Probability 1

Hadrien Montanelli

1 Events and probability

Arranging distinguishable objects Suppose that we have n distinguishable objects. There are

$$n! = n \times (n-1) \times \ldots \times 2 \times 1 \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

ways to order them.

Arranging objects that are not all distinguishable Suppose that we have n objects of k different types, and that type i appears m_i times, with $m_1 + \ldots + m_k = n$. There are

$$\frac{n!}{m_1! \times \ldots \times m_k!}$$

ways to order them.

Binomial coefficient If there are just two types of objects, the expression above is a *binomial coefficient*,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Binomial theorem and identities

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j},$$

$$\binom{n+1}{j+1} = \sum_{k=j}^{n} \binom{k}{j}.$$

Probability space A *probability space* is a triple $(\Omega, \mathcal{F}, \mathrm{P})$ where Ω is the *sample space*, $\mathcal{F}$ is a collection of subsets of Ω called *events* satisfying axioms F_1–F_3 below, and P is a probability measure, a function $\mathrm{P} : \mathcal{F} \to [0,1]$ satisfying axioms P_1–P_3 below.

F_1: $\emptyset \in \mathcal{F}$.

F_2: If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

F_3: If $\{A_i, i \in I\}$ is a finite or countably infinite collection of members of $\mathcal{F}$, then $\cup_{i\in I} A_i \in \mathcal{F}$.

P_1: For all $A \in \mathcal{F}$, $\mathrm{P}(A) \geq 0$.

P_2: $\mathrm{P}(\Omega) = 1$.

P_3: If $\{A_i, i \in I\}$ is a finite or countably infinite collection of members of $\mathcal{F}$, and $A_i \cap A_j = \emptyset$ for $i \neq j$, then $\mathrm{P}(\cup_{i\in I} A_i) = \sum_{i\in I} \mathrm{P}(A_i)$.

Conditional probability Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. If $A, B \in \mathcal{F}$ and $\mathrm{P}(B) > 0$, then the *conditional probability* of A given B is

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

Independence A family of events $\mathcal{A} = \{A_i : i \in I\}$ is *independent* if

$$\mathrm{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathrm{P}(A_i),$$

for all finite subsets J of I.

Law of total probability Suppose $\{B_1, B_2, \ldots\}$ is a partition of Ω by sets from $\mathcal{F}$, such that $\mathrm{P}(B_i) > 0$ for all $i \geq 1$. Then, for any $A \in \mathcal{F}$,

$$\mathrm{P}(A) = \sum_{i\geq 1} \mathrm{P}(A|B_i)\mathrm{P}(B_i).$$

Bayes' thereom Suppose $\{B_1, B_2, \ldots\}$ is a partition of Ω by sets from $\mathcal{F}$, such that $\mathrm{P}(B_i) > 0$ for all $i \geq 1$. Then, for any $A \in \mathcal{F}$ such that $\mathrm{P}(A) > 0$,

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(A|B_k)\mathrm{P}(B_k)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|B_k)\mathrm{P}(B_k)}{\sum_{i\geq 1}\mathrm{P}(A|B_i)\mathrm{P}(B_i)}.$$

Some useful rules for probabilities

$$\begin{array}{ll}
(i') & A \cup A^c = \Omega, \\
(ii') & A = (A \cap B^c) \cup (A \cap B), \\
(iii') & A \cup B = (A \cap B^c) \cup B.
\end{array}$$

$$\begin{array}{ll}
(i) & \mathrm{P}(A) + \mathrm{P}(A^c) = 1, \\
(ii) & \mathrm{P}(A) = \mathrm{P}(A \cap B^c) + \mathrm{P}(A \cap B), \\
(iii) & \mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B), \\
(iii*) & \mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) \quad \text{(disjoint)}, \\
(iv) & \mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A), \\
(iv*) & \mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \quad \text{(independent)}, \\
(v) & \mathrm{P}(A) = \mathrm{P}(A|B)\mathrm{P}(B) + \mathrm{P}(A|B^c)\mathrm{P}(B^c).
\end{array}$$

