Statistics 1

D. González Sánchez
Licenciatura en Economía
CIDE

Fall 2022

One can describe Statistics as the mathematical discipline whose purpose is to use empirical data generated by a random phenomenon, in order to make inferences about some deterministic characteristics of the phenomenon while simultaneously quantifying the uncertainty inherent in these inferences.

Panaretos [4, p. xiii]

Contents

1 Basics of probability

See Mendenhall et al. [9, Chapter 2] or Lefebvre [3, Chapter 2].

2 Discrete random variables

In this section, we assumed that each sample space Ω is endowed with a probability P defined on some $\sigma-$filed $\mathscr{F}$. See Appendix B.

A random variable $X : \Omega \to \mathbb{R}$ is **discrete** if R_X is at most countable.

Given a discrete r.v. X, define the **probability mass function (pmf)**

$$p_X(x) := P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\}), \qquad x \in R_X.$$

Notice that

(a) $p_X(x) \geq 0$ for each x and

(b) $\sum_{x \in R_X} p_X(x) = 1$.

Remark 2.1. Given a function $\pi : S \to \mathbb{R}$, where S is at most countable, that satisfies properties (a) and (b), we can find a discrete r.v. $X : \Omega \to \mathbb{R}$ and a probability P such that $S = R_X$ and

$$\pi(x) = P(X = x) \qquad \forall x \in S.$$

$\diamond$

2.1 Mean, variance, and standard deviation

Definition 2.2. *Let $X : \Omega \to \mathbb{R}$ be a discrete r.v., define the* **expectation (mean or expected value)** *of X as*

$$E(X) = \sum_{x \in R_X} x p_X(x),$$

whenever the series is absolutely convergent.

The r.v. X with values in $\mathbb{N}$ and pmf

$$p_X(n) = \frac{1}{n(n+1)}, \qquad n \in \mathbb{N},$$

does not have finite expectation.

Theorem 2.3. *Let $X : \Omega \to \mathbb{R}$ be a discrete r.v. and $g : \mathbb{R} \to \mathbb{R}$. Then the expectation of $Y = g(X)$ is*

$$E[Y] = \sum_{x \in R_X} g(x) p_X(x),$$

that is

$$\sum_{y \in R_Y} y p_Y(y) = \sum_{x \in R_X} g(x) p_X(x).$$

Proof. Notice that $R_X = \cup_{y \in R_Y}\{x \in R_X \mid g(x) = y\}$ and

$$\{\omega \in \Omega \mid Y(\omega) = y\} = \bigcup_{x \in R_X,\ g(x)=y} \{\omega \in \Omega \mid X(\omega) = x\}, \qquad y \in R_Y.$$

Then

$$\begin{aligned}
\sum_{x \in R_X} g(x) p_X(x) &= \sum_{y \in R_Y} \left[\sum_{x \in R_X,\ g(x)=y} g(x) P(X = x) \right] \\
&= \sum_{y \in R_Y} y \left[\sum_{x \in R_X,\ g(x)=y} P(X = x) \right] \\
&= \sum_{y \in R_Y} y P(Y = y).
\end{aligned}$$

□

Proposition 2.4. *Let $X : \Omega \to \mathbb{R}$ be a discrete r.v. and $g_1, g_2 : \mathbb{R} \to \mathbb{R}$.*

(a) *If X is constant, say $X = c$, then $E(X) = c$.*

(b) *If $a \in \mathbb{R}$, then $E[aX] = aE(X)$.*

(c) $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

(d) *If $a, c \in \mathbb{R}$, then $E[aX + c] = aE(X) + c$.*

Proof. Parts (a) and (b) follow from the definition of expectation. Part (c) is a direct consequence of Theorem 2.3. Finally, (d) follows from (a), (b), and (c). □

The **variance** σ_X^2 of the r.v. X is defined as

$$\sigma_X^2 := E[(X - \mu)^2]$$

where $\mu_X = E(X)$. The variance of X is also denoted $\mathrm{var}(X)$.

Proposition 2.5. *Let $X : \Omega \to \mathbb{R}$ be a discrete r.v.*

(a) $\mathrm{var}(X) = E(X^2) - \mu_X^2$.

(b) *If $a \in \mathbb{R}$, then $\mathrm{var}(aX) = a^2 \mathrm{var}(X)$.*

(c) *If $c \in \mathbb{R}$, then $\mathrm{var}(X + c) = \mathrm{var}(X)$.*

The **standard deviation** σ_X of a discrete r.v. X is defined as

$$\sigma_X = \sqrt{\mathrm{var}(X)}.$$

Theorem 2.6. *Let X be a discrete r.v.*

(a) **(Markov's inequality)**. *If $X \geq 0$ and $a > 0$, then*

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

(b) *If* $c \in \mathbb{R}$, $\varepsilon > 0$, *and* $m > 0$, *then*

$$P(|X - c| \geq \varepsilon) \leq \frac{E(|X - c|^m)}{\varepsilon^m}.$$

(c) **(Chebyshev's inequality)**. *If* μ_X *and* σ_X^2 *are finite, then*

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}.$$

Proof. (a) Since $X \geq 0$,

$$E(X) \geq \sum_{x \in R_X,\ x \geq a} x p_X(x).$$

Then $E(X) \geq aP(X \geq a)$ and hence Markov's inequality follows.

(b) Part (a) and the equality $\{\omega \in \Omega \mid |X(\omega) - c|^m \geq \varepsilon^m\} = \{\omega \in \Omega \mid |X(\omega) - c| \geq \varepsilon\}$ yield (b).

(c) Chebyshev's inequality follows from (b) with $m = 2$.

