Random Variables and Their Distributions

Bo Li, Spring 2019

(I) Let \((\Omega, \mathfrak{A}, P)\) be a probability space (i.e., \(\Omega \neq \phi\) is a set, called the sample space and \(w \in \Omega\) is called a sample point or elementary event; \(\mathfrak{A}\) is a \(\sigma\)-algebra, consisting of subsets of \(\Omega\) and each member in \(\mathfrak{A}\) is called an event; and \(P\) is the probability measure).

Definition. A (real-valued, scalar) random variable (RV) is a mapping \(X : \Omega \to \mathbb{R}\) such that

\[
\{X \leq x\} := \{w \in \Omega : X(w) \leq x\} \in \mathfrak{A} \quad \forall x \in \mathbb{R}
\]

Equivalently, \(X(B) := \{w \in \Omega : X(w) \in B\} \in \mathfrak{A}\)

for any \(B \subseteq \mathbb{R}\), where \(\mathfrak{B}\) is the Borel \(\sigma\)-algebra of \(\mathbb{R}\)

Definition. The expectation (or mean) of a RV \(X : \Omega \to \mathbb{R}\)

\[
E(X) = \int X \, dP
\]

is if the integral exists.

The variance of \(X : \Omega \to \mathbb{R}\) is

\[
\text{Var}(X) = E((X - E(X))^2) = 0 \quad \text{if} \quad E|X|^2 < \infty
\]

Properties (Prove all these!)

1. \(E(aX + bY) = aE(X) + bE(Y)\), \(a, b \in \mathbb{R}, X, Y: \text{RVs}\)

2. \(\text{Var}(X) = E(X^2) - (E(X))^2\) (Important)

3. \(\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)\), if \(X, Y\) are independent

Also, \(E1 = 1\), \(EX \geq 0\) if \(X \geq 0\) on \(\Omega\).
Definition: A RV $X: \mathcal{X} \to \mathbb{R}$ is discrete if the range of $X$, $X(\mathcal{X}) = \{X(\omega) : \omega \in \Omega\}$, is finite or countably infinite.
If $X$ is a discrete RV, then there exist finitely many or countably infinitely many distinct $x_1, x_2, \ldots \in \mathbb{R}$ such that
\[ X(\omega) = \{ x_1, x_2, \ldots \} . \]
For each $i$, let $\mathcal{N}_i = \{ \omega \in \Omega : X(\omega) = x_i \} = X^{-1}(\{ x_i \})$.
Then $\mathcal{N} = \bigcup_i \mathcal{N}_i$, $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$ ($i \neq j$).
\[ X = \sum_i x_i \mathbb{1}_{\mathcal{N}_i} . \]

Here and below, for any $A \subseteq \Omega$, $\mathbb{1}_A : \mathcal{N} \to \mathbb{R}$ is the indicator function of $A$,
\[ \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A . \end{cases} \]

Definition: The probability mass function (PMF), or the discrete probability distribution function (DPDF), or simply discrete density function (DDF), of a discrete RV $X : \Omega \to \mathbb{R}$ is a function $f_X : \mathbb{R} \to [0, 1]$ defined by
\[ f_X(x) = \mathbb{P}(X = x) . \]

Proposition

1. If $X(\Omega) = \{ x_1, x_2, \ldots \}$ then
   \[ F_X(x) = \sum_{x_i \leq x} f(x_i) . \]
2. In general
   \[ f_X(x) = F_X(x) - F_X(x^-) . \]

Try to prove these except 4.

On the next page.
Proposition: The PMF \( f_x: \mathbb{R} \to [0, 1] \) of a discrete RV \( X \) with \( X(x_i) = \{x_1, x_2, \ldots \} \) satisfies

1. \( f_x(x_i) > 0 \iff x_i \in \{x_1, x_2, \ldots \} \).
2. \( \sum_i f_x(x_i) = 1 \).

Moreover, 1, 2 characterize a PMF.

Calculations of the mean and variance of a discrete RV \( X \) can be done through its PMF \( f_x \).

