A review of probability & statistics

Based of earlier versions by Andrea Kraus, Erwan Koch, Victor Panaretos & Shahin Tavakoli

Probability space

- A set $\Omega \neq \emptyset$
  - ✓ $\omega \in \Omega$: elementary events
  - ✓ subsets $A \subseteq \Omega$: events
- σ-algebra (also called σ-field) $\mathcal{A}$ of events
  - ✓ $(\Omega, \mathcal{A})$: measurable space
  - ✓ $A \in \mathcal{A}$ measurable sets
- Probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$
  - ✓ $(\Omega, \mathcal{A}, \mathbb{P})$ probability space
  - ✓ $\mathbb{P}(A)$ probability of event A
  - ✓ union bound: $\mathbb{P}(\cup_{i\in\mathbb{N}} A_i) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i) \ldots$
  - ✓ ... with equality for disjoint sets
- Why a σ-algebra?
  - ✓ if $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are defined, so are $\mathbb{P}(A^c)$, $\mathbb{P}(A \cap B)$, $\mathbb{P}(A \cup B)$, $\mathbb{P}(A \setminus B)$, ...

Conditional probability, independence

- Let $A, B \in \mathcal{A}$ such that $\mathbb{P}(B) > 0$
  - ✓ conditional probability $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
- Let $A, B_1, B_2, \ldots \in \mathcal{A}$, $B_1, B_2, \ldots$ disjoint, $\cup_i B_i = \Omega$,
  - ✓ law of total probability: $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)$
  - ✓ Bayes theorem: $\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}$ if $\mathbb{P}(A) > 0$
- The events $A_1, A_2, \dots \in \mathcal{A}$ are independent iff for any finite sub-collection $A_{i_1}, \ldots, A_{i_k}$:
  $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \mathbb{P}(A_{i_2}) \times \ldots \times \mathbb{P}(A_{i_k})$

Random variables

Random variables with values in $\mathbb{R}$

- probability space $(\Omega, \mathcal{A}, \mathbb{P})$, sample space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
  - ✓ (real-valued-) random variable is a measurable mapping $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ elementary event $\omega \in \Omega \rightsquigarrow$ a numerical value $X(\omega) \in \mathbb{R}$
- measurability
  - ✓ preimages of Borel sets are measurable, i.e. for every $B \in \mathcal{B}(\mathbb{R})$, $[X \in B] = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$
  - $\Rightarrow$ the probabilities $\mathbb{P}(X \in B)$ are well defined

Distribution of a random variable

- **Distribution function**
  - ✓ Definition: $F_X(x) = \mathbb{P}(X \le x), x \in \mathbb{R}$
  - ✓ Properties:
    - nondecreasing, right-continuous, left-limits (cadlag)
    - $\lim_{x\to\infty} F_X(x) = 1$, $\lim_{x\to-\infty} F_X(x) = 0$
    - $P(X > x) = 1 - F_X(x)$, $P(X = x) = F_X(x) - F_X(x-)$
- **Pushforward measure**
  - ✓ Definition: $\mathbb{P}_X(B) = \mathbb{P}(X \in B), B \in \mathcal{B}$
  - ✓ $\mathbb{P}_X$ is a probability measure on the sample space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
  - ✓ $\mathbb{P}_X$ is the pushforward of $\mathbb{P}$ by the function X
- **Computing probabilities:**
  - ✓ for $B \in \mathcal{B}$:
    $\mathbb{P}(X \in B) = \int_{[X\in B]} \mathrm{d}\,\mathbb{P}(\omega) = \int_B \mathrm{d}\,\mathbb{P}_X(x) = \int_B \mathrm{d}F_X(x)$
  - ✓ for $B = (a, b]$: $\mathbb{P}(X \in B) = F_X(b) - F_X(a)$

Density

- Radon–Nikodym theorem

  If $\nu \ll \mu$ then there is a measurable function ϕ such that $\nu(A) = \int_A \phi(u) d\mu(u)$
- ϕ is denoted $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$
- ϕ is called the Radon–Nikodym derivative
- ϕ is nonnegative, unique μ-a.e.