$$\begin{array}{ll}
(i) & \mathrm{P}(A_1 \cup \ldots \cup A_n) = 1 - \mathrm{P}(A_1^c \cap \ldots \cap A_n^c), \\
(iii) & \mathrm{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathrm{P}(A_i) \\
& \quad - \sum_{i<j} \mathrm{P}(A_i \cap A_j) + \sum_{i<j<k} \mathrm{P}(A_i \cap A_j \cap A_k) \\
& \quad - \ldots + (-1)^{n+1}\mathrm{P}(\cap_{1\leq i\leq n} A_i), \\
(iii*) & \mathrm{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathrm{P}(A_i) \text{ (disjoint)}, \\
(iv) & \mathrm{P}(A_1 \cap \ldots \cap A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\ldots \\
& \qquad \times \mathrm{P}(A_n|A_1 \cap \ldots \cap A_{n-1}), \\
(iv*) & \mathrm{P}(A_1 \cap \ldots \cap A_n) = \prod_{i=1}^{n} \mathrm{P}(A_i) \text{ (independent)}, \\
(v) & \mathrm{P}(A) = \sum_{i\geq 1} \mathrm{P}(A|B_i)\mathrm{P}(B_i).
\end{array}$$

2 Discrete random variables

Discrete random variable A *discrete random variable* X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a real-valued function $X : \Omega \to \mathrm{Im}X \subset \mathbb{R}$, such that $\{\omega \in \Omega : X(\omega) = k\} \in \mathcal{F}$ for each $k \in \mathrm{Im}X$, and $\mathrm{Im}X = \{X(\omega) : \omega \in \Omega\}$ is a finite or countable subset of $\mathbb{R}$.

Probability mass function (pmf) The *probability mass function* of a discrete random variable X is the function $p_X : \mathrm{Im}X \to [0, 1]$ defined by

$$p_X(k) = \mathrm{P}(X = k) = \mathrm{P}(\{\omega \in \Omega : X(\omega) = k\}).$$

Expected value The *expected value* (or *mean*) of a discrete random variable X is

$$\mathrm{E}(X) = \sum_{k \in \mathrm{Im}X} k\mathrm{P}(X = k).$$

Properties of expected value

$$\begin{aligned} &\mathrm{E}(aX + b) = a\mathrm{E}(X) + b, \\ &\mathrm{E}(h(X)) = \sum_{k \in \mathrm{Im}X} h(k)\mathrm{P}(X = k). \end{aligned}$$

Variance For a discrete random variable X, the *variance* of X is defined by

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

Property of variance

$$\mathrm{var}(aX + b) = a^2\mathrm{var}(X).$$

Uniform distribution $\mathrm{U}(n)$

- Models a situation where a finite number n of values are equally likely to be observed (e.g., a fair dice).
- *pmf and pgf* $(1 \le k \le n;\ n \ge 1)$*:*

$$\begin{aligned} &p_X(k; n) = \mathrm{P}(X = k; n) = \frac{1}{n}, \\ &G_X(s; n) = \frac{s - s^{n+1}}{n(1 - s)}. \end{aligned}$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{n + 1}{2}, \quad \mathrm{var}(X) = \frac{n^2 - 1}{12}.$$

Binomial distribution $\mathrm{Bin}(n, p)$

- Expresses the successes in a sequence of n independents experiments, each asking a yes-no question, and each with its own boolean-valued outcome: success/yes/one (with probability p) and failure/no/zero (with probability $1-p$). The case $n = 1$ is called the *Bernoulli distribution.*
- *pmf and pgf* $(0 \le k \le n;\ n \ge 1,\ 0 \le p \le 1)$*:*

$$\begin{aligned} &p_X(k; n, p) = \mathrm{P}(X = k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k}, \\ &G_X(s; n, p) = (1 - p + ps)^n. \end{aligned}$$

