

# Statistics Lecture

Claudio Ortelli <sup>1</sup>

<sup>1</sup>Finance Institute  
Università della Svizzera italiana

Advanced Learning and Research Institute, 2011

# Continuous random variables

## Exponential distribution

The exponential distribution find its application in reliability theory and queuing theory. The following random variables are often modeled as exponential:

- 1 Time between two successive job arrivals to a file server (often called **interarrival time**).
- 2 Service time at a server in a queuing network; the server could be a resource such as a CPU, an I/O device, or a communication channel.
- 3 Time to failure (lifetime) of a component.
- 4 Time required to repair a component that has malfunctioned.

**Remark:** The choice of the exponential distribution to model the stochastic structure of the upper described variables is an assumption and not a given fact! Experimental verification of the distributional assumption will be therefore necessary before to relying on the results of the analysis.

# Continuous random variables

## The memoryless property of the exponential distribution

Let  $X \sim \text{Exp}(\lambda)$  be the lifetime of a component. Suppose we have observed that it has already been operating for  $t$  hours.

- What is the distribution of the remaining (residual) lifetime  $Y = X - t$ ?

Let the conditional probability of  $Y \leq y$ , given that  $X > t$ , be denoted by  $G_Y(y|t)$ . For  $y \geq 0$

$$\begin{aligned} G_Y(y|t) &= P(Y \leq y | X > t) = \frac{P(\{Y \leq y\} \text{ and } \{X > t\})}{P(X > t)} \\ &= \frac{P(\{X \leq y + t\} \text{ and } \{X > t\})}{P(X > t)} = \frac{P(t < X \leq y + t)}{P(X > t)} \\ &= \frac{\exp(-\lambda t)(1 - \exp(-\lambda y))}{\exp(-\lambda t)} = 1 - \exp(-\lambda y). \end{aligned}$$

# Continuous random variables

The memoryless property of the exponential distribution

## Result:

The conditional distribution  $G_Y(y|t)$  does not depend on  $t$  and is identical to the distribution of  $X$ , i.e.  $Exp(\lambda)$ .

## Interpretation:

The distribution of the remaining life does not depend on how long the component has been operating, i.e. the component does not age (it is as good as new). Therefore, the exponential distribution is not suited to model components or devices that gradually deteriorate.

# Continuous random variables

## The reliability and failure rate

Let the random variable  $X$  be the lifetime (or time to failure) of a component.

### Definition

The **reliability**  $R(t)$  of the component is the probability that the component survives until some time  $t$ , i.e.

$$R(t) = P(X > t) = 1 - F_X(t)$$

$F_X(t)$  is often called the **unreliability** of the component.

The conditional probability that the component **does not** survive for an additional interval of duration  $x$  given that it has survived until time  $t$  is equal to

$$G_Y(x|t) = \frac{P(t < X \leq t+x)}{P(X > t)} = \frac{F_X(t+x) - F_X(t)}{R(t)}$$

# Continuous random variables

## The reliability and failure rate

### Definition

The instantaneous failure rate  $h(t)$  is defined to be

$$h(t) = \lim_{x \rightarrow 0} \frac{1}{x} G_Y(x|t) = \lim_{x \rightarrow 0} \frac{F_X(t+x) - F_X(t)}{xR(t)},$$

so that

$$h(t) = \frac{f_X(t)}{R(t)}.$$

Alternate terms for  $h(t)$  are *hazard rate*, *force of mortality*, *intensity rate*, *conditional failure rate* or **failure rate**.

Interpretation:

- $h(t)\Delta t$  represents the conditional probability that a component having survived to age  $t$  will fail in the interval  $(t, t + \Delta t]$ .

# Continuous random variables

## The reliability and failure rate

- $f_X(t)\Delta t$  is the *unconditional* probability while  $h(t)\Delta t$  is a conditional probability.

Next theorem shows the connection between reliability and failure rate.

### Theorem

$$R(t) = \exp\left(-\int_0^t h(x)dx\right)$$

### Proof.

$$\int_0^t h(x)dx = \int_0^t \frac{f_X(x)}{R(x)}dx = \int_0^t \frac{-R'(x)}{R(x)}dx = -\ln(R(t))$$

using the fact that  $R'(x) = -f_X(x)$  and the boundary condition  $R(0) = 1$ .  $\square$

# Continuous random variables

## The reliability and failure rate

### Definition

The cumulative hazard is defined to be

$$H(t) = \int_0^t h(x) dx$$

Then, reliability can also be written as  $R(t) = \exp(-H(t))$ .

### Definition

The conditional reliability  $R_t(y)$  is the probability that the component survives an additional interval of duration  $y$  given that it has survived until time  $t$ .

$$R_t(y) = \frac{R(t+y)}{R(t)} \quad (1)$$



# Continuous random variables

## The reliability and failure rate

Assume a component does not age stochastically, i.e. the survival probability over an additional time interval  $y$  is the same regardless of the age  $t$  of the component:

$$R_t(y) = R_s(y) \text{ for all } t, s \geq 0.$$

Thus, for  $s = 0$

$$R_t(y) = R_0(y) = \frac{R(y)}{R(0)} = R(y),$$

so that from  $R_t(y) = R(t+y)/R(t)$  we have

$$R(t+y) = R(t)R(y).$$

In particular we obtain

$$\frac{R(t+y) - R(y)}{t} = \frac{(R(t) - 1)R(y)}{t} = \frac{(R(t) - R(0))R(y)}{t}.$$

# Continuous random variables

## The reliability and failure rate

Taking the limit as  $t \rightarrow 0$

$$\begin{aligned}R'(y) &= R'(0)R(y) \\ R(y) &= \exp(yR'(0)) = \exp(-\lambda y)\end{aligned}$$

which shows that the lifetime  $X \sim \text{Exp}(\lambda)$ .

If a component has exponential lifetime distribution it follows that

- 1 A replacement policy of used components based on the lifetime of the components is useless.
- 2 In estimating mean life and reliability the age of the observed components are of no concern. The number of hours of observed live and the number of observed failures are of interest.

# Continuous random variables

## The reliability and failure rate

### Definition

Increasing (decreasing) failure rate distribution

Let  $X$  be the lifetime of a component and  $F_X(t)$  the corresponding distribution function. If its failure rate  $h(t)$  is an increasing (decreasing) function of  $t$  for  $t \geq 0$  then  $F_X$  is an Increasing (Decreasing) Failure Rate distribution: IFR (DFR) distribution.

# Continuous random variables

## The reliability and failure rate

The behavior of the failure rate  $h(t)$  as a function of age is known as the *mortality curve*, *hazard function*, *life characteristic* or *lambda characteristic*.

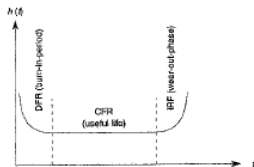


Figure 3.6. Failure rate as a function of time

# Continuous random variables

## The reliability and failure rate

### Exercise

- 1 The failure rate of a certain component is  $h(t) = \lambda_0 t$ , where  $\lambda_0 > 0$  is a given constant. Determine the reliability,  $R(t)$ , of the component. Repeat for  $h(t) = \lambda_0 t^{1/2}$ .
- 2 The failure rate of a computer system for onboard control of a space vehicle is estimated to be the following function of time:

$$h(t) = \alpha \mu t^{\alpha-1} + \beta \gamma t^{\beta-1}.$$

Derive an expression for the reliability  $R(t)$  of the system. Plot  $h(t)$  and  $R(t)$  as functions of time with parameter values  $\alpha = \frac{1}{4}$ ,  $\beta = \frac{1}{7}$ ,  $\mu = 0.0004$  and  $\gamma = 0.0007$ .

# Continuous random variables

## Hypoexponential Distribution

The hypoexponential distribution is used to model processes that can be divided into sequential phases such that the time the process spends in each phase is independent and exponentially distributed.

- Service times for input-output operations in a computer system often follow this distribution.