□

2.2 Common discrete distributions

2.2.1 Bernoulli distribution

Let $S \subseteq \Omega$ be an event such that $P(S) = p$. A *Bernoulli trial* consists of two possible outcomes: *success* S or *failure* S^c. Define the **Bernoulli random variable**

$$B(\omega) = \begin{cases} 1 & \text{if } \omega \in S, \\ 0 & \text{if } \omega \in S^c. \end{cases}$$

Thus $p_B(1) = p$ and $p_B(0) = 1 - p$. The distribution of B is called **Bernoulli distribution**; it is also said that B has Bernoulli distribution with parameter p. We use the notation

$$B \sim \mathrm{Ber}(p).$$

Notice that

$$E(B) = p$$

and

$$\mathrm{var}(B) = p(1 - p).$$

2.2.2 Binomial distribution

A *Binomial experiment* with parameters (n, p) consists of n *independent* Bernoulli trials with parameter p. A **Binomial random variable** Y gives the number of successes of a Binomial experiment. We use the notation

$$Y \sim \mathrm{Bin}(n, p).$$

Then $R_Y = \{0, 1, \ldots, n\}$ and the probability mass function is

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad 0 \leq k \leq n.$$

Lemma 2.7. *Let $m \in \mathbb{N}$ and $x, y \in \mathbb{R}$.*

(a) **(Newton's Binomial)** *For* $(x+y)^m = \sum_{k=0}^{m} \binom{m}{k} x^k y^{m-k}$.

(b) *For* $1 \le k \le m$,

$$k\binom{m}{k} = m\binom{m-1}{k-1}.$$

(c) *For* $2 \le k \le m$,

$$k^2\binom{m}{k} = m(m-1)\binom{m-2}{k-2} + k\binom{m}{k}.$$

Proof. Newton's Binomial is well known. Equalities (b) and (c) follow from direct calculations. □

Notice that $\sum_{k=0}^{n} P(Y=k) = 1$ due to Newton's Binomial and the equality $1 = [p + (1-p)]^n$.

Proposition 2.8. *If $Y \sim Bin(n,p)$, then*

$$E(Y) = np \quad \textit{and} \quad \mathrm{var}(Y) = np(1-p).$$

Proof. It follows from Lemma 2.7. Proposition 3.5(a) is useful to compute the variance. □

2.2.3 Geometric distribution

Consider the number G of independent Bernoulli trials, with parameter p, until we obtain the *first success*. Thus

$$R_G = \{1, 2, 3, \dots\}$$

and G is called **Geometric random variable**. We use the notation $G \sim \mathrm{Geo}(p)$. The probability mass function is given by

$$P(G=k) = (1-p)^{k-1}p, \qquad k = 1, 2, \dots.$$

Lemma 2.9. *Let $0 < r < 1$. Then*

(a) $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$,

(b) $\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2}$, *and*

(c) $\sum_{k=1}^{\infty} k^2 r^{k-1} = \frac{1+r}{(1-r)^3}$.

Proposition 2.10. *If $G \sim \mathrm{Geo}(p)$, then*

$$E(G) = \frac{1}{p} \quad \textit{and} \quad \mathrm{var}(G) = \frac{1-p}{p^2}.$$

Proof. It follows from Lemma 2.9. □

2.2.4 Poisson distribution

Let Y be a discrete r.v. such that

$$P(Y = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots,$$

where $\lambda > 0$. This r.v. is said to have **Poisson distribution**, written as $Y \sim \text{Poi}(\lambda)$.

Poisson random variables are used to count the (random) number of events that occur in a given interval of time.

Recall that, for each $\lambda \in \mathbb{R}$,

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}. \tag{1}$$

Proposition 2.11. *If $Y \sim \text{Poi}(\lambda)$, then*

$$E(Y) = \lambda \quad \textit{and} \quad \text{var}(Y) = \lambda.$$

Proof. The expectation is easily obtained from (1). For the variance, we use 1 again to get

$$E[Y(Y-1)] = \lambda^2.$$

Thus $E(Y^2) - \lambda = \lambda^2$ and hence $\text{var}(Y) = \lambda$ because of Proposition 3.5(a). □

2.3 Moment-generating functions

The $k-$**th moment about the origin** of the r.v. X is given by

$$\mu_k := E(X^k)$$

and the $k-$**th moment about the mean** μ of X is

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

Definition 2.12. *Given a discrete r.v. X, the* **moment-generating function** *m_X of X is given by*

$$m_X(t) = E(e^{tX})$$

for each t such that $E(e^{tX}) < \infty$.

Observe that

$$\begin{aligned}
E(e^{tX}) &= E\left(1 + tX + \frac{t^2X^2}{2} + \frac{t^3X^3}{3!} + \frac{t^4X^4}{4!} + \ldots\right) \\
&= 1 + tE(X) + \frac{t^2E(X^2)}{2} + \frac{t^3E(X^3)}{3!} + \frac{t^4E(X^4)}{4!} + \ldots \\
&= 1 + t\mu + \frac{t^2}{2}\mu_2 + \frac{t^3}{3!}\mu_3 + \frac{t^4}{4!}\mu_4 + \ldots, \qquad (2)
\end{aligned}$$

whenever the function m_X is well defined on a neighborhood about the origin. Further, if m_X has derivatives $m_X^{(k)}$ at $t = 0$, then we can compare the Taylor expansion of m_X and (2) to conclude that

$$\mu_k = m^{(k)}(0), \qquad k = 1, 2, \ldots.$$

Proposition 2.13. *Let $B \sim \mathrm{Bin}(n,p)$, $G \sim \mathrm{Geo}(p)$, and $X \sim \mathrm{Poi}(\lambda)$. Then*

$$m_B(t) = (pe^t + 1 - p)^n, \qquad t \in \mathbb{R},$$

$$m_G(t) = \frac{pe^t}{1-(1-p)e^t}, \qquad (1-p)e^t < 1,$$

and

$$m_X(t) = e^{\lambda(e^t-1)}, \qquad t \in \mathbb{R}.$$

Proof. The moment-generating functions are obtained by direct calculations. □

2.4 Exercises

Solve Exercises 3.15, 3.29, 3.37, 3.40, 3.41, 3.70, 3.71, 3.88, 3.130, 3.155 in Wackerly et al. [9].

