5. CONTINUOUS RANDOM VARIABLES

5.1. Definitions and basic properties.

Definition 5.1. A random variable $X : S \to \mathbb{R}$ is **continuous** if there is a **non-negative integrable function** f **defined on** $\mathbb{R}$ such that, for any proper subset B of $\mathbb{R}$ (Borel set),

$$P(X \in B) = \int_B f(x)dx.$$

Here, f is called the **probability density function** (or briefly p.d.f.) of X.

Remark 5.1. Let X be a continuous random variable with p.d.f. f. If $B = [a, b]$, then

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

Remark 5.2. Note that $\int_{-\infty}^{\infty} f(x)dx = 1$.

Example 5.1. For any $a \in \mathbb{R}$,

$$P(X = a) = P(X \in [a, a]) = \int_a^a f(x)dx = 0.$$

This implies

$$P(X \in (a, b)) = P(X \in (a, b]) = P(X \in [a, b)) = P(X \in [a, b]).$$

Let $B \subset \mathbb{R}$ be a Borel set. The above concept does not imply $P(X \in B) = P(X \in B \cup \partial B)$, where ∂B is the boundary of B. To see a counterexample, let $B = \mathbb{Q}$. Then, $\partial B = \mathbb{R}$ and $P(X \in B \cup \partial B) = 1$. However, by writing $B = \{a_1, a_2, ...\}$ with $a_i \neq a_i$ for $i \neq j$, we have

$$P(X \in B) = \sum_{i=1}^{\infty} P(X = a_i) = 0.$$

Definition 5.2. Given a continuous random variable X with p.d.f. f, the **cumulative distribution function** (or, briefly, the distribution function) of f is defined by

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx.$$

Remark 5.3. According to the fundamental theorem of calculus, if F is the distribution function of a continuous random variable with p.d.f. f, then F is continuous. Moreover, if f is continuous at a, then F is differentiable at a and $F'(a) = f(a)$. It's worthwhile to emphasize that the converse of the above discussion is not necessarily true.

Example 5.2. Suppose that the amount of time in hours that a computer functions before breaking down is a continuous random variable X with p.d.f. f given by

$$f(x) = \begin{cases} ce^{-x/10} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

To find the constant c, observe that

$$1 = \int_{-\infty}^{\infty} f(x)dx = c\int_0^{\infty} e^{-x/10}dx = 10c.$$

This implies $c = 1/10$. Then, the distribution function F is given by

$$F(a) = 0, \quad \forall a < 0, \quad F(a) = \int_0^a \frac{e^{-x/10}}{10}dx = 1 - e^{-a/10}, \quad \forall a \geq 0.$$

Definition 5.3. Let X be a continuous random variable with p.d.f. f.

(1) The **expectation** of X is defined by

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

provided $xf(x)$ is **absolutely integrable**, i.e. $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$.

(2) Assuming that $E(X)$ exists, the **variance** of X is defined to be $\mathrm{Var}(X) = E[(X - EX)^2]$.

Remark 5.4. If X is nonnegative but $xf(x)$ is not integrable, then we write $E(X) = \infty$.

Remark 5.5. Let X be continuous with p.d.f. f and d.f. F. If f has at most finitely many discontinuous points, then $dF(x) = f(x)dx$ and this implies

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xdF(x),$$

where the second integral refers to the Riemann-Stieltjes integral. It is remarkable that the second integral of the expectation applies for any random variable, including the discrete one.

Lemma 5.1. *Let Y be a nonnegative continuous random variable with d.f. F. Then,*

$$E(Y) = \int_0^{\infty} P(Y > y)dy = \int_0^{\infty} [1 - F(y)]dy.$$

Proof. Let f be the p.d.f. of Y. By Fubini's theorem, we have

$$E(Y) = \int_0^{\infty} yf(y)dy = \int_0^{\infty}\int_0^{y} f(y)dxdy = \int_0^{\infty}\int_x^{\infty} f(y)dydx = \int_0^{\infty} P(Y > x)dx.$$

□

Remark 5.6. The above lemma in fact holds for any nonnegative random variable.

Proposition 5.2. *Let X be a continuous random variable with p.d.f. f and let g be a real-valued (measurable) function. Then,*

(1) *For any (Borel) set $B \subset \mathbb{R}$,*

$$P(g(X) \in B) = \int_{g^{-1}(B)} f(x)dx,$$

where $g^{-1}(B) = \{x | g(x) \in B\}$.

