Sample solutions to exam 22-08-16

Fråga 1:

a) Let X be a discrete random variable which only takes the values 0, 1 and 2 with positive probability. We know that $P(X = 0) = P(X = 2) = 0.2$. Calculate the standard deviation of X. (2p)

b) Let Y be a random variable with probability density function

$$f(y) = \frac{1}{2}y \qquad \text{for } 0 \leqslant y \leqslant 2$$

and $f(y) = 0$ otherwise. Determine $P(Y < 1)$ and $E[Y]$. (2p)

c) Calculate the covariance between the random variables Y and $1/Y$. (2p)

Solution:

a) As $P(X = 0) + P(X = 1) + P(X = 2) = 1$, we have $P(X = 1) = 0.6$. So

$$E[X] = 0.2 \cdot 0 + 0.6 \cdot 1 + 0.2 \cdot 2 = 1 \text{ and } E[X^2] = 0.2 \cdot 0^2 + 0.6 \cdot 1^2 + 0.2 \cdot 2^2 = 1.4,$$

so $V[X] = E[X^2] - E[X]^2 = 0.4$ and $D[X] = \sqrt{V[X]} \approx 0.632$.

b)

$$P(Y < 1) = \int_0^1 \frac{1}{2} y dy = [\frac{1}{4}y^2]_0^1 = \frac{1}{4} - 0 = \frac{1}{4}$$

$$E[Y] = \int_0^2 y f(y) dy = \int_0^2 \frac{1}{2} y^2 dy = [\frac{1}{6}y^3]_0^2 = \frac{8}{6} - 0 = \frac{4}{3}$$

c) $C(Y, 1/Y) = E[Y \cdot 1/Y] - E[Y]E[1/Y] = E[1] - \frac{4}{3}E[1/Y] = 1 - \frac{4}{3}E[1/Y]$. Note that

$$E[1/Y] = \int_0^2 \frac{1}{y} f(y) dy = \int_0^2 \frac{1}{2} dy = \frac{1}{2} \cdot 2 = 1,$$

so $C(Y, 1/Y) = 1 - \frac{4}{3} \cdot 1 = -\frac{1}{3}$

Fråga 2: We know that 1% of the population has a certain disease, and there is a test for this disease. If a person has the disease, then the test is positive with probability 95%. If a person does not have the disease, the test is negative with probability 99%.

a) You select a random person from the population, and test them. What is the probability that the test is positive? (1p)

b) You select a random person from the population and test them. The test is positive. What is the probability that they have the disease? (2p)

c) How high would the prevalence of the disease in the population need to be so that this probability is at least 90%? In other words, what percentage of the population needs to have the disease so that, if you randomly select a person and test them and the test is positive, then the probability that they really have the disease is at least 90%? (2p)

Solution:

a) Let D be the event that the randomly sampled person has the disease, and T the event that the test is positive. Then $P(D) = 0.01$, $P(D^*) = 0.99$, $P(T|D) = 0.95$ and $P(T|D^*) = 1 - 0.99 = 0.01$. By the law of total probability,

$$P(T) = P(D)P(T|D) + P(D^*)P(T|D^*) = 0.01 \cdot 0.95 + 0.99 \cdot 0.01 = 0.0194$$

b) By Bayes' Theorem,

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.01 \cdot 0.95}{0.0194} \approx 0.490$$

c)Let p be the proportion of people with the disease in the population, so now $P(D) = p$, $P(D^*) = 1 - p$. The formula in b) becomes

$$P(D|T) = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(D^*)P(T|D^*)} = \frac{0.95p}{0.95p + 0.01(1-p)} = \frac{0.95p}{0.94p + 0.01} \geqslant 0.9$$

This gives $0.95p \geqslant 0.9(0.94p+0.01) = 0.846p+0.009$, so $0.104p \geqslant 0.009$, so $p \geqslant 0.009/0.105 \approx 0.086 = 8.6\%$.

Fråga 3: On a social media platform, users can upload videos. The distribution of the file sizes is unknown, but the company behind the social media platform knows from experience that the average file size is 500MB and the standard deviation is 300MB.

Suppose that 150 users each upload a video, independently from each other. Approximate the probability that 80GB is enough storage space to store all of their videos. When answering this question, introduce a suitable model and define appropriate random variables. (6p)

Solution:

Let $X_1, \ldots, X_{150}$ be the file sizes of the video uploads (in MB), then by our assumptions $\mu = E[X_i] = 500$, $\sigma^2 = V[X_i] = 300^2$, and the random variables are independent. We want to estimate the probability

$$P(X_1 + \cdots + X_{150} \leqslant 80 \cdot 1000) = P(X_1 + \cdots + X_{150} \leqslant 80000)$$

Let $S_{150} = X_1 + \ldots X_{150}$. By the central limit theorem, we can approximate

$$\frac{S_{150} - 150\mu}{\sqrt{150}\sigma}$$

by $Z \sim N(0,1)$. So

$$P(S_{150} \leqslant 80000) = P\left(\frac{S_{150} - 150\mu}{\sqrt{150}\sigma} \leqslant \frac{80000 - 150\mu}{\sqrt{150}\sigma}\right) \approx P\left(Z \leqslant \frac{80000 - 150 \cdot 500}{\sqrt{150} \cdot 300}\right)$$
$$\approx P(Y \leqslant 1.36) = \Phi(1.36) \approx 0.9131$$

using the table for $\Phi(\cdot)$.