- Relevance for random variables:
- if $P_X \ll \mu$, then
  - ✓ $f_X = \frac{dP_X}{d\mu}$ is called the density of the distribution P_X
  - ✓ $P(X \in B) = P_X(B) = \int_B f_X(x) d\mu(x)$

Special Cases: continuous vs. discrete random variables

- continuous r. v.: $\mu =$ the Lebesgue measure
  - ✓ $P_X(B) = P(X \in B) = \int_B f_X(x)dx$
  - ✓ $F_X(x) = \int_{-\infty}^{x} f_X(u)du$
  - ✓ $F_X(x)$ continuous
  - ✓ $f_X(x) = \frac{d}{\mathrm{d}x} F_X(x)$
  - ✓ Intuition: "$f_X(x)dx \approx \mathbb{P}(x \leq X \leq x + \mathrm{d}x)$"
  - ✓ But: $f_X(x) \neq \mathbb{P}(X = x) = 0$
  - ✓ Can be $f_X(x) > 1$ for some x
- discrete r. v.: $\mu =$ the counting measure on $S = \{x_1, x_2, \dots\}$
  - ✓ $\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \sum_{j: x_j \in B} f_X(x_j)$
  - ✓ $F_X(x) = \sum_{j: x_j \leq x} f_X(x_j)$
  - ✓ $F_X(x)$ piecewise constant with possible jumps at $x_1, x_2, \dots$
  - ✓ $f_X(x) = \mathbb{P}(X = x)$ for every $x \in S$
  - ✓ $f_X(x) \leq 1$ for every $x \in S$
  - ✓ $\mathbb{P}(X = x)$ can be nonzero only for $x \in S$

Transformations of random variables

- $X \sim F_X$ continuous, $Y = h(X) \sim F_Y \Rightarrow F_Y = ?$
- h strictly monotonic:

$$F_Y(y) = P(h(X) \le y) = \begin{cases} P(X \le h^{-1}(y)) = F_X(h^{-1}(y)) & \text{if } h \text{ increasing,} \\ P(X \ge h^{-1}(y)) = 1 - F_X(h^{-1}(y)) & \text{if } h \text{ decreasing} \end{cases}$$

- h strictly monotonic and differentiable:
  differentiate to obtain $f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}$
- Special case:
  If $U \sim \text{Unif}(0,1)$ and F is a distribution function then $X = F^{-1}(U) \sim F$ (where $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\}$)

Expectation

- Definition:
  $\mathbb{E}\,X = \int_\Omega X(\omega)\,\mathrm{d}\,\mathbb{P}(\omega) = \int_\mathbb{R} x\,\mathrm{d}\,\mathbb{P}_X(x) = \int_\mathbb{R} x\,\mathrm{d}F_X(x)$
- Expectation is linear: $\mathbb{E}(X_1 + \alpha X_2) = \mathbb{E}(X_1) + \alpha X_2$
- If $X \geq 0$ a.s., we have $\mathbb{E}\,X = \int_0^\infty (1 - F_X(x))\,\mathrm{d}x$ (exercise)
- If $\mathbb{P}_X \ll \mu$ with $\frac{\mathrm{d}\,\mathbb{P}_X}{\mathrm{d}\mu} = f_X$ then $\mathbb{E}\,X = \int_\mathbb{R} x\,f_X(x)\,\mathrm{d}\mu(x)$
  - ✓ For continuous variables: $\mathbb{E}\,X = \int_\mathbb{R} x\,f_X(x)\,\mathrm{d}x$
  - ✓ For discrete variables: $\mathbb{E}\,X = \sum_j x_j\,f_X(x_j) = \sum_j x_j\,\mathbb{P}(X = x_j)$
- If h is nice
  $$\mathbb{E}\,h(X) = \int_\Omega h(X(\omega))d\,\mathbb{P}(\omega) = \int_\mathbb{R} h(x)d\,\mathbb{P}_X(x) = \int_\mathbb{R} h(x)dF_X(x)$$
  $$\left(= \int_\mathbb{R} h(x)f_X(x)d\mu(x)); \quad \text{or} \quad = \sum_j h(x_j)\,\mathbb{P}(X = x_j)\right)$$