(2) *The expectation of $g(X)$ is*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Proof. The first identity is obvious from the definition. For (2), let $Y = g(X)$ and F_Y be the distribution function of Y. By (1), we have

$$1 - F_Y(y) = P(Y > y) = P(g(X) > y) = \int_{g^{-1}(y,\infty)} f(x)dx,$$

By Remark 5.5, one has

$$E(Y) = \int_{-\infty}^{\infty} ydF_Y(y) = \int_{-\infty}^{0} ydF_Y(y) + \int_0^{\infty} ydF_Y(y).$$

Consider the latter term in the above equation.

$$\begin{aligned}\int_0^\infty y dF_Y(y) &= \int_0^\infty \int_0^y dz dF_Y(y) = \int_0^\infty \int_z^\infty dF_Y(y) dz = \int_0^\infty \left(F_Y(y)\Big|_{y=z}^{y=\infty} \right) dz \\ &= \int_0^\infty (1 - F_Y(z)) dz = \int_0^\infty \int_{g^{-1}((z,\infty))} f(x) dx dz = \iint_D f(x) dx dz,\end{aligned}$$

where $D = \{(x,z) | 0 < z < g(x)\}$. Again, by Fubini's theorem,

$$\int_0^\infty y dF_Y(y) = \int_{g^{-1}((0,\infty))} \int_0^{g(x)} f(x) dz dx = \int_{g^{-1}((0,\infty))} g(x) f(x) dx.$$

In a similar way, one may compute

$$\begin{aligned}\int_{-\infty}^0 y dF_Y(y) &= -\int_{-\infty}^0 \int_y^0 dz dF_Y(y) = -\int_{-\infty}^0 \int_{-\infty}^z dF_Y(y) dz = -\int_{-\infty}^0 \left(F_Y(y)\Big|_{y=-\infty}^{y=z} \right) dz \\ &= -\int_{-\infty}^0 F_Y(z) dz = -\int_{-\infty}^0 \int_{g^{-1}((-\infty,z])} f(x) dx dz = -\iint_E f(x) dx dz,\end{aligned}$$

where $E = \{(x,z) | g(x) \le z < 0\}$. Hence,

$$\int_{-\infty}^0 y dF_Y(y) = -\int_{g^{-1}((-\infty,0))} \int_{g(x)}^0 f(x) dz dx = \int_{g^{-1}((-\infty,0))} g(x) f(x) dx.$$

Combining both cases leads to the desired identity. □

Proposition 5.3. *Let X be a continuous random variable and a, b be real numbers.*

1. $E(aX + b) = aE(X) + b$.
2. $\mathrm{Var}(X) = E(X^2) - (EX)^2$.
3. $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.

Proof. (1) comes immediately from the linearity of integration. For (2) and (3), assuming that $\mu = E(X)$ exists, one has

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

This implies

$$\begin{aligned}\mathrm{Var}(aX + b) &= E(aX + b)^2 - [E(aX + b)]^2 \\ &= E(a^2X^2 + 2abX + b^2) - [a^2\mu^2 + 2ab\mu + b^2] \\ &= a^2 \mathrm{Var}(X).\end{aligned}$$

□

5.2. **Uniform random variables.**

Definition 5.4. Let $\alpha < \beta$. A random variable is **uniform over the interval** (α, β) if its p.d.f. satisfies

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{for } x \in (\alpha, \beta), \\ 0 & \text{otherwise.} \end{cases}$$

Remark 5.7. If X is uniform over (α, β), then

$$P(X \in B) = \frac{1}{\beta - \alpha} \int_B dx = \frac{\ell(B)}{\beta - \alpha}, \quad \forall B \subset (\alpha, \beta),$$

where $\ell(B)$ is the length of B. In particular, if $(a,b) \subset (\alpha,\beta)$, then $P(X \in (a,b)) = (b-a)/(\beta-\alpha)$. The distribution function of X is given by

$$F(a) = P(X \le a) = \begin{cases} 0 & \text{if } a \le \alpha \\ (a-\alpha)/(\beta-\alpha) & \text{if } \alpha < a < \beta \\ 1 & \text{if } a \ge \beta \end{cases}.$$

Proposition 5.4. *If X is uniform over (α,β), then $E(X) = (\alpha+\beta)/2$ and $\mathrm{Var}(X) = (\alpha-\beta)^2/12$.*

