# Lesson - Probability and random variables

Year: University Year 1

Probability gives a precise language for uncertainty. This chapter builds from the axioms to conditional probability, random variables and the binomial, Poisson and normal distributions.

**Starter activity**

Roll two dice twenty times and record whether you get a double. Before you start, predict how many doubles you expect. Compare your result with the rest of the group: why are the results different, and why do the averages get closer to the prediction as more results are pooled? Then ask: if the first die shows a six, does that change the chance of a double?

## Probability foundations

**Definition: probability space**

A **probability space** consists of a sample space $\Omega$, a collection of events (subsets of $\Omega$) and a probability $P$ satisfying the Kolmogorov axioms: $P(A) \geq 0$, $P(\Omega) = 1$, and $P(A \cup B) = P(A) + P(B)$ for disjoint events (extended to countable unions).

**Rule: consequences of the axioms**

$P(A^c) = 1 - P(A)$, and the **addition rule** is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

**Example 1**

**If $P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cap B) = 0.2$, find $P(A \cup B)$.**

$0.5 + 0.4 - 0.2 = 0.7$.

## Conditional probability, independence and Bayes' theorem

**Definition: conditional probability and independence**

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ for $P(B) > 0$. Events are **independent** if $P(A \cap B) = P(A)P(B)$, equivalently $P(A|B) = P(A)$.

**Theorem: total probability and Bayes' theorem**

For a partition $\{A_i\}$: $P(B) = \sum P(B|A_i)P(A_i)$ and $P(A_i|B) = \dfrac{P(B|A_i)P(A_i)}{P(B)}$.

**Method: Bayes problems**

1. Draw a tree diagram with the prior probabilities on the first branches.
2. Add the conditional probabilities on the second branches.
3. Find $P(B)$ by adding the branches that end in $B$.
4. Divide the branch you want by $P(B)$.

**Example 2**

**A disease affects 1% of people. A test detects it 95% of the time but gives a false positive for 5% of healthy people. Find $P(D|+)$.**

$P(D|+) = \frac{0.95\times0.01}{0.95\times0.01+0.05\times0.99} = \frac{0.0095}{0.0590} \approx 0.161$. A positive test means only a 16% chance of disease, because the disease is rare.

**Example 3**

**With the numbers of Example 1, check independence and find $P(A|B)$.**

$P(A|B) = \frac{0.2}{0.4} = 0.5 = P(A)$, and $0.5 \times 0.4 = 0.2$, so $A$ and $B$ are independent. For counting, $\binom{10}{3} = 120$ committees of 3 can be chosen from 10 people.

## Random variables, expectation and variance

**Definition: expectation and variance**

A **random variable** $X$ assigns a number to each outcome. For a discrete $X$ with pmf $p(x) = P(X = x)$: $E[X] = \sum xp(x)$ and $\operatorname{Var}(X) = E[X^2] - (E[X])^2$. For a continuous $X$ with density $f$: $P(a, \int f = 1$, $E[X] = \int x f(x)\,dx$.

**Rule: linearity**

$E[aX + b] = aE[X] + b$ and $\operatorname{Var}(aX + b) = a^2 \operatorname{Var} X$. Expectation is always additive; variance is additive when the variables are independent.

**Example 4**

**For $P(0) = 0.2$, $P(1) = 0.5$, $P(2) = 0.3$, find $E[X]$, $\operatorname{Var} X$, $E[3X + 2]$ and $\operatorname{Var}(3X + 2)$.**

$E[X] = 1.1$, $E[X^2] = 1.7$, $\operatorname{Var} X = 1.7 - 1.21 = 0.49$. Then $E[3X + 2] = 5.3$ and $\operatorname{Var}(3X + 2) = 9 \times 0.49 = 4.41$.

**Example 5**

**A density is $f(x) = kx$ on $[0, 2]$. Find $k$, $E[X]$ and $P(X$.**

$\int_0^2 kx\,dx = 2k = 1$, so $k = \frac{1}{2}$. $E[X] = \int_0^2 \frac{x^2}{2}dx = \frac{4}{3}$ and $P(X$.

## The binomial distribution

**Rule: binomial distribution**

If $X$ counts successes in $n$ independent trials with success probability $p$, then $X \sim B(n, p)$ with $P(X = r) = \binom{n}{r}p^r(1-p)^{n-r}$, $E[X] = np$ and $\operatorname{Var} X = np(1-p)$.