Moments

- Definition: the k^{th} moment: $\mathbb{E}\, X^k$
- Central moments
  - ✓ $\mu_k = \mathbb{E}[(X - \mathbb{E}\, X)^k]$
- Variance
  - ✓ $\operatorname{Var} X = \mathbb{E}[(X - \mathbb{E}\, X)^2]$
  - ✓ $\operatorname{Var}(aX + b) = a^2 \operatorname{Var} X$
  - ✓ (Variance is like a squared norm. . .)
- Moment generating function
  - ✓ $M_X(t) = \mathbb{E} \exp\{tX\}$
  - ✓ if $M_X(t)$ is finite for $|t| < b$ for some $b > 0$ then $\mathbb{E}\, X^k = M_X^{(k)}(0)$
  - ✓ $M_X(t) = M_Y(t) < \infty$ around a neighbourhood of 0 $\implies X$ and Y have the same distribution

Random vectors; collections of random variables

Random Vectors

Equivalent Definitions

- random vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$
  - ✓ a vector of random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$
  - ✓ a random variable with values in $(\mathbb{R}^d, \mathcal{B}^d)$

Distribution

- $\mathbb{P}_{\mathbf{X}}$ is a probability measure on $(\mathbb{R}^d, \mathcal{B}^d)$:

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) = \mathbb{P}((X_1, \ldots, X_d)^\top \in B), \quad \text{where } B \in \mathcal{B}^d$$

- distribution function

$$F_{\mathbf{X}}(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

- $\lim_{x_j \to -\infty} F_{\mathbf{X}}(x_1, \ldots, x_d) = 0$ for any $j \in \{1, \ldots, d\}$
- $\lim_{x_1 \to +\infty, \ldots, x_d \to +\infty} F_{\mathbf{X}}(x_1, \ldots, x_d) = 1$

Density

- if $\mathbb{P}_{\mathbf{X}}$ is absolutely continuous w.r.t. a measure μ on $(\mathbb{R}^d, \mathcal{B}^d)$, then the joint density $f_{\mathbf{X}}(x_1, \ldots, x_d)$ is the Radon–Nikodym derivative of $\mathbb{P}_{\mathbf{X}}$ w.r.t. μ

- continuous case $(\mu = \lambda \times \cdots \times \lambda)$
  - ✓ $F_{\mathbf{X}}(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{\mathbf{X}}(u_1, \ldots, u_d)\, \mathrm{d}u_1 \ldots \mathrm{d}u_d$
  - ✓ $f_{\mathbf{X}}(x_1, \ldots, x_d) = \frac{\partial^d}{\partial x_1 \ldots \partial x_d} F_{\mathbf{X}}(x_1, \ldots, x_d)$

- discrete case $(\mu = \gamma_{S_1} \times \cdots \times \gamma_{S_d})$
  - ✓ $f_{\mathbf{X}}(u_1, \ldots, u_d) = P(X_1 = u_1, \ldots, X_d = u_d)$.
  - ✓ In this case the density is sometimes called probability mass function

- there could be more complicated cases . . .

Marginal distributions

- $\mathbf{X} = (X_1, \ldots, X_d)^\top = (\mathbf{Y}^\top, \mathbf{Z}^\top)^\top$
  where $\mathbf{Y} = (X_1, \ldots, X_r)^\top$ and $\mathbf{Z} = (X_{r+1}, \ldots, X_d)^\top$

- marginal distribution function of $\mathbf{Y}$:

$$F_{\mathbf{Y}}(\mathbf{y}) = \lim_{x_{r+1}\to\infty, \ldots, x_d\to\infty} F_{\mathbf{X}}(\mathbf{x}), \qquad \mathbf{y} \in \mathbb{R}^r$$

- if $\mathbf{X}$ has a density $f_{\mathbf{X}}(\mathbf{x})$ w.r.t. $\mu_{\mathbf{Y}} \times \mu_{\mathbf{Z}}$

$\Rightarrow$ then the marginal density of $\mathbf{Y}$ w.r.t. $\mu_{\mathbf{Y}}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}^{d-r}} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})\, \mathrm{d}\mu_{\mathbf{Z}}(\mathbf{z}), \qquad \mathbf{y} \in \mathbb{R}^r$$

