Statistics in Geophysics: Probability Theory II

Steffen Unkel

Department of Statistics
Ludwig-Maximilians-University Munich, Germany

Random variables

- In many experiments it is easier to deal with a summary variable than with the original probability structure.
- Example: In an opinion poll, we might to decide to ask 50 people whether they agree or disagree with a certain issue.
- The sample space for this experiment has 2^{50} elements.
- Define a variable $X =$ number of 1s recorded out of 50.
- The sample space for X is the set of integers $\{0, 1, 2, \ldots, 50\}$.

Random variables

Definition:

A random variable is a function from a sample space Ω into the real numbers.

- We have also defined a new sample space (the range of the random variable).
- Suppose we have a sample space $\Omega = \{\omega_1, \dots, \omega_n\}$ with a probability function P and we define a random variable X with range $\mathcal{X} = \{x_1, \dots, x_m\}$.

Random variables

- We define an induced probability function P_X on $\mathcal{X}$ as follows:

$$\mathrm{P}_X(X = x_i) = \mathrm{P}(\{\omega_j \in \Omega : X(\omega_j) = x_i\}) \ ,$$

  for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

- We will simply write $\mathrm{P}(X = x_i)$ rather than $\mathrm{P}_X(X = x_i)$.

Example: Tossing a fair coin three times

X: number of heads obtained in the three tosses.

ω	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(\omega)$	3	2	2	2	1	1	1	0

Table: Enumeration of the value of X for each point in the sample space.

x	0	1	2	3
$\mathrm{P}(X = x)$	$1/8$	$3/8$	$3/8$	$1/8$

Table: Induced probability function on $\mathcal{X}$.

For example, $\mathrm{P}(X = 1) = \mathrm{P}(\{\text{HTT,THT,TTH}\}) = \frac{3}{8}$.

Distribution function

Definition:

The cumulative distribution function or cdf of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = \mathrm{P}(X \leq x), \quad \text{for all } x \ .$$

Example (Tossing three coins): The cdf of X is

$$F_X(x) = \begin{cases} 0 & \text{if } \infty < x < 0 \\ \frac{1}{8} & \text{if } 0 \leq x < 1 \\ \frac{1}{2} & \text{if } 1 \leq x < 2 \\ \frac{7}{8} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } 3 \leq x < \infty \ . \end{cases}$$

Properties of a cdf

The function $F_X(x)$ is a cdf if and only if the following three conditions hold:

1. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.
2. $F_X(x)$ is a monotone, non-decreasing function of x.
3. $F_X(x)$ is continuous from the right; that is,

$$\lim_{0<h\to 0} F_X(x+h) = F_X(x) \ .$$

A cdf can have jumps or it can be continuous. An example of a continuous cdf is the function

$$F_X(x) = \frac{1}{1+e^{-x}} \ .$$

Density and mass functions

Definition:

A random variable X is continuous if $F_X(x)$ is a continuous function of x. A random variable is discrete if $F_X(x)$ is a step function of x.

Associated with a random variable X and its cdf $F_X(x)$ is another function, called either the probability density function (pdf) or probability mass function (pmf).

Probability mass function

Definition:

The probability mass function (pmf) of a discrete random variable X is given by

$$f_X(x) = \mathrm{P}(X = x) \quad \text{for all } x \ .$$

Hence, for positive integers a and b with $a \leq b$, we have

$$\mathrm{P}(a \leq X \leq b) = \sum_{k=a}^{b} f_X(k) \ .$$

As a special case of this we get $\mathrm{P}(X \leq b) = F_X(b)$.

Probability density function

- A pmf gives us "point probabilities" and we can sum over the values of the pmf to get the cdf.
- The analogous procedure in the continuous case is to substitute integrals for sums:

$$\mathrm{P}(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(t)\, \mathrm{d}t \ .$$

- If $f_X(x)$ is continuous, we have the further relationship

$$\frac{d}{dx} F_X(x) = f_X(x) \ .$$

Probability density function

Definition:

The probability density function (pdf) $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_x(t)\,\mathrm{d}t \quad \text{for all } x \ .$$

Since $\mathrm{P}(X = x) = 0$,

$$\mathrm{P}(a < X < b) = \mathrm{P}(a < X \le b) = \mathrm{P}(a \le X < b) = \mathrm{P}(a \le X \le b).$$