Suppose \( X(x_i) = \{x_1, x_2, \ldots \} \). Then

1. \( E(X) = \sum_i x_i f_x(x_i) = \sum_i x_i \cdot x_i \cdot f_x(x_i) \)
2. \( E(g(X)) = \sum_i g(x_i) \cdot f(x_i) \)

Here, we assume the sums, if they are infinite, converge absolutely.

(III) Definition: A RV \( X: \mathbb{R} \to \mathbb{R} \) is continuous (i.e., \( X \) is a continuous RV) if there exists \( f_x: \mathbb{R} \to [0, \infty) \) s.t.

\[
F_x(x) = \int_{-\infty}^{x} f_x(t) \, dt \quad \forall x \in \mathbb{R}.
\]

We call \( f_x \) is probability density function (PDF) of the RV. Or, just the density

Proposition: \( f_x(x) = F_x'(x) \) at \( x \) at which \( f_x \) is cont.

\[
f_x(x) > 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_x(x) \, dx = 1.
\]
\( P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(t) \, dt \)
\( \forall a, b \in \mathbb{R}, \ a < b \)

\( P(X \in B) = \int_B f_X(x) \, dx \) for any Borel set \( B \subset \mathbb{R} \)

where \( \mathcal{B} \) is the Borel \( \sigma \)-algebra of \( \mathbb{R} \)

\( P(x = x) = 0 \quad \forall x \in \mathbb{R} \)

Moreover, any \( h : \mathbb{R} \rightarrow [0, \infty) \), with \( \int_{-\infty}^{\infty} h(x) \, dx = 1 \),
can serve as a PDF for some continuous RV.

For a continuous RV, \( X \), we can calculate its expectation and variance using the PDF, \( f_X \). We have
\[
E(X) = \int_{-\infty}^{\infty} x f_X(x) \, dx
\]
\[
E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \quad \text{if } g \text{ is a continuous function}
\]

Here, we assume the integrals converge absolutely.

(IV) Examples of Distributions of Discrete RVs

1. Bernoulli variables and Bernoulli distribution.
   A discrete RV, \( X \), is called a Bernoulli variable if \( X(\{0,1\}) = \{0,1\} \).

Let \( p = P(X = 1) \) and \( q = P(X = 0) = 1 - p \). (0 < \( p \) < 1)

Then the PMF, \( f_X \), is given by
\[
f_X(x) = 0 \text{ if } x \neq 0,1 \quad f_X(0) = q \quad f_X(1) = p
\]
This is the Bernoulli distribution: Bernoulli (\( p \))
Note: The CDF is \( F_X(x) = \sum_{k \leq x} f_X(k) = \begin{cases} 0 & \text{if } x < 0, \\ 1-p & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x \geq 1. \end{cases} \)

We have 
\[
E(X) = p, \quad \text{Var}(X) = p(1-p).
\]

If \( A \subseteq \Omega \), then the indicator variable \( I_A \)

is a Bernoulli variable.

\[
E(I_A) = P(I_A = 1) = P(A) \quad \text{Var}(I_A) = P(A)(1-P(A))
\]

We write \( X \sim \text{Bernoulli}(p) \) to indicate that \( X \) is a Bernoulli variable and \( P(X = 1) = p \).

2. **Binomial distribution** \( B(n, p) \) \((n \in N, 0 < p < 1)\)

A discrete RV \( X : \mathbb{Z} \rightarrow \{0, 1, \ldots, n\} \) is said to have the binomial distribution with parameters \( n, p \), written \( X \sim B(n, p) \), if the PMF \( f_X(x) : \mathbb{R} \rightarrow [0, 1] \) is given by

\[
f_X(x) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{if } k \in \{0, 1, \ldots, n\}
\]

It is clear that \( f_X \geq 0 \) and

\[
\sum_{k=0}^{n} f_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1
\]

The CDF is 
\[
F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } 0 \leq x < n, \\ 1 & \text{if } x \geq n. \end{cases}
\]

If \( X \sim B(n, p) \), we say \( X \) is a binomial variable.

We have for \( X \sim B(n, p) \).

\[
E(X) = np \quad \text{and} \quad \text{Var}(X) = np(1-p) = npq
\]

The CDF is 
\[
F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } 0 \leq x < n, \\ 1 & \text{if } x \geq n. \end{cases}
\]
3. Poisson variable and Poisson distribution $P(W)$ with parameter $\lambda > 0$.