3 Continuous random variables

Let X be a random variable, on the probability space $(\Sigma, \mathscr{F}, P)$ (see Appendix B), and distribution F. We say that X is a **continuous random variable** if F is a continuous function.

In this section we deal with a subclass of continuous random variables for which there is an integrable **probability density function** (pdf) (or simply **density**) $f : \mathbb{R} \to [0, \infty)$, that is,

$$P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \qquad x \in \mathbb{R}.$$

In particular, by the Fundamental Theorem of Calculus,

$$F'(x) = f(x), \qquad x \in \mathbb{R},$$

whenever f is continuous at x.

Given a r.v. X with a continuous density f, we have

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a < X < b) \\ &= \int_{a}^{b} f(x)dx \end{aligned}$$

for $a < b$.

3.1 Mean and variance

Definition 3.1. *Let $X : \Omega \to \mathbb{R}$ be a r.v. with density f, define the* **expectation (mean or expected value)** *of X as*

$$E(X) = \int_{\mathbb{R}} x f(x)dx,$$

whenever $\int_{\mathbb{R}} |x| f(x)dx < \infty$.

Example 3.2. The distribution of a r.v. X with density

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad x \in \mathbb{R},$$

is known as **Cauchy distribution**. The r.v. X has undefined expectation. ◇

A proof of Theorem 3.3 can be found in Rosenthal [7, Proposition 6.2.3]. The proof of Proposition 3.4 is analogous to that of Proposition 2.4.

Theorem 3.3. *Let $X : \Omega \to \mathbb{R}$ be a r.v. with density f. If $h : \mathbb{R} \to \mathbb{R}$, then*

$$E[h(X)] = \int_{\mathbb{R}} h(x) f(x)dx,$$

whenever the integral is well defined.

Proposition 3.4. *Let $X : \Omega \to \mathbb{R}$ be a r.v. with density f and $g_1, g_2 : \mathbb{R} \to \mathbb{R}$. Assume the expectations below are well defined.*

(a) *If X is constant, say $X = c$, then $E(X) = c$.*

(b) *If* $a \in \mathbb{R}$, *then* $E[aX] = aE(X)$.

(c) $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

(d) *If* $a, c \in \mathbb{R}$, *then* $E[aX + c] = aE(X) + c$.

The **variance** σ_X^2 of the r.v. X, with density f, is defined as

$$\sigma_X^2 := E[(X - \mu_X)^2] = \int_{\mathbb{R}} (x - \mu_X)^2 f(x) dx$$

where $\mu_X = E(X)$. The variance of X is also denoted $\mathrm{var}(X)$. The **standard deviation** σ_X of X is defined as

$$\sigma_X = \sqrt{\mathrm{var}(X)}.$$

Proposition 3.5. *Let* $X : \Omega \to \mathbb{R}$ *be a continuous r.v., with density* f. *Assume the mean and variance of* X *are finite.*

(a) $\mathrm{var}(X) = E(X^2) - \mu_X^2$.

(b) *If* $a \in \mathbb{R}$, *then* $\mathrm{var}(aX) = a^2 \mathrm{var}(X)$.

(c) *If* $c \in \mathbb{R}$, *then* $\mathrm{var}(X + c) = \mathrm{var}(X)$.

Proof. The assertions follow from properties of the integral. □

Theorem 3.6. *Let* X *be a continuous r.v. with density* f. *Assume the expectations below are finite.*

(a) **(Markov's inequality)**. *If* $X \geq 0$ *and* $a > 0$, *then*

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

(b) *If* $c \in \mathbb{R}$, $\varepsilon > 0$, *and* $m > 0$, *then*

$$P(|X - c| \geq \varepsilon) \leq \frac{E(|X - c|^m)}{\varepsilon^m}.$$

(c) **(Chebyshev's inequality)**. *If* μ_X *and* σ_X^2 *are finite, then*

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}.$$

Proof. (a) Since $X \geq 0$,

$$E(X) \geq \int_a^\infty x f(x) dx.$$

Then $E(X) \geq aP(X \geq a)$ and hence Markov's inequality follows.

(b) Part (a) and the equality $\{\omega \in \Omega \mid |X(\omega) - c|^m \geq \varepsilon^m\} = \{\omega \in \Omega \mid |X(\omega) - c| \geq \varepsilon\}$ yield (b).

(c) Chebyshev's inequality follows from (b) with $m = 2$.

□

3.2 Common continuous distributions

3.2.1 Uniform distribution

The r.v. U has **uniform distribution** on the interval $[a, b]$, with $a < b$, if it has density of the form

$$f(u) = \begin{cases} \frac{1}{b-a} & \text{if } a \le u \le b, \\ 0 & \text{otherwise.} \end{cases}$$

We use the notation $U \sim \text{Unif}([a, b])$. Thus

$$F(u) = \begin{cases} 0 & \text{if } u < a, \\ \frac{u-a}{b-a} & \text{if } a \le u \le b, \\ 1 & \text{if } u > b. \end{cases}$$

Further, $E(U) = (a + b)/2$ and $\text{var}(U) = (b - a)^2/12$. Finally, the moment-generating function is

$$m_U(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{(b-a)t} & \text{if } t \neq 0, \\ 1 & \text{if } t = 0. \end{cases}$$

3.2.2 Exponential distribution

Let $\beta > 0$. If the r.v. Y has density

$$f(y) = \begin{cases} \beta e^{-\beta y} & \text{if } y \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

then we say that Y has **exponential distribution**, denoted $Y \sim \text{Exp}(\beta)$. The distribution becomes

$$F(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 - e^{-\beta y} & \text{if } y \ge 0. \end{cases}$$

Integrating by parts, we see that

$$E(Y) = \int_0^\infty e^{-\beta y} = \frac{1}{\beta}.$$

Further, integrating by parts again, we have $E(Y^2) = 2/\beta^2$. Hence

$$\text{var}(Y) = \frac{1}{\beta^2}.$$

The moment-generating function m_Y is defined for $t < \beta$,

$$m_Y(t) = \frac{\beta}{\beta - t}.$$

3.2.3 Normal distribution

Let $\mu \in \mathbb{R}$ and $\sigma > 0$. We say that Z has **normal distribution** with parameters (μ, σ) whenever the density is

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right), \qquad z \in \mathbb{R}. \tag{3}$$

We use the notation $Z \sim N(\mu, \sigma^2)$. In particular, the **standard normal distribution** happens when $\mu = 0$ and $\sigma = 1$.