- marginals DO NOT determine the joint distribution

Quantile function

Given a random variable X (taking values in $\mathbb{R}$) and a probability $\alpha \in (0, 1)$, the α-quantile function is supposed to be the number X such that $F_X(x) = \mathbb{P}(X \leq x) = \alpha$.
Since such x might not exist or might fail to be unique, we define the quantiles by the quantile function

$$F_X^-(\alpha) = \inf\{t \in \mathbb{R} : F_X(t) \geq \alpha\}.$$

F_X^- is left-continuous and $F_X(F_X^-(\alpha)) \geq \alpha$ for all $\alpha \in (0.1)$.
If F_X is continuous (e.g., X is conitnuous) then $F_X(F_X^-(\alpha)) = \alpha$.
if F_X is continuous and strictly increasing on the set $\{t : 0 < F_X(t) < 1\}$ (e.g., X has strictly positive density on a (possibly unbounded) interval I), then F_X is invertible and $F_X^- = F_X^{-1}$ is continuous.

Independence

- general definition:
  $X_1, X_2, \ldots$ (on the same probability space) independent
  - ✓ iff $\forall k = 1, 2, \ldots$, the events $[X_{i_1} \in B_{i_1}], [X_{i_2} \in B_{i_2}], \ldots, [X_{i_k} \in B_{i_k}]$ are independent for any finite sub-collection of Borel sets $B_{i_1}, B_{i_2}, \ldots, B_{i_k}$
  - ✓ iff the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots$ are independent

- $X_1, \ldots, X_d$ are independent iff

$$F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = F_{X_1}(x_1) \times \ldots \times F_{X_d}(x_d)$$

(for all $x_1, \ldots, x_d \in \mathbb{R}$)

- if densities exist: $X_1, \ldots, X_d$ are independent iff

$$f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = f_{X_1}(x_1) \times \ldots \times f_{X_d}(x_d)$$

($[\mu_1 \times \cdots \times \mu_d]$-a.e.)

Expectation, Covariance

- $\operatorname{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y)]$
  - ✓ $\operatorname{Cov}(aX + b, cY + d) = ac \operatorname{Cov}(X, Y)$
  - ✓ Bilinear: similar to a scalar product...
- $\operatorname{Corr}(X, Y) = \operatorname{Cov}(X, Y)/\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}$
  - ✓ $\operatorname{Corr}(aX + b, cY + d) = \operatorname{sgn}(ac) \operatorname{Corr}(X, Y)$
  - ✓ $\operatorname{Corr}(X, Y) = \pm 1 \implies \mathbb{P}(X = a \pm bY) = 1$
- $\operatorname{Var}(\sum_{i=1}^{d} X_i) = \sum_i \operatorname{Var} X_i + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j)$
- if X, Y are independent, then
  - ✓ $\mathbb{E}\,XY = \mathbb{E}\,X\,\mathbb{E}\,Y$
  - ✓ $\operatorname{Cov}(X, Y) = 0$ (converse is false!)
  - ✓ $\operatorname{Var}(X \pm Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$

Convolution

- X_1, X_2 continuous random variables with joint density $f_{(X_1,X_2)}$
- $Z = X_1 + X_2$
  - ✓ Z has density $f_Z(z) = \int_{\mathbb{R}} f_{(X_1,X_2)}(u, z-u)du$
  - ✓ if X_1, X_2 independent, $f_Z(z) = \int_{\mathbb{R}} f_{X_1}(u) f_{X_2}(z-u)du$
    i.e., $f_Z = f_{X_1} * f_{X_2}$ (convolution)

Transformations of random vectors

- continuous $\mathbf{X} = (X_1, \ldots, X_d)^\top$ with density $f_\mathbf{X}$
- $h : \mathbb{R}^d \to \mathbb{R}^d$
- $\mathbf{Y} = h(\mathbf{X}) = h(X_1, \ldots, X_d)$
- $\mathbb{P}(\mathbf{X} \in A) = 1$ for some open set $A \subset \mathbb{R}^d$
- $h : A \to h(A)$ is one-to-one, has continuous partial derivatives and $\mathbf{J}_h(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in A$
- then the density of $\mathbf{Y} = h(X_1, \ldots, X_d)$ is

$$f_\mathbf{Y}(\mathbf{y}) = \begin{cases} f_\mathbf{X}(h^{-1}(\mathbf{y}))|\mathbf{J}_h(h^{-1}(\mathbf{y}))|^{-1} & \\ \quad = f_\mathbf{X}(h^{-1}(\mathbf{y}))|\mathbf{J}_{h^{-1}}(\mathbf{y})|, & \mathbf{y} \in h(A), \\ 0 & \text{otherwise} \end{cases}$$

