Bayesian statistics, simulation and software

Module 1: Course intro and probability brush-up

Jesper Møller and Ege Rubak

Department of Mathematical Sciences
Aalborg University

Course outline

- Course consists of 12 half-days – modules of only 3 hours and 15 minutes each – of lectures and practicals. Expect you work hard on your own – otherwise it may be hard to pass! Solutions to (perhaps all) exercises are available, but use them modestly.
- **To pass**: Active participation in at least 10 of 12 modules plus a satisfactory solution of the exercise considered at the last module (where you will be informed about the details to whom and when the solution should be send).

Today

- **1. module**: Probability brush-up.
- **2. module**: Introduction to R software.

Probability brush-up

Setup: Perform an "experiment".

State space $\Omega =$ the set of all possible outcomes of the experiment.

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

Example: Trip to the casino – what is the relevant state space?

Depends on the types of events...

Examples of events:

- At least three wins on "even"out of five trials: $\Omega =??$ (Yes, $\Omega = \{$even, not even$\}^5$.)
- $\text{Temperature inside the casino at noon} \in [25, 26]$. (Maybe $\Omega = [18, 30]$ (degrees in C).)

Probability

Notation: Probability of an event A is denoted $P(A)$. **Basic properties:**

- $0 \leq P(A) \leq 1$.
- $P(\Omega) = 1$.
- If $A_1, A_2, \ldots$ are pairwise disjoint events ($A_i \cap A_j = \emptyset$ for $i \neq j$), then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Consequences:

- A^C denotes A's complement, i.e. $A \cap A^C = \emptyset$ and $\Omega = A \cup A^C$. So $P(A) + P(A^C) = P(A \cup A^C) = 1$ and hence

$$P(A^C) = 1 - P(A).$$

- For any events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Example: A fair coin is tossed 10 times. What is the probability of any outcome?

Answer: 2^{-10} since all 2^{10} possible outcomes are equally likely.

What is the probability of at least one head?

Answer: $1 - P(\text{all tail}) = 1 - 2^{-10}$.

What is the probability of at least one head and at least one tail?

Answer: $P(\text{at least one head}) + P(\text{at least one tail})] -$
$P(\text{at least one head or at least one tail}) = 2[1 - 2^{-10}] - 1 = 1 - 2^{-9}$.

Note that $\Omega = \{\text{head, tail}\}^{10}$ but we didn't explicitly state that... often we just do probability calculations without stating the state space.

Law of total probability

Breaks a probability into a sum of probabilities...: For any events A and B,

$$P(A) = P(B \cap A) + P(B^C \cap A).$$

Extension: Split Ω into pairwise disjoint sets

$$B_1, B_2, \dots,$$

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

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots = \bigcup_{n=1}^{\infty} (B_n \cap A).$$

Then $(B_i \cap A) \cap (B_j \cap A) = \emptyset$ for $i \neq j$, so

$$P(A) = \sum_{n=1}^{\infty} P(B_n \cap A).$$

Conditional probability

For events $A, B \subseteq \Omega$ with $P(B) > 0$, the conditional probability of A given B is

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

Can be rewritten as

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

and so we obtain...

Bayes' Theorem

Bayes' theorem

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

Notice that we have "reversed" the conditioning.

Since

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

we can reformulate Bayes' theorem as

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

Example: Test for a rare disease

Events: I=infected $\qquad I^C$=uninfected

Z=positive test $\qquad Z^C$=negative test

Known:

- $P(I) = 0.1\%$
- $P(Z|I) = 92\%$ (true positive)
- $P(Z|I^C) = 4\%$ (false positive)

Question:

- Given a positive test, what is the probability of having the disease?

It is $P(I|Z) \approx 2.5\%$ (which is far from $P(Z|I)$) because

$$P(I|Z) = \frac{P(Z|I)P(I)}{P(Z|I)P(I) + P(Z|I^C)P(I^C)} = \frac{0.92 \times 0.001}{0.92 \times 0.001 + 0.04 \times (1 - 0.001)}$$

Independence

Two events A and B are independent if and only if

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

Consequences:

- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$ provided $P(B) > 0$.
- $P(B|A) = P(B)$ provided $P(A) > 0$.
- A and B^C are independent.
- A^C and B are independent.
- A^C and B^C are independent.

Example:

Events: I=infected $\quad I^C$=uninfected

Z=positive test $\quad Z^C$=negative test

Known probabilities:

- $P(I) = p \in (0,1)$
- $P(Z|I) = q$ (true positive)
- $P(Z|I^C) = r$ (false positive)

Fact: Z and I are independent if and and only if $P(Z) = q = r$. However, as we want q to be much larger than r, there will be dependence.