Remark 3.7. From the so-called Gaussian integral (see Theorem A.8)

$$\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi},$$

we can see that $\int_{-\infty}^{\infty} f(z)dz = 1$, where f is given by (3). In particular,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 1.$$

$\diamond$

Theorem 3.8. *Let* $Z \sim N(\mu, \sigma^2)$. *Then* $E(Z) = \mu$ *and* $\mathrm{var}(Z) = \sigma^2$.

Proof. In order to compute the expectation, set the change of variable $x = (z - \mu)/\sigma$ to obtain

$$E(Z) = \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_{-\infty}^{\infty} \frac{\sigma}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx.$$

The first integral equals μ. For the second integral, notice that $\int_0^{\infty} x e^{-\frac{x^2}{2}} dx$ is finite because

$$\begin{aligned} \int_0^b x e^{-\frac{x^2}{2}} dx &= 1 - e^{-b^2/2} \\ &\to 1 \end{aligned}$$

as $b \to \infty$. Since the integrand is an odd function,

$$\int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = 0.$$

Then $E(Z) = \mu$.

We now compute the variance

$$\begin{aligned} \mathrm{var}(Z) &= E[(Z - \mu)^2] \\ &= \int_{-\infty}^{\infty} (z - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{z - \mu}{\sigma}\right)^2\right) dz \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx, \end{aligned} \tag{4}$$

where the latter equality follows from the change of variable $x = \frac{z-\mu}{\sigma}$. On the other hand, integrate both sides of the equality

$$\frac{d}{dx}(x e^{-x^2/2}) = -x^2 e^{-x^2/2} + e^{-x^2/2}$$

on the interval $[-b, b]$, then let $b \to \infty$ to obtain

$$0 = -\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx + \int_{-\infty}^{\infty} e^{-x^2/2} dx.$$

This equality of integrals and (4) imply (see Remark 3.7)

$$\mathrm{var}(Z) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = \sigma^2.$$

$\square$

Theorem 3.9. *The moment-generating function m_Z of $Z \sim N(\mu, \sigma^2)$ is given by*

$$m_Z(t) = e^{\mu t + \sigma^2 t^2/2}, \qquad t \in \mathbb{R}.$$

Proof. The conclusion follows from the equality

$$tz - \frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2 = \mu t + \frac{\sigma^2 t^2}{2} - \frac{1}{2}\left(\frac{z - (\sigma^2 t + \mu)}{\sigma}\right)^2.$$

$\square$

3.2.4 Gamma distribution and its particular cases

A continuous r.v. Y has **gamma distribution** when its density is of the form

$$f(y) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where Γ is the function defined in Appendix A.1. We use the notation $Y \sim \text{Gamma}(\alpha, \beta)$ Notice that

$$\begin{aligned} \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-\beta y} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\beta^\alpha} t^{\alpha-1} e^{-t} dt \\ &= 1. \end{aligned}$$

where we have used the change of variable $t = \beta y$. Thus (5) defines a pdf.

Remark 3.10. A particular case of the gamma distribution happens when $\alpha = 1$, this is the exponential distribution. Another two particular cases are given in the following definition.

Definition 3.11. *Let $k \in \mathbb{N}$ and $\beta > 0$.*

(a) *The r.v. Y whose pdf is given by*

$$f(y) = \begin{cases} \frac{\beta^k}{(k-1)!} y^{k-1} e^{-\beta y} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

is said to have **Erlang distribution**.

(b) *The r.v. Y has* **chi-squared distribution** *with k degrees of freedom, written as $Y \sim \chi^2_k$, when the pdf is of the form*

$$f(y) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2-1} e^{-y/2} & \text{if } y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Theorem 3.12. *The moment-generating function m_Y of $Y \sim \text{Gamma}(\alpha, \beta)$ is given by*

$$m_Y(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \qquad t < \beta.$$

Proof. To compute the corresponding integral, let the change of variable $u = (\beta - t)y$, thus

$$\begin{aligned} m_Y(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-(\beta-t)y} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)(\beta-t)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u} du \\ &= \frac{\beta^\alpha}{(\beta-t)^\alpha}, \end{aligned}$$

whenever $\beta - t > 0$. □

Theorem 3.13. *If $Y \sim \mathrm{Gamma}(\alpha, \beta)$, then $E(Y) = \alpha/\beta$ and $\mathrm{var}(Y) = \alpha/\beta^2$.*

Proof. By differentiating m_Y, we have $m'_Y(0) = \alpha/\beta$ and $m''_Y(0) = \alpha(\alpha+1)/\beta^2$. Then $E(Y) = \alpha/\beta$ and

$$\begin{aligned} \mathrm{var}(Y) &= E(Y^2) - (\alpha/\beta)^2 \\ &= \alpha(\alpha+1)/\beta^2 - \alpha^2/\beta^2 \\ &= \alpha/\beta^2. \end{aligned}$$

□

3.2.5 Beta distribution

The r.v. Y follows a **beta distribution** with parameters (α, β) whenever its pdf has the form

$$f(y) = \begin{cases} \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} & \text{if } 0 \le y \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where B is the beta function—see Appendix A.2. We use the notation $Y \sim \mathrm{Beta}(\alpha, \beta)$.