A discrete RV $X$ is a Poisson variable if it has a Poisson distribution with some parameter $\lambda > 0$, written $X \sim \text{Poisson}(\lambda)$ or $P(\lambda)$.

\[ X: \lambda \rightarrow \{0, 1, 2, \ldots \} \]

\[ f_X(x) = \begin{cases} 0 & \text{if } x \notin \{0, 1, 2, \ldots \} \\ \frac{\lambda^x e^{-\lambda}}{x!} & \text{if } x \in \{0, 1, 2, \ldots \} \end{cases} \]

Note that

\[ \sum_{k=0}^{\infty} f_X(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1. \]

We have

\[ F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sum_{k=0}^{x-1} \frac{\lambda^k e^{-\lambda}}{k!} & \text{if } 0 \leq x \leq \infty \end{cases} \]

\[ F_X(x) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} e^\lambda = \lambda \]

$E(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Prove this.

4. Geometric variable and geometric distribution with parameter $p \in (0, 1)$.

We say a discrete RV $X: \omega \rightarrow \{1, 2, \ldots \}$ is a geometric variable, with the geometric distribution with parameter $p \in (0, 1)$, written $X \sim \text{Geometric}(p)$ if the PMF of $X$ is

\[ f_X(x) = \begin{cases} 0 & \text{if } x \notin \{1, 2, \ldots \} \\ \sum_{n=0}^{x-1} p(1-p)^n & \text{if } x \in \{1, 2, \ldots \} \end{cases} \]

\[ f_X(x) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots \]

We have

\[ E(X) = \frac{1}{p} \quad \text{and} \quad \text{Var}(X) = \frac{1-p}{p^2}. \]

Prove these!
5. The discrete uniform distribution over \( \{1, 2, \ldots, n\} \).

A discrete RV \( X: \mathbb{R} \to \{1, 2, \ldots, n\} \) with such a distribution has the PMF

\[
  f_X(x) = \begin{cases} 
    0 & \text{if } x \notin \{1, 2, \ldots, n\} \\
    \frac{1}{n} & \text{if } x \in \{1, 2, \ldots, n\}.
  \end{cases}
\]

The CDF is

\[
  F_X(x) = \begin{cases} 
    0 & \text{if } x < 0 \\
    \frac{\lfloor x \rfloor}{n} & \text{if } 0 \leq x < n \\
    1 & \text{if } x \geq n
  \end{cases}
\]

where \( \lfloor x \rfloor \) is the largest integer \( \leq x \).

(V) Examples of Distributions of Continuous RVs

1. The uniform distribution on \([a, b]\) \((a, b \in \mathbb{R}, a < b)\)

A continuous RV \( X: \mathbb{R} \to \mathbb{R} \) with such a distribution, written \( X \sim U[a, b] \), if

\[
  F_X(x) = \begin{cases} 
    0 & \text{if } x \leq a \\
    \frac{x-a}{b-a} & \text{if } a < x < b \\
    1 & \text{if } x \geq b
  \end{cases}
\]

\[
  f_X(x) = \begin{cases} 
    0 & \text{if } x < a \text{ or } x > b \\
    \frac{1}{b-a} & \text{if } a \leq x \leq b
  \end{cases}
\]

Check these:

\[
  \mathbb{E}(X) = \frac{1}{2}(a+b).
\]

\[
  \text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{(b-a)^2}{12}
\]
2. The normal distribution (or Gaussian distribution) \( N(\mu, \sigma^2) \) \((\mu \in \mathbb{R}, \sigma > 0)\)
A continuous RV \( X \) that has such a distribution, means that
\[
f_X(x) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (-\infty < x < \infty)
\]
\[
F_X(x) = \int_{-\infty}^{x} f_X(t) dt, \quad \forall x \in \mathbb{R}.
\]
Verify that \( \int_{-\infty}^{\infty} f_X(x) dx = 1 \).

Verify there! \( E(X) = \mu \)
\( \text{Var}(X) = \sigma^2 \)

The standard normal distribution is \( N(0, 1) \).

