

## Maths Boot Camp — Session 4

### Probability, Bayes' Rule, and Random Variables

Formula & Definition Cheat Sheet — keep this to hand during class

#### 1. Counting

| Concept                                                               | Formula                                        |
|-----------------------------------------------------------------------|------------------------------------------------|
| Fund. Theorem of Counting                                             | $m \times n$                                   |
| Factorial                                                             | $n! = n(n-1) \cdots 1$                         |
| Permutations ( $n$ obj., $r$ at a time; <i>order matters</i> )        | $P(n, r) = \frac{n!}{(n-r)!}$                  |
| Combinations ( $n$ obj., $r$ at a time; <i>order doesn't matter</i> ) | $C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$ |

*Rule of thumb:* arranging in sequence  $\rightarrow$  permutations; picking a group  $\rightarrow$  combinations. E.g. 5 messages, pick 3 in order:  $P(5, 3) = 60$ ; pick 3, order irrelevant:  $C(5, 3) = 10$ .

#### 2. Probability Basics

$0 \leq P(A) \leq 1$ ;  $P(S) = 1$  (sample space has total probability 1).

**Axioms (Kolmogorov):**

- **Non-negativity:**  $P(A) \geq 0$
- **Normalisation:**  $P(S) = 1$
- **Countable additivity:** mutually exclusive  $E_1, E_2, \dots$ :  
 $P(\bigcup E_i) = \sum P(E_i)$

**Joint & conditional:**

- **Joint:**  $P(A \cap B)$  — both happen
- **Independent:**  $P(A \cap B) = P(A)P(B)$ ; mutually exclusive:  $P(A \cap B) = 0$
- **General union:**  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

**Conditional probability** —  $A$  given  $B$  has occurred:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

If independent:  $P(A | B) = P(A)$ .

**Law of Total Probability** — partition  $\{B_1, \dots, B_n\}$  of  $S$  (mutually exclusive, exhaustive):

$$P(A) = \sum_{i=1}^n P(A | B_i) P(B_i)$$

**Bayes' Theorem** — updates a belief given new evidence:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

**Prior**  $P(A)$  · **Likelihood**  $P(B|A)$  · **Marginal**  $P(B)$   $\rightarrow$  **Posterior**  $P(A|B)$ .

*Derivation:* both  $P(A|B)P(B)$  and  $P(B|A)P(A)$  equal  $P(A \cap B)$  — set equal, solve for  $P(A|B)$ .

*Worked (medical test):*  $P(D)=0.01$ ,  $P(+|D)=0.99$ ,  $P(+|no D)=0.05$ .  $P(+)=0.0594 \Rightarrow P(D|+) \approx 0.167$  — rare disease  $\Rightarrow$  most positives are false alarms.

**Boole's Inequality** — upper bound on a union (events may overlap):

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

Tight when events are near mutually exclusive; loose when they overlap a lot.

### 3. Random Variables & Distributions

**Levels of measurement:** **Nominal** (no order, e.g. party) → **Ordinal** (ordered, no fixed scale, e.g. Likert) → **Interval** (consistent differences, arbitrary zero, e.g. °C) → **Ratio** (true zero, e.g. crime rate).

**Random variable:** numeric outcome of a random process. **Discrete:** listable values (e.g. seats won). **Continuous:** any value in a range (e.g. vote share).

**PMF** (discrete):  $P(X=x) = p(x)$ , with  $0 \leq p(x) \leq 1$  and  $\sum p(x) = 1$ .

**PDF** (continuous):  $f(x) \geq 0$ ,  $\int_{-\infty}^{\infty} f(x) dx = 1$ .  $P(X=x) = 0$  for any single point — integrate over an interval to get a probability.

**Expectation (mean):** weighted average of all possible values.

$$\mathbb{E}[Y] = \sum_y y p(y) \text{ (discrete),} \quad \mathbb{E}[Y] = \int_{-\infty}^{\infty} y f(y) dy \text{ (cts.)}$$

E.g. fair die:  $\mathbb{E}[Y] = \frac{1}{6}(1 + \dots + 6) = 3.5$ .

**Variance:**  $\text{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2]$  — average squared deviation from the mean.

**Standard deviation:**  $\sigma_Y = \sqrt{\text{Var}(Y)}$  — same units as  $Y$ .

E.g. scores 50, 60, 70 (mean 60):  $\text{Var} = \frac{1}{3}(100+0+100) = 66.67$ ,  $\sigma \approx 8.16$ .

**Covariance & correlation:**

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)], \quad \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$\text{Cov} > 0$ : move together;  $< 0$ : opposite directions.  $\rho \in [-1, 1]$  — closer to  $\pm 1$  means a stronger linear relationship.

**Normal distribution:** symmetric, bell-shaped; mean = median = mode.

**68-95-99.7 Rule:**  $\approx 68\%$  of data within  $1\sigma$  of the mean,  $\approx 95\%$  within  $2\sigma$ ,  $\approx 99.7\%$  within  $3\sigma$ .

Standard normal ( $\mu=0, \sigma^2=1$ ):  $f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$ .

**Binomial shortcuts** (for  $X \sim \text{Binomial}(n, p)$ ):  $\mathbb{E}[X] = np$ ,  $\text{Var}(X) = np(1-p)$ .

---

*Dr. Marta Antonetti — Maths Boot Camp — Department of Government, LSE*