Unfortunately, there is not a closed-form expression for the moment-generating function of the beta distribution.

Proposition 3.14. *Let $Y \sim \mathrm{Beta}(\alpha, \beta)$. Then*

$$E(Y) = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \mathrm{var}(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Proof. It follows from direct calculations and the relation

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

proved in Theorem A.7. □

3.3 Exercises

Solve Exercises 4.59, 4.62, 4.63, 4.79, 4.88, 4.96, 4.105, 4.123, 4.133, 4.137, and 4.146 in Wackerly et al. [9]. The following websites could be useful

https://homepage.divms.uiowa.edu/~mbognar/
https://college.cengage.com/nextbook/statistics/wackerly_966371/student/html/

A Gamma and beta functions

A.1 The gamma function

Definition A.1. *For each positive real number x, the* **gamma function** *is given by*

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt. \tag{9}$$

To see why the improper Riemann integral (9) exists, notice first that

$$t^{x-1}e^{-t} \leq t^{x-1} \qquad 0 < t \leq 1.$$

On the other hand, for each $x > 0$, $t^{x-1}e^{-t/2} \to 0$ as $t \to \infty$, thus

$$t^{x-1}e^{-t} \leq M_x e^{-t/2}, \qquad t \geq 1,$$

for some constant M_x. Therefore the integral (9) is finite for each $x > 0$.

Lemma A.2. *Let $p, q \in (1, \infty)$. If $\frac{1}{p} + \frac{1}{q} = 1$, then for each $\alpha, \beta \geq 0$,*

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \tag{10}$$

Proof. Notice that $h : [0, \infty) \to \mathbb{R}$ given by

$$h(x) = (1-t) + tx - x^t, \qquad 0 < t < 1,$$

attains its minimum at $x = 1$. Then

$$0 \leq (1-t) + tx - x^t, \qquad \forall x > 0.$$

In particular, for $x = a/b$,

$$a^t b^{1-t} \leq ta + (1-t)b.$$

The conclusion of the lemma follows from the change of variables $t = 1/p$, $\alpha = a^t$, and $\beta = b^{1-t}$. $\square$

Lemma A.3 (Hölder's inequality)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions. Then*

$$\left|\int_a^b fg\right| \leq \left(\int_a^b |f|^p\right)^{1/p} \left(\int_a^b |g|^q\right)^{1/q}. \tag{11}$$

Proof. If $f \equiv 0$ or $g \equiv 0$, then the inequality trivially holds. When f and g are not identically zero, we have

$$\|f\|_p := \left(\int_a^b |f|^p\right)^{1/p} > 0$$

and

$$\|g\|_q := \left(\int_a^b |g|^q\right)^{1/q} > 0.$$

By Lemma A.2

$$\frac{|f(t)|}{\|f\|_p}\frac{|g(t)|}{\|g\|_q} \leq \frac{1}{p}\left(\frac{|f(t)|}{\|f\|_p}\right)^p + \frac{1}{q}\left(\frac{|g(t)|}{\|g\|_q}\right)^q \qquad \forall t \in [a, b].$$

By integrating the latter inequality,

$$\begin{aligned}\frac{1}{\|f\|_p\|g\|_q}\int_a^b |f(t)g(t)|\,dt &\le \frac{1}{p(\|f\|_p)^p}\int_a^b |f(t)|^p dt + \frac{1}{q(\|g\|_q)^q}\int_a^b |g(t)|^q dt \\ &= \frac{1}{p}+\frac{1}{q} \\ &= 1,\end{aligned}$$

and hence the required inequality follows. □

Hölder's inequality holds for *integrable* functions (not necessarily continuous) and also for *improper* integrals, see Aliprantis and Burkinshaw [1, Theorem 31.3].

Theorem A.4. *The gamma function satisfies*

(a) $\Gamma(x+1) = x\Gamma(x)$ *for every* $x > 0$,

(b) $\Gamma(n+1) = n!$ *for each* $n = 1, 2, 3, \ldots$, *and*

(c) $\log(\Gamma)$ *is convex on the interval* $(0, \infty)$.

Proof. (a) Integration by parts yields

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty t^x e^{-t} dt \\ &= 0 + x\int_0^\infty t^{x-1}e^{-t}dt \\ &= x\Gamma(x).\end{aligned}$$

(b) From (9), we see that $\Gamma(1) = 1$. By (a),

$$\begin{aligned}\Gamma(2) &= 1, \\ \Gamma(3) &= 2\cdot 1, \\ \Gamma(4) &= 3\cdot 2\cdot 1.\end{aligned}$$

By induction, we conclude $\Gamma(n+1) = n!$.

(c) Let $p, q \in (1, \infty)$ satisfy $\frac{1}{p}+\frac{1}{q} = 1$. It suffices to show that

$$\log\Gamma\left(\frac{x}{p}+\frac{y}{q}\right) \le \frac{\log\Gamma(x)}{p} + \frac{\log\Gamma(y)}{q} \qquad \forall\, x, y \ge 0. \tag{12}$$

Thus

$$\begin{aligned}\Gamma\left(\frac{x}{p}+\frac{y}{q}\right) &= \int_0^\infty t^{\frac{x}{p}+\frac{y}{q}-1}e^{-t}dt \\ &= \int_0^\infty \left[t^{\frac{x}{p}-\frac{1}{p}}e^{-\frac{t}{p}}\right]\left[t^{\frac{y}{q}-\frac{1}{q}}e^{-\frac{t}{q}}\right]dt \\ &\le \left[\int_0^\infty t^{x-1}e^{-t}dt\right]^{1/p}\left[\int_0^\infty t^{y-1}e^{-t}dt\right]^{1/q} \qquad \text{by Hölder's inequality} \\ &= [\Gamma(x)]^{1/p}[\Gamma(y)]^{1/q}.\end{aligned}$$

Then (4) follows because $\log(\cdot)$ is increasing.