A two stage hypoexponential random variable  $X \sim \text{Hypo}(\lambda_1, \lambda_2)$  has pdf and distribution function equal to

$$f(t) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (\exp(-\lambda_1 t) - \exp(-\lambda_2 t)), \quad t > 0$$

$$F(t) = 1 - \frac{\lambda_2}{\lambda_2 - \lambda_1} \exp(-\lambda_1 t) + \frac{\lambda_1}{\lambda_2 - \lambda_1} \exp(-\lambda_2 t)$$

# Continuous random variables

## Erlang Distribution

When  $r$  sequential phases have identical exponential distribution the resulting density is known as  $r$ -stage Erlang and is given by

$$f(t) = \frac{\lambda^r t^{r-1} \exp(-\lambda t)}{(r-1)!} \text{ with } t > 0, \lambda > 0, r = 1, 2, \dots$$

$$F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \text{ with } t \geq 0, \lambda > 0, r = 1, 2, \dots$$

# Continuous random variables

## Hyperexponential Distribution

Suppose that a process consists of alternate phases, i.e. during any single experiment the process experiences one and only one of the many alternate phases, and these phases have exponential distributions. The overall distribution is then hyperexponential with density and distribution functions given by

$$f(t) = \sum_{i=1}^k \alpha_i \lambda_i \exp(-\lambda_i t) \text{ with } t > 0, \lambda_i > 0, \sum_{i=1}^k \alpha_i = 1$$
$$F(t) = \sum_i \alpha_i (1 - \exp(-\lambda_i t)) \quad t \geq 0$$



# Continuous random variables

## Weibull Distribution

The Weibull distribution is the most widely used parametric family of failure distributions. It has been used to describe

- fatigue failure
- electronic component failure
- ballbearing failure

The reason is that by a proper choice of the shape parameter  $\alpha$  we can obtain an IFR, DFR or constant failure rate distribution. The corresponding density and distribution functions are given by

$$\begin{aligned}f(t) &= \lambda \alpha t^{\alpha-1} \exp(-\lambda t^{\alpha}) \\F(t) &= 1 - \exp(-\lambda t^{\alpha})\end{aligned}$$

where  $t \geq 0$ ,  $\lambda > 0$  and  $\alpha > 0$ .

# Continuous random variables

## Pareto Distribution

The Pareto (also known as double-exponential, hyperbolic or power-law) distribution has been used to model

- the amount of CPU time consumed by an arbitrary process
- the Web file size on the Internet servers
- the thinking time of the web browser
- the number of data bytes in FTP bursts
- the access frequency of Web traffic

The density and distributions functions are given by

$$f(x) = \alpha k^\alpha x^{-\alpha-1} \quad x \geq k, \quad k > 0, \quad \alpha > 0$$

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

## Exercise

- ① Show that the failure rate  $h(t)$  of the hypoexponential distribution has the property

$$\lim_{t \rightarrow +\infty} h(t) = \min\{\lambda_1, \lambda_2\}.$$

- ② Show that a two-stage Erlang pdf is the limiting case of two-stage hypoexponential pdf. In other words, show that

$$\lim_{t \rightarrow +\infty} \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 t} - e^{-\lambda_2 t}) = \lambda_2^2 t e^{-\lambda_2 t}.$$

(Hint: Use l'Hôpital's rule.)

- ③ The CPU time requirement of a typical program measured in minutes is found to follow a three-stage Erlang distribution with  $\lambda = 0.5$ . What is the probability that the CPU demand of a program will exceed 1 minute?

# Functions of random variables

- Let  $X$  be a continuous random variable with distribution function  $F_X$ ,  $\Psi$  a function and

$$Y = \Psi(X).$$

- Under regularity conditions on  $\Psi$ ,  $Y$  is a random variable!
- Continuity or stepwise continuity of  $\Psi$  are sufficient conditions for  $Y$  to be a random variable.

## Example: Quadratic cost function.

Let  $X$  denote a measurement error. We assume a quadratic cost function, i.e.  $Y = \Psi(X) = X^2$ . The random variable  $Y$  has a distribution function  $F_Y$  which depends on  $F_X$  and  $\Psi$ .

- To derive  $F_Y$  simply compute the preimage of the event  $C = (-\infty, y]$ . In fact by definition of  $F_Y$

$$\begin{aligned}F_Y(y) &= P(Y \leq y) = P(Y \in C) \\&= P(\Psi(X) \in C) \\&= P(\Psi^{-1}(C)) \\&= P_X(B)\end{aligned}$$

where the set  $B$  is the preimage of  $C$ , i.e.

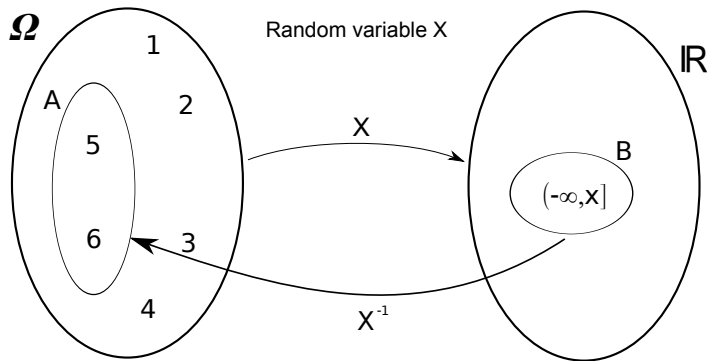
$$B = \Psi^{-1}(C) = \{x \in \mathbb{R} \mid \Psi(x) \in C\}.$$

- The preimage function of  $\Psi$  is a set function (the arguments are subsets of  $\mathbb{R}$ ) and is defined even if  $\Psi$  is not one-to-one.

# Functions of random variables

Probability space  $(\Omega, \mathcal{G}, P)$

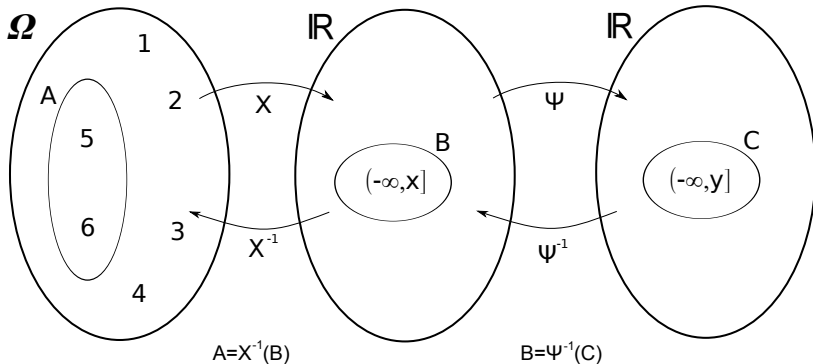
Probability space  $(\mathbb{R}, \mathcal{G}', P_x)$



Induced probability  $P_x(B) := P(A)$

# Functions of random variables

Probability space  $(\Omega, \mathcal{G}, P)$     Probability space  $(\mathbb{R}, \mathcal{G}', P_X)$     Probability space  $(\mathbb{R}, \mathcal{G}', P_Y)$



Induced probability  $P_Y(C) = P_X(B) = P(A)$

## Example continued

Because  $Y = X^2$  is always positive  $F_Y(y) = 0$  whenever  $y \leq 0$ . When  $y > 0$  it follows  $Y \leq y \Leftrightarrow -\sqrt{y} \leq X \leq \sqrt{y}$  so that  $F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$ . If the random variable  $X$  has a density function we can differentiate the last expression to obtain the density function of  $Y$ :

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})], & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

## Exercise

*Let  $X$  be uniformly distributed on  $(0, 1)$  and define  $Y = -\lambda^{-1} \ln(1 - X)$  where  $\lambda > 0$  is a parameter. Show that  $Y$  has an exponential distribution with parameter  $\lambda$ .*



## Theorem

Let  $X$  be a continuous random variable with density function  $f_X$  satisfying

