, \sigma^2)$ and has the density $f(x) = (\sqrt{2\pi}\sigma)^{-1} e^{-(x-\mu)^2/2\sigma^2}$. For $\int_{-\infty}^{\infty} f(x) dx = 1$ see [Sch].

More examples for probability distributions can be found in [Sch] and in appendix A.

Both discrete and continuous distributions are special cases of a more general case. A basic knowledge of measure theory, which can be found in [Bau], is needed to define the general probability distribution.

Definition 2.1. Let Ω be an arbitrary set with a σ -algebra \mathfrak{A} . Let μ be a measure on \mathfrak{A} . A nonnegative μ -integrable function f with $\int f d\mu = 1$ is called μ -**density**

of the probability measure P with

$$P(A) = \int_A f d\mu. \quad (2.1)$$

The Theorem of Radon-Nikodym (see [Bau]) yields for continuous distributions $f = dP/d\lambda^1$, where λ^1 is the one-dimensional Lebesgue measure, and for discrete distributions $f = dP/d\mu$, where μ is the counting measure on Ω . Inserting this into (2.1) gives us

$$P(A) = \int_A dP = \int \mathbf{1}_A dP.$$

3. RANDOM VARIABLES AND RANDOM VECTORS

In an experiment the complete result is very often not important or not interesting. For example, when throwing n coins we might not necessarily be interested in the sequence of heads and numbers $\omega = (\omega_1, \dots, \omega_n)$ but in the number of heads $X(\omega) = |\{\omega_i \mid \omega_i \text{ is head}\}|$ or in the average of throwing a head $\bar{X}(\omega) = X(\omega)/n$.

Definition 3.1. Let $(\Omega_1, \mathfrak{A}_1, P)$ be a probability space and \mathfrak{A}_2 a σ -algebra over the set Ω_2 . A function $X : \Omega_1 \rightarrow \Omega_2$ is called **random variable** if

$$X^{-1}(A) \in \mathfrak{A}_1$$

for all $A \in \mathfrak{A}_2$. If X is a random variable we write

$$X : (\Omega_1, \mathfrak{A}_1) \rightarrow (\Omega_2, \mathfrak{A}_2).$$

If $\Omega_2 \subseteq \mathbb{R}^n$ with $n \geq 2$, then X is also called **random vector**. For $A \in \mathfrak{A}_2$ we define

$$P^X(A) = P(X^{-1}(A)) = P(X \in A) = P(\{\omega \in \Omega_1 \mid X(\omega) \in A\}).$$

Lemma 3.2. *The function P^X is a probability distribution on \mathfrak{A}_2 . This gives us the new probability space $(\Omega_2, \mathfrak{A}_2, P^X)$.*

Proof. We have $P^X(\Omega_2) = P(X^{-1}(\Omega_2)) = P(\Omega_1) = 1$, which is the first condition of a probability distribution.

Let $(A_n)_{n \in \mathbb{N}} \in \mathfrak{A}_2^{\mathbb{N}}$ pairwise disjoint. Then we have

$$\begin{aligned} P^X\left(\sum_{n \in \mathbb{N}} A_n\right) &= P\left(X^{-1}\left(\sum_{n \in \mathbb{N}} A_n\right)\right) = P\left(\sum_{n \in \mathbb{N}} X^{-1}(A_n)\right) = \\ &= \sum_{n \in \mathbb{N}} P(X^{-1}(A_n)) = \sum_{n \in \mathbb{N}} P^X(A_n). \end{aligned}$$

□

If P^X is $b(n, p)$ distributed we can also write $X \sim b(n, p)$. We use the same notation for the other distributions introduced in Section 2 and appendix A.

4. INDEPENDENCE

Definition 4.1. Let $(\Omega, \mathfrak{A}, P)$ be a probability space. The events A_1, \dots, A_n are called independent if

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

for all $1 \leq i_1 < \dots < i_k \leq n$.

The sequence $(A_n)_{n \in \mathbb{N}} \in \mathfrak{A}^{\mathbb{N}}$ is called independent if A_1, \dots, A_n is independent for all n .