Example

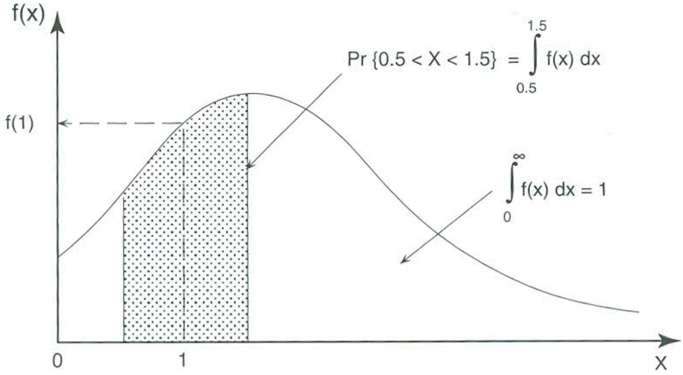


Figure: Hypothetical pdf for a non-negative random variable X.

Properties of a pdf (or pmf)

A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

1. $f_X(x) \geq 0$ for all x.
2. $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$ (pdf).

Remark: In the sequel, we will frequently have to state that a random variable has a certain distribution. We will make such a statement by giving either the cdf or the pdf (or pmf) of the random variable of interest.

Mean of a random variable

Definition:

Let X be a random variable. The expected value or mean of X, denoted by $\mathrm{E}(X)$ (or μ_X), is (provided that the sum or integral exists):

$$\text{(i)} \qquad \mathrm{E}(X) = \sum_{x \in \mathcal{X}} x\, f_X(x) = \sum_{x \in \mathcal{X}} x\, \mathrm{P}(X = x) \ ,$$

if X is discrete, or

$$\text{(ii)} \qquad \mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, \mathrm{d}x \ ,$$

if X is continuous.
If $\mathrm{E}(X) = \infty$, we say that $\mathrm{E}(X)$ does not exist.

Expected value of a function of a random variable

Let $g(x)$ be a function of a random variable X. Then (provided that the sum or integral exists),

$$\text{(i)} \qquad \mathrm{E}(g(X)) = \sum_{x \in \mathcal{X}} g(x)\, f_X(x) \ ,$$

if X is discrete, or

$$\text{(ii)} \qquad \mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, \mathrm{d}x \ ,$$

if X is continuous.

Properties of expected values

If X is any random variable, then (as long as the expectations exist):

1. $\mathrm{E}(c) = c$ for a constant c.
2. $\mathrm{E}(c\, g(X)) = c\mathrm{E}(g(X))$.
3. $\mathrm{E}(c_1\, g_1(X) + c_2\, g_2(X)) = c_1\mathrm{E}(g_1(X)) + c_2\mathrm{E}(g_2(X))$.
4. $\mathrm{E}(g_1(X)) \leq \mathrm{E}(g_2(X))$ if $g_1(x) \leq g_2(x)$ for all x.

Variance of a random variable

Definition:

Let X be a random variable. The variance of X, denoted by $\mathrm{Var}(X)$ (or σ_X^2), is

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

The positive square root of $\mathrm{Var}(X)$ is the standard deviation of X.

Linear transformations of random variables

Assume X is a random variable with mean μ_X and variance σ_X^2. If $Y = aX + b$, where a and b are any constants, then

$$\mu_Y = a\mu_X + b, \quad \sigma_Y^2 = a^2\sigma_X^2, \quad \sigma_y = |a|\sigma_X \ .$$

Introduction

- Statistical distributions are used to model populations.
- We usually deal with a family of distributions, which is indexed by one or more parameters.
- Here, we catalog some of the frequently occurring probability laws and examine the assumed chance mechanisms that lead to their usage.
- This presentation is by no means comprehensive in its coverage of statistical distributions!

Binomial distribution

- The binomial distribution is based on the idea of a Bernoulli trial.
- A random variable X is defined to have a Bernoulli distribution, denoted by $X \sim \mathcal{B}(\pi)$, if

$$X = \begin{cases} 1 & \text{with probability } \pi \\ 0 & \text{with probability } 1-\pi \end{cases}, \qquad (0 \le \pi \le 1) \ .$$

- Consider a sequence of n identical, independent Bernoulli trials, $X_1, X_2, \ldots, X_n$, each with success probability π.
- $X_1, \ldots, X_n$ are independent if $P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i)$.