- $f_X > 0$  for  $x \in I \subset \mathbb{R}$  and
- $f_X = 0$  for  $x \notin I$

and let  $\Phi$  be a differentiable and monotone real valued function with domain  $I$ . Then  $Y = \Phi(X)$  is a continuous random variable with density function

$$f_Y(y) = \begin{cases} f_X [\Phi^{-1}(y)] \left| \frac{\partial}{\partial y} (\Phi^{-1})(y) \right|, & y \in \Phi(I), \\ 0, & \text{otherwise.} \end{cases}$$

## Examples

1) Let  $\Phi$  be the distribution function  $F$  of the random variable  $X$  with density function  $f$  (we need to assume that  $F$  has the previous properties of continuity and differentiability) and define  $Y = F(X)$ . The random variable  $Y$  has density given by

$$f_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

2) Assume  $Y = aX + b$ , i.e.  $Y$  is a affine linear transformation of  $X$ . By the previous Theorem we have that ( $I$  is the set over which  $f_X \neq 0$ )

$$f_Y(y) = \begin{cases} \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), & y \in aI + b, \\ 0, & \text{otherwise.} \end{cases}$$

## Exercise

- 1 *In the second part of the previous Example assume that  $X \sim N(\mu, \sigma^2)$  and derive the density function of  $Y$ . What do you observe?*
- 2 *Let as before  $X$  be  $N(\mu, \sigma^2)$  and assume  $Y = e^X$ . Derive the density and distribution function of  $Y$ .*
- 3 *Show that if  $X$  has the  $k$ -stage Erlang distribution with parameter  $\lambda$ , then  $Y = 2\lambda X$  has the chi-square distribution with  $2k$  degrees of freedom.*

# Jointly distributed random variables

**Definition.** If  $\mathbf{X} = (X_1, \dots, X_n)$  is a  $n$ -dimensional random variable ( $n \geq 1$ ), its *joint distribution function* is defined by

$$F_{\mathbf{X}}(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n], \text{ for } x_i \in \mathbb{R}, i = 1, \dots, n.$$

In the case that  $\mathbf{X}$  is a multivariate continuous random variable, then

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(y_1, \dots, y_n) dy_1 \dots dy_n,$$

where  $f_{\mathbf{X}}(x_1, \dots, x_n)$  is the *joint probability density function* of  $\mathbf{X}$ , that is, by definition,

$$(i) f_{\mathbf{X}}(x_1, \dots, x_n) \geq 0 \text{ for all } \mathbf{x} \in \mathbb{R}^n \text{ and } (ii) \int_{\mathbb{R}^n} f_{\mathbf{X}}(y_1, \dots, y_n) dy_1 \dots dy_n = 1.$$

In most cases we are interested only in the distribution of a subset of the  $n$  variables. For this reason, we introduce the *marginal density function*.

# Jointly distributed random variables

**Definition.** If  $\mathbf{X} = (X_1, \dots, X_n)$  is a continuous  $n$ -dimensional random vector, the distribution function of any subset of  $X_1, \dots, X_n$  is continuous and its probability density function is called a **marginal density function**. In particular, the **marginal density** of  $X_1, \dots, X_i$  for  $i \in \{1, \dots, n\}$  equals

$$f_{X_1, \dots, X_i}(x_1, \dots, x_i) = \int_{\mathbb{R}^{n-i}} f_{\mathbf{X}}(x_1, \dots, x_i, y_{i+1}, \dots, y_n) dy_{i+1} \dots dy_n.$$

In the bivariate case  $n = 2$  we have that

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, y) dy \quad \text{and} \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_1, X_2}(y, x_2) dy. \end{aligned}$$

In the case  $n = 3$  we have that, for example,

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2, X_3}(x_1, y_2, y_3) dy_2 dy_3 \quad \text{or} \\ f_{X_1, X_3}(x_1, x_3) &= \int_{-\infty}^{\infty} f_{X_1, X_2, X_3}(x_1, y, x_3) dy. \end{aligned}$$

# Jointly distributed random variables

**Example.** Let us consider a three-dimensional vector  $\mathbf{X} = (X_1, X_2, X_3)$  having the following joint density function

$$f_{\mathbf{X}}(x_1, x_2, x_3) = \begin{cases} 6x_1x_2^2x_3, & \text{if } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq \sqrt{2} \\ 0, & \text{otherwise.} \end{cases}$$

We show that  $f_{\mathbf{X}}(x_1, x_2, x_3)$  is a probability density function. First note that  $f_{\mathbf{X}}(x_1, x_2, x_3) \geq 0$  for all  $(x_1, x_2, x_3) \in \mathbb{R}^3$ . Next, since

$$\int_{\mathbb{R}^3} f_{\mathbf{X}}(x_1, x_2, x_3) dx_1 dx_2 dx_3 = \int_0^{\sqrt{2}} \int_0^1 \int_0^1 6x_1x_2^2x_3 dx_1 dx_2 dx_3 = 1,$$

by definition  $f_{\mathbf{X}}(x_1, x_2, x_3)$  is a probability density function.

We then compute the marginal densities of  $X_1$ ,  $X_2$ ,  $X_3$  and  $(X_1, X_3)$ .

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, y_2, y_3) dy_2 dy_3 \\ &= \begin{cases} \int_0^{\sqrt{2}} \int_0^1 6x_1y_2^2y_3 dy_2 dy_3 = 2x_1, & \text{if } 0 \leq x_1 \leq 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

# Jointly distributed random variables

Similarly,

$$\begin{aligned}f_{X_2}(x_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1, x_2, y_3) dy_1 dy_3 \\&= \begin{cases} \int_0^{\sqrt{2}} \int_0^1 6y_1 x_2^2 y_3 dy_1 dy_3 = 3x_2^2, & \text{if } 0 \leq x_2 \leq 1 \\ 0, & \text{otherwise,} \end{cases} \\f_{X_3}(x_3) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1, y_2, x_3) dy_1 dy_2 \\&= \begin{cases} \int_0^1 \int_0^1 6y_1 x_2^2 y_3 dy_1 dy_2 = x_3, & \text{if } 0 \leq x_3 \leq \sqrt{2} \\ 0, & \text{otherwise,} \end{cases}\end{aligned}$$

and

$$\begin{aligned}f_{X_1, X_3}(x_1, x_3) &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, y_2, x_3) dy_2 \\&= \begin{cases} \int_0^1 6x_1 y_2^2 x_3 dy_2 = 2x_1 x_3, & \text{if } 0 \leq x_1 \leq 1, 0 \leq x_3 \leq \sqrt{2} \\ 0, & \text{otherwise.} \end{cases}\end{aligned}$$

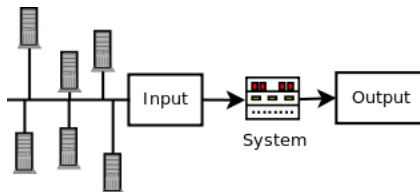
Note that we can test whether the calculations are correct verifying that  $\int_{\mathbb{R}} f_{X_i}(x_i) dx_i = 1$  for all  $i = 1, 2, 3$  and  $\int_{\mathbb{R}^2} f_{X_1, X_3}(x_1, x_3) dx_1 dx_3 = 1$  since all these density functions must be probability density functions.

