PROV I MATHEMATIK
SANNOLIKHET OCH STATISTIK DV, 1MS321

VALENTIN GARINO

UPPSALA UNIVERSITET
2024-05-28

Hjälpmedel: Räknedosa, formel- och tabellsamling och engelsk-svensk ordlista för kursen 1MS321. Goda resonemang och motiveringar vägs in vid poängsättning. För betygen 3, 4, resp 5 krävs normalt minst 18, 25, resp 32 poäng.

1. Question 1

We throw a coin three times in a row (the faces are numbered 0 and 1 and are equally likely) and write X_1, X_2, X_3 the independent outcomes of each trial. Consider the following events:

- $A = \{X_1 = 0\}$
- $B = \{X_1 + X_2 + X_3 = 3\}$
- $C = \{X_1 + X_2 + X_3 \textit{ is an even number}\}$

(1) (2p) What is the law of $X_1 + X_2 + X_3$? Compute $P(C)$.
(2) (2p) Are A and B independent?
(3) (2p) Are A and C independent?

Solution

(1) Since X_1, X_2, X_3 are independent $\mathcal{B}\left(\frac{1}{2}\right)$ bernoullis variables, $X := X_1 + X_2 + X_3$ is Binomial with parameters 3 and $\frac{1}{2}$. Moreover,

$$\begin{aligned} P(C) &= P(X \in \{0,2\}) = P(\{X = 0\} \cup \{X = 2\}) \\ &= P(X = 0) + P(X = 2) \\ &= \binom{0}{3}\frac{1}{2^3} + \binom{2}{3}\frac{1}{2^3} \\ &= \frac{1}{8} + \frac{3}{8} = \frac{1}{2} \end{aligned}$$

(2) A, B are independent if and only if $P(A \cap B) = P(A)P(B)$. We have

$$P(A) = P(X_1 = 0) = \frac{1}{2},\ P(B) = P(X = 3) = \frac{1}{8}.$$

Moreover, $A \cap B = \{X_1 = 0\} \cap \{X_1 = 1, X_2 = 1, X_3 = 1\} = \emptyset$, so $P(A \cap B) = 0$ and A, B are not independent.

(3) We have $P(A)P(C) = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}$. Moreover,

$$A \cap C = \{X_1 = 0\} \cap \{X \textit{ is even}\} = \{X_1 = 0\} \cap \{X_2 + X_3 \textit{ is even}\}$$

Then, since X_1, X_2, X_3 are independent and $X_2 + X_3$ is binomial with parameters $2, \frac{1}{2}$,

$$P(A \cap B) = P(X_1 = 0)P(X_2 + X_3 \in \{0, 2\}) = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}.$$

A, C are independent.

2. Question 2

Consider the random variable X whith density function given by

$$f_X(x) := \begin{cases} 1 - \frac{x}{2} & \text{if } x \in [0, 2] \\ 0 & \text{else.} \end{cases}$$

(1) (1p) Draw a graphical representation of f_X.
(2) (1p) Compute the cumulative distribution function F_X of X. What is $P(X \leq 1)$?
(3) (3p) Compute $\mathbb{E}[X]$ and $V(X)$.
(4) (2p) Compute $Cov(X, X^2)$.

Solution

(1)
(2)

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u)du \\ &= \int_{-\infty}^{x} \left(1 - \frac{u^2}{2}\right) 1_{0 \leq u \leq 2} du \end{aligned}$$

$$= \begin{cases} 0 & \text{if } x < 0 \\ x - \frac{x^2}{4} & \text{if } x \in [0, 2] \\ 1 & \text{else.} \end{cases}$$

Then, $P(X \leq 1) = F_X(1) = \frac{1}{4}$

(3)

$$\mathbb{E}[X] = \int_0^2 x f_X(x)dx = \int_0^2 \left(x - \frac{x^2}{2}\right) dx = \left[\frac{x^2}{2}\right]_0^2 - \left[\frac{x^3}{6}\right]_0^2 = \frac{2}{3}.$$

$$\mathbb{E}[X^2] = \int_0^2 x^2 f_X(x)dx = \int_0^2 \left(x^2 - \frac{x^3}{2}\right) dx = \left[\frac{x^3}{3}\right]_0^2 - \left[\frac{x^4}{8}\right]_0^2 = \frac{2}{3}.$$

$$V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{3} - \frac{4}{9} = \frac{2}{9}.$$

(4)

$$\begin{aligned} Cov(X, X^2) &= \mathbb{E}[X.X^2] - \mathbb{E}[X]\mathbb{E}[X^2] \\ &= \int_0^2 x^3 \left(1 - \frac{x}{2}\right) dx - \frac{4}{9} \\ &= \int_0^2 \left(x^3 - \frac{x^4}{2}\right) dx - \frac{4}{9} \\ &= \left[\frac{x^4}{4}\right]_0^2 - \left[\frac{x^5}{10}\right]_0^2 - \frac{4}{9} \\ &= \frac{16}{45}. \end{aligned}$$

3. Question 3

We consider two populations of cars on the highway: blue cars and yellow cars. We assume that each vehicle runs at constant speed. Furthermore, we assume that the speed (in km/h) of a blue car taken at random follows a normal distribution $\mathcal{N}(90, 400)$ and the speed of a yellow car taken at random follows a normal distribution $\mathcal{N}(130, 900)$.