Binomial distribution

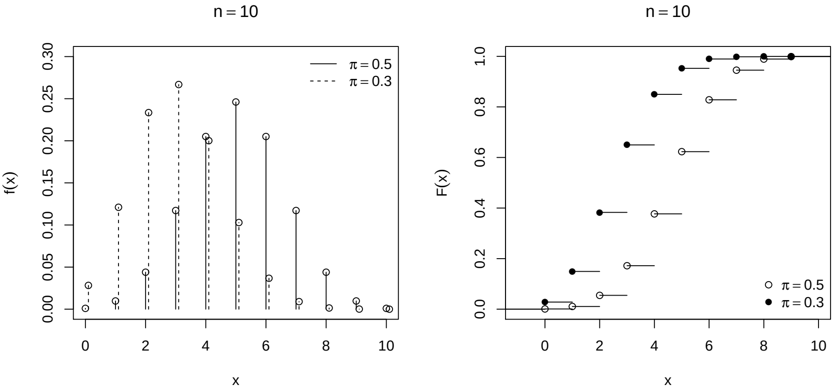
- Let X count the number of successes observed in a sequence of n identical and independent Bernoulli trials, that is, $X := \sum_{i=1}^{n} X_i$.

- Then, X has a binomial distribution, denoted by $X \sim \mathcal{B}(n, \pi)$, with pmf

$$f_X(x) = \mathrm{P}(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \ , \text{ for } x = 0, 1, \ldots, n.$$

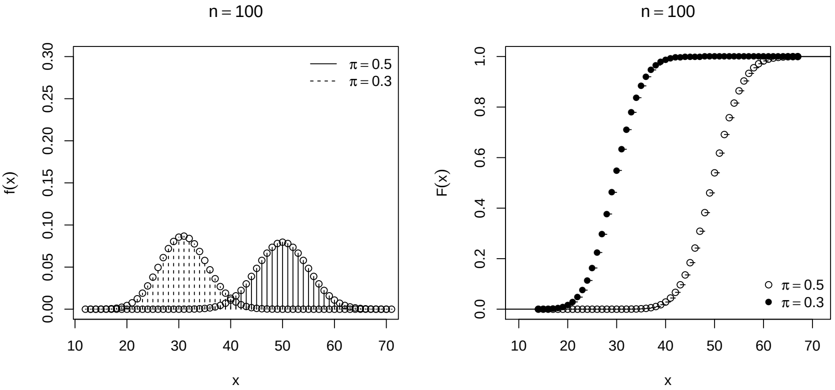
- If $X \sim \mathcal{B}(n, \pi)$, then $\mathrm{E}(X) = n\pi$ and $\mathrm{Var}(X) = n\pi(1 - \pi)$.

Binomial distribution



Pmf (left) and cdf (right) for $X \sim \mathcal{B}(n, \pi)$ with $n = 10$ and two different choices of π.

Binomial distribution



Pmf (left) and cdf (right) for $X \sim \mathcal{B}(n, \pi)$ with $n = 100$ and two different choices of π.

Poisson distribution

- The Poisson distribution describes the number of events occurring in a certain time interval and so pertains to data on counts.

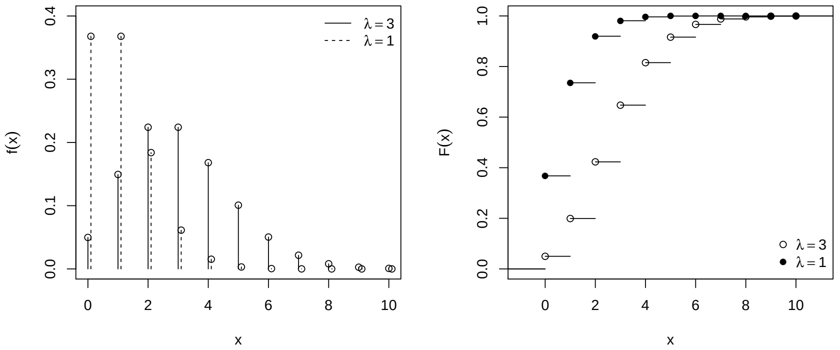
- A random variable X is defined to have a Poisson distribution, denoted by $X \sim \mathcal{P}(\lambda)$, if the pmf of X is given by

$$f_X(x) = \mathrm{P}(X = x) = \frac{\lambda^x}{x!} \cdot \exp(-\lambda) \ , \quad \text{for } x = 0, 1, \dots \ ,$$

  where the parameter $\lambda > 0$ is called the intensity and has physical dimensions of occurrences per unit time.