# Simulation of random variables

## Motivation

- We are interested in simulating a model of a real system (network, electronic device, ...).
- The output  $Y$  of the model depends on a stochastic input, i.e. a random variable  $X$  with known distribution function  $F_X$ :

$$Y = g(X).$$





# Simulation of random variables

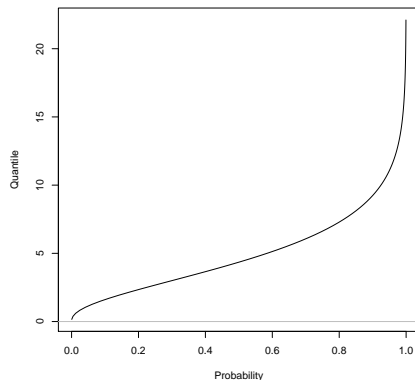
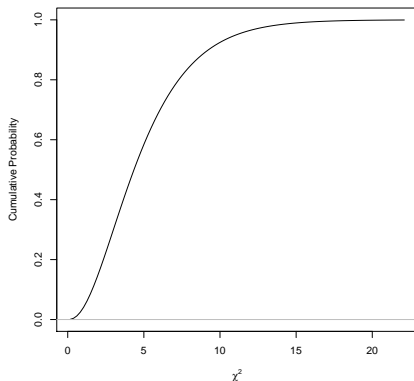
- The model is too complex in order to analytically derive the probabilistic properties of the output, i.e.  $F_Y$ .
- Idea: simulate a possible outcome  $x$  of the input  $X$  and evaluate the corresponding outcome  $y = g(x)$  of the output. Repeat the experiment  $N$  times and analyse the results.
- The simulated values of the input must be drawn from the distribution of  $X$ .
- Question: how is it possible to simulate independent realizations from a given distribution  $F_X$ ?
- Answer: different methods available. The simplest of them requires simulating from the uniform distribution on the open interval  $(0,1)$ , i.e.  $U(0,1)$ . In Matlab use the function “rand”.

# Simulation of random variables

## Inverse Transform Method

If the distribution function  $F_X$  is continuous and strictly increasing then  $F_X^{-1} : (0, 1) \rightarrow \mathbb{R}$  exists.

Example: Chi-Squared Distribution with 5 degrees of freedom



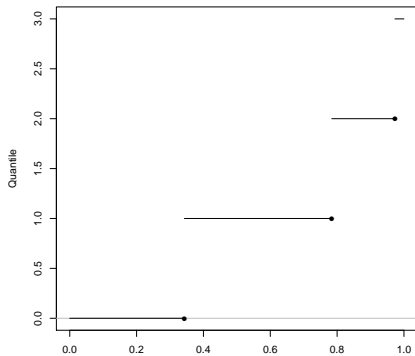
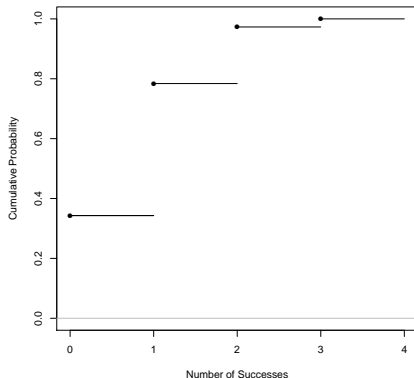
# Simulation of random variables

For a general distribution function not necessarily strictly increasing define

$$F_X^{-1}(p) = \inf\{x : p \leq F_X(x)\} \quad 0 < p < 1.$$

It then follows that  $F_X^{-1}(p) \leq x \iff p \leq F_X(x)$ .

Example: Binomial distribution  $Bin(n = 3, p = 0.3)$



# Simulation of random variables

## Inverse Transform Method

### Theorem

*[Inverse Transform Method] Let  $U$  a continuous  $\text{Unif}(0,1)$  distributed random variable. The random variable  $Y = F_X^{-1}(U)$  has distribution function  $F_X$ .*

### Proof.

By definition

$$F_Y(c) := P(Y \leq c) = P(F_X^{-1}(U) \leq c).$$

But the last equality is equivalent to (see previous slide)

$$P(U \leq F_X(c)) = F_X(c).$$



# Simulation of random variables

## Inverse Transform Method

From the previous theorem we derive the following *two steps* simulation algorithm

- 1 Simulate a realization  $u$  from a  $\text{Unif}(0,1)$  random variable  $U$ .
- 2 Compute  $x = F_X^{-1}(u)$ .

## Example

Simulation of  $Y \sim \text{Exp}(\lambda)$

- The distribution function is  $F_Y(x) = 1 - \exp(-\lambda x)$
- Compute  $F_Y^{-1}(p) = -\frac{1}{\lambda} \ln(1-p)$
- Sample a random draw  $u$  from  $U \sim \text{Unif}(0,1)$
- Set  $y = -\frac{1}{\lambda} \ln(1-u)$

# Simulation of random variables

## Inverse Transform Method

Remark: if  $U \sim \text{Unif}(0,1)$  then  $1 - U$  is also  $\text{Unif}(0,1)$ . Therefore we can also write  $Y = -\frac{1}{\lambda} \ln(U)$ .

# Simulation of random variables

## Inverse Transform Method

For a discrete random variable  $Y$  with probability mass function  $P(Y = x_i) = p_i, i = 1, \dots, m$  consider the following algorithm:

- 1 Generate a  $\text{Unif}(0,1)$  random variable  $U$
- 2 Compute  $Y$  as follows

$$Y = x_j \quad \text{if} \quad \sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^j p_i.$$

i.e.  $Y = x_j$  if  $F_Y(x_{j-1}) < U \leq F_Y(x_j)$ .

# Simulation of random variables

## Inverse Transform Method

### Example

Let  $Y$  have the following mass function

$p_1$	$p_2$	$p_3$	$p_4$
0.1	0.2	0.4	0.3

The simulation algorithm is the following:

- 1 If  $u \leq p_1 = 0.1$  then  $y = x_1$ . Stop.
- 2 if  $u \leq p_1 + p_2 = 0.3$  then  $y = x_2$ . Stop.
- 3 if  $u \leq p_1 + p_2 + p_3 = 0.7$  then  $y = x_3$ . Stop.
- 4  $y = x_4$ . Stop.

This algorithm is correct but inefficient. In fact, probabilities of one, two, ... comparisons are equal to the probabilities of  $x_1, x_2, \dots$ , respectively. The expected number of comparisons is

$$p_1 + 2p_2 + 3p_3 + 4p_4 = 2.9.$$



# Simulation of random variables

## Inverse Transform Method

We can improve efficiency by sorting the values of  $x_i$  by decreasing order of probabilities  $p_i$ 's:  $x_3, x_4, x_2$  and  $x_1$ .

- ① If  $u \leq p_3 = 0.4$  then  $y = x_3$ . Stop.
- ② if  $u \leq p_3 + p_4 = 0.7$  then  $y = x_4$ . Stop.
- ③ if  $u \leq p_3 + p_4 + p_2 = 0.9$  then  $y = x_2$ . Stop.
- ④  $y = x_1$ . Stop.

The expected number of comparisons is now equal to

$$p_3 + 2p_4 + 3p_2 + 4p_1 = 2.$$

# Simulation of random variables

## Exercises

The Laplace distribution is a continuous distribution with density function

$$f(x; \mu, b) = \begin{cases} \frac{1}{2b} \exp\left(-\frac{\mu-x}{b}\right) & \text{if } x < \mu \\ \frac{1}{2b} \exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases}$$

where  $\mu$  is the mean and  $b$  a scale parameter.

- 1 Plot the density, distribution and quantile functions of the Laplace distribution with parameter  $\mu = 1$  and  $b = 0.5$ .
- 2 Using the previous values of  $\mu$  and  $b$  simulate  $N = 1000$  independent realizations of a Laplace distributed random variable  $Y$ .
- 3 Plot the histogram of the simulated random variables and compare it with the density function of  $Y$ .
- 4 Plot the empiric distribution function of the simulated sample and compare it with the distribution function of  $Y$ .

# Simulation of random variables

## Transform methods

The theorem on *Distributions of functions of continuous random variables* allows us to generate random variables by means of ad hoc transformations of  $\text{Unif}(0,1)$  random variables. The following theorem generalizes the previous theorem to the multivariate case.