3. The exponential distribution with parameter \( \lambda > 0 \), denoted \( \text{Exp}(\lambda) \).
\( X \sim \text{Exp}(\lambda) \); \( X \) is a continuous RV with the exponential distribution \( \text{Exp}(\lambda) \).
It means:
\[
f_X(x) = \begin{cases} 
0 & \text{if } x < 0 \\
\lambda e^{-\lambda x} & \text{if } x \geq 0
\end{cases}
\]
\[
F_X(x) = \begin{cases} 
0 & \text{if } x < 0 \\
1 - e^{-\lambda x} & \text{if } x \geq 0
\end{cases}
\]
\[
E(X) = \frac{1}{\lambda}, \quad \text{since } E(X) = \int_{0}^{\infty} x f_X(x) dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}
\]
\( \text{Var}(X) = \frac{1}{\lambda^2} \).
4. The Gamma Distribution $\Gamma(\lambda, t)$ ($\lambda, t > 0$)

A RV $X$ has such a distribution, $X \sim \Gamma(\lambda, t)$, if

$$f_X(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} \quad \text{if } x \geq 0$$

$$f_X(x) = 0 \quad \text{if } x < 0$$

where $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} \, dx$ is the Gamma function.

$t = 1$. This is $\text{Exp}(\lambda)$.

5. The Beta Distribution $B(a, b)$ ($a, b > 0$)

A RV $X$ has such a distribution, $X \sim B(a, b)$, if

$$f_X(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \quad 0 \leq x \leq 1$$

$$f_X(x) = 0 \quad \text{if } x < 0 \text{ or } x > 1$$

where $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} \, dx$

is the Beta function. $B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$.

6. The Weibull Distribution $W(d, \beta)$ ($d, \beta > 0$)

$X \sim W(d, \beta)$:

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ d \beta x^{\beta-1} e^{-d x^\beta} & \text{if } x \geq 0 \end{cases}$$

7. The Cauchy Distribution

$$f_X(x) = \frac{1}{\pi (1 + x^2)} \quad (x \in \mathbb{R})$$

$N = E(X) = \int_{-\infty}^{\infty} x f_X(x) \, dx$ diverges.
(VI) Joint Distributions for Multiple RVs

Definition Let $X_1, X_2, \ldots, X_n$ be $n$ real-valued scalar RVs. The joint cumulative distribution function (joint CDF) or simply, the joint distribution of these RVs is $F = F_{X_1, \ldots, X_n} : \mathbb{R}^n \rightarrow [0, 1]$ defined by

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \quad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

Here, $(X_1 \leq x_1, \ldots, X_n \leq x_n) = \{ \omega \in \Omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \ldots, X_n(\omega) \leq x_n \} = \bigcap_{k=1}^n \{ \omega : X_k(\omega) \leq x_k \}$.

Definition If $X_1, \ldots, X_n : \Omega \rightarrow \mathbb{R}$ are discrete RVs, then the joint probability mass function (joint PMF) or the joint density (discrete) function of these RVs is $f = f_{X_1, \ldots, X_n} : \mathbb{R}^n \rightarrow [0, 1]$, given by

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^n \{ \omega : X_j(\omega) = x_j \} \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$ 

Calculate the mean/expectation of a function of multiple discrete RVs.

Proposition

If $X_1, \ldots, X_n$ are real-valued, scalar, discrete RVs, then for any continuous function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(X_1, \ldots, X_n)$ is also a discrete RV. Moreover, if $g(X_1, \ldots, X_n) = \bigoplus_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) \frac{f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)}{f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)}$,
if the sum is finite, or countably infinite which converges absolutely.

**Definition** Let $X_1, \ldots, X_n : \mathcal{X} \to \mathbb{R}$ be continuous RVs, with the joint distribution (i.e., the joint CDF)

$$F = F_{X_1, \ldots, X_n} : \mathbb{R}^n \to [0,1].$$

Suppose

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

for some integrable function $f = f_{X_1, \ldots, X_n} : \mathbb{R}^n \to [0, \infty)$

then we call $f_{X_1, \ldots, X_n}$ the joint PDF for $X_1, \ldots, X_n$.