□

The following theorem gives three properties that characterize the gamma function. A proof of Theorem A.5, also known as Bohr-Mollerup Theorem, can be found in Rudin [8, Teorema 8.19].

Theorem A.5. *Let* $G : (0,\infty) \to (0,\infty)$ *satisfy*

(a) $G(x+1) = xG(x)$ *for every* $x > 0$,

(b) $G(1) = 1$, *and*

(c) $\log(G)$ *is convex on the interval* $(0,\infty)$.

Then $G(x) = \Gamma(x)$ *for every* $x > 0$.

A.2 The beta function

Definition A.6. *For each pair* (x,y) *of positive numbers, define the* **beta function** *as*

$$B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}dt. \tag{13}$$

Theorem A.7. *The following equality holds for each pair* (x,y) *of positive numbers*

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{14}$$

Proof. Put $x = 1$ in (13) to have

$$B(1,y) = \frac{1}{y}. \tag{15}$$

Notice also that

$$\begin{aligned} B(x+1,y) &= \int_0^1 t^x(1-t)^{y-1}dt \\ &= \int_0^1 \left(\frac{t}{1-t}\right)^x (1-t)^{x+y-1}dt \\ &= \left.\frac{t^x(1-t)^y}{x+y}\right|_1^0 + \int_0^1 \frac{x}{x+y}(1-t)^{y-1}t^{x-1}dt, \end{aligned}$$

hence

$$B(x+1,y) = \frac{x}{x+y}B(x,y). \tag{16}$$

In addition, for each y

$$\begin{aligned} B\left(\frac{x_1}{p}+\frac{x_2}{q},y\right) &= \int_0^1 t^{\frac{x_1}{p}+\frac{x_2}{q}-1}(1-t)^{y-1}dt \\ &= \int_0^1 t^{\frac{x_1-1}{p}}(1-t)^{\frac{y-1}{p}} t^{\frac{x_2-1}{q}}(1-t)^{\frac{y-1}{q}}dt \\ &\le \left[\int_0^1 t^{x_1-1}(1-t)^{y-1}dt\right]^{\frac{1}{p}} \left[\int_0^1 t^{x_2-1}(1-t)^{y-1}dt\right]^{\frac{1}{q}}, \end{aligned}$$

the latter inequality follows from Hölder's inequality. Equivalently,

$$\log B\left(\frac{x_1}{p}+\frac{x_2}{q},y\right) \le \frac{\log B(x_1,y)}{p} + \frac{\log B(x_2,y)}{q},$$

which implies that $\log B(\cdot, y)$ is convex.

Let $G : (0, \infty) \to (0, \infty)$ be given by

$$G(x) := \frac{\Gamma(x+y)}{\Gamma(y)} B(x, y).$$

Equations (15) and (16) imply

(a) $G(x+1) = xG(x)$ for every $x > 0$,

(b) $G(1) = 1$,

further, since $\log B(\cdot, y)$ is convex, we conclude that

(c) $\log(G)$ is convex on $(0, \infty)$.

By Theorem A.5,

$$\frac{\Gamma(x+y)}{\Gamma(y)} B(x, y) = \Gamma(x),$$

which is equivalent to (13). □

A.3 The Gaussian integral

The integral in the following theorem is known as **Gaussian integral**.

Theorem A.8. $\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi}$.

Proof. Consider the integral (13) and set $t = \sin^2 \theta$, thus

$$\begin{aligned} B(x, y) &= \int_0^1 t^{x-1}(1-t)^{y-1} dt \\ &= \int_0^{\pi/2} (\sin^2 \theta)^{x-1}(1-\sin^2 \theta)^{y-1} 2 \sin \theta \cos \theta \, d\theta \\ &= 2 \int_0^{\pi/2} (\sin \theta)^{2x-1}(\cos \theta)^{2y-1} d\theta. \end{aligned}$$

We now use Theorem A.7 with $x = y = \frac{1}{2}$ to obtain

$$\frac{[\Gamma(1/2)]^2}{\Gamma(1)} = 2 \int_0^{\pi/2} 1 \, d\theta.$$

Then

$$\Gamma(1/2) = \sqrt{\pi}. \tag{17}$$

On the other hand, the change of variable $t = s^2$ in the gamma function yields

$$\begin{aligned} \Gamma(x) &= \int_0^{\infty} t^{x-1} e^{-t} dt \\ &= 2 \int_0^{\infty} s^{2x-1} e^{-s^2} ds. \end{aligned}$$

In particular, for $x = \frac{1}{2}$, we have $\Gamma(1/2) = 2 \int_0^{\infty} e^{-s^2} ds$. Finally, from (17), we conclude that

$$\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi}.$$

□

B Probability spaces

B.1 The family of events

In this appendix, Ω denotes a nonempty set called **sample space**.

Definition B.1. *A family $\mathscr{F}$ of subsets of Ω is called a $\sigma-$**field** if*

(1) $\emptyset \in \mathscr{F}$,

(2) $A \in \mathscr{F} \implies A^c \in \mathscr{F}$.

(3) $A_n \in \mathscr{F}, n \in \mathbb{N}, \implies \cup_{n=1}^{\infty} A_n \in \mathscr{F}$.

*The members of the $\sigma-$field $\mathscr{F}$ (also called $\sigma-$**algebra**) are called **events**.*

Due to property (3) in the definition of $\sigma-$field, $\mathscr{F}$ is said to be *closed under countable unions*. In fact, $\mathscr{F}$ is also closed under *finite* unions because

$$A_1 \cup \cdots \cup A_n = A_1 \cup \cdots \cup A_n \cup \emptyset \cup \emptyset \cup \cdots .$$

Further, $\mathscr{F}$ is closed under countable intersections since

$$\cap_{n\in\mathbb{N}} B_n = (\cup_{n\in\mathbb{N}} B_n^c)^c.$$

Example B.2. The smallest $\sigma-$field of a sample space Ω is

$$\mathscr{F} = \{\emptyset, \Sigma\}.$$

On the other hand, the largest $\sigma-$field of Ω is the *power set* that consists of all the subsets of Ω. When Ω is a finite or countably infinite set, the usual $\sigma-$field for Ω is the power set. $\diamond$

It can be shown that the intersection of $\sigma-$fields of Ω is also a $\sigma-$field (Exercise A.**??**), thus the following definition makes sense.

