Probability and Statistics

Giuliano Casale

Department of Computing, Imperial College London

List of Probability Formulas

Key formulas: Events and Probability

- Outcomes, sample space, events
- σ-algebra and axioms of probability

  Axiom 1. $\forall E \in \mathcal{F},\ 0 \leq \mathrm{P}(E) \leq 1$;
  Axiom 2. $\mathrm{P}(S) = 1$;
  Axiom 3. For *mutually exclusive events* $E_1, E_2, \ldots \in \mathcal{F}$

$$\mathrm{P}\left(\bigcup_i E_i\right) = \sum_i \mathrm{P}(E_i).$$

- Independent events

$$\mathrm{P}(E \cap F) = \mathrm{P}(E)\mathrm{P}(F)$$

Key formulas: Events and Probability

- Conditional probability (for $P(F) \neq 0$)

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(E \cap F)}{\mathrm{P}(F)}$$

- Conditional independence

$$\mathrm{P}(E_1 \cap E_2|F) = \mathrm{P}(E_1|F)\mathrm{P}(E_2|F)$$

- Bayes theorem

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(F|E)\mathrm{P}(E)}{\mathrm{P}(F)}$$

- Law of total probability

$$\mathrm{P}(E) = \sum_i \mathrm{P}(E|F_i)\mathrm{P}(F_i)$$

Key formulas: Random Variables

- Probability space, Random variables (rv.s)
- Induced probabilities

$$\mathrm{P}_X(X \leq x) \equiv \mathrm{P}(S_x)$$

with $S_x = \{s \in S | X(s) \leq x\}$.

- Support of a random variable

$$\mathrm{supp}(X) \equiv \{x \in \mathbb{R} \mid \exists s \in S \text{ s.t. } X(s) = x\}$$

- Cumulative distribution function

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

- Interval probabilities

$$\mathrm{P}_X(a < X \leq b) = F_X(b) - F_X(a).$$

Key formulas: Discrete rv.s

- Discrete rv.s:

$$X \text{ is discrete } \iff \text{supp}(X) \text{ is countable.}$$

- Probability mass function (pmf)

$$p(x_i) = \mathrm{P}_X(X = x_i) = F_X(x_i) - F_X(x_{i-1}),$$

- pmf properties:
  1. $0 \leq p(x) \leq 1, \forall x \in \mathbb{R};$
  2. $\sum_x p(x) = 1.$

Key formulas: Expectation

- Expectation of rv X

$$\mathrm{E}(X) = \sum_x x\, p(x)$$

- Expectation of a function of a rv

$$\mathrm{E}(g(X)) = \sum_x g(x)\, p(x)$$

- Expectation of a linear transformation

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b, \qquad \forall a, b \in \mathbb{R}$$

- Expectation of a sum of rv.s

$$\mathrm{E}(g(X) + h(X)) = \mathrm{E}(g(X)) + \mathrm{E}(h(X))$$

Key formulas: Moments

- Raw moments: $E[X^n]$
- Mean: $\mu = E[X]$
- Variance: $\sigma^2 = Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Standard deviation: $\sigma = \mathrm{sd}(X) = \sqrt{\mathrm{Var}_X(X)}$
- Coefficient of Variation: $c = \sigma/\mu$
- Skewness

$$\gamma_1 = \frac{\mathrm{E}[(X-\mu)^3]}{\sigma^3} = \frac{\mathrm{E}[(X-\mathrm{E}(X))^3]}{\mathrm{sd}(X)^3}.$$

- Variance of a linear transformation:

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

- Sum of n random variables: $S_n = X_1 + \ldots + X_n$
- Moments of a sum of i.i.d. rv.s[1]

$$\mathrm{E}[S_n] = n\mu, \qquad \mathrm{Var}\,(S_n) = n\sigma^2$$

- Moments of a sum of independent, but not identically distributed, rv.s

$$\mathrm{E}[S_n] = \sum_{i=1}^{n} E[X_i], \qquad \mathrm{Var}\,(S_n) = \sum_{i=1}^{n} Var(X_i)$$

---

[1] The formula for $\mathrm{E}[S_n]$ still holds when $n \to \infty$ as

$$\lim_{n\to\infty} E[S_n/n] = \mu$$

a celebrated result called the (Weak) *Law or the Large Numbers (LLN)*.