- *Mean and variance:*

$$\mathrm{E}(X) = np, \quad \mathrm{var}(X) = np(1 - p).$$

- *Sum:* If $X_i \sim \mathrm{Bin}(m_i, p)$, $1 \le i \le n$, are independent, then

$$\sum_{i=1}^{n} X_i \sim \mathrm{Bin}\left(\sum_{i=1}^{n} m_i, p\right).$$

- *Approximation:*

$$\begin{aligned} &\mathrm{Bin}(n, p) \to \mathrm{N}(np, np(1 - p)) \quad \text{as } n \to \infty \ (p \text{ fixed}), \\ &\mathrm{Bin}(n, p) \to \mathrm{Po}(np) \quad \text{as } n \to \infty,\ p \to 0 \ (np \text{ fixed}). \end{aligned}$$

Poisson distribution $\mathrm{Po}(\lambda)$

- Expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event.
- *pmf and pgf* $(k \ge 0;\ \lambda > 0)$*:*

$$\begin{aligned} &p_X(k; \lambda) = \mathrm{P}(X = k; \lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, \\ &G_X(s; \lambda) = e^{\lambda(s-1)}. \end{aligned}$$

- *Mean and variance:*

$$\mathrm{E}(X) = \mathrm{var}(X) = \lambda.$$

- *Sum:* If $X_i \sim \mathrm{Po}(\lambda)$, $1 \le i \le n$, are independent, then

$$\sum_{i=1}^{n} X_i \sim \mathrm{Po}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Geometric distribution $\mathrm{Geom}(p)$

- Models the number of independent trials needed until we see the first success, where p is the probability of success on a single trial.
- *pmf and pgf* $(k \ge 1;\ 0 \le p \le 1)$*:*

$$\begin{aligned} &p_X(k; p) = \mathrm{P}(X = k; p) = p(1 - p)^{k-1}, \\ &G_X(s; p) = \frac{ps}{1 - s(1 - p)}. \end{aligned}$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{1}{p}, \quad \mathrm{var}(X) = \frac{1 - p}{p^2}.$$

- *Sum:* If $X_i \sim \mathrm{Geom}(p)$, $1 \le i \le n$, are independent, then

$$\sum_{i=1}^{n} X_i \sim \mathrm{NegBin}(n, p).$$

Negative binomial distribution $\mathrm{NegBin}(n, p)$

- Models the number of independent trials needed until we see n successes, where p is the probability of success on a single trial. The case $n = 1$ corresponds to the geometric distribution.
- *pmf and pgf* ($k \geq n$; $n \geq 1$, $0 \leq p \leq 1$):

$$p_X(k; n, p) = \mathrm{P}(X = k; n, p) = \binom{k-1}{n-1} p^n (1-p)^{k-n},$$
$$G_X(s; n, p) = \left(\frac{ps}{1 - s(1-p)}\right)^n.$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{n}{p}, \quad \mathrm{var}(X) = \frac{n(1-p)}{p^2}.$$

Conditional distribution Suppose that B is an event such that $\mathrm{P}(B) > 0$. Then, the *conditional distribution* (or *conditional pmf*) of X given B is

$$\mathrm{P}(X = k | B) = \frac{\mathrm{P}(\{X = k\} \cap B)}{\mathrm{P}(B)}.$$

Conditional expectation Suppose that B is an event such that $\mathrm{P}(B) > 0$. Then, the *conditional expectation* of X given B is

$$\mathrm{E}(X|B) = \sum_{k \in \mathrm{Im}X} k \mathrm{P}(X = k|B).$$

Partition theorem for expectations Suppose there is a partition $\{B_1, B_2, \ldots\}$ of Ω by sets from $\mathcal{F}$, such that $\mathrm{P}(B_i) > 0$ for all $i \geq 1$. Then,

$$\mathrm{E}(X) = \sum_{i \geq 1} E(X|B_i) \mathrm{P}(B_i).$$

Joint probability mass function ($n = 2$) Given two discrete random variables X and Y, their *joint probability mass function* is

$$p_{X,Y}(k, \ell) = \mathrm{P}(\{X = k\} \cap \{Y = \ell\}) = \mathrm{P}(X = k, Y = \ell).$$

