# A Summary of Random Variables[1]

**Starting point:** A probability space $(\Omega, \mathcal{F}, \mathbb{P})$; $\Omega$ is the sample space, $\mathcal{F}$ is the collection of all events, $\mathbb{P}$ is the probability (also known as the probability measure or probability function).

**Definition:** a `random variable` $\xi$ is a *measurable* real-valued function on $\Omega$. That is, for every $\omega \in \Omega$, $\xi(\omega)$ is a real number, and, for all $-\infty < a < b < +\infty$, $\{\omega : a < \xi(\omega) < b\} \in \mathcal{F}$.

**Fact:** If $\xi$ is a random variable and $g = g(x)$ is a measurable real-valued function (continuous or Riemann integrable on every bounded interval will work), then $g(\xi)$ is a random variable.

**Definition:** the `cumulative distribution function` (cdf) of the random variable $\xi$ is the function

$$F_\xi(x) = \mathbb{P}(\xi \le x).$$

The following properties of the cdf follow from the properties of the probability measure $\mathbb{P}$:

- $\lim_{x\to-\infty} F_\xi(x) = 0$, $\lim_{x\to+\infty} F_\xi(x) = 1$, and, for $x < y$, $F_\xi(x) \le F_\xi(y)$;
- $F_\xi(x) = F_\xi(x^+)$, $F_\xi(x^-)$ exists, and $F_\xi(x) - F_\xi(x^-) = \mathbb{P}(\xi = x)$.[2]

**Definition:** For $\alpha \in (0,1)$, the $100\alpha\%$-quantile (or percentile) is a number $\xi_\alpha$ such that $\mathbb{P}(\xi \le \xi_\alpha) = \alpha$. The function $\alpha \mapsto \xi_\alpha = \min\{x \in \mathbb{R} : F_\xi(x) \ge \alpha\}$ is called the `quantile function`. If the inverse function $F_\xi^{-1}$ exists, then $\xi_\alpha = F_\xi^{-1}(\alpha)$.

`Quartiles` correspond to $\alpha = 0.25, 0.5, 0.75$, and $\xi_{0.5}$ is called the `median`. There is a alternative definition of the median, as a number $m_\xi$ such that $\mathbb{P}(\xi \le m_\xi) \ge 1/2$ and $\mathbb{P}(\xi \ge m_\xi) \ge 1/2$; with this definition, $m_\xi$ might not be unique.

**Definition:** random variable $\xi$ is called

- `discrete` if $\xi \in \{a_1, a_2, \ldots\}$; the collection of numbers $p_\xi(k) = \mathbb{P}(\xi = a_k)$ is called `probability mass function` (pmf) of $\xi$.
- `continuous` if $F_\xi$ is a continuous function for all $x \in \mathbb{R}$ (equivalently, $\mathbb{P}(\xi = x) = 0$ for every $x \in \mathbb{R}$;
- a continuous random variable $\xi$ is called `absolutely continuous` if there exists an non-negative integrable function $f_\xi = f_\xi(x)$, called the `probability density function` (pdf) of $\xi$, such that, for all $x \in \mathbb{R}$,
$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(y)\, dy.$$
- a continuous random variable is called `singular` if it is not absolutely continuous. The main example is a random variable that is uniform on the *Cantor set* in $[0,1]$; the corresponding cdf is sometimes called the Cantor (or devil's) staircase.

**Fact:** every random variable has a *unique* representation as a sum of a discrete, an absolutely continuous, and a singular random variables.

**Definition:** The `expected value` $\mathbb{E}(\xi)$ of a random variable $\xi$ is the number

$$\mathbb{E}(\xi) = -\int_{\infty}^{0} F_\xi(x)\, dx + \int_{0}^{+\infty} \big(1 - F_\xi(x)\big)\, dx,$$

provided at least one of the integrals is finite.

The following properties of the expected value follow from the definition, but not always in an easy way; because of that, other equivalent definitions exist.

- $\mathbb{E}(a\xi + b\eta) = a\mathbb{E}(\xi) + b\mathbb{E}(\eta)$ [$a, b$ are real numbers; $\xi, \eta$ are random variables];
- $\mathbb{E}(1) = 1$;
- If $\xi \le \eta$, then $\mathbb{E}(\xi) \le \mathbb{E}(\eta)$; in particular, if $|\xi| \le c$, then $\mathbb{E}|\xi| \le c$.

---

[1] Sergey Lototsky, USC

[2] $f(x^\pm) = \lim_{0<\varepsilon\to 0} f(x \pm \varepsilon)$.