With this definition we can formulate and prove the following

Theorem 4.2 (Borel-Cantelli). *Let $(\Omega, \mathfrak{A}, P)$ be a probability space and $(A_n)_{n \in \mathbb{N}} \in \mathfrak{A}^{\mathbb{N}}$.*

- *If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(\limsup_{n \rightarrow \infty} A_n) = 0$.*
- *If $(A_n)_{n \in \mathbb{N}}$ is independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$ then $P(\limsup_{n \rightarrow \infty} A_n) = 1$.*

Proof. Consider

$$\lim_{k \rightarrow \infty} P\left(\bigcup_{n=k}^{\infty} A_n\right) \leq \lim_{k \rightarrow \infty} \sum_{n=k}^{\infty} P(A_n) = 0.$$

Lemma 1.4 yields

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) = \lim_{k \rightarrow \infty} P\left(\bigcup_{n=k}^{\infty} A_n\right) = 0.$$

To prove the second statement consider $(A_n^c)_{n \in \mathbb{N}}$, which is independent, as $(A_n)_{n \in \mathbb{N}}$ is independent as well. From this follows

$$\begin{aligned} P\left(\limsup_{n \rightarrow \infty} A_n\right) &= 1 - P\left(\liminf_{n \rightarrow \infty} A_n^c\right) = \\ &= 1 - \lim_{k \rightarrow \infty} P\left(\bigcap_{n=k}^{\infty} A_n^c\right) = 1 - \lim_{k \rightarrow \infty} \prod_{n=k}^{\infty} (1 - P(A_n)) = 1. \end{aligned}$$

To see the last equality let $p_n = P(A_n)$ for $n \in \mathbb{N}$. If $p_n = 1$ for infinitely many n this equality is trivial. So let $p_n < 1$ for all $n \geq k$ for some $k \in \mathbb{N}$. Therefore, as $\log x \leq x - 1$ for positive x , we have

$$\prod_{n=k}^{\infty} (1 - p_n) = \exp\left(\sum_{n=k}^{\infty} \log(1 - p_n)\right) \leq \exp\left(-\sum_{n=k}^{\infty} p_n\right) = 0,$$

as $\sum p_n = \infty$. □

Consider, for example, throwing a coin infinitely often. Then $\Omega = \{0, 1\}^{\mathbb{N}}$. We want to know the probability to throw infinitely often a “1”. Let $A_n =$

$\{\omega \in \Omega \mid \omega_i = 1\}$, which are independent. Therefore we have $P(\limsup_{n \rightarrow \infty} A_n) = 1$, as $\sum_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} \frac{1}{2} = \infty$.

Similarly to the independence of events we can define the independence of random variables $(X_n)_{n \in \mathbb{N}}$. Such a sequence is called independent if the events $(X_n^{-1}(A_n))_{n \in \mathbb{N}}$ are independent for every series $(A_n)_{n \in \mathbb{N}}$ with $A_n \in X_n(\mathfrak{A})$. In [Sch] it is proved that if the above statement holds for events A_n from intersection stable generators of the $X_n(\mathfrak{A})$ then $(X_n)_{n \in \mathbb{N}}$ is still independent. Therefore we have the following easy to check criteria for independence for discrete and continuous distributions.

The sequence $(X_n)_{n \in \mathbb{N}}$ of discrete random variables is independent iff

$$P\left(\bigcap_{n \in \mathbb{N}} X_n = x_n\right) = \prod_{n \in \mathbb{N}} P(X_n = x_n)$$

for all x_n .

The sequence $(X_n)_{n \in \mathbb{N}}$ of continuous random variables is independent iff

$$P\left(\bigcap_{n \in \mathbb{N}} X_n \leq x_n\right) = \prod_{n \in \mathbb{N}} P(X_n \leq x_n)$$

for all x_n .

5. EXPECTED VALUE AND VARIANCE

Often it is of high interest to know the mean expected value of a random variable, for example when calculation the fair price for a game or forecasting prices on the stock market. In this article we will only examine the expected value of random variables $X : (\Omega, \mathfrak{A}) \rightarrow (\mathbb{R}^1, \mathfrak{B}^1)$, although it is possible to define the expected value also for complex random variables and random vectors.

Definition 5.1. If the expression

$$EX = \int X dP = \int_{\mathbb{R}^1} id dP^X$$

exists it is called the **expected value** of X .

For discrete random variables follows immediately $EX = \sum_{n=1}^{\infty} x_n P(X = x_n)$ and for continuous random variables, $EX = \int_{\mathbb{R}^1} xf(x) dx$.

The following example demonstrates that the expected value must not necessarily exist. Consider a series of an infinitely often thrown coin. If the coin shows “1” in the n^{th} throw we get 2^n £. The game ends with the first “0” being thrown.