### Theorem

Assume that  $X = (X_1, \dots, X_n)$  is a random vector with joint density function  $f_X(x_1, \dots, x_n)$  and  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$  a one-to-one and continuously differentiable function. Define  $Y = (Y_1, \dots, Y_n) = g(X)$ . The joint density function of  $Y$  is then equal to

$$f_Y(y) = f_X(g^{-1}(y)) |J(g^{-1}(y))|$$

where

$$J(g^{-1}(y)) = \det(M_J) = \det \left[ \frac{\partial x_i(y)}{\partial y_j} \right]_{i=1, n; j=1, n}$$

# Simulation of random variables

## Transform methods

### Example

Let  $X \sim N(\mu, \Sigma)$  be a  $2 \times 1$  random vector and consider the affine transformation

$$Y = b + A X$$

where  $b$  is a deterministic  $2 \times 1$  vector and  $A$  a  $2 \times 2$  invertible matrix. We then have

$$X = A^{-1}(Y - b); M_J = A^{-1}$$

and  $f_Y(y)$  is equal to

$$\frac{1}{\sqrt{2\pi \det(\Sigma)}} \exp\left(-\frac{1}{2}(A^{-1}(Y - b) - \mu)' \Sigma^{-1}(A^{-1}(Y - b) - \mu)\right) |\det(A^{-1})|$$

$$\frac{1}{\sqrt{2\pi \det(A \Sigma A')}} \exp\left(-\frac{1}{2}(Y - b - A\mu)' (A \Sigma A')^{-1}(Y - b - A\mu)\right)$$

# Simulation of random variables

## Transform methods

### Example continued

Looking at the density function of  $Y$  we note that  $Y \sim N(\tilde{\mu}, \tilde{\Sigma})$  with  $\tilde{\mu} = b + A\mu$  and  $\tilde{\Sigma} = A\Sigma A'$ .

### Exercise

$U_1$  and  $U_2$  are two independent  $\text{Unif}(0,1)$  distributed random variables. Define  $Y = (Y_1, Y_2)$  with

$$Y_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \text{ and } Y_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

Show that  $Y \sim N(0, I)$ , i.e.  $Y_1$  and  $Y_2$  are independent standard normal distributed random variables.

### Exercise

*The joint density of the random variables  $X_1$  and  $X_2$  is*

$$f(x_1, x_2) = 2 \exp(-x_1) \exp(-x_2), \text{ for } 0 < x_1 < x_2 < \infty$$

*and  $f(x_1, x_2) = 0$  otherwise.*

*We define the transformation*

$$Y_1 = 2X_1, Y_2 = X_2 - X_1.$$

*Find the joint density of  $Y_1$  and  $Y_2$ . Are  $Y_1$  and  $Y_2$  independent?*

# Expectation

## Definition

The expectation,  $E[X]$ , of a random variable  $X$  is defined by

$$E[X] = \begin{cases} \sum_i x_i p(x_i), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{if } X \text{ is continuous,} \end{cases}$$

provided the relevant sum or integral is absolutely convergent, i.e.

$$\sum_i |x_i| p(x_i) < \infty \text{ and } \int_{-\infty}^{\infty} |x| f(x) dx < \infty.$$

## Example

Assume  $X$  is Binomial distributed with  $n = 5$  and  $p = 0.5$ . Then

$$E[X] = \sum_i x_i p(x_i) = 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 2 \cdot \frac{10}{32} + 3 \cdot \frac{10}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{1}{32} = 2.5$$

The expected value need not correspond to a possible value of  $X$ !

The expected value is a weighted average and it denotes the “center” of a probability mass or density function in the sense of a center of gravity.



# Expectation

Let  $X$  be a random variable, and define  $Y = g(X)$ . Suppose we want to compute  $E[Y]$ . In order to apply the definition of  $E[Y]$  we need to derive the *pmf* (or the *pdf*) of  $Y$ . An easier method is to use the following result

$$E[Y] = E[g(X)] = \begin{cases} \sum_i g(x_i) p_X(x_i), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx, & \text{if } X \text{ is continuous,} \end{cases}$$

provided the sum or the integral is absolutely convergent, i.e.

$$\sum_i |g(x_i)| p_X(x_i) < \infty \text{ or } \int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty.$$

# Expectation

**Theorem. Properties of Expectation.** *If  $c$  is a constant and  $g(X)$ ,  $g_1(X)$  and  $g_2(X)$  are functions whose expectations exist, then*

(i)  $E[c] = c$ ;

(ii)  $E[cg(X)] = cE[g(X)]$ ;

(iii)  $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$ ;

(iv)  $E[g_1(X)] \leq E[g_2(X)]$  if  $g_1(x) \leq g_2(x) \forall x$ ;

(v)  $|E[g(X)]| \leq E[|g(X)|]$ ;

(vi)  $E[|g(X)|] = 0$  then it follows that  $g(x) = 0$  for all  $x$  with positive probability.

(vii) If  $X$  is a random variable with distribution function  $F(x)$  then  $E[X]$  exists if and only if  $\int_0^\infty (1 - F(x))dx$  and  $\int_{-\infty}^0 F(x)dx$  are finite, in which case

$$E[X] = \int_0^\infty (1 - F(x))dx - \int_{-\infty}^0 F(x)dx.$$

**Example.** Suppose that  $X$  has the following distribution (called **Laplace distribution**)

$$f_X(x) = \frac{1}{2} \exp(-|x|/2), \quad -\infty < x < +\infty.$$

Then the average value  $E[X]$  exists since

$$\int_0^\infty x/2 \exp(-|x|/2)dx = \int_0^\infty x/2 \exp(-x/2)dx = 2.$$

Similarly,  $\int_{-\infty}^0 x/2 \exp(-|x|/2)dx = -2$  and

$$E[X] = \int_{\mathbb{R}} x f_X(x)dx = \int_{-\infty}^0 f_X(x)dx + \int_0^\infty x f_X(x)dx = -2 + 2 = 0.$$

## Definition

A special case is the power function  $g(X) = X^k$ ,  $k = 1, 2, 3, \dots$ .  $E(X^k)$  is known to be the  $k$ -th moment of the random variable  $X$ . The first moment, i.e.  $k = 1$ , is the ordinary expectation of  $X$ .

Sometimes it is useful to center the origin of measurement, i.e. to work with powers of  $X - E[X]$ .

## Definition

The  $k$ -th central moment of the random variable  $X$ ,  $\mu_k$ , is defined as

$$\mu_k = E[(X - E[X])^k].$$

The second central moment  $\mu_2$  is called the *Variance* of the random variable  $X$ , typically denoted by  $\sigma^2$ , is a measure of dispersion. It measures the amount by which the R.V.  $X$  deviates from its expected value.

## Definition

The variance of a random variable  $X$  is

$$\sigma^2 = \begin{cases} \sum_i (x_i - E[X])^2 p(x_i), & \text{discrete case,} \\ \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx, & \text{continuous case.} \end{cases}$$

- The variance is a sum of squares and therefore is a nonnegative number.
- The square root of the variance is denoted by  $\sigma$  and is called the standard deviation.

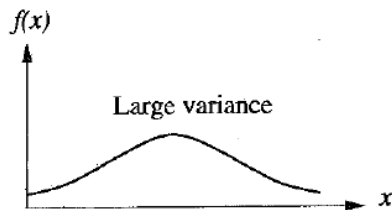


Figure 4.2. The pdf of a diffuse distribution

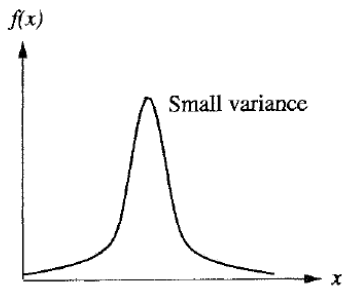


Figure 4.1. The pdf of a "concentrated" distribution

Figure: Small and large variance.

**Theorem. Properties of the Variance.** Let  $X, X_i, i = 1, \dots, n$ , be *independent* discrete random variables with finite second moment, and  $a, b$  two constants. Then

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

$$(ii) \text{Var}(aX + b) = a^2 \text{Var}(X);$$

$$(iii) \text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

Note that relaxing the assumption of independence property (iii) does not hold any more (see the following sections).

We define two special measures based on  $\mu_3$  and  $\mu_4$ :

- $n = 3$ ,  $\mu_3$  divided by the standard deviation cubed is called *skewness* of the distribution of  $X$  and is often denoted

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{E[(X - E[X])^3]}{\sigma(X)^3}.$$

If the distribution of  $X$  is *symmetric about  $\mu$* , then  $\alpha_3(X) = 0$ .

