

1 Probability Basics

Introduction

An (RL) agent repeatedly acts without knowing exactly what will happen next. An action may lead to several possible states, a reward may be noisy, and even the agent's own policy may randomize its choice. Probability provides a language for describing these possibilities. It also tells us how to average experience without confusing one observed outcome with the behavior of the underlying system. There are three objects to keep separate. An **outcome** is one possible result of an experiment. The **sample space** Ω contains all possible outcomes. An **event** is a subset of Ω . If a die is rolled, then $\Omega = \{1, 2, 3, 4, 5, 6\}$, while the event “the result is even” is $\{2, 4, 6\}$. A probability measure assigns a number $\mathbb{P}(A)$ to each event A so that

$$\mathbb{P}(A) \geq 0, \quad \mathbb{P}(\Omega) = 1,$$

and probabilities add over disjoint events. These axioms are simple, but almost every rule used below follows from them. When information becomes available, probabilities change. If $\mathbb{P}(B) > 0$, the **conditional probability** of A given B is

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.1)$$

Rearranging this definition gives the multiplication rule

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

Events A and B are **independent** when $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Conditional probability and independence answer different questions. Conditioning asks how information changes a probability. Independence says that learning one event does not change the probability of the other. Suppose $B_1, \dots, B_m$ are disjoint and together cover Ω . Splitting A according to which B_i occurred gives the **law of total probability**,

$$\mathbb{P}(A) = \sum_{i=1}^m \mathbb{P}(A | B_i)\mathbb{P}(B_i). \quad (1.2)$$

Combining this identity with the multiplication rule gives Bayes' formula:

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j)\mathbb{P}(B_j)}{\sum_{i=1}^m \mathbb{P}(A | B_i)\mathbb{P}(B_i)}. \quad (1.3)$$

Example 1.1;(An alarm)

A component has prior failure probability 0.02. An alarm has sensitivity 0.90 and false-alarm probability 0.05. If F denotes failure and A denotes an alarm, then

$$\mathbb{P}(F | A) = \frac{0.90(0.02)}{0.90(0.02) + 0.05(0.98)} \approx 0.269.$$

The posterior failure probability is about 27%. The alarm is informative, but the low prior failure rate still matters.

Random variables and distributions

An outcome is often too detailed for the question being asked. A **random variable** X converts each outcome into a number. Formally, X is a function from Ω to $\mathbb{R}$. The randomness lies in the outcome selected by the experiment; once the outcome is fixed, the value of X is fixed as well. A random variable is **discrete** when its possible values form a finite or countable set. The number of faulty servers, the reward in a finite game, and the number of arrivals during one minute are discrete. A random variable is **continuous** when its possible values range over an interval or another uncountable set. Waiting time and temperature are standard examples.

Probability mass functions

For a discrete random variable, the **probability mass function** is

$$p_X(x) = \mathbb{P}(X = x).$$

It must satisfy $p_X(x) \geq 0$ and $\sum_x p_X(x) = 1$. A histogram places a bar of height $p_X(x)$ at every possible value. The bar height represents probability mass, not merely how often the value appeared in one finite data set. Three discrete families occur often. A Bernoulli random variable records one success and has $\mathbb{P}(X = 1) = p$. A binomial random variable counts successes in n independent Bernoulli trials:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

A Poisson random variable is commonly used for counts in a fixed interval:

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

The choice of a family is a modelling decision. Its assumptions should be checked against the mechanism that produces the data.

Probability density functions

For a continuous random variable, probability is described by a **density** f_X . It satisfies

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x) dx = 1,$$

and probabilities are areas:

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x) dx. \quad (1.4)$$

The value $f_X(x)$ is not itself a probability. A density can exceed one, provided its total area is one. Moreover, $\mathbb{P}(X = x) = 0$ for every fixed x when X has a density. This does not mean that X takes no value; it means that individual points have zero width and hence zero area. The uniform, exponential, normal, gamma, triangular, and truncated normal distributions shown in the lecture slides illustrate how different shapes encode different kinds of uncertainty. A uniform density treats every equal-length subinterval alike. An exponential model is often used for positive waiting times. A normal density is symmetric around its mean. No shape is universally correct; the context determines which model is reasonable.