Note that $f_{X_1, \ldots, X_n} \geq 0$ and $\int_{\mathbb{R}^n} f_{X_1, \ldots, X_n} \, dx_1 \cdots dx_n = 1$.

**Proposition**

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \frac{\partial^n F_{X_1, \ldots, X_n}(x_1, \ldots, x_n)}{\partial x_1 \cdots \partial x_n}$$

if the partial derivative exists.

$$P(\mathbf{x} \in B) = \int_{\mathbf{x} \in \mathbb{R}^n} f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

if $\mathbf{x} = (x_1, \ldots, x_n)$, $B \subseteq \mathbb{R}^n$ a Borel set.

In particular,

$$P(a_1 \leq x_1 \leq b_1, \ldots, a_n \leq x_n \leq b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{X_1, \ldots, X_n} \, dx_1 \cdots dx_n.$$
Proposition. If $X_1, \ldots, X_n : \mathbb{R} \rightarrow \mathbb{R}$ are continuous RVs with the joint density $f_{X_1, \ldots, X_n} : \mathbb{R}^n \rightarrow [0, \infty)$, and $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a nice function (e.g., smooth, integrable in $\mathbb{R}^n$, etc.), then $g(X_1, \ldots, X_n)$ is a continuous RV, and

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

$$= \int_{\mathbb{R}^n} g(x) f_X(x_1) dx_1$$

where $x = (x_1, \ldots, x_n)$.

Marginal distributions

Definition. Marginal distributions are just distributions for a single RV from a group of RVs.

Let $X_1, \ldots, X_n : \mathbb{R} \rightarrow \mathbb{R}$ be RVs. The marginal distribution function for the $i$th RV $X_i$ is given by

$$F_{X_i}(x) = \mathbb{P}(X_i < x_i) = \mathbb{P}(x_1 < x_1, \ldots, x_i < x_i, \ldots, x_n < x_n)$$

$$= \lim_{x_j \to +\infty, \ldots, x_{i-1} \to +\infty} F_{X_1, \ldots, X_n}(x_1, \ldots, x_i, \ldots, x_n)$$

where $F_{X_1, \ldots, X_n} : \mathbb{R}^n \rightarrow [0, 1]$ is the joint distribution function for $X_1, \ldots, X_n$.

$n=2, X=X_1, Y=X_2$

$$F_X(x) = F_{x_X}(x, \infty), \quad F_Y(y) = F_{x_Y}(\infty, y) \quad \forall x, y \in \mathbb{R}$$
If \( X, Y \) are discrete RVs, then the marginal probability mass functions \( f_X, f_Y \) are given by
\[
    f_X(x) = \sum_y f_{X,Y}(x, y),
    \]
\[
    f_Y(y) = \sum_x f_{X,Y}(x, y),
\]
where \( f_{X,Y}(x, y) = P(X = x, Y = y) \) is the joint mass function.

If \( X, Y \) are continuous RVs, then the marginal PDFs are
\[
    f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy,
    \]
\[
    f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx,
\]
where \( f_{X,Y} \) is the joint PDF (the joint density) of \( X, Y \).

(VII) **Independence**

**Definition** Let \((\Omega, \mathcal{A}, P)\) be a probability space.
Let \( A, B \in \mathcal{A} \) (two events) with \( P(B) > 0 \).
The conditional probability of \( A \) given \( B \) (i.e., the probability \( A \) occurs given that \( B \) occurs) is
\[
    P(A|B) = \frac{P(A \cap B)}{P(B)}. \]
Lemma
1. If \( P(B) > 0 \) then and \( P(B^c) > 0 \) then
\[
P(A) = P(A \cap B)P(B) + P(A \cap B^c)P(B^c)
\]
2. If \( B_1, \ldots, B_n \) is a partition of \( \Omega \) (i.e., all \( B_j \in \Omega, j = 1, \ldots, n \), \( B_j \cap B_k = \emptyset, 1 \leq j < k \leq n \)), and \( P(B_j) > 0 \) \( (j = 1, \ldots, n) \), then
\[
P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i) \quad \forall A \in \Omega.
\]

Definition
Two events \( A, B \in \Omega \) are independent
if \( P(A \cap B) = P(A)P(B) \).