- $n = 4$ ,  $\mu_4$  divided by  $\sigma(X)^4$  is called *kurtosis*, is denoted by  $\alpha_4$  and is used as a measure of how "*heavy*" the tails of a distribution are.



## Exercise

- 1) Starting from the definition of  $\sigma^2$  use the property of linearity of the expected value to prove that  $\text{Var}[X] = E[X^2] - (E[X])^2$ .
- 2) Define the function  $g : \mathbb{R} \rightarrow [0, \infty]$ ,  $c \mapsto E[(X - c)^2]$ . Show that  $c_{\min} = E[X]$  is the minimum of the function  $g$ .

## Examples and Exercises.

**Example.** Let  $X$  be a random variable having a “semitriangular” distribution given by (for some  $a > 0$ )

$$f_X(x) = \begin{cases} 2(a-x)/a^2, & \text{if } 0 \leq x \leq a \\ 0, & \text{otherwise.} \end{cases}$$

First of all, note that this is a probability density function since

$$(i) f_X(x) \geq 0 \quad \forall x, \text{ and } (ii) \int_{\mathbb{R}} f_X(x) dx = \int_0^a \left( \frac{2}{a} - \frac{2x}{a^2} \right) dx = 2 - 1 = 1.$$

We compute  $E[X]$  and  $Var[X]$ .

$$E[X] = \int_{\mathbb{R}} x f_X(x) dx = \int_0^a \left( \frac{2x}{a} - \frac{2x^2}{a^2} \right) dx = \frac{x^2}{a} \Big|_0^a - \frac{2x^3}{3a^2} \Big|_0^a = a - \frac{2}{3}a = \frac{1}{3}a,$$

$$E[X^2] = \int_{\mathbb{R}} x^2 f_X(x) dx = \int_0^a \left( \frac{2x^2}{a} - \frac{2x^3}{a^2} \right) dx = \frac{2x^3}{3a} \Big|_0^a - \frac{x^4}{2a^2} \Big|_0^a = \frac{2}{3}a^2 - \frac{1}{2}a^2 = \frac{1}{6}a^2.$$

$$\text{Therefore } Var(X) = E[X^2] - E[X]^2 = a^2/6 - a^2/9 = a^2/18.$$

As an exercise, compute the median, skewness  $\alpha_3$  and kurtosis  $\alpha_4$  of  $X$ .

**Example. Mean and Variance of the uniform distribution.** Let us consider  $X$  uniformly distributed on  $(b, c)$ , i.e.  $f_X(x) = 1/(c-b)$  if  $x \in (b, c)$  and 0 otherwise. Then

$$E[X] = \int_b^c \frac{x}{c-b} dx = \frac{x^2}{2(c-b)} \Big|_b^c = \frac{c^2 - b^2}{2(c-b)} = \frac{b+c}{2};$$
$$E[X^2] = \int_b^c \frac{x^2}{c-b} dx = \frac{x^3}{3(c-b)} \Big|_b^c = \frac{c^3 - b^3}{3(c-b)} = \frac{c^2 + bc + b^2}{3}.$$

Hence  $Var(X) = E[X^2] - E[X]^2 = (c-b)^2/12$ .

Let us also compute the median  $m$  of  $X$ .  $m$  must satisfy

$$P[X \leq m] = \int_b^m \frac{1}{c-b} dx = \frac{m-b}{c-b} = 0.5.$$

Thus,  $m = (b+c)/2 = E[X]$ .

**Exercise.** Compute higher order moments of  $X$ . In particular, compute skewness and kurtosis of  $X$ .

**Example.** An urn contains 4 red balls, 1 yellow ball, and 3 green balls. Balls are drawn successively at random *without replacement*, until a green ball is drawn. Let  $X$  be the number of draws required. We compute  $E[X]$  and  $\sigma^2(X)$ .

The possible values of  $X$  are 1, 2, 3, 4, 5, 6 with corresponding probabilities  $p_X(1) = 3/8, p_X(2) = 5/8 \cdot 3/7, p_X(3) = 5/8 \cdot 4/7 \cdot 3/6, p_X(4) = 5/8 \cdot 4/7 \cdot 3/6 \cdot 3/5, p_X(5) = 5/8 \cdot 4/7 \cdot 3/6 \cdot 2/5 \cdot 3/4, p_X(6) = 5/8 \cdot 4/7 \cdot 3/6 \cdot 2/5 \cdot 1/4$ . Hence

$$E[X] = 1 \cdot 3/8 + 2 \cdot 15/56 + 3 \cdot 5/28 + 4 \cdot 3/28 + 5 \cdot 3/56 + 6 \cdot 1/56 = 9/4$$

and

$$E[X^2] = 1 \cdot 3/8 + 4 \cdot 15/56 + 9 \cdot 5/28 + 16 \cdot 3/28 + 25 \cdot 3/56 + 36 \cdot 1/56 = 27/4.$$

Therefore

$$\text{Var}(X) = \sigma^2(X) = E[X^2] - E[X]^2 = 27/4 - (9/4)^2 = 27/4 - 81/16 = 27/16 = 1.6875.$$

**Example. Mean and Variance of the binomial distribution.** To begin, let  $X$  be Bernoulli distributed with success parameter  $p$ . Then

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p \text{ and } E[X^2] = 0^2 \cdot (1 - p) + 1^2 \cdot p = p.$$

Consequently,  $Var(X) = p - p^2 = p(1 - p)$ .

Now, we know that  $Y = X_1 + \dots + X_n \sim B(n, p)$  for  $n$  independent Bernoulli variables  $X_i$  with same success parameter  $p$ . Thus, the expectation and variance of the binomial variable  $Y$  equal

$$E[Y] = \sum_{i=1}^n E[X_i] = np \text{ and } Var[Y] = \sum_{i=1}^n Var(X_i) = np(1 - p).$$

**Exercise.** 1. Derive the results for the expectation and variance of the binomial distribution *directly* from its definition.

2. Find  $E[X]$  and  $Var(X)$  for the following distribution

$$p_X(x) = \binom{5}{x} 0.4^x 0.6^{5-x} \text{ if } x \in \{0, 1, 2, 3, 4, 5\}, \text{ and } 0 \text{ otherwise.}$$

**Example. Mean and Variance of the Poisson distribution.** If  $X$  has the Poisson distribution with parameter  $\lambda$ . Then

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x \exp(-\lambda) \lambda^x / x! = \exp(-\lambda) \sum_{x=1}^{\infty} \lambda^x / (x-1)! \\ &= \exp(-\lambda) \cdot \lambda \sum_{x=1}^{\infty} \lambda^{x-1} / (x-1)! = \exp(-\lambda) \lambda \exp(\lambda) = \lambda; \\ E[X^2] &= \sum_{x=0}^{\infty} x^2 \exp(-\lambda) \lambda^x / x! = \sum_{x=0}^{\infty} (x(x-1) + x) \exp(-\lambda) \lambda^x / x! \\ &= \sum_{x=2}^{\infty} \exp(-\lambda) \lambda^x / (x-2)! + \sum_{x=0}^{\infty} x \exp(-\lambda) \lambda^x / x! = \lambda^2 + \lambda. \end{aligned}$$

Thus  $Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$ .

**Exercise.** Compute skewness and kurtosis of the Poisson distribution with parameter  $\lambda$ . What is the skewness of the Poisson distribution with expectation equal to 2?

Trick: Compute  $E[X^3]$  as  $E[X(X-1)(X-2)] + 3E[X(X-1)] + E[X]$ .

## Example: Mean and Variance of the normal distribution.

Let we consider  $X \sim N(\mu, \sigma^2)$ . We will show that  $E[X] = \mu$  and  $Var(X) = \sigma^2$ . In fact

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{\infty} \frac{\sigma y + \mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}y^2} \sigma dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \left( -e^{-\frac{1}{2}y^2} \right) \Big|_{-\infty}^{\infty} + \mu \cdot 1 = \mu \end{aligned}$$

and

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{\infty} \frac{(\sigma y + \mu)^2}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}y^2} \sigma dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}y^2} dy + \frac{2\sigma\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy + \mu^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left( (-ye^{-\frac{1}{2}y^2}) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \right) + 0 + \mu^2 \cdot 1 = \sigma^2 + \mu^2. \end{aligned}$$

Thus,  $\text{Var}(X) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$ .