Marginal distribution ($n = 2$) Given two discrete random variables X and Y with joint pmf $p_{X,Y}$, the *marginal distribution* of X is

$$p_X(k) = \sum_{\ell \in \mathrm{Im}Y} p_{X,Y}(k, \ell),$$

while the *marginal distribution* of Y is

$$p_Y(\ell) = \sum_{k \in \mathrm{Im}X} p_{X,Y}(k, \ell).$$

Independence ($n = 2$) Two discrete random variables X and Y with joint pmf $p_{X,Y}$ are *independent* if

$$p_{X,Y}(k, \ell) = p_X(k) p_Y(\ell),$$

for all $k \in \mathrm{Im}X$ and $\ell \in \mathrm{Im}Y$.

Expected value ($n = 2$) Given two discrete random variables X and Y with joint pmf $p_{X,Y}$, and any function $h : \mathbb{R}^2 \to \mathbb{R}$, $h(X, Y)$ is also a discrete random variable and has expected value

$$\mathrm{E}(h(X, Y)) = \sum_{k \in \mathrm{Im}X} \sum_{\ell \in \mathrm{Im}Y} h(k, \ell) p_{X,Y}(X = k, Y = \ell).$$

Properties of expected value ($n = 2$)

$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y),$$
$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) \quad \text{(independent)}.$$

Variance and covariance ($n = 2$)

$$\mathrm{var}(h(X, Y)) = \mathrm{E}(h(X, Y)^2) - \mathrm{E}(h(X, Y))^2,$$
$$\mathrm{cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Properties of variance ($n = 2$)

$$\mathrm{var}(aX + bY) = a^2 \mathrm{var}(X) + b^2 \mathrm{var}(Y) + 2ab\mathrm{cov}(X, Y),$$
$$\mathrm{var}(aX + bY) = a^2 \mathrm{var}(X) + b^2 \mathrm{var}(Y) \quad \text{(uncorrelated)}.$$

Note that independence implies uncorrelatedness, but the converse is not always true.

Sum of discrete random variables ($n = 2$) Given two discrete random variables X and Y with joint pmf $p_{X,Y}$, the pmf of the sum $Z = X + Y$ is given by

$$p_Z(\ell) = \sum_{k \in \mathrm{Im}X} p_{X,Y}(k, \ell - k).$$

Joint probability mass function ($n \geq 2$) We can define *multivariate joint probability mass functions* analogously,

$$p_{X_1, \ldots, X_n}(k_1, \ldots, k_n) = \mathrm{P}(X_1 = k_1, \ldots, X_n = k_n).$$

Independence ($n \geq 2$) A family $\{X_i : i \in I\}$ of discrete random variables are *independent* if for all finite $J \subseteq I$ and all collections $\{A_i, i \in J\}$ of subsets of $\mathbb{R}$,

$$\mathrm{P}\left(\bigcap_{i \in J} \{X_i \in A_i\}\right) = \prod_{i \in J} \mathrm{P}(X_i \in A_i).$$

Properties of expected value and variance ($n \geq 2$)

$$\mathrm{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathrm{E}(X_i),$$
$$\mathrm{var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathrm{var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \mathrm{cov}(X_i, X_j),$$
$$\mathrm{var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathrm{var}(X_i) \quad \text{(uncorrelated)}.$$

Independent and identically distributed (iid) Suppose that $X_1, X_2, \ldots$ are independent discrete random variables that all have the same distribution. Then, we say that $X_1, X_2, \ldots$ are *independent and identically distributed.*

3 Difference equations and random walks

First-order linear recurrences The solution of

$$u_{n+1} = au_n + b$$

is of the form $u_n = v_n + w_n$, where $v_n = Aa^n$ is the general solution of the homogeneous equation, while $w_n = B$ is a particular solution to the equation.