A family of events \( A_i \in \Omega \) (i.e., \( i \in I \)) is independent if
\[
P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)
\]
for any finite subset \( J \) of \( I \).

Definition
Let \( X_1, X_2, \ldots, X_n : \Omega \rightarrow \mathbb{R} \) be RVs. They are independent if
\[
P(\bigcap_{i=1}^{n} \{X_i \leq b_i\}) = \prod_{i=1}^{n} P(\{X_i \leq b_i\})
\]
for any Borel sets \( B_1, \ldots, B_n = \mathbb{R} \).

Theorem
Proposition
The RVs \( X_1, \ldots, X_n : \Omega \rightarrow \mathbb{R} \) are independent \( \iff \) (if and only if)
\[
F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \forall x_1, \ldots, x_n \in \mathbb{R}
\]
\[
\iff
F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)
\]
\[
\iff [f_{X_1, \ldots, X_n} = f_{X_i} \text{ can be discrete or continuous densities}].
\]
Theorem. If $X_1, \ldots, X_n$ are independent RVs and $g : \mathbb{R}^k \to \mathbb{R}$ and $h : \mathbb{R}^k \to \mathbb{R}$ are two functions, where $1 \leq k \leq n-1$, then $g(X_1, \ldots, X_k)$ and $h(X_{k+1}, \ldots, X_n)$ are independent RVs.

Theorem. Let $X_1, \ldots, X_n$ be independent RVs. Then

1. $\mathbb{E}\left( \prod_{i=1}^{n} X_i \right) = \prod_{i=1}^{n} \mathbb{E}(X_i)$ provided that each $\mathbb{E}(X_i) < \infty$.

2. $\text{Var}\left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i)$ provided that each $\text{Var}(X_i) < \infty$.

Note. If $X, Y$ are discrete RVs, then

$X(x) = \{x_1, x_2, \ldots\}$,

$Y(x) = \{y_1, y_2, \ldots\}$.

Then $X = \sum_i x_i 1_{A_i}, A_i = \{X = x_i\}$,

$Y = \sum_j y_j 1_{B_j}, B_j = \{Y = y_j\}$.

$X, Y$ are independent $\iff A_i, B_j$ are independent for any $i, j$.

Some skipped subject.

Move this part on $\text{cov}(X, Y) \to$ right before the def. of CDF on page 2.

Definition. The covariance of two RVs $X$ and $Y$ is

$\text{cov}(X, Y) = \mathbb{E}(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$.

The correlation (coefficient) of $X$ and $Y$ is

$p(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$
We have

1. \( \text{cov}(X, Y) = E(XY) - E(X)E(Y) \)

2. \( |\rho(X, Y)| \leq 1 \) holds \( \iff P(Y = aX + b) = 1 \) for some \( a, b \in \mathbb{R} \).

Note 3 follows from the Cauchy-Schwarz inequality:

\[ \left[ E(XY) \right]^2 \leq E(X^2)E(Y^2) \]

Move this part to right before (VII) Independence, page 14.

The bivariate normal distribution for a pair of RVs \( X \) and \( Y \):

\[ f(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2} \left( \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} \right)} \]

where \( \sigma_x, \sigma_y > 0, 0 < \rho < 1, \mu_x, \mu_y \in \mathbb{R} \).

We have \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1 \) \( \Rightarrow f \geq 0 \).

Prove these:

1. \( X \) is \( N(\mu_x, \sigma_x^2) \) and \( Y \) is \( N(\mu_y, \sigma_y^2) \).
2. The correlation between \( X \) and \( Y \) is \( \rho \).
3. \( X, Y \) are independent \( \iff \rho = 0 \).

The standard bivariate normal distribution for \( X \) and \( Y \) is:

\[ f(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} e^{-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{2 \rho xy}{\sigma_x \sigma_y} \right)} \]

where, \( \mu_x = \mu_y = 0, \sigma_x = \sigma_y = 1 \).