(1) (2p) What is the probability that a blue car taken at random runs faster than 120 km/h? Use the quantile table of the normal distribution to help yourself.

(2) We organise a competition between the blue and yellow cars: we select at random $2n$ independent blue cars and n independent yellow cars. All of them drive for one hour. At the end, whichever team has travelled the most cumulated distance wins.

   (a) (2p) What is the distribution of the cumulative distance travelled by all the $2n$ cars in the blue team? Same question for the yellow team.

   (b) (2p) Compute the probability that the yellow team is winning for n=5.

Solution

(1) Let X_1 be the distribution of the speed of the blue cars. We have $X \sim \mathcal{N}(90, 400)$.

$$P(X_1 > 120) = P\left(\frac{X_1 - 90}{\sqrt{400}} > \frac{120 - 90}{\sqrt{400}}\right) = P\left(Z > \frac{3}{2}\right)$$

with $Z \sim \mathcal{N}(0, 1)$. Thus, by reading the quantile table, we have

$$= P(X_1 > 120) = 1 - \phi\left(\frac{3}{2}\right) = 0.0668$$

(2) (a) $2n$ blue cars with speed X_i where X_i are independent $\mathcal{N}(90, 400)$. In one hour, the car i travel a distance $X_i * 1 = X_i$ (in km).

From the course, we know that

$$X = \sum_{i=1}^{2n} X_i \sim \mathcal{N}(2n * 90, 2n * 400) = \mathcal{N}(180n, 800n).$$

Similarly, the distance Y traveled by the n yellow cars is $Y \sim \mathcal{N}(130n, 900n)$.

(b) We have that $X - Y$ follows a normal distribution $\mathcal{N}(50n, 1700n)$. Then,

$$\begin{aligned} P(X < Y) = P(X - Y < 0) &= P\left(\frac{X - Y - 50n}{\sqrt{1700n}} < \frac{-50n}{\sqrt{1700n}}\right) \\ &= P\left(Z < -\sqrt{n}\frac{50}{\sqrt{1700}}\right) \\ &= 1 - \phi\left(\sqrt{n}\frac{50}{\sqrt{1700}}\right). \end{aligned}$$

For $n = 5$, we get $n\frac{50}{\sqrt{1700}} = 2,711$, and then $P(X < Y) = 0.0034$

4. Question 4

We consider a biased dice (with faces numbered from 1 to 6): faces numbered from 1 to 5 have same probability and face numbered 6 has a probability $p > 0$ unknown. We want to estimate p. We roll the dice n times and denote by X_i each outcome. Consider the estimator

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i=6}.$$

(where $1_{X_i=6}$ equals 1 if $X_i = 6$ and 0 otherwise). In other words, this estimator counts the proportion of dice rolls that lands on the face numbered 6.

(1) (2p) Let X be the outcome of one dice roll. What is the probability mass function of X (expressed as a function of p)? What is the distribution of $1_{X=6}$?

(2) (3p) We assume that $n = 500$ and $\hat{p} = 0.1$. Compute a confidence interval at level 95% for the parameter p and justify your answer. You can use the formulary to help yourself.

Solution

- We have that $1_{X=6} = 1$ if $X = 6$ and 0 otherwise. As a consequence, $1_{X_1=6}$ is a Bernoulli variable with parameter p. Morevoer, $P(\{1, 2, 3, 4, 5\}) = 5p_X(1) = 1 - p$. $p_X(x) = p$ if $x = 6$ and $p_X(x) = \frac{1-p}{5}$ if $x \in \{1, 2, 3, 4, 5\}$.

- $n = 500, \hat{p} = 0.1$ implies that $n\hat{p}(1-\hat{p}) > 10$. As a consequence, we can use the normal approximation to compute the confidence interval (see formulary p3):

$$I_p = \left[\hat{p} - \lambda_{0.025}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + \lambda_{0.025}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] = [0.0737, 1.263]$$

5. Question 5

Consider two normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y \sim \mathcal{N}(2\mu - 1, \sigma^2)$. Assume that X and Y are independent and that we want to estimate the unknown parameter μ (we assume that σ^2 is known).