The cumulative distribution function

The **cumulative distribution function**, or CDF, works for every random variable:

$$F_X(x) = \mathbb{P}(X \leq x). \quad (1.5)$$

It is nondecreasing, approaches 0 as $x \rightarrow -\infty$, approaches 1 as $x \rightarrow \infty$, and is right-continuous. It also gives interval probabilities through

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

For a discrete variable, the CDF is a step function whose jump at x equals $p_X(x)$. For a continuous variable with density f_X , it is the accumulated area

$$F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

and $F'_X(x) = f_X(x)$ wherever the derivative exists.

Example 1.2;(The CDF of a fair die)

For a fair die, $F_X(x) = 0$ when $x < 1$, and

$$F_X(x) = \frac{k}{6} \quad \text{when } k \leq x < k+1, \quad k = 1, \dots, 5,$$

with $F_X(x) = 1$ for $x \geq 6$. The jump at each integer is $1/6$. In particular,

$$\mathbb{P}(2 < X \leq 5) = F_X(5) - F_X(2) = \frac{5}{6} - \frac{2}{6} = \frac{1}{2}.$$

Expectation, variance, and covariance

The distribution contains the full probabilistic description of a random variable. In practice, a few numerical summaries are often useful. The most important is the **expected value**,

$$\mathbb{E}[X] = \sum_x x p_X(x) \quad \text{or} \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (1.6)$$

according as X is discrete or continuous. Expectation is a probability-weighted average over hypothetical repetitions. It need not be a possible value of X . A fair die has mean 3.5, although no roll can equal 3.5. More generally, if g is a function, then

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x) \quad \text{or} \quad \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (1.7)$$

This rule computes the expectation of a transformed variable without first deriving its new distribution. Expectation is linear. Whenever the expectations exist,

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c. \quad (1.8)$$

Independence is not required. This simple fact is especially useful with indicator variables. If $I_A = \mathbf{1}_{\{A\}}$ equals one when A occurs and zero otherwise, then $\mathbb{E}[I_A] = \mathbb{P}(A)$. Consequently, the expected number

of events that occur equals the sum of their individual probabilities, even when the events are dependent. The mean describes location but not spread. The **variance** is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (1.9)$$

The standard deviation $\sqrt{\text{Var}(X)}$ has the same units as X . A variance of zero means that X is constant with probability one. A large variance indicates that values far from the mean receive appreciable probability, although variance alone does not describe the complete shape of the distribution.

Proof. Write $\mu = \mathbb{E}[X]$. Expanding the square and applying linearity gives

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

This proves the second expression in (1.9). □

For constants a, b ,

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \quad \text{Var}(aX + b) = a^2 \text{Var}(X).$$

Adding b shifts the distribution without changing its spread. Multiplying by a scales standard deviation by $|a|$ and variance by a^2 .

Example 1.3;(Mean and variance of a die)

For a fair die,

$$\mathbb{E}[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{7}{2}, \quad \mathbb{E}[X^2] = \frac{1^2 + 2^2 + \dots + 6^2}{6} = \frac{91}{6}.$$

Therefore

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}, \quad \text{sd}(X) = \sqrt{\frac{35}{12}} \approx 1.708.$$

Two random variables may move together. Their **covariance** is

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

The correlation is $\text{Corr}(X, Y) = \text{Cov}(X, Y)/(\sigma_X \sigma_Y)$ when both standard deviations are positive. Independence implies zero covariance when the required moments exist, but zero covariance does not usually imply independence. The variance of a sum is

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y). \quad (1.10)$$

Thus variances add for independent random variables, or more generally for uncorrelated ones.

Remark 1.1;(Averages and nonlinear functions)

In general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$. For a convex function g , using Jensen's inequality gives $g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$. This distinction matters whenever an objective contains a maximum, a square, a logarithm, or another nonlinear transformation.