**Exercise:** Let  $X \sim N(\mu, \sigma^2)$ . Find the skewness and kurtosis of  $X$ . In particular, compute these values when  $\mu = 0$  and  $\sigma = 1$ .



**Exercise:** Let  $Y = aX^2 + b$  for  $a, b$  two constants. Compute the mean and variance of  $Y$  for the case

- 1  $X \sim N(0, 1)$ ;
- 2  $X \sim N(\mu, \sigma^2)$ ;
- 3  $X$  exponentially distributed with parameter  $\lambda$ .
- 4 A particular container of one brand of a certain food is labeled as containing 500 grams of food. Suppose it is known that the weight of the food in a container selected at random is distributed  $N(500, 25)$ . A container is considered to be underweight if its net food weight is less than 98% of the label claim, in this case less than 490 grams. If 1000 containers are chosen at random, how many should we expect to be underweight?
- 5 Let  $X \sim N(\mu, \sigma^2)$ . Let  $Y = \exp(X)$ .  $Y$  is called a lognormal random variable. Find  $E[Y]$  and  $Var(Y)$ .

# Expectation based on multiple Random Variables

Let  $X_1, X_2, \dots, X_n$  be  $n$  random variables defined on the same probability space and define  $Y = \Phi(X_1, X_2, \dots, X_n)$ . Then

$$\begin{aligned} E[Y] &= E[\Phi(X)] \\ &= \begin{cases} \sum_{x_1} \dots \sum_{x_n} \Phi(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n), & \text{descrete case} \\ \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \Phi(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n, & \text{contin. case} \end{cases} \end{aligned}$$

**Remark:** It is not necessary to derive the pmf (descrete case) or the pdf (cont. case) of the random variable  $Y$  in order to compute its expected value.

# Expectation based on multiple Random Variables

Recall the linearity property of expectation:

- ①  $E[\lambda X] = \lambda E[X]$ , where  $\lambda \in \mathbb{R}$ ,
- ②  $E[X + Y] = E[X] + E[Y]$ .

If function  $\Psi$  is linear, we then obtain that

$$E[Y] = E[\Phi(X)] = E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E[X_i].$$

## Theorem

*If  $X$  and  $Y$  are independent random variables, then  $E[XY] = E[X]E[Y]$ . The converse does not hold, i.e. it is possible that two random variables  $X$  and  $Y$  satisfy  $E[XY] = E[X]E[Y]$  without being independent.*

# Expectation based on multiple Random Variables

## Definition

Let  $X$  and  $Y$  be two random variable. The covariance between  $X$  and  $Y$  is defined to be

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

The covariance is a measure of linear dependence between two random variables. If  $\text{Cov}(X, Y) = 0$  we say that  $X$  and  $Y$  are uncorrelated.

- The covariance is a *bilinear* operator. In fact let  $X, Y$  and  $Z$  be two random variables with existing expectation and  $\lambda \in \mathbb{R}$ . Then
  - 1  $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$ ,
  - 2  $\text{Cov}(\lambda X, Z) = \lambda \text{Cov}(X, Z)$ ,
  - 3  $\text{Cov}(X, Z) = \text{Cov}(Z, X)$ .

General rule: freeze the first (second) argument and consider the covariance an expectation in the second (first) argument.

# Expectation based on multiple Random Variables

**Remark:** Zero covariance does not imply that the two random variables are independent!

## Example

Let  $X$  be uniformly distributed over the interval  $(-1, 1)$  and let  $Y = X^2$  so that  $Y$  is completely dependent on  $X$ . Noting that for all odd values of  $k > 0$ , the  $k$ -th moment  $E[X^k] = 0$ , we have (see part 2 of next Exercise for the equality  $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$ )

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = E[X^3] - 0E[Y] = 0.$$

# Expectation based on multiple Random Variables

## Exercise

- 1 Starting from the definition of Cov use the property of linearity of the expected value to prove property 1. and 2. of the covariance.
- 2 Using the linearity of the expected value show that 
$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y].$$
- 3 Show that if  $X$  and  $Y$  are two independent random variables, then 
$$\text{Cov}(X, Y) = 0.$$
- 4 Show that  $\text{Cov}(X, X) = \text{Var}(X)$ . Hint: simply start from the definition of Cov.
- 5 Show that 
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$
- 6 Let  $X \sim N(0, 1)$  and  $Y$  a random variable independent from  $X$  and such that  $P(Y = 1) = P(Y = -1) = 0.5$ . Finally define  $Z = X \cdot Y$ . Show that 
$$\text{Cov}(X, Z) = 0.$$
 Are  $X$  and  $Z$  independent random variables?

## Definition

**Correlation:** the correlation coefficient  $\rho(X, Y)$  between the random variables  $X$  and  $Y$  is defined by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

**Remark:** the correlation coefficient satisfies the following inequalities:  
 $-1 < \rho(X, Y) < 1$ .

# Distribution and moments of a random variable

## Transform methods: definitions

- Transform methods are transformations of the probability mass function (discrete case) or the density function (continuous case).
- They are particular useful to compute moments of a distribution and in problems involving sums of independent random variables.

### Definition

The moment generating function (MGF)  $M_X(\theta)$ , abbreviated  $M(\theta)$ , of the random variable  $X$  is defined by

$$M(\theta) = E[\exp(X\theta)]$$

provided the expectation exists ( $M(\theta)$  may not exist for all  $\theta \in \mathbb{R}$ ).



# Distribution and moments of a random variable

## Transform methods: definitions

### Definition

The characteristic function of a random variable  $X$  is given by

$$N_X(\tau) = N(\tau) = E[\exp(iX\tau)] = M_X(i\tau) \text{ where } i = \sqrt{-1}.$$

Note that  $N_X(\tau)$  is always defined for any  $X$  and all  $\tau$ .

### Definition

Let  $X$  be a nonnegative continuous random variable. The Laplace - Stieltjes transform of  $X$  is

$$L_X(s) = L(s) = M_X(-s) = \int_0^{\infty} \exp(-sx) f(x) dx.$$

# Distribution and moments of a random variable

Transform methods: definition and theorems

## Definition

Let  $X$  be a discrete nonnegative integer-valued random variable. The  $z$  transform (or probability generating function) of  $X$  is defined as

$$G_X(z) = G(z) = E[z^X] = M_X(\ln(z)) = \sum_{i=0}^{\infty} p_X(i)z^i.$$

## Theorem

*Affine transformation. Let  $Y = aX + b$ . Then*

$$M_Y(\theta) = \exp(b\theta)M_X(a\theta)$$

# Distribution and moments of a random variable

## Transform methods: theorems

### Theorem

*[The Convolution Theorem] Let  $X_1, X_2, \dots, X_n$  be mutually independent random variables. Define  $Y = \sum_{i=1}^n X_i$ . If  $M_{X_i}(\theta)$  exists for all  $i$ , then  $M_Y(\theta)$  exists, and*

$$M_Y(\theta) = \prod_{i=1}^n M_{X_i}(\theta).$$

### Theorem

*[Uniqueness Theorem] If  $M_X(\theta) = M_Y(\theta)$  for all  $\theta$ , then  $F_X = F_Y$ , i.e.  $X$  and  $Y$  have the same distribution.*

# Distribution and moments of a random variable

## Transform methods: theorems

### Theorem

*[Moment generating property of the MGF] Let  $X$  be a random variable such that all moments exist. Then*

$$E[X^k] = \frac{\partial^k M_X}{\partial \theta^k} \Big|_{\theta=0} \quad k = 1, 2, \dots$$

### Proof.

$$\exp(X\theta) = 1 + X\theta + \frac{X^2\theta^2}{2!} + \dots + \frac{X^k\theta^k}{k!} + \dots$$

Taking expectation on both sides

$$M_X(\theta) = E[\exp(X\theta)] = 1 + E[X]\theta + \frac{E[X^2]\theta^2}{2!} + \dots + \frac{E[X^k]\theta^k}{k!} + \dots$$



# Distribution and moments of a random variable

## Transform methods: theorems

The corresponding properties for the characteristic function  $N_X$ , the Laplace - Stieltjes transform  $L_X$  and the  $z$  transform  $G_X$  are

$$E[X^k] = (-i)^k \frac{\partial^k N_X}{\partial \tau^k} \Big|_{\tau=0} \quad k = 0, 1, \dots$$

$$E[X^k] = (-1)^k \frac{\partial^k L_X}{\partial s^k} \Big|_{s=0} \quad k = 0, 1, \dots$$

$$E\left[\frac{X!}{(X-k)!}\right] = \lim_{\bar{z} \uparrow 1} \frac{\partial^k G_X}{\partial z^k} \Big|_{z=\bar{z}} \quad k = 0, 1, \dots$$

respectively, where  $\left[\frac{X!}{(X-k)!}\right] = X(X-1)\dots(X-k+1)$ .