Proof. Note that

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} xdx = \frac{\alpha+\beta}{2}$$

and

$$E(X^2) = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} x^2dx = \frac{\alpha^2+\alpha\beta+\beta^2}{3}.$$

Then, the variance is given by $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = (\alpha-\beta)^2/12$. □

Example 5.3. Suppose that a bus arrives at a specified stop every 20 minutes after 7 A.M. A passenger arrives at the stop uniformly between 7 and 8. If X denotes the time that the passenger has to wait for the bus, then $X \in [0,20)$. Let Y be the time after 7 the passenger arrives the stop. Then, Y is uniform over $(0,60)$ and, for $a \in (0,20)$

$$\begin{aligned} P(X \le a) &= P(Y \in \{0\} \cup [20-a,20] \cup [40-a,40] \cup [60-a,60]) \\ &= P(Y=0) + P(Y \in [20-a,20]) \\ &\qquad + P(Y \in [40-a,40]) + P(Y \in [60-a,60]) \\ &= \frac{3a}{60} = \frac{a}{20}. \end{aligned}$$

Clearly, $P(X \le 0) = P(Y \in \{0,20,40,60\}) = 0$ and $P(X \le 20) = 1$. If F is the distribution of X, then

$$F(a) = \begin{cases} 0 & \text{if } a \le 0 \\ a/20 & \text{if } a \in (0,20) \\ 1 & \text{if } a \ge 20 \end{cases}.$$

This means that X is uniform over $(0,20)$ and, hence, $E(X) = 10$ and $\mathrm{Var}(X) = 100/3$.

5.3. Normal random variables.

Definition 5.5. A random variable X is a **normal random variable** with parameters (μ,σ^2) if its p.d.f. satisfies

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2} \quad \forall x \in \mathbb{R}.$$

If $\mu = 0$ and $\sigma = 1$, X is called a standard normal random variable.

Remark 5.8. To see the function f defined above is a p.d.f., we set $y = (x-\mu)/\sigma$. Then, $dy = \sigma^{-1}dx$ and

$$\int_{\mathbb{R}} f(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}dy.$$

Using the above identity, it remains to show that

$$\int_{-\infty}^{\infty} e^{-y^2/2}dy = \sqrt{2\pi}$$

or, equivalently,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y^2+z^2)/2}dydz = \left(\int_{-\infty}^{\infty} e^{-y^2/2}dy\right)^2 = 2\pi.$$

Consider the polar coordinate in $\mathbb{R}^2$

$$y = r\cos\theta, \quad z = r\sin\theta.$$

Note that the Jacobian determinant of the above change of variables is

$$J(r,\theta) = \det\left(\frac{\partial(y,z)}{\partial(r,\theta)}\right) = \det\begin{pmatrix} y_r & y_\theta \\ z_r & z_\theta \end{pmatrix} = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r.$$

This implies

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y^2+z^2)/2}dydz = \int_0^{\infty}\int_0^{2\pi} J(r,\theta)e^{-r^2/2}d\theta dr = \int_0^{\infty}\int_0^{2\pi} re^{-r^2/2}d\theta dr = 2\pi.$$

Remark 5.9. Let X be a standard normal random variable and let $\mu \in \mathbb{R}$ and $\sigma > 0$. Then, the random variable $Y = \mu + \sigma X$ is a normal random variable with parameters (μ, σ^2). To see the details, let F_X, F_Y be the distribution functions of X, Y with p.d.f. f_X, f_Y. Then, for $a \in \mathbb{R}$,

$$F_Y(a) = P(Y \le a) = P(\mu + \sigma X \le a) = P(X \le \tfrac{a-\mu}{\sigma}) = \int_{-\infty}^{(a-\mu)/\sigma} f_X(x)dx.$$

Hence,

$$f_Y(a) = \frac{d}{da}F_Y(a) = f_X(\tfrac{a-\mu}{\sigma}) \cdot \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma}e^{-(a-\mu)^2/2\sigma^2}.$$

Consequently, if $Z = a + bY$, then $Z \sim a + b(\mu + \sigma X) = a + b\mu + b\sigma X$, i.e. Z is a normal random variable with parameters $(a + b\mu, b^2\sigma^2)$.

Remark 5.10. Observe that if $X \sim N(\mu, \sigma)$ and F is the distribution function of the standard normal random variable, then the distribution function F_X of X satisfies $F_X(a) = F(\frac{a-\mu}{\sigma})$.