Joint distributions and conditional expectation

Several random quantities are often observed together. For discrete X and Y , the **joint mass function** is

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y).$$

Summing out one variable gives a marginal distribution:

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad p_Y(y) = \sum_x p_{X,Y}(x, y).$$

For continuous variables, a joint density $f_{X,Y}$ satisfies

$$\mathbb{P}((X, Y) \in C) = \iint_C f_{X,Y}(x, y) dx dy,$$

and the marginal density is obtained by integration, for example

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

The variables are independent exactly when their joint distribution factors into the product of their marginals. For discrete variables with $\mathbb{P}(Y = y) > 0$,

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad \mathbb{E}[X | Y = y] = \sum_x x p_{X|Y}(x | y).$$

For continuous variables with $f_Y(y) > 0$, the corresponding formulas are

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad \mathbb{E}[X | Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx.$$

As y varies, the quantity $\mathbb{E}[X | Y = y]$ varies. Therefore $\mathbb{E}[X | Y]$ is itself a random variable: after Y is observed, it gives the appropriate conditional mean.

Theorem 1.1;(The law of iterated expectation)

Whenever $\mathbb{E}[|X|] < \infty$,

$$\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]. \quad (1.11)$$

In words, averaging the conditional averages recovers the overall average.

Proof. For discrete variables,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_y \mathbb{E}[X | Y = y] p_Y(y) \\ &= \sum_y \sum_x x p_{X|Y}(x | y) p_Y(y) \\ &= \sum_x x \sum_y p_{X,Y}(x, y) = \sum_x x p_X(x) = \mathbb{E}[X]. \end{aligned}$$

The continuous proof replaces the sums by integrals. □

Conditional expectation also separates two sources of variation:

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]). \quad (1.12)$$

The first term is the average uncertainty remaining after Y is known. The second measures how much the conditional mean changes across values of Y .

Example 1.4;(Two actions with the same mean reward)

Suppose an agent chooses a safe action with probability 0.6 and a risky action with probability 0.4. The safe action gives reward 1. The risky action gives 3 or -1 , each with conditional probability $1/2$. Both actions have conditional mean reward 1, so the tower property gives $\mathbb{E}[R] = 1$. They do not have the same risk. The safe action has conditional variance 0, while the risky action has conditional variance 4. Hence

$$\text{Var}(R) = 0.6(0) + 0.4(4) + \text{Var}(1) = 1.6.$$

An expected reward reports the centre of a distribution; it does not report the possible consequences around that centre.

Averages of repeated observations

Expectations become useful in data analysis because sample averages often approach them. Let $X_1, X_2, \dots$ be independent and identically distributed random variables with mean μ , and define

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

Theorem 1.2;(Weak law of large numbers)

If $\text{Var}(X_i) = \sigma^2 < \infty$, then for every $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \longrightarrow 0. \quad (1.13)$$

Thus the sample average converges in probability to the population mean.

Proof. Independence gives

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

Chebyshev's inequality now yields

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0.$$

□

The law of large numbers is a statement about accuracy, not about the shape of the sampling distribution. It says that large averages concentrate near μ . The **central limit theorem** describes the scale and approximate shape of the remaining error.

Theorem 1.3;(Central limit theorem)

If the X_i are i.i.d. with mean μ and variance $0 < \sigma^2 < \infty$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1). \quad (1.14)$$

Equivalently, for large n ,

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right), \quad \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2).$$

Unlike the short proof of the weak law, a full proof of the central limit theorem ris beyond the scope of this chapter. The typical error in a sample average is of order $1/\sqrt{n}$, not $1/n$. Reducing the standard error by a factor of two therefore requires about four times as many independent observations. For repeated fair-die rolls, $\mu = 7/2$ and $\sigma^2 = 35/12$. Hence

$$\bar{X}_n \approx N\left(\frac{7}{2}, \frac{35}{12n}\right).$$

The number “30 or 40 observations” is only a rule of thumb. A highly skewed or heavy-tailed distribution may require a much larger sample, and dependence between observations can invalidate the i.i.d. calculation. If σ is estimated by the sample standard deviation s , the CLT motivates the approximate large-sample interval

$$\bar{X}_n \pm 1.96 \frac{s}{\sqrt{n}}.$$

The quantity $s/\sqrt{n}$ is the **standard error**. It measures uncertainty in the estimated mean, whereas s measures variability among individual observations.