Applied Statistics - First test

Stéphane Rivaud

October 2024

1 Probabilities

1.1 Independence of events (2 points)

Let Ω be the event space and $A, B \subset \Omega$ such that A and B are independent, i.e.

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

Let us denote by A^c and B^c the complementary sets of A and B respectively, i.e. $A^c = \Omega \setminus A$ and $B^c = \Omega \setminus B$. Show that A^c and B^c are independent, i.e.

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c) \cdot \mathbb{P}(B^c)$$

*(**Hint**: Recall that for two sets A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$)*

1.2 Conditional probabilities (3 points)

There are three cards. The first is green on both sides, the second is red on both sides and the third is green on one side and red on the other. We choose a card at random and we see one side (also chosen at random). If the side we see is green, what is the probability that the other side is also green?

***Hint**: Many people intuitively answer 1/2. Show that the correct answer is 2/3. Use the definition of conditional probability, i.e.*

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

2 Random Variables

This section deal with Binomial random variables. A random variable Y follows a Binomial distribution with parameters n and p if its probability mass function is:

$$f_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

If $X_1, \cdots, X_n$ are i.i.d. random variables following a Bernoulli distribution with parameter n, then $Y = \sum_{i=1}^n X_i$ follows a Binomial distribution. A random variable following a Binomial distribution thus represent the potential gain after n coin toss where the probability of winning each round is p.

Recall that for a Binomial random variable Y with parameters n and p:

$$\mathbb{E}[Y] = np, \quad \mathrm{Var}(Y) = np(1-p)$$

2.1 A game example (3 points)

Suppose we play a game where we start with c dollars. On each play of the game you either double or halve your money, with equal probability. The goal is to compute your expected fortune after n trial.

1. Let X be a discrete random variable following a Binomial distribution with parameters n and $p = \frac{1}{2}$, i.e.

$$\mathbb{P}(X = k) = \binom{n}{k} \frac{1}{2^k} \left(1 - \frac{1}{2}\right)^{n-k} = \binom{n}{k} \frac{1}{2^n} \quad \text{for } k \geqslant 0.$$

   Let Y be the gain after n trials. Show that $Y = c \cdot 2^{2X-n}$.

2. Show that the expected fortune after n trial is

$$\mathbb{E}[Y] = c \left(\frac{5}{4}\right)^n$$

   Hint: *Use the binomial formula*

$$(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k$$

2.2 A random walk example (4 points)

A particle starts at the origin of the real line and moves along the line in jumps of one unit. At each timestep, the particle move one unit to the left with probability p or one unit to the right with probability $1-p$. Let Y_n be the position of the particle after n jump.

1. Show that

$$Y_n = n - 2X$$

   where X follows a Binomial distribution with parameters n and p.

   Note: *pay attention to the sign.*

2. Show that

$$\mathbb{E}[Y_n] = n(1-2p)$$

3. Using the definition of the variance, namely

$$\operatorname{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2]$$

Show that $\operatorname{Var}(Y) = 4np(1-p)$ with calculus.

4. Alternatively, we can use the properties of the variance operator to derive this result.

    (a) Show that for any random variable X and constant b, we have

    $$\operatorname{Var}(X + b) = \operatorname{Var}(X)$$

    (b) Show that for any random variable X and constant a, we have

    $$\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$$

    (c) Use the above properties to derive $\operatorname{Var}(Y)$.

3 Parameter Estimation (13 points)

Let $X_1, \cdots, X_n$ be i.i.d. random variables following a uniform distribution on the interval $[a, b] \subset \mathbb{R}$, i.e. they have density

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

1. Derive the value $\mathrm{CDF}_{X_i}(x)$ for $x < a$, $a \leqslant x \leqslant b$ and $x > b$.

   Note: *pay attention to the boundary values when you integrate.*

2. Compute the expected value of X_i, i.e. compute $\mathbb{E}[X_i]$.

3. Show that the second order moment is equal to

   $$\mathbb{E}[X_i^2] = \frac{a^2 + ab + b^2}{3}$$

   Hint: *develop the polynomial $(x-y)(x^2 + xy + y^2)$ and use the resulting identity to simplify your result.*

4. Show that the variance is equal to

   $$\operatorname{Var}(X_i) = \frac{(a-b)^2}{12}$$

5. Write down the likelihood function $\mathcal{L}(a, b | X_1, \cdots, X_n)$.

   Note: *similarly to question 1., there might be distinct cases depending on the value of $X_1, \cdots, X_n$.*

6. Show that for $a \leqslant \min(X_1, \cdots, X_n) \leqslant \max(X_1, \cdots, X_n) \leqslant b$

$$\ell(a, b|X_1, \cdots, X_n) = -n \log(b - a)$$

7. Show that the functions $a \mapsto \ell(a, b|X_1, \cdots, X_n)$ and $b \mapsto \ell(a, b|X_1, \cdots, X_n)$ do not have critical point.

8. Recall that the MLE estimator is the quantity that maximizes the likelihood. Show that the MLE estimators for a and b are:

$$\hat{a} = \min(X_1, \cdots, X_n), \quad \hat{b} = \max(X_1, \cdots, X_n)$$

9. Let $Y = \max(X_1, \cdots, X_n)$ be the Maximum Likelihood Estimator for $\hat{b}$. Show that

$$\mathrm{CDF}_Y(y) = F_{X_i}(y)^n \quad \text{for } a \leqslant y \leqslant b$$

Hint: *Make use of the fact that X_i are independent.*

10. Let's denote the PDF of Y by $f_Y(y)$. Derive the value of $f_Y(y)$ for $a \leqslant y \leqslant b$.

11. Show that the bias of Y as an estimator of b is equal to:

$$\begin{aligned} \mathrm{Bias}(Y) &= \mathbb{E}[Y] - b \\ &= \frac{a - b}{n + 1} \end{aligned}$$

Hint: $y(y - a)^{n-1} = (y - a)^n - a(y - a)^{n-1}$