Fråga 4: For an integer $n \geqslant 1$, let $x_1, \ldots, x_n$ be independent samples from $X \sim N(\mu, 2\sigma^2)$, where μ and $\sigma > 0$ are unknown parameters. Independently, let y be a sample from $Y \sim N(\mu, \sigma^2)$, where μ and σ are the same unknown parameters. Let $\bar{x}$ denote the sample mean of $x_1, \ldots, x_n$.

a) Show that both $\hat{\mu}_1 = \frac{3\bar{x}+y}{4}$ and $\hat{\mu}_2 = \frac{\bar{x}+3y}{4}$ are unbiased estimators for μ. (2p)

b) For which values of n is $\hat{\mu}_1$ more effective than $\hat{\mu}_2$? (3p)

Solution:

a) By linearity of the expectation we have $E[Y] = \mu$ and $E[\bar{X}] = \frac{1}{n}E[X_1 + \cdots + X_n] = \frac{1}{n} \cdot n\mu = \mu$, so $E[\hat{\mu}_1] = \frac{1}{4}(3E[\bar{X}]) + E[Y]) = \frac{1}{4}(3\mu + \mu)$, as well as $E[\hat{\mu}_2] = \frac{1}{4}(E[\hat{X}] + 3E[Y]) = \frac{1}{4}(\mu + 3\mu) = \mu$, so both estimators are unbiased.

b) We have $V[Y] = \sigma^2$ and $V[\bar{X}] = \frac{1}{n^2}(n \cdot 2\sigma^2) = 2\sigma^2/n$, so $V[\hat{\mu}_1] = \frac{1}{16}(9V[\bar{X}] + V[Y]) = \frac{1}{16}(18\sigma^2/n + \sigma^2)$ and $V[\hat{\mu}_2] = \frac{1}{16}(V[\bar{X}] + 9V[Y]) = \frac{1}{16}(2\sigma^2/n + 9\sigma^2)$. So $\hat{\mu}_1$ more effective than $\hat{\mu}_2$ if $V[\hat{\mu}_1] < V[\hat{\mu}_2]$, which means

$$\frac{1}{16}(18\sigma^2/n + \sigma^2) < \frac{1}{16}(2\sigma^2/n + 9\sigma^2), \quad \text{so} \quad \frac{18}{n} + 1 < \frac{2}{n} + 9, \quad \text{so} \quad \frac{16}{n} < 8$$

which gives $n > 16/8 = 2$.

Fråga 5: An email provider wants to mark some emails as important. The engineering team designing the marking algorithm claims that 15% of all emails get marked as important. As a test, the company sends 200 emails to an email account. Of these, 44 are marked as important.

a) With good motivation, give an approximate 95% confidence interval for the probability that an email is marked as important. What can you say about the engineering team's claim? (4p)

b) What changes about the confidence interval, and your assessment of the team's claim, if you set the confidence level to 99% instead? (2p)

Solution:

a) We want a confidence interval for p, the proportion of emails that gets marked as important. We can give an approximate confidence interval for this:

$$I_p = \left[\frac{x}{n} \pm \lambda_{\alpha/2} d(\hat{p})\right]$$

where $x = 44$, $n = 200$, $\alpha = 0.05$, $\hat{p} = x/n = 44/200 = 0.22$, $d(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0.22 \cdot 0.78/200} \approx 0.02929$. (Note that $n\hat{p}(1-\hat{p}) = 200 \cdot 0.22 \cdot 0.78 = 34.32 > 10$, so we can make this approximation.)

Plugging in these values, and $\lambda_{0.025} \approx 1.96$ from the table, gives

$$I_p \approx [0.22 \pm 0.0574] \approx [0.163, 0.277].$$

This interval does not contain 0.15, so we can reject the engineer team's claim with a confidence level of 99%.

b) If we instead have $\alpha = 0.01$, then we use $\lambda_{0.005} \approx 2.58$ instead and get the longer interval

$$I_p \approx [0.22 \pm 0.0756] \approx [0.144, 0.296].$$

This interval contains 0.15, so we cannot reject the engineering team's claim at this confidence level — the confidence interval is compatible with their claim.