Definition B.3. *Let $\mathscr{C}$ be a nonempty collection of subsets of Ω. The $\sigma-$**field generated by** $\mathscr{C}$ is the smallest $\sigma-$field in Ω containing $\mathscr{C}$, i.e.,*

$$\sigma(\mathscr{C}) := \cap\{\mathcal{F} \mid \mathscr{C} \subseteq \mathcal{F},\ \mathcal{F} \text{ is a } \sigma-\text{field}\}.$$

Example B.4. Let $\Omega = \mathbb{R}$. Consider the collection $\mathscr{C}$ of bounded open intervals

$$\mathscr{C} = \{(a,b) \mid a,b \in \mathbb{R},\ \ a < b\}.$$

The $\sigma-$field $\sigma(\mathscr{C})$ is known as the **Borel** $\sigma-$**field** of $\mathbb{R}$, also denoted $\mathcal{B}(\mathbb{R})$. The events of $\mathcal{B}(\mathbb{R})$ are called **Borel events** or **Borel sets**. We now present some events of the Borel $\sigma-$field $\mathcal{B}(\mathbb{R})$:

(a) Intervals of the form (a,∞) are Borel sets since

$$(a,\infty) = \cup_{n\in\mathbb{N}}(a, a+1).$$

Similarly for the event $(-\infty, b)$.

(b) Intervals of the form $(-\infty, a]$ are also events since

$$(-\infty, a] = (a, \infty)^c.$$

Similarly for the event $[b, \infty)$.

(c) Intervals of the form $[a, b)$, with $a < b$, because

$$[a, b) = [(-\infty, a) \cup [b, \infty)]^c.$$

Similarly, $(a, b]$ and $[a, b]$ are also events.

Countable unions and intersections of the above intervals are also Borel sets. However, not very subset of $\mathbb{R}$ is an event of the Borel $\sigma-$field (see, for instance, Rana [5, p. 113]). $\diamond$

Example B.5. Given a sample space Ω and a $\sigma-$field $\mathscr{F}$, we can endow any nonempty event S in Ω with the *induced* $\sigma-$field

$$\{S \cap A \mid A \in \mathscr{F}\}.$$

As a particular case, the induced Borel $\sigma-$field of $[0, 1]$, denoted $\mathcal{B}([0, 1])$, consists of all the Borel subsets of $[0, 1]$, that is,

$$\mathcal{B}([0, 1]) = \{A \in \mathcal{B}(\mathbb{R}) \mid A \subseteq [0, 1]\}.$$

$\diamond$

Definition B.6. *The events $E_1, E_2, E_3, \dots$ in $\mathscr{F}$ are* **disjoint** *if*

$$E_j \cap E_i = \emptyset \quad \text{for each} \;\; i \neq j.$$

B.2 Probability measures

Definition B.7. *Let $\mathscr{F}$ be a $\sigma-$field of subsets of Ω. A* **probability measure** *is any function $P : \mathscr{F} \to [0, 1]$ such that*

(a) $P(\Omega) = 1$, $P(\emptyset) = 0$, *and*

(b) *P is* **countably additive** *on $\mathscr{F}$, that is, for disjoint events $E_1, E_2, E_3, \dots$ in $\mathscr{F}$,*

$$\cup_{j=1}^{\infty} E_j \in \mathscr{F} \quad \Rightarrow \quad \mu(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} \mu(E_j).$$

We can observe that any probability measure is finitely additive. Thus

$$B = (B \setminus A) \cup (B \cap A) \quad \Rightarrow \quad P(B \setminus A) = P(B) - P(B \cap A).$$

In particular, $P(A^c) = 1 - P(A)$. Notice also that, $A \subseteq B$ implies $P(A) \leq P(B)$. With similar arguments we can prove the following properties of P.

Proposition B.8. *Let $\mathscr{F}$ be a $\sigma-$field and $P : \mathscr{F} \to \mathbb{R}$. Then for any events $A_1, A_2, \dots, A_n$ in $\mathscr{F}$,*

(a) $P(A_2 - A_1) = P(A_2) - P(A_2 \cap A_1)$,

(b) $P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{j=1}^{n} P(A_j)$.

One of the most important probability measures is the so-called **Lebesgue measure** which extends the *length* of intervals to a wider family of sets. Lebesgue's measure λ assigns a real number $\lambda(B)$ to each Borel set $B \in \mathcal{B}(\mathbb{R})$.

A **probability space** is a triple $(\Omega, \mathscr{F}, P)$ where $\mathscr{F}$ is a $\sigma-$field of subsets of Ω and $P : \mathscr{F} \to [0,1]$ is a probability measure.

Some examples of probability spaces are

(a) $(N, 2^N), P$ where N es finite, the $\sigma-$field 2^N is the *power set* of N, and P is any pmf on N.

(b) In part (a), N can be replaced by a countably infinite set.

(c) $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ is the Lebesgue's measure.

B.3 General random variables

Definition B.9. Let $(\Omega, \mathscr{F}, P)$ a probability space. A **random variable** is any *measurable* function $X : \Omega \to \mathbb{R}$, i.e.,

$$X^{-1}((-\infty, a]) \in \mathscr{F} \qquad \forall a \in \mathbb{R}.$$

Recall that $X \leq a$ is a shorter notation for the event

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega \mid X(\omega) \leq a\}.$$

The **probability distribution** $F_X : \mathbb{R} \to \mathbb{R}$ of X is defined as

$$F_X(x) := P(X \leq a).$$

If there is no confusion, we simply write F instead of F_X.