Here, \( \text{cov}(X, Y) = \rho \).
Definition Let $\mu=(\mu_1,\ldots,\mu_n) \in \mathbb{R}^n$, $V=(v_{ij})_{n \times n}$ be an $n \times n$, real, symmetric, nonsingular matrix. RVs $X_1, X_2, \ldots, X_n : \mathcal{X} \to \mathbb{R}$ has the multivariate normal distribution (or multivariate normal distribution), written $\mathcal{N}(\mu, V)$, if their joint density function $f = f_{X_1,\ldots,X_n}$ is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(V)}} e^{-\frac{1}{2}(x-\mu)^T V^{-1} (x-\mu)}$$

for all $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$.

Call $V=(v_{ij})$ the covariance matrix.

Theorem If $X=(X_1,\ldots,X_n)$ is $\mathcal{N}(\mu, V)$ then

1. $E(X) = \mu$, i.e., $\langle E(X_i) \rangle = \mu_i$, $i=1,2,\ldots,n$.
2. For $V=(v_{ij})$, we have

$$v_{ij} = \text{cov}(X_i,X_j), \quad i,j=1,2,\ldots,n.$$

(VII) Distributions of Functions of RVs and Sums of RVs

If $X_1,\ldots,X_n : \mathcal{X} \to \mathbb{R}$ are RVs, then $X=(X_1,\ldots,X_n) : \mathcal{X} \to \mathbb{R}^n$ is a vector-valued RV. The distribution of this RV $X$ is the joint distribution of $X_1,\ldots,X_n$ with the density $f_X(x) = f_{X_1,\ldots,X_n}(x)$, $X=(x_1,\ldots,x_n) \in \mathbb{R}^n$.

Let $T: \mathbb{R}^n \to \mathbb{R}^n$ be a $C^1$-map with the Jacobian that is invertible from $\text{Range } T \to \mathbb{R}^n$. We denote by $J=T(y_1,\ldots,y_n) = J(y)$ the Jacobian.
of $T$, the inverse of $T$,

$$T(x_1, x_2, \ldots, x_n) = (y_1, \ldots, y_n) \iff x_i = x_i(y_1, \ldots, y_n),$$

$$J(y) = \left| \frac{\partial (x_1, \ldots, x_n)}{\partial (y_1, \ldots, y_n)} \right| = \det \left( \frac{\partial x_i}{\partial y_j} \right).$$

**Theorem.** Under these assumptions, the joint density function for $Y = (y_1, \ldots, y_n) = T(X) = T(x_1, \ldots, x_n)$ is

$$f_Y(y) = \begin{cases} f_X(T^{-1}(y)) \left| J(y) \right| & \text{if } y \in \text{Range}(T) \\ 0 & \text{otherwise} \end{cases}$$

**Theorem.** Let $X, Y : \mathbb{R} \rightarrow \mathbb{R}$ be two RVs. Having the joint density

(1) Let $Z = X + Y$. Then $Z$ has the density

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) \, dx \quad \forall z \in \mathbb{R}.$$ 

(2) If $X, Y$ are independent, then

$$f_{X+Y} = f_X \ast f_Y \quad \text{(convolution)}.$$ 

**Some examples of distributions of sums of RVs.**

(1) Let $X_1, \ldots, X_n : \mathbb{R} \rightarrow \mathbb{R}$ be independent RVs. Assume $X_i$ is Gamma $(\alpha_i, b)$ distributed, $i = 1, \ldots, n$. $\alpha_i > 0$, $b > 0$. Then $X_1 + \ldots + X_n$ is Gamma $(\sum_{i=1}^{n} \alpha_i, b)$ distributed.
(2) If a RV $Z : \mathcal{N}(0, 1) \rightarrow \mathbb{R}$ is $N(0, 1)$ distributed, then $Z^2$ is Gamma ($\frac{1}{2}, \frac{1}{2}$) distributed.

(3) If $Z_1, \ldots, Z_n$ are i.i.d. $N(0, 1)$ RVs, then $Z_1^2 + \cdots + Z_n^2$ is Gamma ($n/2, 1/2$) distributed. (This is also called the chi-squared distribution with degree $n$, denoted $\chi^2_n$).

(4) If $X_1, \ldots, X_n$ are i.i.d. with common distribution Bernoulli ($p$), then $X_1 + \cdots + X_n$ is $\text{Bin}(n, p)$ distributed.