Key formulas: Notable discrete distributions

Distribution	*rv*	*pmf*	μ	σ^2
Bernoulli(p)	$X \in \{0,1\}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial(n,p)	$X \in \{0,\dots,n\}$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Geometric(p)	$X \in \{1,2,\dots\}$	$p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$
Poisson(λ)	$X \in \{0,1,\dots\}$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$	λ	λ
Uniform($1,n$)	$X \in \{1,\dots,n\}$	$\dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$

Key formulas: Continuous Random Variables

- Probability density function (pdf):

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

where $f_X(u) \geq 0$. Unlike pmfs, pdf values are not probabilities.

- The fundamental theorem of calculus implies

$$f_X(x) = \frac{d}{dx}F_X(x) = F'_X(x).$$

- Interval probability:

$$\mathrm{P}_X(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$$

- For countable sets:

$$\mathrm{P}_X(X \in \{x_1, x_2, \ldots\}) = 0$$

Key formulas: Moments of continuous rv.s

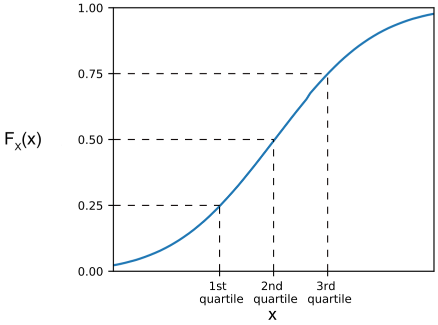
Similarly defined as for discrete rv.s, but based on densities:

- $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$
- $\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
- $\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \mu_X^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$

Linearity of expectation still applies:

- $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b,$
- $\mathrm{E}(g(X) + h(X)) = \mathrm{E}(g(X)) + \mathrm{E}(h(X))$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X), \qquad \forall a, b \in \mathbb{R}$

Key formulas: Quantiles and percentiles



- Lower quartile = 1st quartile = 25th percentile = 0.25-quantile.
- Median = 2nd quartile = 50th percentile = 0.50-quantile.
- Upper quartile = 3rd quartile= 75th percentile = 0.75-quantile.

(CDF inversion is also important to generate random samples.)

Key formulas: Notable continuous distributions

Distribution	*rv*	*pdf*	μ	σ^2
$\text{Exp}(\lambda)$	X	$\lambda e^{-\lambda x}$, $x > 0$	λ^{-1}	λ^{-2}
$N(\mu, \sigma^2)$	X	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty \leq x \leq +\infty$	μ	σ^2
$\text{Lognormal}(\mu, \sigma^2)$	$Y = e^X$ $X \sim N(\mu, \sigma^2)$	$\frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}}$, $y > 0$	$e^{\mu+\sigma^2/2}$	$e^{\sigma^2} - 1$
$U(a, b)$	X	$\frac{1}{b-a}$, $a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

Key formulas: Central limit theorem

- Moment generating function (mgf): $M_X(t) = E[e^{tX}]$.
- Characteristic function: $\phi_X(t) = M_X(it) = E[e^{itX}]$.
- Mgf of a sum of independent random variables $S_n = \sum_{i=1}^{n} X_i$:

$$M_{S_n}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

- Central limit theorem

$$\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n}\sigma} \sim \mathrm{N}(0,1).$$

Key formulas: Joint Random Variables

- Sample space S shared by two or more rv.s, e.g. $S = S_1 \times S_2$.
- The induced probability $\mathrm{P}_Z(\cdot)$ specified from semi-open box

$$S_{xy} = \{s \in S | X(s) \leq x \text{ and } Y(s) \leq y\}$$

  The idea generalizes to more than two rv.s.
- Joint CDF

$$F(x, y) = \mathrm{P}_Z((-\infty, x], (-\infty, y]) = \mathrm{P}(S_{xy})$$

- Interval probabilities:

$$\mathrm{P}_Z(x_1 < X \leq x_2, y_1 < Y \leq y_2) =$$
$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1).$$

  This also gives $\mathrm{P}_Z(X = x_2, Y = y_2)$ for discrete joint rv.s.