Conditional distribution

- $\mathbf{X} = (X_1, \ldots, X_d)^\top = (\mathbf{Y}^\top, \mathbf{Z}^\top)^\top$
  where $\mathbf{Y} = (X_1, \ldots, X_r)^\top$, $\mathbf{Z} = (X_{r+1}, \ldots, X_d)^\top$
- joint density $f_{\mathbf{X}}(\mathbf{y}, \mathbf{z})$ w.r.t. $\mu_{\mathbf{Y}} \times \mu_{\mathbf{Z}}$
- marginal density $f_{\mathbf{Z}}(\mathbf{z})$ w.r.t. $\mu_{\mathbf{Z}}$
- the conditional density of $\mathbf{Y}$ given $\mathbf{Z}$ is a measurable function $f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z})$ satisfying

$$\mathbb{P}(\mathbf{Y} \in B, \mathbf{Z} \in C) = \int_C \left[\int_B f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}) d\mu_{\mathbf{Y}}(\mathbf{y}) \right] f_{\mathbf{Z}}(\mathbf{z}) d\mu_{\mathbf{Z}}(\mathbf{z})$$

- computational formula:

$$f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}) = \begin{cases} f_{\mathbf{X}}(\mathbf{y}, \mathbf{z}) / f_{\mathbf{Z}}(\mathbf{z}), & \text{if } f_{\mathbf{Z}}(\mathbf{z}) \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

- More general cases require disintegration (nontrivial!)

Conditional expectation and variance

- $\mathbf{X} = (\mathbf{Y}^\top, \mathbf{Z}^\top)^\top$
- conditional expectation of $S = S(\mathbf{X}) = S(\mathbf{Y}, \mathbf{Z})$ given $\mathbf{Z} = \mathbf{z}$

$$\mathbb{E}[S|\mathbf{Z} = \mathbf{z}] = \int_{\mathbb{R}^{d-r}} S(\mathbf{y}, \mathbf{z}) f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}) d\mu_{\mathbf{Y}}(\mathbf{y})$$

- $\mathbb{E}[S|\mathbf{Z}]$ is a random variable (a function of $\mathbf{Z}$)
- $\mathbb{E}(\mathbb{E}[S|\mathbf{Z}]) = \mathbb{E}\, S$
- conditional variance $\mathrm{Var}[S|\mathbf{Z}] = \mathbb{E}[(S - \mathbb{E}[S|\mathbf{Z}])^2|\mathbf{Z}]$
- $\mathrm{Var}\, S = \mathrm{Var}(\mathbb{E}[S|\mathbf{Z}]) + \mathbb{E}(\mathrm{Var}[S|\mathbf{Z}])$
- general definition (conditioning on a σ-algebra): $\mathbb{E}[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable r.v. satisfying $\mathbb{E}\{1_A\, \mathbb{E}[X|\mathcal{G}]\} = \mathbb{E}\{1_A X\}$ for all $A \in \mathcal{G}$

Inequalities

Markov. Let Y be a non-negative random variable. Then

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}\, Y}{t}, \qquad \forall t \geq 0.$$

Chebishev. Let X be any random variable with finite first moment. Then

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\operatorname{Var} X}{t^2}, \qquad \forall t \geq 0.$$

Jensen. Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex with $\mathbb{E}\,|\phi(X)| + \mathbb{E}\,|X| < \infty$. Then $\phi(\mathbb{E}\, X) \leq \mathbb{E}\, \phi(X)$.

Monotonicity and covariance. If $\mathbb{E}[X^2] < \infty$ and $g : \mathbb{R} \to \mathbb{R}$ is nondecreasing with $\mathbb{E}[g^2(X)] < \infty$, then $\operatorname{Cov}[X, g(X)] \geq 0$.