- If $X \sim \mathcal{P}(\lambda)$, then $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$.

Poisson distribution



Pmf (left) and cdf (right) for $X \sim \mathcal{P}(\lambda)$ with $\lambda = 1$ and $\lambda = 3$.

Example: Annual Hurricane Landfalls on the U.S. coastline

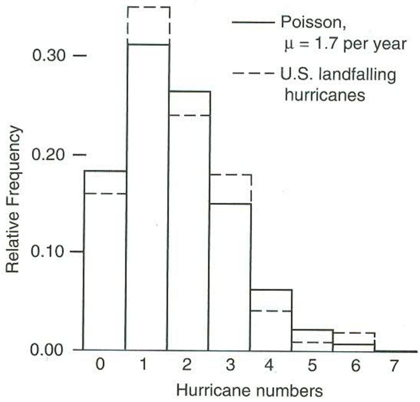
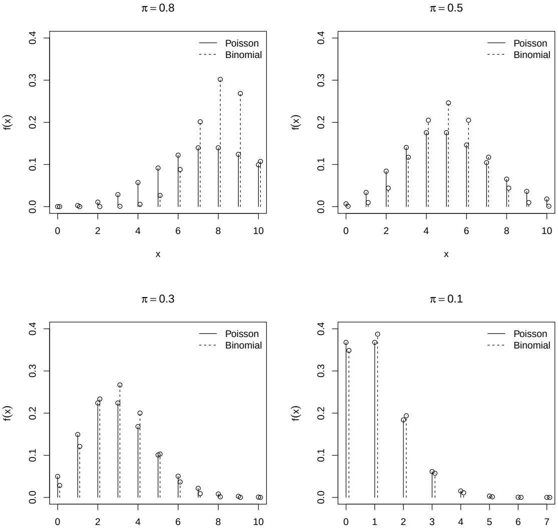
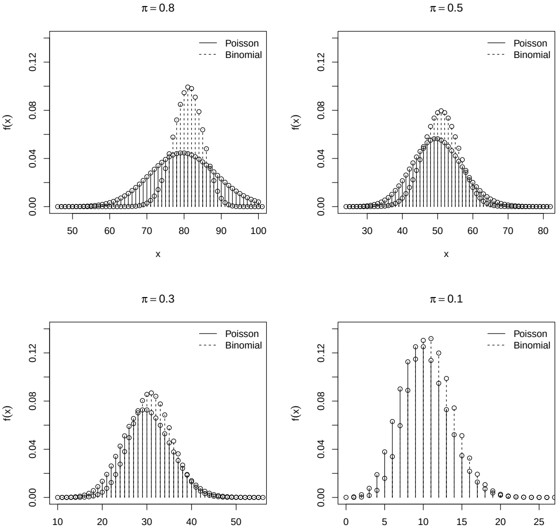


Figure: Histogram of annual numbers of U.S. landfalling hurricanes for 1899-1998 (dashed), and fitted Poisson distribution with $\lambda = 1.7$ (solid).

Approximation of the binomial distribution



Approximation of the binomial distribution



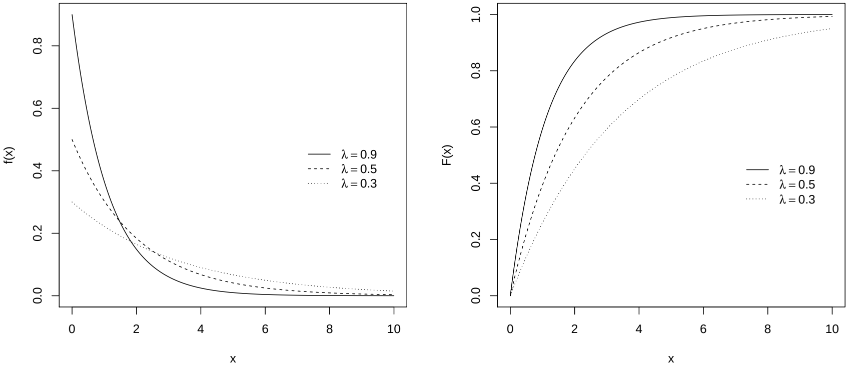
Exponential distribution

- The exponential distribution can be used to model lifetimes.
- If X is a continuous random variable with non-negative range, which has pdf