**Example 6**

**For $X \sim B(10, 0.3)$ find $P(X = 3)$. What is the chance of at least one success in 5 trials with $p = 0.2$?**

$P(X = 3) = 120(0.3)^3(0.7)^7 \approx 0.2668$. And $1 - 0.8^5 = 0.6723$.

## The Poisson distribution

**Rule: Poisson distribution**

For rare events at a constant average rate $\lambda$ per interval, $X \sim \mathrm{Po}(\lambda)$ with $P(X = r) = \dfrac{e^{-\lambda}\lambda^r}{r!}$ and $E[X] = \operatorname{Var} X = \lambda$. It approximates $B(n, p)$ when $n$ is large and $p$ small, with $\lambda = np$.

**Example 7**

**For $\lambda = 3$ find $P(X = 2)$ and $P(X = 0)$.**

$\frac{9e^{-3}}{2} \approx 0.2240$ and $e^{-3} \approx 0.0498$. For $n = 200$, $p = 0.01$, $\lambda = 2$, $P(X = 1) \approx 2e^{-2} = 0.2707$, matching the binomial value.

## The normal distribution

**Rule: standardising**

$X \sim N(\mu, \sigma^2)$ has a bell-shaped density symmetric about $\mu$. Standardise $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ and use $\Phi(z) = P(Z \le z)$. Key values: $\Phi(1) = 0.8413$, $\Phi(2) = 0.9772$, $\Phi^{-1}(0.95) = 1.645$. Sums of independent normals are normal, with means and variances added.

### Example 8

**For $X \sim N(100, 15^2)$ find $P(X > 115)$, $P(85$ and the 95th percentile.**

$1 - \Phi(1) = 0.1587$; $\Phi(2) - \Phi(-1) = 0.8186$; $100 + 1.645 \times 15 \approx 124.7$. If $X \sim N(5, 4)$ and $Y \sim N(3, 9)$ are independent then $X + Y \sim N(8, 13)$.

### Watch out

- Confusing independent with mutually exclusive events.
- Forgetting that variance involves the square of constants.
- Adding standard deviations instead of variances.
- Neglecting the base rate in Bayes problems.

## Using the distributions

### Key fact: approximating one distribution by another

Exact binomial calculations become heavy when $n$ is large. If $n$ is large and $p$ small (roughly $n \geq 50$ and $np$), $B(n, p)$ is close to $\mathrm{Po}(np)$, because rare events in many trials behave like random arrivals. If $np$ and $n(1-p)$ are both larger than about 5, $B(n, p)$ is close to $N(np,\, np(1-p))$, and a Poisson distribution with large $\lambda$ is close to $N(\lambda, \lambda)$. When a discrete variable is approximated by a continuous one, use a continuity correction: $P(X \leq k)$ becomes $P(Y$ and $P(X \geq k)$ becomes $P(Y > k - 0.5)$.

### Method: normal probabilities and percentiles

1. Write the distribution clearly, for example $X \sim N(\mu, \sigma^2)$, and note whether the second parameter is the variance or the standard deviation.
2. Standardise: $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.
3. Sketch the bell curve and shade the region you need; use symmetry $\Phi(-z) = 1 - \Phi(z)$.
4. For a percentile, find $z$ with $\Phi(z) =$ the required proportion, then return to $x = \mu + z\sigma$.
5. Check that the answer is sensible: probabilities between 0 and 1, percentiles on the correct side of the mean.

### Worked example 9: Bayes with a table of natural frequencies

**5% of emails are spam. A filter flags 98% of spam and 1% of genuine emails. What proportion of flagged emails are spam?**

Think of 10 000 emails: 500 are spam, of which 490 are flagged; 9 500 are genuine, of which 95 are flagged. Altogether 585 emails are flagged, and 490 of them are spam.

So $P(\text{spam} \mid \text{flagged}) = \dfrac{490}{585} \approx 0.838$. The same answer comes from Bayes' theorem: $\dfrac{0.05 \times 0.98}{0.05 \times 0.98 + 0.95 \times 0.01} = \dfrac{0.049}{0.0585}$. Natural frequencies make it easy to see that the false alarms come from the large genuine group.

### Worked example 10: Variance of a sum

**$X$ and $Y$ are independent with $E[X] = 3$, $\operatorname{Var}(X) = 4$, $E[Y] = 5$, $\operatorname{Var}(Y) = 9$. Find the mean and variance of $2X - Y + 1$.**

By linearity, $E[2X - Y + 1] = 2 \times 3 - 5 + 1 = 2$. For the variance the constant has no effect, and for independent variables $\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y)$. So $\operatorname{Var}(2X - Y + 1) = 4 \times 4 + 1 \times 9 = 25$ and the standard deviation is 5. Note that the minus sign disappears when squared: subtracting $Y$ adds its variance.