If $a = 1$, then $w_n = Bn$. When instead of b, one has a polynomial of degree d, one may choose w_n to be a polynomial of the same degree d.

Second-order linear recurrences The solution of

$$u_{n+1} + au_n + bu_{n-1} = c$$

is of the form $u_n = v_n + w_n$ where $v_n = A\lambda_1^n + B\lambda_2^n$ is the general solution of the homogeneous equation, while $w_n = C$ is a particular solution to the equation. The scalars λ_1 and λ_2 are the solutions of

$$r^2 + ar + b = 0.$$

If $\lambda_1 = \lambda_2 = \lambda$, then $v_n = (An + B)\lambda^n$. When instead of c, one has a polynomial of degree d, one may choose w_n to be a polynomial of the same degree d.

Gambler's ruin A gambler repeatedly plays a game in which he wins \$1 with probability p and loses \$1 with probability $q = 1 - p$ (independently at each play). He will leave the casino if he loses all his money or if his fortune reaches $\$M$.

The probability that he leaves with nothing if his initial fortune is $\$n$ is

$$u_n = pu_{n+1} + qu_{n-1}, \quad 1 \le n \le M - 1,$$

with $u_0 = 1$ and $u_M = 0$. This yields, for $p \ne 1/2$,

$$u_n = \frac{\left(\frac{1-p}{p}\right)^n - \left(\frac{1-p}{p}\right)^M}{1 - \left(\frac{1-p}{p}\right)^M}, \quad 0 \le n \le M.$$

(For $p = 1/2$, one has $u_n = 1 - n/M$, $0 \le n \le M$.)

Let X be the number of plays before the gambler's fortune his either 0 or M. Then, $e_n = \mathrm{E}(X)$ is

$$\begin{aligned} e_n &= p\mathrm{E}(X|\text{won first play}) + q\mathrm{E}(X|\text{lost first play}), \\ &= p(1 + e_{n+1}) + q(1 + e_{n-1}), \quad 1 \le n \le M - 1, \end{aligned}$$

with $e_0 = e_M = 0$. This gives, for $p \ne 1/2$,

$$e_n = \left(\frac{M}{p-q}\right)\frac{1 - \left(\frac{q}{p}\right)^n}{1 - \left(\frac{q}{p}\right)^M} - \frac{n}{p-q}, \quad 0 \le n \le M.$$

(For $p = 1/2$, one has $e_n = n(M - n)$, $0 \le n \le M$.)

4 Probability generating functions

Probability generating function (pgf) Let X be a non-negative integer-valued random variable. The *probability generating function* of X is

$$G_X(s) = \mathrm{E}(s^X) = \sum_{k=0}^{\infty} s^k \mathrm{P}(X = k).$$

The probability mass function p_X of X is uniquely determined by G_X.

Independence Let X and Y be non-negative integer-valued random variables with pgfs G_X and G_Y. If X and Y are independent, then

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

Expected value and variance Let X be a non-negative integer-valued random variable with pgf G_X. Then,

$$\mathrm{E}(X) = G_X'(1), \quad \mathrm{var}(X) = G_X''(1) + G_X'(1) - G_X'(1)^2.$$

Sum of iid random variables Let $X_1, X_2, \ldots$ be iid non-negative integer-valued random variables with pgf G_X. Let N be another non-negative integer-valued random variable, independent of $X_1, X_2, \ldots$, and with pgf G_N. Then, the pgf of $\sum_{i=1}^N X_i$ is $G_N \circ G_X$.

Branching processes Suppose we have a population, and each individual in the population lives a unit time and, just before dying, gives birth to a random number of children in the next generation. This number of children has pmf $p(i)$, $i \ge 0$, called the *offspring distribution*.