# Distribution and moments of a random variable

## Transform methods: theorems and examples

Finally, let  $X$  be a discrete nonnegative integer-valued random variable with  $z$  transform  $G_X$ . The probability mass function of  $X$  can be recovered by taking derivatives of  $G_X$ :

$$p_k = P(X = k) = \frac{1}{k!} \frac{\partial^k G_X}{\partial z^k} \Big|_{z=0}$$

Examples: Let  $X$  be exponentially distributed with parameter  $\lambda$ . Then

$$f_X(x) = \lambda \exp(-\lambda x), \quad x > 0.$$

$$\begin{aligned} L_X(s) &= \int_0^{\infty} \exp(-sx) \lambda \exp(-\lambda x) dx \\ &= \frac{\lambda}{s + \lambda} \int_0^{\infty} (\lambda + s) \exp(-(\lambda + s)x) dx \\ &= \frac{\lambda}{s + \lambda}. \end{aligned}$$

# Distribution and moments of a random variable

## Transform methods: examples

Example (continued):

$$E[X] = (-1) \frac{\partial L_X}{\partial s} \Big|_{s=0} = (-1) \frac{-\lambda}{(\lambda + s)^2} \Big|_{s=0} = \frac{1}{\lambda}.$$

$$E[X^2] = \frac{\partial^2 L_X}{\partial s^2} \Big|_{s=0} = \frac{2\lambda}{(\lambda + s)^3} \Big|_{s=0} = \frac{2}{\lambda^2}.$$

Example: Let  $X$  be a  $n$  trials Binomial distributed random variable with probability of success  $p$ . The  $z$  transform of  $X$  is by definition

$$\begin{aligned} G_X(z) &= E(z^X) = \sum_{k=0}^n z^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= (pz + 1 - p)^n \end{aligned}$$

# Distribution and moments of a random variable

## Transform methods: exercises

Exercise:

Let  $X$  be a Bernoulli distributed random variable with probability of success  $p$ .

- 1 Compute the MGF  $M_X$ .
- 2 Compute skewness and kurtosis of  $X$ .

Exercise:

Let  $X_1, X_2, \dots, X_n$  a sequence of independent Bernoulli distributed random variables.

- 1 Compute the moment generating function of  $Y = \sum_{i=1}^n X_i$ .
- 2 Show that  $M_Y$  is the MGF of a Bernoulli  $(n, p)$  distributed random variable.



# Distribution and moments of a random variable

## Transform methods: exercises

Exercise:

Let  $X$  be a standard normally distributed random variable.

- 1 Compute the MGF of  $X$ .
- 2 Compute the Kurtosis of  $X$ .
- 3 Define  $Y = \sigma X + \mu$ . Derive the MGF of  $Y$  and compute its expected value and variance.

Exercise:

Let  $X$  be a geometric distributed random variable with probability mass function  $p_X(i) = p(1-p)^i, i = 1, 2, \dots$

- 1 Compute the  $z$  transform of  $X$ .
- 2 Compute the skewness of  $X$ .

# Distribution and moments of a random variable

Transform methods: exercises

Exercise:

Let  $X$  be a continuous  $\text{Unif}(a, b)$  distributed random variable with  $0 \leq a < b$ .

- 1 Compute the MGF and the Laplace - Stieltjes transform of  $X$ .
- 2 Compute skewness and kurtosis of  $X$ .

# Mean time to failure

Let  $X$  denote the lifetime of a component. The reliability of the component is defined to be  $R(t) = P(X > t) = 1 - F(t)$  so that  $R'(t) = -f(t)$ .

## Definition

The expected life or the mean time to failure (MTTF) of the component is given by

$$E[X] = \int_0^{\infty} tf(t)dt = - \int_0^{\infty} tR'(t)dt.$$

Because of Property *vii*) of the Expectation

$$E[X] = \int_0^{\infty} (1 - F(t))dt - \int_{-\infty}^0 F(t)dt$$

we obtain the alternative expression

$$E[X] = \int_0^{\infty} R(t)dt.$$

# Mean time to failure

More generally, using partial integration, we derive the following general result.

$$\begin{aligned} E[X^k] &= \int_0^{\infty} t^k f(t) dt \\ &= - \int_0^{\infty} t^k R'(t) dt \\ &= -t^k R(t) \Big|_0^{\infty} + \int_0^{\infty} k t^{k-1} R(t) dt. \end{aligned}$$

Because  $\lim_{t \rightarrow \infty} t^k R(t) = 0$  we obtain the final formula

$$E[X^k] = \int_0^{\infty} k t^{k-1} R(t) dt.$$

In particular

$$\text{Var}(X) = \int_0^{\infty} 2tR(t)dt - \left[ \int_0^{\infty} R(t)dt \right]^2.$$

# Mean time to failure: Series System

The system reliability of a series system with  $n$  independent components is

$$R(t) = \prod_{i=1}^n R_i(t)$$

where  $R_i(t)$  denotes the reliability of the  $i$ th component. The MTTF of a series system is much smaller than the MTTF of its components. If  $X_i$  denotes the lifetime of component  $i$  and  $X$  the system lifetime, since  $0 \leq R_i(t) \leq 1$ ,

$$R_X(t) = \prod_{j=1}^n R_{X_j}(t) \leq R_{X_i}(t) \forall i$$

so that

$$\begin{aligned} E[X] &= \int_0^{\infty} R_X(t) dt \leq \int_0^{\infty} R_{X_i}(t) dt \forall i \\ &\leq E[X_i] \forall i \\ &\leq \min_i \{E[X_i]\} \end{aligned}$$

# Mean time to failure: Parallel System

Because  $X = \max\{X_1, X_2, \dots, X_n\}$  and the components are independent

$$F_X(t) = \prod_{i=1}^n F_{X_i}(t) = \prod_{i=1}^n [1 - R_{X_i}(t)].$$

The system reliability of a parallel system with  $n$  independent components is

$$R_X(t) = 1 - F_X(t) = 1 - \prod_{i=1}^n [1 - R_{X_i}(t)] \geq 1 - [1 - R_{X_i}(t)] \quad \forall i.$$

This implies that the reliability of a parallel redundant system is larger than that of any of its components. Therefore

$$E[X] = \int_0^{\infty} R_X(t) dt \geq \max_i \{E[X_i]\}.$$

# Mean time to failure: Standby Redundancy

- The system has one component operating and  $(n - 1)$  unpowered spares.
- The failure rate of an operating component is  $\lambda$ . A cold spare does not fail.
- The switching equipment is failure free.
- Define  $X_i$  to be the lifetime of the  $i$ th component from the point it is put into operation until its failure.

The system lifetime  $X$  is

$$X = \sum_{i=1}^n X_i,$$

$X$  has an  $n$ -stage Erlang distribution and

$$E[X] = \frac{n}{\lambda} \text{ and } Var[X] = \frac{n}{\lambda^2}.$$