Proof. $\mathbb{E}(X - \mathbb{E}\, X)(g(X) - g(\mathbb{E}\, X) + g(\mathbb{E}\, X) - \mathbb{E}\, g(X)) =$
$[\mathbb{E}(X - \mathbb{E}\, X)][g(\mathbb{E}\, X) - \mathbb{E}\, g(X)] + \mathbb{E}(X - \mathbb{E}\, X)(g(X) - g(\mathbb{E}\, X))$,
where the first term is zero and the second nonnegative since $g \uparrow$.

Some distributions

Bernoulli distribution

A random variable X follows a Bernoulli distribution with parameter $p \in [0, 1]$, denoted $X \sim \text{Bern}(p)$, if

1. it takes values in $\{0, 1\}$ (almost surely)
2. it has probability mass function $f(x; p) = p\mathbf{1}\{x = 1\} + (1 - p)\mathbf{1}\{x = 0\}$.

The mean, variance and moment generating functions of $X \sim \text{Bern}(p)$ are

$$\mathbb{E}[X] = p, \qquad \text{Var}[X] = p(1 - p), \qquad M(t) = 1 - p + pe^t.$$

Binomial distribution

A random variable X follows a Binomial distribution with parameters $p \in [0,1]$ and $n \in \mathbb{N}$, denoted $X \sim \mathrm{B}(n,p)$, if

1. it takes values in $\{0, 1, 2, \ldots, n\}$,
2. it has probability mass function $f(x; p, n) = \binom{n}{x} p^x (1-p)^{n-x}$.

The mean, variance and moment generating functions of $X \sim \mathrm{Binom}(n,p)$ are

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}[X] = np(1-p), \qquad M(t) = (1 - p + pe^t)^n.$$

If $X = \sum_{i=1}^n Y_i$ with $Y_i \overset{iid}{\sim} Bern(p)$, then $X \sim \mathrm{B}(n,p)$.

Geometric distribution

A random variable X follows a geometric distribution with parameter $p \in (0, 1]$, denoted $X \sim \text{Geom}(p)$, if

1. it takes values in $\{0\} \cup \mathbb{N}$,
2. it has probability mass function $f(x; p) = (1 - p)^x p$.

The mean, variance and moment generating functions of $X \sim \text{Geom}(p)$ are

$$\mathbb{E}[X] = \frac{1-p}{p}, \quad \text{Var}[X] = \frac{1-p}{p^2}, \quad M(t) = \frac{p}{1-(1-p)e^t}, \; t < \log\frac{1}{1-p}.$$

If $\{Y_i\}_{i \geq 1}$ are such that $Y_i \overset{iid}{\sim} Bern(p)$ and $X = \min\{k \in \mathbb{N} : Y_k = 1\} - 1$, then $X \sim Geom(p)$.

Negative binomial distribution

A random variable X follows a negative binomial distribution with parameters $p \in (0, 1]$ and $r > 0$, denoted $X \sim \text{NegBin}(r, p)$, if

1. it takes values in $\mathcal{X} = \{0\} \cup \mathbb{N}$,
2. it has probability mass function
$$f(x; p, r) = \binom{x + r - 1}{x} (1 - p)^x p^r.$$

The mean, variance and moment generating functions of $X \sim \text{NegBin}(r, p)$ are

$$\mathbb{E}[X] = r\frac{1 - p}{p}, \quad \text{Var}[X] = r\frac{1 - p}{p^2}, \quad M(t) = \frac{p^r}{[1 - (1 - p)e^t]^r}, \ t < -\text{lo}$$

If $r \in \mathbb{N}$ and $X = \sum_{i=1}^{r} Y_i$ with $Y_i \overset{iid}{\sim} Geom(p)$ then $X \sim \text{NegBin}(r, p)$.

Poisson distribution

A random variable X follows a negative binomial distribution with parameter $\lambda > 0$, denoted $X \sim \text{Poisson}(\lambda)$, if

1. it takes values in $\mathcal{X} = \{0\} \cup \mathbb{N}$,
2. it has probability mass function $f(x; \lambda) = e^{-\lambda} \dfrac{\lambda^x}{x!}$.