A proof of the following theorem can be found in Resnick [6, pp. 33-34].

Theorem B.10. *Let $(\Omega, \mathscr{F}, P)$ be a probability space and a random variable $X : \Omega \to \mathbb{R}$. Then its probability distribution $F : \mathbb{R} \to \mathbb{R}$ satisfies the following properties*

(a) *if $x < y$, then $F(x) \leq F(y)$,*

(b) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$,

(c) *F is right-continuous at any $a \in \mathbb{R}$, that is,* $\lim_{x \downarrow a} F(x) = F(a)$.

If $F : \mathbb{R} \to \mathbb{R}$ satisfies properties (a), (b), and (c) of Theorem B.10, then F is called a **distribution function**.

A converse of Theorem B.10 is useful to simulate random variables by means of he so-called *inverse transformation (or inversion) method*. This method uses the so-called **quantile function** Q given in the following theorem.

Theorem B.11. *Let U be a random variable with uniform distribution on the interval $[0,1]$. Given a distribution function F, define the quantile function*

$$Q(u) := \inf F^{-1}([u,1]) = \inf\{x \in \mathbb{R} \mid u \leq F(x)\}, \quad u \in (0,1). \tag{18}$$

Then $Q \circ U$ has probability distribution F.

Proof. To see that $Q \circ U$ indeed has distribution F, we note the following properties for each $u \in (0,1)$ and $a \in \mathbb{R}$.

(a) From the definition of Q, we conclude that Q is nondecreasing.

(b) $Q(u) \in F^{-1}([u,1])$, that is, $Q(u) = \min F^{-1}([u,1])$. To prove this assertion, observe that for each $n \in \mathbb{N}$, there exists $x_n \in F^{-1}([u,1])$ such that

$$Q(u) \leq x_n < Q(u) + \frac{1}{n}.$$

Then $F(x_n) \geq u$ for each n and $x_n \downarrow Q(u)$. Since F is right-continuous, in particular at $Q(u)$,

$$\begin{aligned} u &\leq \lim_{n\to\infty} F(x_n) \\ &= F(Q(u)), \end{aligned}$$

that is, $Q(u) \in F^{-1}([u,1])$.

(c) $Q(F(a)) \leq a$. This inequality follows from the definition (18) of $Q(F(a))$ because

$$a \in \{x \in \mathbb{R} \mid F(a) \leq F(x)\}.$$

(d) $Q^{-1}((-\infty, a]) = (0, F(a)]$ or, equivalently, $Q(u) \leq a$ if and only if $u \leq F(a)$.

Let us show first that $Q(u) \leq a$ implies $u \leq F(a)$. Notice that (b) allows us to assert the existence of some x_u such that $Q(u) = x_u$ and

$$F(x_u) \geq u.$$

Then $x_u \leq a$ and, by the monotonicity of F, $F(x_u) \leq F(a)$. But $u \leq F(x_u)$, hence $u \leq F(a)$.

Conversely, $u \leq F(a)$ implies $Q(u) \leq a$. Indeed, by part (a) and (c),

$$Q(u) \leq Q(F(a)) \leq a.$$

Finally, the theorem follows from (d) because

$$\begin{aligned} P(Q \circ U \leq a) &= P(Q[U(\omega)] \leq a) \\ &= P(U(\omega) \leq F(a)) \\ &= F(a). \end{aligned}$$

□

Remark B.12. The function Q is sometimes called the **quantile function** of F. If there exists the inverse function F^{-1} of F, then $Q = F^{-1}$.

Theorem B.13. *Let $F : \mathbb{R} \to \mathbb{R}$ be a distribution function. Then there exists a probability space $(\Omega, \mathscr{F}, P)$ and a random variable $X : \Omega \to \mathbb{R}$ such that*

$$P(X \leq a) = F(x) \qquad x \in \mathbb{R}.$$

Proof. Let $\Omega = (0,1)$, $\mathcal{B}((0,1))$, and $P = \lambda$. Given F, consider the corresponding quantile function Q. Define

$$X = Q \circ U,$$

where $U : (0,1) \to (0,1)$ is the uniform random variable. Then, by Theorem B.11, X has probability distribution F. $\square$

Remark B.14. Given a random variable X, the **moment-generating function** is

$$m_X(t) = E(e^{tX})$$

whenever it is finite on a neighborhood of the origin. If Y is another random variable and

$$m_X(t) = m_Y(t)$$

on a neighborhood of the origin, then $F_X(x) = F_Y(x)$ for every $x \in \mathbb{R}$. A proof of this fact can be found in Billingsley [2, Section 30] $\diamond$

References

[1] C. D. Aliprantis and O. Burkinshaw, *Principles of real analysis*, Academic Press, Inc., San Diego, CA, third ed., 1998.

[2] P. Billingsley, *Probability and measure*, John Wiley & Sons, Inc., New York, third ed., 1995.

[3] M. Lefebvre, *Applied probability and statistics*, Springer, New York, 2006.

[4] V. M. Panaretos, *Statistics for mathematicians. A rigorous first course*, Compact Textbooks in Mathematics, Birkhäuser/Springer, [Cham], 2016.

[5] I. K. Rana, *An introduction to measure and integration*, vol. 45 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2002.

[6] S. I. Resnick, *A probability path*, Birkhäuser Boston, Inc., Boston, MA, 1999.

[7] J. S. Rosenthal, *A first look at rigorous probability theory*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second ed., 2006.

[8] W. Rudin, *Principles of mathematical analysis*, International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third ed., 1976.

[9] D. D. Wackerly, W. Mendenhall III, and R. L. Scheaffer, *Mathematical Statistics with Applications*, Duxbury Advanced Series, seventh edition ed., 2008.