Answers to Questions

Probability and Statistics Refresher

1. Random variables

A **random variable** assigns a real number to each outcome of a random experiment.

$$X : \Omega \longrightarrow \mathbb{R}, \qquad \omega \longmapsto X(\omega).$$

- Ω is the sample space of possible outcomes.
- On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X must be measurable:

  $$\{\omega : X(\omega) \le x\} \in \mathcal{F} \qquad \text{for every } x \in \mathbb{R}.$$

  This ensures that probabilities such as $\mathbb{P}(X \le x)$ are defined.

Example: toss a coin and set $X(H) = 1$, $X(T) = 0$.

2. Probability distributions: discrete variables

The **distribution** of X specifies the probabilities of its possible values. For a discrete variable, its **probability mass function (PMF)** is

$$p(x) = \mathbb{P}(X = x), \qquad p(x) \geq 0, \qquad \sum_x p(x) = 1.$$

Example: for the score X on a fair six-sided die,

$$p(x) = \frac{1}{6} \quad (x = 1, \ldots, 6), \qquad p(x) = 0 \quad \text{otherwise.}$$

Hence

$$\mathbb{P}(X \leq 2) = p(1) + p(2) = \frac{1}{3}.$$

2. Probability distributions: densities

For a variable with a **probability density function (PDF)** f,

$$f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1, \qquad \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

- Probability is the **area under the density** over an interval.
- A single point has probability $\mathbb{P}(X = x) = 0$.

Example: if $X \sim \mathrm{Uniform}(0, 1)$, then

$$\mathbb{P}(0.2 \leq X \leq 0.5) = \int_{0.2}^{0.5} 1\,dx = 0.3.$$

A general distribution can be described by its cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x)$, whether or not it has a density.

3. Confidence intervals

A **confidence interval** uses sample data D to estimate an unknown parameter θ:

$$[L(D), U(D)].$$

An exact $100(1-\alpha)\%$ confidence procedure satisfies

$$\mathbb{P}_\theta(L(D) \le \theta \le U(D)) = 1 - \alpha \qquad \text{for every } \theta.$$

- The parameter θ is fixed. The endpoints vary across samples.
- A 95% procedure covers θ in 95% of repeated samples in the long run.
- Conservative procedures have coverage at least $1 - \alpha$.

3. Confidence intervals: a numerical example

Suppose $X_1, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ observations, with known standard deviation σ.

A 95% confidence interval for μ is approximately

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

Example: $\bar{x} = 10$, $\sigma = 2$, $n = 100$ gives

$$10 \pm 1.96\frac{2}{10} = 10 \pm 0.392, \qquad [9.608,\ 10.392].$$

The 95% refers to the coverage of the **procedure** across repeated samples. Once we observe the data, this particular interval either contains μ or does not.

4. Expectation

The **expectation** is a probability-weighted average of possible values.

For a variable with density f,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

For a discrete variable with PMF p,

$$\mathbb{E}[X] = \sum_{x} x p(x).$$

Example: for a fair die,

$$\mathbb{E}[X] = \frac{1+2+3+4+5+6}{6} = 3.5.$$

These formulas give a finite mean when $\mathbb{E}[|X|] < \infty$. The mean need not be a possible observed value.

5. Variance

The **variance** measures the average squared distance from the mean $\mu = \mathbb{E}[X]$:

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

For a variable with density f,

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

- Variance is nonnegative and has squared units.
- The standard deviation $\sigma = \sqrt{\mathrm{Var}(X)}$ has the same units as X.

Example: if X equals 0 or 2 with equal probabilities, then $\mu = 1$ and $\mathrm{Var}(X) = \frac{1}{2}(0-1)^2 + \frac{1}{2}(2-1)^2 = 1$.

6. Normal distribution with mean 1 and variance 4

For $X \sim N(\mu, \sigma^2)$, the density is

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

Substituting $\mu = 1$ and $\sigma^2 = 4$ gives

$$\boxed{f(x) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{(x-1)^2}{8}\right).}$$

- The mean is $\mathbb{E}[X] = 1$.
- The variance is $\mathrm{Var}(X) = 4$, so the standard deviation is $\sigma = 2$.
- Standardising gives $Z = (X-1)/2 \sim N(0,1)$.

7. Independent events

Events A and B are **independent** if and only if

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

When $\mathbb{P}(B) > 0$, this is equivalent to

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

Knowing that B occurred leaves the probability of A unchanged.

Example: toss two independent fair coins. Let A be “the first coin is heads” and B be “the second coin is heads”.

$$\mathbb{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}.$$

8. Pearson correlation coefficient

For variables with finite, strictly positive variances,

$$\boxed{\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y},} \qquad \sigma_X = \sqrt{\mathrm{Var}(X)}, \quad \sigma_Y = \sqrt{\mathrm{Var}(Y)}.$$

- $-1 \le \rho_{X,Y} \le 1$. Correlation has no units.
- A positive value indicates a positive linear association.
- A negative value indicates a negative linear association.
- A zero value indicates no linear association, but dependence may remain.

Example: if $Y = 2X + 3$ and $0 < \mathrm{Var}(X) < \infty$, then $\rho_{X,Y} = 1$.

9. Covariance

The **covariance** measures how two variables vary together:

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

Equivalently, when the second moments are finite,

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

- Positive covariance: deviations from the means tend to have the same sign.
- Negative covariance: deviations tend to have opposite signs.
- Independence implies zero covariance when the second moments are finite.

Example: if $Y = 2X + 3$, then $\mathrm{Cov}(X, Y) = 2\,\mathrm{Var}(X)$.

10. Bayes' formula

For events A and B with $\mathbb{P}(B) > 0$,

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.}$$

Here the displayed conditional $\mathbb{P}(B \mid A)$ also assumes $\mathbb{P}(A) > 0$.

Example: choose a fair coin or a double-headed coin, each with probability $1/2$. After one toss gives heads, what is the probability of having chosen the double-headed coin?

$$\mathbb{P}(H) = 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}, \qquad \mathbb{P}(D \mid H) = \frac{1 \cdot \frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}.$$

The observed head increases the probability of D from $1/2$ to $2/3$.