Let X_n be the size of the population in generation n ($X_0 = 1$), and $C_i^{(n)}$ be the number of children of the ith individual in generation $n \ge 0$, so that we may write

$$X_{n+1} = C_1^{(n)} + \ldots + C_{X_n}^{(n)}.$$

Note that $C_1^{(n)}, \ldots, C_{X_n}^{(n)}$ are iid random variables. Let $G(s) = \sum_{i=0}^{\infty} s^i p(i)$ and $G_n(s) = \mathrm{E}(s^{X_n})$. Then,

$$G_{n+1}(s) = G_n(G(s)) = G(G_n(s)).$$

Moreover, if $\mu = \sum_{i=1}^{\infty} ip(i)$ is the mean number of children of a single individual, then

$$\mathrm{E}(X_n) = \mu^n.$$

The *extinction probability* is the smallest non-negative solution of $s = G(s)$.

5 Continuous random variables

Random variable A *random variable* X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a real-valued function $X : \Omega \to \mathbb{R}$, such that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. Discrete/continuous random variables are particular cases.

Cumulative distribution function (cdf) The *cumulative distribution function* of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = \mathrm{P}(X \leq x).$$

Continuous random variable A *continuous random variable* is a random variable whose cdf satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du,$$

where $f_X : \mathbb{R} \to \mathbb{R}$ is ≥ 0 and $\int f_X = 1$.

Probability density function (pdf) The function f_X defined in the previous paragraph is called the *probability density function* of X, and satisfies

$$\frac{dF_X(x)}{dx} = f_X(x), \quad \mathrm{P}(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

Expected value The *expected value* of a continuous random variable X is

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

Properties of expected value

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b,$$
$$\mathrm{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x)dx.$$

Variance For a continuous random variable X,

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

Property of variance

$$\mathrm{var}(aX + b) = a^2\mathrm{var}(X).$$

Uniform distribution $\mathrm{U}(a, b)$

- Models a situation where infinitely many uncountable values are equally likely to be observed.
- *pdf and cdf* ($a \leq x \leq b$; $a < b$)*:*

$$f_X(x; a, b) = \frac{1}{b - a},$$
$$F_X(x; a, b) = \frac{x - a}{b - a}.$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{a + b}{2}, \quad \mathrm{var}(X) = \frac{(b - a)^2}{12}.$$

Exponential distribution $\mathrm{Exp}(\lambda)$

- It is the probability distribution of the time between events in a process in which events occur continuously and independently at a constant average rate.
- *pdf and cdf* ($x \geq 0$; $\lambda \geq 0$)*:*

$$f_X(x; \lambda) = \lambda e^{-\lambda x},$$
$$F_X(x; \lambda) = 1 - e^{-\lambda x}.$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{1}{\lambda}, \quad \mathrm{var}(X) = \frac{1}{\lambda^2}.$$

Gamma distribution $\mathrm{Gamma}(\alpha, \lambda)$

- *pdf* ($x \geq 0$; $\alpha > 0$, $\lambda \geq 0$)*:*

$$f_X(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\lambda x}.$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{\alpha}{\lambda}, \quad \mathrm{var}(X) = \frac{\alpha}{\lambda^2}.$$

Beta distribution $\mathrm{Beta}(\alpha, \beta)$

- *pdf* ($0 \leq x \leq 1$; $\alpha > 0$, $\beta > 0$)*:*

$$f_X(x; \alpha, \beta) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha - 1}(1 - x)^{\beta - 1}.$$

- *Mean and variance:*

$$\mathrm{E}(X) = \frac{\alpha}{\alpha + \beta}, \quad \mathrm{var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Normal distribution $\mathrm{N}(\mu, \sigma^2)$

- The normal distribution is useful because of the central limit theorem, which states that averages of samples of observations of random variables independently drawn from independent distributions converge in distribution to the normal. Physical quantities that are expected to be the sum of many independent processes (such as measurement errors) often have distributions that are nearly normal.
- *pdf and cdf* ($x \in \mathbb{R}$; $\mu \in \mathbb{R}, \sigma^2 > 0$)*:*

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$
$$F_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-t^2/2}dt = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