Let $(X_n)_{n \in \mathbb{N}}$ be independent identically $b(1, \frac{1}{2})$ distributed on $\{0, 1\}^{\mathbb{N}}$. For $\omega \in \Omega$ let $Y(\omega)$ be the first position at which a “0” is thrown. Hence we have $P(Y = n) = 1/2^n$. Let X denote the amount of money we win, that is $X =$

$2^Y - 1$ £. The expected amount of money to win from this game would therefore be

$$\begin{aligned} EX &= \sum_{n=0}^{\infty} (2^n - 1) P(X = 2^n - 1) = \sum_{n=0}^{\infty} (2^n - 1) P(Y = n) = \\ &= \sum_{n=0}^{\infty} (2^n - 1) \frac{1}{2^n} = \sum_{n=0}^{\infty} \left(1 - \frac{1}{2^n}\right) = \infty \text{ £.} \end{aligned}$$

We say for an statement A to be **P-almost sure** and denote this by $[P]$ if there exists a set $N \in \mathfrak{A}$ such that $P(N) = 0$ and A is true for all $\omega \in N^c$. With this notation we have the following

Lemma 5.2. *If EX and EY exist then*

- (1) $E(\alpha X + \beta Y) = \alpha EX + \beta EY$ for all $\alpha, \beta \in \mathbb{R}^1$,
- (2) $E\alpha = \alpha$ for all $\alpha \in \mathbb{R}^1$,
- (3) $EX \leq EY$ if $X \leq Y$ $[P]$ and
- (4) $EX = EY$ if $X = Y$ $[P]$.

These properties follow directly from the general properties of the measure integral, which can be found in [Bau] and are basically the same as for sums and Riemann integrals.

Another value of interest is how much a random variable is expected to differ from its expected value.

Definition 5.3. If $Var X = E(X - EX)^2$ exists it is called the **variance** of X . We call $\sqrt{Var X}$ the **standard deviation** of X .

Some examples for the expected value and the variance are in the tables in appendix A.

6. CONCLUSION

In this article we have been introduced to probability theory with almost no references to measure theory. Measure and integration theory provides the key to understand probability theory. The first two chapters of [Bau] give a very good introduction into this topic. Further studies of probability theory could include taking a closer look at the expected value and the variance, proving the Chebyshev inequality

$$P(|X - a| \geq \varepsilon) \leq \frac{Var X}{\varepsilon^2}$$

and examining the convergence properties of sequences of random variables. In [Sch] we can find a wide variety of topics about probability theory such as conditional probabilities, densities of random vectors, sums of random variables, characteristic functions, the Central Limit Theorem, the Laws of Large Numbers, stochastic processes and martingals. It provides the knowledge to go deeper into

special topics like simulated annealing and genetic algorithms, which are a very good alternative in approaching np complete problems like “travelling salesman” and find “approximately optimal” solutions, or stochastic differential equations and Itô integrals, which can be used to forecast prices on the stock market.

APPENDIX A. OVERVIEW OF SOME PROBABILITY DISTRIBUTIONS

A.1. Discrete Distributions.

Name	binomial distribution
Notation	$X \sim b(n, p)$
Density	$p_i = \binom{n}{i} p^i (1-p)^{n-i}$ for $1 \leq i \leq n$
Expected Value	$EX = np$
Variance	$Var X = np(1-p)$
Name	geometrical distribution
Notation	$X \sim geo(p)$
Density	$p_i = p(1-p)^{i-1}$ for $i \in \mathbb{N}$
Expected Value	$EX = \frac{1}{p}$
Variance	$Var X = \frac{1-p}{p^2}$
Name	Poisson distribution
Notation	$X \sim Poi(a)$
Density	$p_i = e^{-a} \frac{a^i}{i!}$ for $i \in \mathbb{N}_0$
Expected Value	$EX = a$
Variance	$Var X = a$
Name	negative binomial distribution
Notation	$X \sim \bar{b}(n, p)$
Density	$p_i = \binom{n+i-1}{i} p^n (1-p)^i$ for $i \in \mathbb{N}_0$
Expected Value	$EX = \frac{n(1-p)}{p}$
Variance	$Var X = \frac{n(1-p)}{p^2}$

A.2. Continuous Distributions.

Name	uniform distribution
Notation	$X \sim U[a, b]$
Density	$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$
Expected Value	$EX = \frac{a+b}{2}$
Variance	$Var X = \frac{(b-a)^2}{12}$
Name	exponential distribution
Notation	$X \sim exp(\lambda)$
Density	$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,\infty)}(x)$
Expected Value	$EX = \frac{1}{\lambda}$
Variance	$Var X = \frac{1}{\lambda^2}$
Name	normal distribution
Notation	$X \sim N(\mu, \sigma^2)$
Density	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$
Expected Value	$EX = \mu$
Variance	$Var X = \sigma^2$
Name	Chi-squared distribution
Notation	$X \sim \chi_1^2$
Density	$f(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} \mathbf{1}_{(0,\infty)}(x)$
Expected Value	$EX = 1$
Variance	$Var X = 3$

REFERENCES

- [Bau] H. Bauer, Maß- und Integrationstheorie (deGruyter 1992)
 [Sch] N. Schmitz, Vorlesungen über Wahrscheinlichkeitstheorie (Teubner 1996)