$$f_X(x) = \lambda \exp(-\lambda x) \ , \quad \text{for } x \geq 0 \ ,$$

  where $\lambda > 0$, then X is defined to have an exponential distribution, denoted by $X \sim \mathcal{E}(\lambda)$.
- If $X \sim \mathcal{E}(\lambda)$, then $\mathrm{E}(X) = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.
- If the number of events in the unit time interval follows a Poisson distribution with mean λ, then the time to the next event is exponentially distributed with mean $1/\lambda$ (Poisson process).

Exponential distribution



Pdf (left) and cdf (right) for $X \sim \mathcal{E}(\lambda)$ with different intensities λ.

Normal distribution

- The normal distribution plays a central role in statistics and has many applications.

- A random variable X, is defined to be normally distributed, denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$, if its density is given by

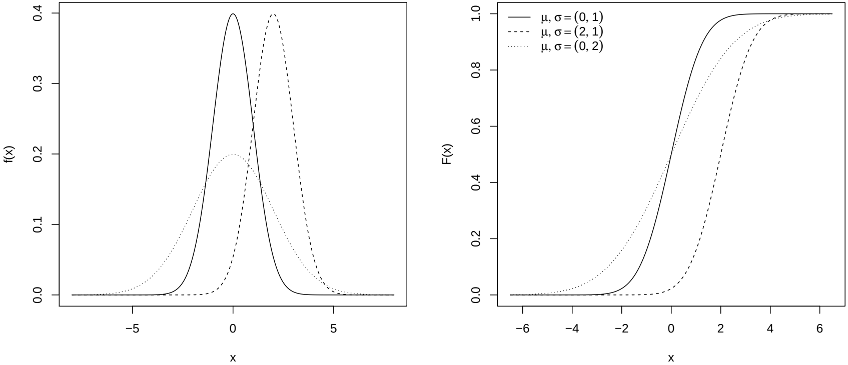
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \quad , \quad -\infty < x < \infty \ ,$$

  where $-\infty < \mu < \infty$ and $\sigma > 0$.

- If X is a normal random variable, then $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

- Integration of the normal density is analytically not tractable.

Normal distribution



Pdf (left) and cdf (right) of the normal distribution for different values of μ and σ.

Standard normal distribution

Normal distribution having $\mu = 0$ and $\sigma = 1$:

The pdf simplifies to

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \quad .$$

$\Phi(z) = \int_{-\infty}^{z} \phi(u) \mathrm{d}u = \mathrm{P}(Z \leq z)$ is the conventional notation for its cdf.

Any Gaussian random variable can be standardized by subtracting its mean and dividing by its standard deviation:

$$z = \frac{x - \mu}{\sigma} \quad .$$

Distributions of functions of a random variable

- Let X be a continuous random variable with density $f_X(x)$ and let $Y = g(X)$ be a strongly monotone and differentiable function.

- The density $f_Y(y)$ of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \underbrace{\left| \frac{dg^{-1}(y)}{dy} \right|}_{g^{-1'}(y)} \,,$$

where the inverse function $g^{-1}(y)$ gives the value of x for which $g(x) = y$.

Lognormal distribution

- Assume $X \sim N(\mu, \sigma^2)$.
- $Y = \exp(X)$ has a log-normal distribution with parameters μ and σ^2, denoted by $Y \sim \mathcal{LN}(\mu, \sigma^2)$, with pdf

$$f_Y(y) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(\ln(y)-\mu)^2}{\sigma^2}\right)}_{f_X(g^{-1}(y))} \cdot \underbrace{\frac{1}{y}}_{\frac{dg^{-1}(y)}{dy}}$$

  for $y > 0$ and zero elsewhere.
- If Y is a log-normal random variable, then

$$\begin{aligned} \mathrm{E}(Y) &= \exp\left(\mu + \frac{1}{2}\sigma^2\right) \\ \mathrm{Var}(Y) &= \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1] \; . \end{aligned}$$

Vector of random variables

- We need to know how to describe and use probability models that deal with more than one random variable at a time (called multivariate models).
- We will focus on bivariate models involving two random variables.
- A bivariate random vector (X, Y) associates an ordered pair of real numbers, that is, a point (x, y), with each experimental outcome.
- Example (tossing two fair dice): With each of the 36 possible outcomes associate two numbers, X and Y, e.g. let $X =$ sum of the two dice and $Y = |$difference of the two dice$|$.