Random variable

Definition: A **random variable (RV)** is a function X from the state space Ω to the real numbers $\mathbb{R}$ (i.e. $X : \Omega \mapsto \mathbb{R}$).

Definition: Its **distribution function**

$$F(x) = P(X \leq x), \quad x \in \mathbb{R},$$

is a non-decreasing function with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Discrete random variable

Definition: A **discrete RV** takes countably many values and has a **probability mass function (pmf)** $\pi(x)$:

- $\pi(x) = P(X = x) \geq 0$ for $x \in \mathbb{R}$ (or just $x \in X(\Omega)$),
- $\sum_x \pi(x) = 1$ (where $\sum_x \ldots$ means $\sum_{x \in \boldsymbol{X}(\Omega)} \ldots$).

Then

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

(where $\sum_{y \leq x} \ldots$ means $\sum_{y \in \boldsymbol{X}(\Omega):\, y \leq x} \ldots$) is a step function.

Example: Binomial distribution

A discrete RV X follows a **binomial distribution** with parameters p and n ($0 \le p \le 1$ 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

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

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

- Perform n independent experiments, each with outcomes "success" or "failure".
- $P($"success"$) = p$ for all experiments.
- Let $X =$ number of successes.
- Then $X \sim B(n, p)$.

Expectation and variance of RV

Definition: The **expectation (or mean value) of a discrete RV** is $\mu = E[X] = \sum_x x\pi(x)$.

Properties:

- $E[h(X)] = \sum_x h(x)\pi(x)$ for functions h.
- $E[a + bX] = a + bE[X]$ for numbers a and b.

Definition: The **variance of a discrete RV** is

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

Property: $Var(a + bX) = b^2 Var(X)$ for numbers a and b.

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

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

Continuous random variable

A RV X with a continuous distribution function is called a **continuous RV** – this implies $P(X = x) = 0$ for all $x \in \mathbb{R}$. It is usually specified by a **probability density function** (pdf) π, that is,

$$\pi(x) \geq 0 \quad \text{and} \quad F(x) = \int_{-\infty}^{x} \pi(y)dy \quad \text{for all } x \in \mathbb{R}.$$

Thus $\pi = F'$ and

- $P(a \leq X \leq b) = \int_a^b \pi(x)dx$ for all numbers $a \leq b$.

Expected value of continuous RV:

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

Variance of continuous RV:

- $\sigma^2 = Var(X) = E[(X - \mu)^2] = \int (x - \mu)^2\pi(x)dx = E[X^2] - \mu^2$.

Remarks

For simplicity we call both a pmf and a pdf for a **density** (it will always be clear whether we consider the density of a discrete or a continuous RV).

Important special case: a probability can be expressed as an expectation. For example, if $-\infty \leq a \leq b \leq \infty$,

$$E[1(a \leq X \leq b)] = P(a \leq X \leq b)$$

where $1(\cdot)$ is the indicator function.

Example: Normal distribution

A RV X follows a **normal distribution with mean μ and precision τ** if it has density/pdf

$$\pi(x) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau(x-\mu)^2}{2}\right), \quad x \in \mathbb{R}.$$

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

Note: X is a continuous RV, $\mu \in \mathbb{R}$, and $\tau = \frac{1}{\mathrm{Var}(X)} > 0$.

Two (or more) continuous RVs

Let X and Y be continuous RVs with **joint pdf/density**

$$\pi(x, y) \geq 0$$

meaning that $P\left((X, Y) \in A\right) = \iint_A \pi(x, y) dx dy$ for any $A \subseteq \mathbb{R}^2$.

Let $\pi_X(x)$ and $\pi_Y(y)$ be the **(marginal) densities** for X and Y, respectively; e.g.

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

We have

$$Eh(X, Y) = \int \int h(x, y)\pi(x, y) dx dy$$

for any real function h (provided the mean exists). For any real numbers a and b,

$$E[aX + bY] = aEX + bEY.$$

Covariance:

$Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EXEY.$

Conditional densities and independence of continuous RVs

The **conditional pdf/density** is

$$\pi_{Y|X}(y|x) = \frac{\pi(x,y)}{\pi_X(x)} \quad \text{if } \pi_X(x) > 0.$$

Definition: X and Y are **independent** if and only if

$$\pi(x,y) = \pi_X(x)\pi_Y(y), \quad x, y \in \mathbb{R},$$

or equivalently

$$\pi_{Y|X}(y|x) = \pi_Y(y) \quad \text{whenever } \pi_X(x) > 0.$$

Independence implies

$$Cov(X,Y) = 0, \quad Var(X+Y) = VarX + VarY.$$

Example: Independent normals

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

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

Similar exposition if we consider independent discrete RVs...

Or when considering discrete and continuous RVs together...