Key formulas: Joint pmf, pdfs, and marginals

- Discrete case: the joint pmf is

$$p(x,y) = \mathrm{P}_Z(X = x, Y = y), \qquad x, y \in \mathbb{R}.$$

Marginals:

$$p_X(x) = \sum_y p(x,y), \qquad p_Y(y) = \sum_x p(x,y).$$

- Continuous case: joint pdf is a non-negative function $f(x,y)$

$$F(x,y) = \int_{t=-\infty}^{y} \int_{s=-\infty}^{x} f(s,t)\,ds\,dt, \qquad x, y \in \mathbb{R}$$

Marginals:

$$f_X(x) = \int_{y=-\infty}^{\infty} f(x,y)dy, \quad f_Y(y) = \int_{x=-\infty}^{\infty} f(x,y)dx.$$

Key formulas: Notable Joint Distributions

- Multinomial distribution with n i.i.d. trials, r possible outcomes, each with probability q_i. Let X_i be the number of experiments that yield outcome i, then:

$$\mathrm{P}(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} q_1^{n_1} q_2^{n_2} \cdots q_r^{n_r}$$

- Multivariate normal distribution with mean $\mu = (\mu_1, \ldots, \mu_n)$ and covariance matrix $\Sigma = [Cov(X_i, X_j)]$. The joint density is:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

Key formulas: Independence and Conditioning

- Independence conditions

$$F(x,y) = F_X(x)F_Y(y),$$

or equivalenty, if and only if pmf or pdf factorize.

- Conditional probability

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}, \qquad x,y \in \mathbb{R},$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \qquad x,y \in \mathbb{R}.$$

where for continuous rv.s we use the density in place of probability since $\mathrm{P}(X = x) = 0$ always.

- Bayes theorem and partition rule extend to joint rv.s using conditional probabilities.

Key formulas: Expectation of joint rv.s

- If X and Y are discrete, we define $\mathrm{E}\big(g(X,Y)\big)$ by
$$\mathrm{E}\big(g(X,Y)\big) = \sum_y \sum_x g(x,y)p(x,y).$$

- If X and Y are jointly continuous, we define $\mathrm{E}\big(g(X,Y)\big)$ by
$$\mathrm{E}\big(g(X,Y)\big) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x,y)f(x,y)dxdy.$$

- If $g(X,Y) = g_1(X) + g_2(Y)$,
$$\mathrm{E}\big(g_1(X) + g_2(Y)\big) = E_X\big(g_1(X)\big) + E_Y\big(g_2(Y)\big).$$

- If $g(X,Y) = g_1(X)g_2(Y)$ and X and Y are *independent*,
$$\mathrm{E}\big(g_1(X)g_2(Y)\big) = \mathrm{E}_X\big(g_1(X)\big)\mathrm{E}_Y\big(g_2(Y)\big).$$

Key formulas: Expectation of joint rv.s

- Covariance:

$$\begin{aligned}\sigma_{XY} = \mathrm{Cov}(X, Y) &= \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathrm{E}[XY] - \mu_X \mu_Y\end{aligned}$$

- Correlation:

$$\rho_{XY} = \mathrm{Cor}(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- Conditional expectation:

$$\mathrm{E}_{Y|X}(Y|x) = \sum_y y \; p_{Y|X}(y|x).$$

$$\mathrm{E}_{Y|X}(Y|x) = \int_{y=-\infty}^{\infty} y \; f_{Y|X}(y|x) dy.$$

- Conditional expectation as a rv:

$$\mathrm{E}_Y(Y) = \mathrm{E}_X(\mathrm{E}_{Y|X}(Y|X))$$

Key formulas: Discrete-Time Markov chains

- State space S
- Initial probability vector π_0
- Memoryless property

$$P(X_{n+1} = j_{n+1} | X_n = j_n, \ldots, X_1 = j_1, X_0 = j_0) = P(X_{n+1} = j_{n+1} | X_n = j_n)$$

- Homogeneous DTMCs: transition probability matrix $R = [r_{ij}]$

$$r_{ij} = P(X_{n+1} = j | X_n = i)$$

- State probability at time n

$$P(X_n = j) = (\pi_0 R^n)_j$$

- Steady-state probability vector ($n \to \infty$)

$$\pi_\infty R = \pi_\infty$$

subject to $\pi_\infty \geq 0$ and $\sum_{i \in S} \pi_{\infty,i} = 1$.