- If $\xi \geq 0$, then $\mathbb{E}(\xi) = \int_0^\infty \mathbb{P}(\xi > x)\,dx$.
- If $A$ is an event, then $1_A$ or $\chi_A$ denotes the `indicator function` of $A$, which is the random variable equal to 1 if $\omega \in A$ and zero if $\omega \notin A$; then $\mathbb{E}(1_A) = \mathbb{P}(A)$.
- $|\mathbb{E}(\xi)| \leq \mathbb{E}|\xi|$; more generally, if $g = g(x)$ is a *convex* function, then $g\big(\mathbb{E}(\xi)\big) \leq \mathbb{E}\big(g(\xi)\big)$: `Jensen's inequality`.
- `The Law/Lemma/Theorem of the unconscious statistician`:

$$\mathbb{E}g(\xi) = \sum_k g(a_k)\mathbb{P}(\xi = a_k) \text{ (discrete } \xi);$$

$$\mathbb{E}g(\xi) = \int_{-\infty}^{+\infty} g(x) f_\xi(x)\,dx \text{ (absolutely continuous } \xi).$$

**Definition:** for a random variable $\xi$,

- `characteristic function` $\varphi_\xi(t) = \mathbb{E}\left(e^{\sqrt{-1}\,t\xi}\right)$, $t \in \mathbb{R}$ [cf. Fourier Transform];
- `moment generating function` $M_\xi(\lambda) = \mathbb{E}\left(e^{\lambda\xi}\right)$, $\lambda \in \mathbb{R}$ [cf. Laplace Transform];
- `moment` of order $k = 1, 2, \ldots$ is $\mathbb{E}\big(\xi^k\big)$;
- `absolute moment` of order $p > 0$ is $\mathbb{E}|\xi|^p$;
- `central moment` of order $k = 2, 3, 4, \ldots$ is $\mathbb{E}\big(\xi - \mu_\xi\big)^k$, where $\mu_\xi = \mathbb{E}(\xi)$. In particular, `variance` $\sigma_\xi^2 = \mathrm{Var}(\xi) = \mathbb{E}\big(\xi - \mu_\xi\big)^2$ is the central moment of order 2; `skewness` is *normalized* central moment of order 3, and `kurtosis` is normalized central moment of order 4. `Standard deviation` $\sigma_\xi$ is the square root of the variance.

**Note:**

- $|\varphi_\xi(t)| \leq 1$ for all $t \in \mathbb{R}$; if $\varphi_\xi(t) = \varphi_\eta(t)$ for all $t \in \mathbb{R}$, then $F_\xi(x) = F_\eta(x)$ for all $x$: the reason for "characteristic";
- $M_\xi(\lambda)$ might be infinite for all $\lambda \neq 0$; if $M_\xi(\lambda)$ is finite for all $\lambda$ near zero, then $M_\xi^{(k)}(0) = \mathbb{E}\big(\xi^k\big)$ ($k$-th derivative at zero is $k$-th moment: the reason for "moment generating");
- $\mathrm{Var}(a\xi + b) = a^2\mathrm{Var}(\xi)$ for all real numbers $a, b$;
- If $\xi$ represents a physical quantity and is measured in physical units (time, distance, etc.), then the mean and the standard deviation of $\xi$ (if defined) are measured in the same units, but skewness and kurtosis are always dimensionless.

**Definition:** Random variables $\xi$ and $\eta$ are called `independent` if the events $\{\omega : \xi(\omega) \leq a\}$ and $\{\omega : \eta(\omega) \leq b\}$ are independent *for all* $a, b \in \mathbb{R}$. `Joint independence` of a countable collection of random variables is defined similarly, via joint independence of the corresponding events. A countable collection $\xi_1, \xi_2, \ldots$ of random variables is called `iid` (independent and identically distributed) if the random variables are jointly independent and have the same cdf.

**Facts:** if $\xi$ and $\eta$ are independent, then $\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t)$, $t \in \mathbb{R}$, and, provided the corresponding objects are finite,

$$\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta),\ \mathrm{Var}(\xi + \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta);\ M_{\xi+\eta}(\lambda) = M_\xi(\lambda)M_\eta(\lambda).$$

**An example:** the Cantor random variable $\zeta$ can be written as $\zeta = 2\sum_{k=1}^{\infty} \frac{\xi_k}{3^k}$, where $\xi_k$, $k \geq 1$, are iid and take values 0 and 1 with probability 1/2 [then $2\xi_k$ takes values 0 or 2, which is consistent with the construction of the Cantor set]. Then we get $\mathbb{E}(\xi_k) = 1/2$, $\mathrm{Var}(\xi_k) = 1/4$, so that

$$\mathbb{E}(\zeta) = \frac{2}{2}\sum_{k=1}^{\infty} 3^{-k} = \frac{1/3}{1 - (1/3)} = \frac{1}{2},\ \mathrm{Var}(\zeta) = \frac{4}{4}\sum_{k=1}^{\infty} 3^{-2k} = \frac{1/9}{1 - (1/9)} = \frac{1}{8}.$$