Proposition 5.5. *Let X be a normal random variable with parameters (μ, σ^2). Then, $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.*

Proof. By the previous example, it suffices to show the case $\mu = 0$ and $\sigma = 1$. In the standard normal case, since $xf(x)$ is an odd function, $E(X) = \int_{-\infty}^{\infty} xf(x)dx = 0$. This implies

$$\mathrm{Var}(X) = E(X^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2e^{-x^2/2}dx.$$

Using the integration by parts with $u = x$ and $v = -e^{-x^2/2}$, one has

$$\sqrt{2\pi}E(X^2) = \int_{-\infty}^{\infty} udv = uv|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} vdu = \int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}.$$

□

Theorem 5.6 (The DeMoivre-Laplace limit theorem)**.** *For $n \ge 1$, let X_n be a binomial random variable with parameters (n, p). Suppose $p \in (0, 1)$. Then,*

$$\lim_{n\to\infty} P\left(a \le \frac{X_n - np}{\sqrt{np(1-p)}} \le b\right) = \Phi(b) - \Phi(a),$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}dx$.

Remark 5.11. The above theorem means $(X_n - E(X_n))/\mathrm{SD}(X_n)$ converges in distribution to a standard normal random variable, which is a specific case of the central limit theorem.

5.4. **Exponential random variables.**

Definition 5.6. A random variable is an **exponential** random variable with parameter $\lambda > 0$ if its p.d.f. satisfies

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 5.12. If X is exponential(λ) with distribution function F, then

$$F(a) = \int_0^a \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_{x=0}^{x=a} = 1 - e^{-\lambda a}.$$

Proposition 5.7. *Let X be an exponential random variable with parameter $\lambda > 0$. Then, $E(X) = 1/\lambda$ and $\mathrm{Var}(X) = 1/\lambda^2$.*

Proof. Note that

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx, \quad E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx.$$

Letting $y = \lambda x$ implies

$$E(X) = \lambda^{-1} \int_0^\infty y e^{-y} dy = \lambda^{-1} \left(y(-e^{-y})\Big|_{y=0}^{y=\infty} + \int_0^\infty e^{-y} dy \right) = \lambda^{-1}$$

and

$$E(X^2) = \lambda^{-2} \int_0^\infty y^2 e^{-y} dy = \lambda^{-2} \left(y^2(-e^{-y})\Big|_{y=0}^{y=\infty} + 2\int_0^\infty y e^{-y} dy \right) = 2\lambda^{-2}.$$

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

Proposition 5.8 (The memoryless property)**.** *Let X be an exponential random variable. Then, X possesses the memoryless property, i.e.*

$$P(X > a + b | X > a) = P(X > b), \quad \forall a, b \geq 0.$$

Conversely, if X is nonnegative random variable satisfying $P(X = 0) < 1$ and the memoryless property, then X is an exponential random variable.

Remark 5.13. This proposition should be compared with geometric random variables.

Proof of Proposition 5.8. The memoryless property of exponential random variables is clear from the following fact

$$P(X > s) = \int_s^\infty \lambda e^{-\lambda x} dx = e^{-\lambda s}, \quad \forall s \geq 0.$$

Next, assume that X is nonnegative and satisfies the memoryless property. By the mathematical induction, one can show that, for any positive integer n and $a > 0$,

$$P(X > na) = P(X > a + (n-1)a) = P(X > a)P(X > (n-1)a) = [P(X > a)]^n.$$

Letting $a = 1/n$ implies $P(X > 1/n) = [P(X > 1)]^{1/n}$ and, hence, for any positive integers n, m,

$$P(X > n/m) = [P(X > 1/m)]^n = [P(X > 1)]^{n/m}.$$

Given $a > 0$, one may choose a decreasing sequence of rational numbers, say $(r_n)_{n=1}^{\infty}$, converging to a. By the right continuity of the distribution function F of X, we obtain

$$\begin{aligned} P(X > a) = 1 - F(a) = 1 - \lim_{n\to\infty} F(r_n) = \lim_{n\to\infty} P(X > r_n) \\ = \lim_{n\to\infty} [P(X > 1)]^{r_n} = [P(X > 1)]^a. \end{aligned}$$