- *Mean and variance:*

$$\mathrm{E}(X) = \mu, \quad \mathrm{var}(X) = \sigma^2.$$

- *Sum:* If $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $1 \leq i \leq n$, are independent, then

$$\sum_{i=1}^n a_i X_i \sim \mathrm{N}\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2\right).$$

Chi-squared distribution $\chi^2(n)$

- Let $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \text{N}(0,1)$. Then, we say that the random variable $X = Z_1^2 + \ldots + Z_n^2 \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$ has a chi-squared distribution with n degrees of freedom.
- *pdf* $(x \geq 0;\ n > 0)$*:*
$$f_X(x;n) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}.$$
- *Mean and variance:*
$$\text{E}(X) = n, \quad \text{var}(X) = 2n.$$
- *Sum:* If $X_i \sim \chi^2(m_i)$, $1 \leq i \leq n$, are independent, then
$$\sum_{i=1}^n X_i \sim \chi^2\left(\sum_{i=1}^n m_i\right).$$

Function of continuous random variable Suppose that X is a continuous random variable with density f_X and that $h : \mathbb{R} \to \mathbb{R}$ is a differentiable function that is strictly increasing. Then $Y = h(X)$ is a continuous random variable with pdf

$$f_Y(y) = f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y).$$

Joint cumulative distribution function ($n = 2$**)** The *joint cumulative distribution function* of a pair of random variables X and Y is the function $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ defined by

$$F_{X,Y}(x,y) = \text{P}(X \leq x, Y \leq y).$$

Jointly continuous random variable ($n = 2$**)** Let X and Y be two random variables such that

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)dudv,$$

where $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ is ≥ 0 and $\iint f_{X,Y} = 1$. Then, X and Y are said to be *jointly continuous*.

Joint probability density function ($n = 2$**)** The function $f_{X,Y}$ in the previous paragraph is called the *joint probability density function* of X and Y, and satisfies

$$\frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = f_{X,Y}(x,y),$$

$$\text{P}(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy.$$

Marginal distribution ($n = 2$**)** Let X and Y be two jointly continuous random variables with joint pdf $f_{X,Y}$. The *marginal distribution* of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$

while the marginal distribution of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

Independence ($n = 2$**)** Two jointly continuous random variables X and Y with joint pdf $f_{X,Y}$ are *independent* if the following holds,

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

for all $x, y \in \mathbb{R}$. This is equivalent to $F_{X,Y} = F_X F_Y$.

Expected value $(n = 2)$ Given two jointly continuous random variables X and Y with joint pdf $f_{X,Y}$, and any function $h : \mathbb{R}^2 \to \mathbb{R}$, $h(X,Y)$ is also a continous random variable and has expected value

$$\text{E}(h(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f_{X,Y}(x,y)dxdy.$$

Sum of continuous random variables ($n = 2$**)** Given two continuous random variables X and Y with joint pdf $f_{X,Y}$, the pdf of the sum $Z = X + Y$ is given by

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

6 Random samples and weak law of large numbers

Random sample Let $X_1, \ldots, X_n$ denote n iid random variables. Then these random variables are said to constitute a *random sample* of size n from the distribution.

Sample mean $\overline{X}_n$ The *sample mean* is defined to be

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

Mean and variance of $\overline{X}_n$ Suppose that $X_1, \ldots, X_n$ form a random sample of size n from a distribution with mean μ and variance σ^2. Then,

$$\text{E}(\overline{X}_n) = \mu, \quad \text{var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Markov's inequality Suppose that $Y \geq 0$ is random variable with finite mean. Then, for any $t > 0$,

$$\text{P}(Y > t) \leq \frac{\text{E}(Y)}{t}.$$

Chebyshev's inequality Suppose that Z is a random variable with finite mean and variance. Then, for any $t > 0$,

$$\text{P}(|Z - \text{E}(Z)| > t) \leq \frac{\text{var}(Z)}{t^2}.$$

Week law of large numbers Suppose that $X_1, \ldots, X_n$ are iid random variables with finite mean μ. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \text{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \epsilon\right) = 0.$$