### Worked example 11: A Poisson model in context

**Typing errors occur at random at an average rate of 1.5 per page. Find the probability that a 4-page essay has at most 3 errors.**

For 4 pages the number of errors is $\operatorname{Po}(6)$. Then $P(X \le 3) = e^{-6}\left(1 + 6 + \dfrac{36}{2} + \dfrac{216}{6}\right) = e^{-6}(1 + 6 + 18 + 36) = 61e^{-6} \approx 0.151$. The key step is to scale the rate to the interval in the question: the Poisson parameter is always the mean number of events in that interval.

### Worked example 12: A normal approximation with continuity correction

**A fair coin is tossed 100 times. Estimate the probability of at least 60 heads.**

$X \sim B(100, 0.5)$ has mean 50 and variance 25, so $X \approx N(50, 25)$. With the continuity correction, $P(X \ge 60) \approx P(Y > 59.5) = P\left(Z > \dfrac{9.5}{5}\right) = P(Z > 1.9) = 1 - 0.9713 = 0.0287$. The exact binomial value is $0.0284$, so the approximation is excellent. Without the correction we would get $P(Z > 2) = 0.0228$, noticeably worse.

### Worked example 13: A continuous random variable

**A random variable has density $f(x) = \frac{3}{8}x^2$ on $[0, 2]$. Find its cumulative distribution function and median.**

For $0 \le x \le 2$, $F(x) = \int_0^x \frac{3}{8}t^2\,dt = \dfrac{x^3}{8}$; also $F(x) = 0$ for $x$ and $F(x) = 1$ for $x > 2$. The median $m$ satisfies $F(m) = \frac{1}{2}$, so $m^3 = 4$ and $m = \sqrt[3]{4} \approx 1.587$. The mean is $\int_0^2 \frac{3}{8}x^3\,dx = \frac{3}{2}$, a little less than the median, because the density is skewed towards

the right end of the interval.

### Worked example 14: Expected value as a long-run average

**A raffle sells 500 tickets at £1. There is one prize of £100 and four prizes of £20. Find the expected gain of a ticket holder.**

Let $G$ be the gain. A ticket wins £100 with probability $\frac{1}{500}$ and £20 with probability $\frac{4}{500}$; otherwise it wins nothing. The expected prize is $\dfrac{100 + 4 \times 20}{500} = \dfrac{180}{500} = £0.36$, so the expected gain is $0.36 - 1 = -£0.64$ per ticket. The organisers keep, on average, 64p from every ticket sold, that is $500 \times 0.64 = £320$ for the whole raffle. An expected value is not a value you will actually see on one ticket: it is the average gain per ticket over many tickets, which is why it is the right tool for judging whether a game is fair.

### Note: modelling assumptions

Every distribution in this chapter comes with assumptions. The binomial model needs a fixed number of independent trials with the same probability of success; the Poisson model needs events that occur singly, independently and at a constant average rate; the normal model is justified for measurements that are the sum of many small independent effects. In applications, always state the assumptions in context and ask whether they are reasonable: calls to a helpline may come in bursts after an advert, and the heights of a mixed group of adults may be better described by two overlapping normal curves. A good habit is to compare the model with data: if the sample mean and variance of a count are very different, a Poisson model is doubtful; if a histogram of measurements is strongly skewed, a normal model is doubtful. Models are simplifications, and their conclusions are only as reliable as their assumptions.

### Watch out: typical mistakes

- Confusing $P(A \mid B)$ with $P(B \mid A)$, the classic error in medical and legal reasoning.
- Assuming that mutually exclusive events are independent: they are not, unless one has probability 0.
- Writing $N(\mu, \sigma)$ when the notation means $N(\mu, \sigma^2)$.
- Forgetting that $\operatorname{Var}(X - Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$ for independent variables.

### Did you know?

Probability began with letters between Pascal and Fermat in 1654 about a gambling problem. Thomas Bayes' famous theorem was published after his death in 1763, and the Poisson distribution appeared in 1837 in a study of court verdicts. Its most famous early use was by Ladislaus Bortkiewicz in 1898, who showed that the number of Prussian soldiers killed by horse kicks each year followed a Poisson distribution very

closely.

**Otto's tip**

Draw a tree diagram for conditional probability questions.