Course outline

- Course consists of 12 half-days
- Each half-day is a mix of lectures and practicals.
- **To pass**: Active participation in 10 of 12 half-days.

- **Today**: Probability brush-up
- **Tomorrow**: Introduction to R software

Probability brush-up

Setup: Perform an experiment

State space Ω — the set of all possible outcomes of the "experiment"

Example:

- Trip to the casino

Event: $A \subseteq \Omega$ — subset of the state space.

Examples:

- At least three wins
- $\text{Temperature inside the casino at noon} \in [25, 26]$

Probability

Notation: Probability of A is denoted $P(A)$.

Properties

- $0 \leq P(a) \leq A$
- $P(A) =$ 'Area of A'
- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $P(A \cup B = P(A) + P(B) - P(A \cap B)$

Complement: A^C is A's complement, i.e.

- $A \cap A^C = \emptyset$
- $A \cup A^C = \Omega$

$$\begin{aligned} P(A \cup A^C) &= P(A) + P(A^C) \\ 1 &= P(A) + P(A^C) \\ P(A^C) &= 1 - P(A) \end{aligned}$$

Law of total probabilty

Split Ω in to disjoint sets

$$B_1, B_2, \ldots, B_n$$

That is $B_i \cap B_j = \emptyset$ for $i \neq j$, and $\cup_{i=1}^{n} = \Omega$.
Consider event A

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_n \cap A)$$

Notice: $(B_i \cap A) \cap (B_:j \cap A) = \emptyset$ for $i \neq j$

$$P(A) = P(B_1 \cap A) + P(B_2 \cap A) + \cdots + P(B_n \cap A)$$

Conditional probability

$A, B \subseteq \Omega$ events

The conditional probability of A given is

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

Can be rewritten as

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Bayes' Theorem

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

Recall

$$\begin{aligned} P(B) &= P(A \cap B + P(A^C \cap B) \\ &= P(A)P(B|A) + P(A^C)P(B|A^C) \end{aligned}$$

Hence

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

Notice that we have "reversed" the conditioning.

Example: Test for rare disease

Events: I=infected I^C=uninfected

Z=postive test Z^C=negative test

Known:

- $P(I) = 0.001$
- $P(Z|I) = 0.92$
- $P(Z|I^C) = 0.04$ (false positive)

Question:

- Given positive test, what is the probability of having the disease? I.e. what is $P(I|Z)$?

Independence

Events A and B are independent if and only if

$$P(A \cap B) = P(A)P(B)$$

Consequences

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

Random variable (RV)

Definition: A **random variable** is a fxunction from state space to the real numbers.

Definition: A **discrete RV** takes countably many values.

Definition: **Probaility function** $\pi(x)$

- $\pi(x) = P(X = x) \geq 0$
- $\sum_x \pi(x) = 1$

Definition: **Distribution function** $F(x)$

$$F(x) = P(X \leq x) = \sum_{y \leq x} \pi(x)$$

Example: Binomial distribution

A random varibale, X, follows a binomial distribution with parameters p and n ($0 \leq p \leq 1$ and $n \in \{1, 2, 3, \ldots\}$), if

$$\pi(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \ldots, n\}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad k! = 1 \cdot 2 \cdot 3 \cdots k.$$

Notation: $X \sim B(n, p)$.

Example: Binomial distribution

A random varibale, X, follows a binomial distribution with parameters p and n ($0 \leq p \leq 1$ and $n \in \{1, 2, 3, \ldots\}$), if

$$\pi(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \ldots, n\}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad k! = 1 \cdot 2 \cdot 3 \cdots k.$$

Notation: $X \sim B(n, p)$.

Interpretation

- Perform n independent experiments, each with outcomes “success” or “failure”
- $P(\text{“success”}) = p$ for all experiments.
- $X =$ number of successes.
- $X \sim B(n, p)$.

Expectation and variance of RV

Definition: **Expectation of a discrete RV** $E(x)$

- $\mu = E[X] = \sum_x x\pi(x)$
- $E[h(X)] = \sum_x h(x)\pi(x)$
- $E[a + bX] = a + bE[X]$

Definition: **Variance of a discrete RV** $Var(X)$

$$\begin{aligned}\sigma^2 = Var[X] &= E[(X - \mu)^2] \\ &= \sum_x (x - \mu)^2 \pi(x) \qquad = E[X^2] = -(E[X])^2\end{aligned}$$

$Var(a + bX) = b^2 Var(X)$

Example: Assume $X \sim B(n, p)$:

- $E[X] = np$
- $Var(X) = np(1 - p)$

Continuous Random Variables

A continuous RV is specified by a **probability density function** (pdf) $\pi(x)$:

- $\pi(x) \geq 0$
- $\int_{-\infty}^{\infty} \pi(x)dx = 1$
- $P(a \leq X \leq b) = \int_a^b \pi(x)dx$

Distribution function: $F(x) = P(X \leq x) = \int_{-\infty}^{x} \pi(y)dy$

Expected value of continuous RV

- $E[X] = \int_{-\infty}^{\infty} x\pi(x)dx$
- $E[h(X)] = \int_{-\infty}^{\infty} h(x)\pi(x)dx$

Important special case

$$E[1[a \leq X \leq B]] = \int_{-\infty}^{\infty} 1[a \leq x \leq B]\pi(x)dx$$
$$= \int_a^b \pi(x)dx = P(a \leq X \leq b)$$

A probability can be expressed as an expectation.

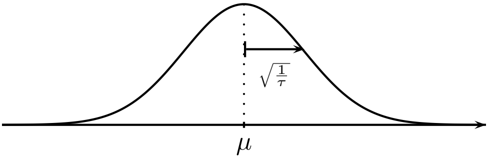
Example: Normal distribution

A random variable X follows a normal distribution with mean μ and precision τ if it has density

$$\pi(x) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{1}{2}\tau(x-\mu)^2\right).$$

Notation: $X \sim \mathcal{N}(\mu, \tau)$.

Note: $\tau = \frac{1}{\mathrm{Var}(X)}$



Independence of continuous RVs

Lat X and Y be continuous RV with joint pdf

$$\pi(x, y)$$

so that $P(X, Y) < inA) = \int_A \pi(x, y) dx dy$.
Then X and Y are independent if and only if

$$\pi(x, y) = \pi(x)\pi(y)m$$

where $\pi(x)$ and $\pi(y)$ are the marginal pdfs, i.e.

$$\pi(x) = \int_{-\infty}^{\infty} \pi(x, y) dy$$

The **conditional pdf** is

$$\pi(y|x) = \frac{\pi(x, y)}{\pi(x)}$$

Example: Independent normals

Assume $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \tau)$ ($iid =$ independent and identially distributed). Then the joint pdf. is

$$
\begin{aligned}
\pi(x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{1}{2}\tau(x_i - \mu)^2\right) \\
&= \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}\tau \sum_{i=1}^{n}(x_i - \mu)^2\right)
\end{aligned}
$$