Joint and marginal distributions

- The two cases we will discuss are those in which (X,Y) is discrete or in which (X, Y) is continuous.

- When (X, Y) is discrete, the joint pmf is

$$f_{X,Y}(x, y) = \mathrm{P}(X = x, Y = y) \ ,$$

  where $f_{X,Y}(x, y) \geq 0$ for all (x, y) and must sum to 1, if we add over all possible observed vectors.

- The marginal pmfs of X and Y, $f_X(x) = \mathrm{P}(X = x)$ and $f_Y(y) = \mathrm{P}(Y = y)$, are given by

$$f_X(x) = \sum_y f_{X,Y}(x, y) \qquad \text{and} \qquad f_Y(y) = \sum_x f_{X,Y}(x, y) \ .$$

Joint and marginal distributions

- The joint cdf of two random variables, $F_{X,Y}(x,y)$ is $F_{X,Y}(x,y) = \mathrm{P}(X \leq x, Y \leq y)$.

- When (X,Y) is continuous, the joint pdf can be defined as the function that satisfies

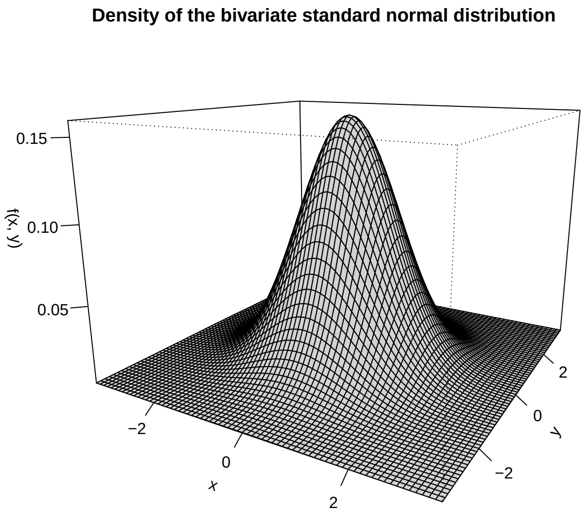
$$F_{X,Y}(x,y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f(u,v)\,\mathrm{d}u\,\mathrm{d}v \quad , \quad \forall\, x,y \in \mathbb{R} \;,$$

  where $f_{X,Y}(x,y) \geq 0$ for all (x,y) and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$.

- The marginal pdfs of X and Y, $f_X(x)$ and $f_Y(y)$, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x \;.$$

Example of a joint density



Conditional distributions

Definition:

When (X, Y) is discrete, the conditional pmf for X given $Y = y$ is

$$f_{X|Y} = \mathrm{P}(X = x|Y = y) = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad .$$

When (X, Y) is continuous, the conditional pdf for X given $Y = y$ is

$$f_{X|Y} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad .$$

Conditional pmf and pdf are defined for any y such that $f_Y(y) > 0$.

Independence

Definition:

Let (X, Y) be a bivariate random vector with joint pdf or pmf $f_{X,Y}(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called independent if, for every $x, y \in \mathbb{R}$,

$$f_{X,Y} = f_X(x) f_Y(y) \ .$$

If X and Y are independent, then

$$f_{X|Y}(x|y) = f_X(x) \qquad \text{and} \qquad f_{Y|X}(y|x) = f_Y(y) \ .$$

Expectation

Definition:

The expected value of a function $g(X, Y)$ of the random vector (X, Y), denoted by $\mathrm{E}(g(X, Y))$, is

$$\mathrm{E}(g(X, Y)) = \sum g(x, y) f_{X,Y}(x, y) \ ,$$

if (X, Y) is discrete, where the summation is over all possible values of (X, Y), and

$$\mathrm{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) \, \mathrm{d}x \, \mathrm{d}y \ ,$$

if (X, Y) is continuous.

Covariance and correlation

The covariance of X and Y is the number defined by

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] = \mathrm{E}(XY) - \mu_X \mu_Y \; .$$

The correlation of X and Y is the number defined by

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} \; , \quad -1 \leq \rho_{XY} \leq 1 \; .$$

The value ρ_{XY} is also called the correlation coefficient. X and Y are called uncorrelated if $\rho_{XY} = 0$; they are positively (negatively) correlated if $\rho_{XY} > 0$ ($\rho_{XY} < 0$).