Note that $P(X > 1) \neq 0$, otherwise

$$P(X = 0) = P(X \le 0) = \lim_{a\downarrow 0} P(X \le a) = 1 - \lim_{a\downarrow 0}[P(X > 1)]^a = 1,$$

which contradicts the assumption. Also, $P(X = 0 \neq 1$, otherwise

$$P(X < \infty) = \lim_{n\to\infty} P(X \le n) = 1 - \lim_{n\to\infty} P(X > n) = 1 - \lim_{n\to\infty} P(X > 1)^n = 0.$$

Consequently, setting $\lambda = -\ln P(X > 1)$ yields

$$\frac{d}{da}F(a) = -\frac{d}{da}(1 - F(a)) = -\frac{d}{da}[P(X > 1)]^a = \lambda e^{-\lambda a}, \quad \forall a > 0.$$

□

Definition 5.7. The **hazard rate** function of a positive continuous random variable with p.d.f. f and c.d.f. F is defined by

$$\lambda(t) = \frac{f(t)}{1 - F(t)}, \quad \forall t > 0.$$

Remark 5.14. To see an interpretation of a hazard rate function, consider the following conditional probability. Note that, for $s, t > 0$,

$$G(s) = P(X \in (t, t+s) | X > t) = \frac{P(t < X < t+s)}{P(X > t)} = \frac{1}{P(X > t)} \int_t^{t+s} f(x)dx.$$

Assuming that f is continuous at t, G has a right derivative at 0 and given by

$$\lim_{s\downarrow 0} \frac{G(s) - G(0)}{s} = \frac{f(t)}{P(X > t)} = \lambda(t).$$

If X denotes the lifetime of some item, then the value of $\lambda(t)$ can be regarded as the instantaneous rate of change of the conditional probability that the lifetime ends immediately after time t given that the item has survived for time t.

Example 5.4. If X is an exponential random variable with parameter $\lambda > 0$, then its hazard rate function is $\lambda(t) = \lambda e^{-\lambda t}/e^{-\lambda t} = \lambda$. This means that the hazard rate is constant regardless of how long the item has survived. Such an observation is consistent with the memoryless property of X.

Consider a positive continuous random variable X with p.d.f. f, c.d.f. F and hazard rate function $\lambda(t)$. Assume that f is continuous. Then, F is differentiable and

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt}\ln(1 - F(t)).$$

This implies, for $a > 0$,

$$\ln P(X > a) = \ln(1 - F(a)) = \ln(1 - F(t))\Big|_{t=0}^{t=a} = -\int_0^a \lambda(t)dt,$$

and, hence,

$$P(X > a) = \exp\left\{-\int_0^a \lambda(t)dt\right\}.$$

Example 5.5. The statement "The death rate of a person who smokes is twice that of a nonsmoker at any age." means: If X, Y are the lifetimes of a smoker and a nonsmoker with hazard rate functions λ_s, λ_n, then $\lambda_s(t) = 2\lambda_n(t)$. This implies that, for $0 < A < B$,

$$\begin{aligned} S =& P(\text{The lifetime reaches } B|\text{The smoker is of age } A) = \\ =& P(X > B)/P(X > A) = \exp\left\{-\int_A^B \lambda_s(t)dt\right\} \end{aligned}$$

and

$$\begin{aligned} N =& P(\text{The lifetime reaches } B|\text{The nonsmoker is of age } A) = \\ =& P(Y > B)/P(Y > A) = \exp\left\{-\int_A^B \lambda_n(t)dt\right\} \end{aligned}$$

Clearly, $S = N^2$.

5.5. **Other continuous random variables.**

Definition 5.8. A random variable has a **gamma** distribution with parameter (α, λ), where $\alpha, \lambda > 0$, if its p.d.f. satisfies

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty e^{-y}y^{\alpha-1}dy$.

Remark 5.15. For $\alpha = 1$, the gamma distribution is exactly the exponential distribution.

Remark 5.16. For $\alpha > 1$, one may apply the integration by parts to derive

$$\begin{aligned} \Gamma(\alpha) &= \int_0^\infty e^{-y}y^{\alpha-1}dy = \int_0^\infty y^{\alpha-1}d(-e^{-y}) \\ &= y^{\alpha-1}(-e^{-y})\Big|_0^\infty - \int_0^\infty (\alpha-1)(-e^{-y})y^{\alpha-2}dy = (\alpha-1)\Gamma(\alpha-1). \end{aligned}$$

Particularly, if $\alpha = n \in \mathbb{N}$, then $\Gamma(n) = (n-1)!$.