The mean, variance and moment generating functions of $X \sim \text{Poisson}(\lambda)$ are

$$\mathbb{E}[X] = \lambda, \qquad \text{Var}[X] = \lambda, \qquad M(t) = \exp\{\lambda(e^t - 1)\}.$$

One can show that $\mathit{Binom}(n, \lambda/n) \to \mathit{Poisson}(\lambda)$ as $n \to \infty$

Uniform distribution

A random variable X follows a uniform distribution with parameters $-\infty < \theta_1 < \theta_2 < \infty$, denoted $X \sim \text{Unif}(\theta_1, \theta_2)$, if it has the density function

$$f_X(x;\theta) = \begin{cases} (\theta_2 - \theta_1)^{-1} & \text{if } x \in (\theta_1, \theta_2), \\ 0 & \text{otherwise.} \end{cases}$$

The mean, variance and moment generating functions of $X \sim \text{Unif}(\theta_1, \theta_2)$ are

$$\mathbb{E}[X] = (\theta_1 + \theta_2)/2,\ \text{Var}[X] = (\theta_2 - \theta_1)^2/12,\ M(t) = \frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)},\ t \neq 0,$$

with $M(0) = 1$.

Exponential distribution

A random variable X follows an exponential distribution with parameter $\lambda > 0$, denoted $X \sim \text{Exp}(\lambda)$, if it has the density function

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

The mean, variance and moment generating functions of $X \sim Exp(\lambda)$ are

$$\mathbb{E}[X] = \lambda^{-1}, \qquad \text{Var}[X] = \lambda^{-2}, \qquad M(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

Gamma and χ^2 distributions

A random variable X follows a gamma distribution with parameters $r > 0$ et $\lambda > 0$, denoted $X \sim \text{Gamma}(r, \lambda)$, if it has the density function

$$f_X(x; r, \lambda) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & \text{si } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

The mean, variance and moment generating functions are

$$\mathbb{E}[X] = r/\lambda, \qquad \text{Var}[X] = r/\lambda^2, \qquad M(t) = \left(\frac{\lambda}{\lambda - t}\right)^r, \quad t < \lambda.$$

The χ^2_ν **distribution** is the particular case $\lambda = 1/2$ and $r = \nu/2$.
If $r \geq 1$ is an integer, then $\Gamma(r) = (r-1)!$.
If $r \in \mathbb{N}$ and $X = \sum_{i=1}^r Y_i$ with $Y_i \overset{iid}{\sim} Exp(\lambda)$ then $X \sim \text{Gamma}(r, \lambda)$.

Normal (or Gaussian) distribution

A random variable X follows a normal distribution with parameters $\mu \in \mathbb{R}$ et $\sigma^2 > 0$, denoted $X \sim \mathsf{N}(\mu, \sigma^2)$, if it has the density function

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \mathbb{R}.$$

The mean, variance and moment generating function of $X \sim \mathsf{N}(\mu, \sigma^2)$ are

$$\mathbb{E}[X] = \mu, \qquad \mathrm{Var}[X] = \sigma^2, \qquad M(t) = \exp\{t\mu + t^2\sigma^2/2\}.$$

If $Z \sim N(0, 1)$, we use the notation $\varphi(z) = f_Z(z)$ and $\Phi(z) = F_Z(z)$ for the corresponding density and distribution functions, which are called standard normal/Gaussian density/distribution functions.

Student t and Fisher–Snedecor distributions

A random variable X is said to follow the **Student** t distribution with parameter $k \in \mathbb{N}$ (called the number of degrees of freedom), denoted $X \sim t_k$, if it can be written as

$$X = \sqrt{n}\frac{\overline{Y}_n - \mu}{S}$$

where $n = k + 1$, $Y_1, \ldots, Y_n \overset{iid}{N} (\mu, \sigma^2)$, $\overline{Y}_n = \sum_{i=1}^n Y_i/n$, and $S^2 = \sum_{i=1}^n (\overline{Y}_i - Y)^2/(n-1)$.
If $k > 1$ then $\mathbb{E}\, X = 0$ and if $k > 2$ then $\operatorname{Var} X = k/(k-2)$.
A random variable X is said to follow the **Fisher–Snedecor** F distribution with parameters $d_1, d_2 > 0$ if it can be written as $X = \frac{X_1/d_1}{X_2/d_2}$, where $X_1 \sim \chi^2_{d_1}$ and $X_2 \sim \chi^2_{d_2}$ are independent.