(1) (2p) Consider the estimator $T_1 = \frac{1}{2}X + \frac{1}{4}(Y+1)$. Show that T is unbiased and compute its variance.
(2) (2p) Consider now the estimator $T = aX + bY + c$ for three parameters a, b, c. Show that T is unbiased if and only if the parameters satisfies the following system of equations:

$$\begin{cases} a + 2b - 1 = 0 \\ b = c \end{cases}$$

Solution

(1) We have

$$\mathbb{E}[T_1] = \frac{1}{2}\mathbb{E}[X] + \frac{1}{4}(\mathbb{E}[Y] + 1) = \frac{1}{2}\mu + \frac{1}{4}(2\mu - 1 + 1) = \mu.$$

T_1 is then unbiased. Furthermore,

$$V(T_1) = \frac{1}{4}\sigma^2 + \frac{1}{16}(\sigma^2) = \frac{5}{16}\sigma^2.$$

(2) We have $\mathbb{E}[T] = a\mu + b(2\mu - 1) + c$. If T is unbiased, we then need to have $(a+2b)\mu + (c-b) = \mu$, which implies $(a+2b-1)\mu + (c-b) = 0$. Since this equation needs to be valid for all $\mu > 0$, this means that $a + 2b - 1 = 0$ and $c = b$. Conversely, if $a + 2b - 1 = 0$ and $c = b$, then $\mathbb{E}[T] = \mu$ and the estimator is unbiased.

6. Question 6

Assume that the police observe the highway and notice that the number of speed excesses over time follows a Poisson process with intensity $\lambda_1 = 1/h$ (one per hour) for cars and $\lambda_2 = 2/h$ for motorbikes.

(1) (2p) What is the probability of no speed excess by cars in the time interval $[0, 2]$? (the units are hours)
(2) (2p) What is the probability of at least three speed excesses in total (taking into account both cars and motorbikes) during the first two hours?

(3) (2p) We assume that the police install a radar that record the license plates of vehicles in speed excess, and then send a fine to the vehicle driver (both cars and motorbikes). However, with probability $p = 0.2$, the license plate is unreadable and the driver cannot be identified. Using the thinning property of Poisson processes, give the probability that at least three drivers are fined during the first two hours.

Solution

(1) Since the number of car speed excesses follows a Poisson process N^1, the number of speed excess in the first two hours is a Poisson variable with parameter $2\lambda_1$:

$$P(N_2^1 = 0) = e^{-2\lambda_1} = e^{-2} = 0.136.$$

(2) Let N^2 be the number of speed excesses by motorbikes. By the addition property, $N^1 + N^2$ is a Poisson process with intensity $\lambda = \lambda_1 + \lambda_2 = 3/h$. Thus, the number of combined speed excess in the first two hours is a Poisson variable with parameter 2λ:

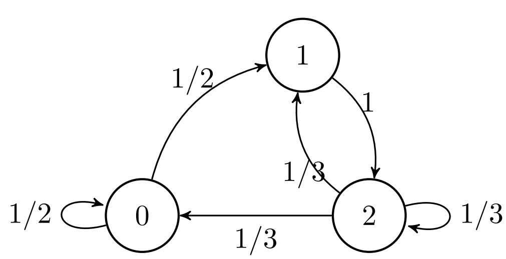
$$\begin{aligned} & P(N_2^1 + N_2^2 \geq 3) \\ = \; & 1 - (P(N_2^1 + N_2^2 = 0) + P(N_2^1 + N_2^2 = 1) + P(N_2^1 + N_2^2 = 2)) \\ = \; & 1 - e^{-6} - 6e^{-6} - 18e^{-6} = 1 - 0.062 = 0.938 \end{aligned}$$

(3) By the Thinning property, the number of fined drivers X_t follows a Poisson process with parameter $(1 - 0.2) * \lambda = 4.8$. Then,

$$P(X_2 = 3) = \frac{(4.8)^3}{6} e^{-4.8} = 0.152.$$

7. Question 7

Consider the following three-state Markov chain with values in $E = \{0, 1, 2\}$



(1) (1p) Write the transition matrix associated with this Markov chain.
(2) (2p) Consider the initial state distribution $p_0 = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$. Compute the probability distribution p_2 after two iterations.
(3) (1p) Does this Markov chain possess an asymptotic distribution? Justify your answer.

(4) (2p) If the answer to the previous question was yes, compute the asymptotic distribution.

Solution

(1) $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$

(2) We know that $p_2 = p_0 P^2$. The computations gives $p_2 = \left(\frac{19}{72}, \frac{19}{72}, \frac{17}{36}\right)$.

(3) The chain is both irreductible and aperiodic, which means that it has a unique stationary distribution, which is also asymptotic.

(4) To identify the stationary distribution, we need to solve the system of equations $\pi = \pi P$, $\pi_0 + \pi_1 + \pi_2 = 1$. We get:

$$= \begin{cases} \frac{\pi_0}{2} + \frac{\pi_2}{3} = \pi_1 \\ \frac{\pi_0}{2} + \frac{\pi_2}{3} = \pi_2 \\ \pi_1 + \frac{\pi_2}{3} = \pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases}$$

From the first and second equations, we get that $\pi_0 = \pi_1 = \frac{2}{3}\pi_2$. Plugging this into the last eqution, we get $\pi_2 = \frac{3}{7}$ and then $\pi_0 = \pi_1 = \frac{2}{7}$.