Proposition 5.9. *Let X be a gamma random variable with parameters (α, λ). Then, $E(X) = \alpha/\lambda$ and $\mathrm{Var}(X) = \alpha/\lambda^2$.*

Proof. The proof is a simple corollary of the above remarks. Let $y = \lambda x$. For the expectation, one has

$$E(X) = \frac{1}{\Gamma(\alpha)}\int_0^\infty x\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}dx = \frac{1}{\lambda\Gamma(\alpha)}\int_0^\infty e^{-y}y^\alpha dy = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

For the variance, note that

$$E(X^2) = \frac{1}{\Gamma(\alpha)}\int_0^\infty x^2\lambda e^{-\lambda x}(\lambda x)^{\alpha+1}dx = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{(\alpha+1)\alpha}{\lambda^2}.$$

This implies $\mathrm{Var}(X) = \alpha/\lambda^2$. □

Definition 5.9. A random variable has a **beta** distribution with parameters (a, b), where $a, b > 0$, if its density function satisfies

$$f(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} & \text{if } x \in (0,1), \\ 0 & \text{otherwise,} \end{cases}$$

where

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

Remark 5.17. Note that $B(a,b) = B(b,a)$. If X is a beta random variable with parameters (a,b) and $Y = 1 - X$, then, for $t \in (0,1)$,

$$\begin{aligned} F_Y(t) = P(Y \le t) = P(X \ge 1-t) &= \frac{1}{B(a,b)} \int_{1-t}^1 x^{a-1}(1-x)^{b-1}dx \\ &= \frac{1}{B(b,a)} \int_0^t (1-y)^{a-1} y^{b-1} dy. \end{aligned}$$

This implies $F_Y'(t) = \frac{1}{B(b,a)} t^{b-1}(1-t)^{a-1}$, i.e. Y has a beta distribution with parameters (b,a).

Lemma 5.10. *Let* Γ *be the gamma function and* $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$. *Then,* $B(a,b)\Gamma(a+b) = \Gamma(a)\Gamma(b)$.

Proof. Observe that

$$\Gamma(a)\Gamma(b) = \iint_D e^{-x} x^{a-1} e^{-y} y^{b-1} dxdy,$$

where $D = \{(x,y) | x > 0,\ y > 0\}$. Set $x = uv$ and $y = u(1-v)$. Then, the Jacobian determinant is given by

$$J(u,v) = \det\left(\frac{\partial(x,y)}{\partial(u,v)}\right) = \det\begin{pmatrix} v & u \\ 1-v & -u \end{pmatrix} = -u$$

and the region D is transformed into

$$D' = \{(u,v) | uv > 0, u(1-v) > 0\} = \{(u,v) | u > 0,\ 0 < v < 1\}.$$

Putting all above together, we obtain

$$\begin{aligned} \Gamma(a)\Gamma(b) &= \int_{D'} |J(u,v)| e^{-u} u^{a+b-2} v^{a-1}(1-v)^{b-1} dudv \\ &= \left(\int_0^\infty e^{-u} u^{a+b-1} du\right)\left(\int_0^1 v^{a-1}(1-v)^{b-1} dv\right) = \Gamma(a+b)B(a,b). \end{aligned}$$

□

Proposition 5.11. *Let* X *be a beta random variable with parameters* (a,b). *Then,* $E(X) = a/(a+b)$ *and* $\mathrm{Var}(X) = ab/[(a+b)^2(a+b+1)]$.

Proof. Observe that

$$E(X) = \frac{1}{B(a,b)} \int_0^1 x \cdot x^{a-1}(1-x)^{b-1}dx = \frac{B(a+1,b)}{B(a,b)}$$

and

$$E(X^2) = \frac{1}{B(a,b)} \int_0^1 x^2 \cdot x^{a-1}(1-x)^{b-1}dx = \frac{B(a+2,b)}{B(a,b)}.$$

Using the identity $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, we obtain

$$E(X) = \frac{\Gamma(a+1)\Gamma(b)/\Gamma(a+b+1)}{\Gamma(a)\Gamma(b)/\Gamma(a+b)} = \frac{a}{a+b}$$

and

$$E(X^2) = \frac{\Gamma(a+2)\Gamma(b)/\Gamma(a+b+2)}{\Gamma(a)\Gamma(b)/\Gamma(a+b)} = \frac{(a+1)a}{(a+b+1)(a+b)}.$$

This implies $\mathrm{Var}(X) = ab/[(a+b)^2(a+b+1)]$. □