Properties of covariance and correlation

The following statements hold:

1. If X and Y are independent random variables, then $\mathrm{Cov}(X, Y) = 0$ and $\rho_{XY} = 0$.
2. If X and Y are any two random variables and a and b are any two constants, then
$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y) \ .$$
3. If X and Y are independent random variables, then
$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \ .$$

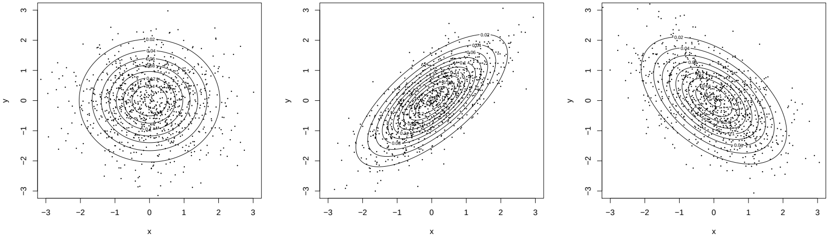
Example: The bivariate standard normal distribution

The bivariate standard normal distribution with parameter ρ ($|\rho| < 1$) has the joint density

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

- The marginal distributions of X and Y are (for any ρ) standard normally distributed.
- The correlation of X and Y is ρ.
- In this case: Uncorrelatedness implies independence.

Example: The bivariate standard normal distribution



Contour plots of the joint density of the bivariate standard normal distribution, obtained using 500 samples for $\rho = 0$ (left), $\rho = 0.7$ (middle) und $\rho = -0.5$ (right).

Sums of random variables

- If X and Y are independent random variables with pmfs or pdfs $f_X(x)$ and $f_Y(y)$, then the pmf or pdf of $Z = X + Y$ is

$$f_Z(z) = \mathrm{P}(X + Y = z) = \sum_y f_X(z - y) f_Y(y) \ ,$$

  if X and Y are discrete and

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) \, \mathrm{d}y \ ,$$

  if X and Y are continuous.
- The function $f_Z(z)$ is called the convolution of $f_X(x)$ and $f_Y(y)$.

- Example: $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and independent, then $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Law of large numbers

Consider independently and identically distributed (i.i.d) random variables $X_1, X_2, \ldots, X_n$ with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$ $(i = 1, \ldots, n)$.

If we define

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

it can be shown that $\mathrm{E}(\bar{X}_n) = \mu$ and $\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\sigma^2$.

The law of large numbers states that

$$\mathrm{P}\left(\lim_{n\to\infty} \left|\bar{X}_n - \mu\right| \geq \epsilon\right) = 0 \ ,$$

for every $\epsilon > 0$.

Central limit theorem

A random variable X with mean $\mu = \mathrm{E}(X)$ and variance $\sigma^2 = \mathrm{Var}(X)$ can be linearly transformed, such that the transformed variable $\tilde{X}$ has zero mean and unit variance:

$$\tilde{X} = \frac{X - \mu}{\sigma} \ .$$

Then,

$$\begin{aligned} \mathrm{E}(\tilde{X}) &= \frac{1}{\sigma}(\mathrm{E}(X) - \mu) = 0 \ , \\ \mathrm{Var}(\tilde{X}) &= \frac{1}{\sigma^2}\mathrm{Var}(X) \quad = 1 \ . \end{aligned}$$

Central limit theorem

Consider i.i.d random variables $X_1, X_2, \ldots, X_n$ with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ $(i = 1, \ldots, n)$.
For the sum $Y_n = X_1 + X_2 + \ldots + X_n$ it holds that

$$\mathrm{E}(Y_n) = n \cdot \mu \quad \text{and} \quad \mathrm{Var}(Y_n) = n \cdot \sigma^2 \ .$$

For the standardised sum

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n} \cdot \sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \ ,$$

it therefore holds that $\mathrm{E}(Z_n) = 0$ und $\mathrm{Var}(Z_n) = 1$.

The central limit theorem states that

1. $Z_n \overset{a}{\sim} \mathcal{N}(0, 1)$.
2. $Y_n \overset{a}{\sim} \mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$.