Fråga 6: We have the following information on a Markov chain $X_0, X_1, X_2, \ldots$ with state space $E = \{0, 1, 2\}$. If the Markov chain is at state 0, with probability $\frac{1}{4}$ it moves to state 1. If the Markov chain is at state 1, it stays where it is with probability $\frac{1}{2}$. If the chain is at state 2, it cannot stay where it is, but has to move to a different state. No matter what state the Markov chain is in, the probability that the next state is 0 is always exactly $\frac{1}{2}$.

a) Determine the transition matrix $\mathbf{P}$ of the Markov chain. (2p)

b) Draw the state transition diagram. (1p)

c) The Markov chain starts at time $t = 0$ in the state 0. Calculate $P(X_2 = 1)$. (1p)

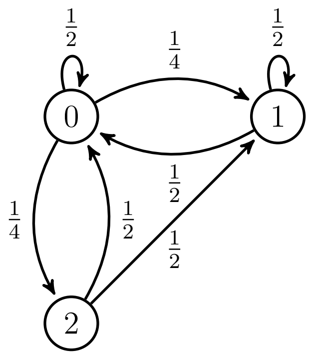
d) Determine the Markov chain's stationary distribution. (You can assume that a unique stationary distribution exists.) (2p)

Solution:

a)

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

b)



c)

$$\pi_0 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad \pi_1 = \pi_0 \mathbf{P} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \end{pmatrix} \text{ and } \pi_2 = \pi_1 \mathbf{P},$$

which has as its second entry $\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8}$, so $P(X_2 = 1) = \frac{3}{8}$.

d) We are looking for a vector $\pi = \begin{pmatrix} a & b & c \end{pmatrix}$ where $a + b + c = 1$ and $\pi \mathbf{P} = \pi$. This gives

$$\frac{1}{2}a + \frac{1}{2}b + \frac{c}{=}a \tag{1}$$

$$\frac{1}{4}a + \frac{1}{2}b + \frac{1}{2}c = b \tag{2}$$

$$\frac{1}{4}a = c \tag{3}$$

From (3) we get $a = 4c$, plugging into (2) gives $c + \frac{1}{2}b + \frac{1}{2}c = b$, which gives $b = 3c$. Plugging into $a + b + c = 1$ we get $8c = 1$, so $c = \frac{1}{8}$, $b = \frac{3}{8}$, $a = 4/8$, so the stationary distribution is given by

$$\pi = \begin{pmatrix} \frac{4}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}.$$

Fråga 7: On your phone, you receive notifications from emails and from texts. On average, you receive 0.5 emails and 2 texts per hour, and the number of emails and texts you receive over time can be assumed to be distributed according to two independent Poisson processes. Introduce a suitable model, defining appropriate random variables and events, to answer the following questions.

a) You forget your phone at home, and will only return 2 hours later. How many notifications do you expect you will receive on your phone? (2p)

b) What is the probability that you will have at least 3 notifications? (1p)

c) You return to see that you have exactly 4 notifications. What is the probability that all of them are texts? (2p)

d) You receive a text at 8:00 in the morning. What is the expected time of the next text? (1p)

Solution:

Let $\{N_A(t) : t \geqslant 0\}$ and $\{N_B(t) : t \geqslant 0\}$ be two independent Poisson processes with intensities $\lambda_A = 0.5$ (hour^{-1}) and $\lambda_B = 2$ (hour^{-1}) which model the incoming texts and emails over time, respectively. Then, by the superposition property, the total number of notifications

$$N_C(t) = N_A(t) + N_B(t), t \geqslant 0$$

is also given by a Poisson process with intensity $\lambda_C = \lambda_A + \lambda_B = 2.5$.

a) We have $N_C(2) = N_C(2) - N_C(0) \sim \text{Po}(2 \cdot 2.5) = \text{Po}(5)$, which has expectation 5.

b) $P(N_C(2) \geqslant 3) = 1 - P(N_C(2) = 0) - P(N_C(2) = 1) - P(N_C(2) = 2) = 1 - e^{-5}(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!}) \approx 0.875$

c) Let M be the event that you have 4 notifications in total, and N the event that you have four text notifications, then we want $P(N|M)$. Note that $N_A(2) \sim \text{Po}(2 \cdot 2) = \text{Po}(4)$ and $N_B(2) \sim \text{Po}(2 \cdot 0.5) = \text{Po}(1)$

$$P(N|M) = \frac{P(N \cap M)}{P(M)} = \frac{e^{-4}\frac{4^4}{4!}e^{-1}\frac{1^0}{0!}}{e^{-5}\frac{5^4}{4!}} = \frac{4^4}{5^4} = 0.4096.$$

(Note that $N \cap M$ holds if and only if $N_A(2) = 4$ *and* $N_B(2) = 0$.)

d) The waiting time between two events of a Poisson process with intensity λ has the distribution $\text{Exp}(a)$ with $a = 1/\lambda$, which has expectation $a = 1/\lambda$. In this case, $\lambda_A = 2$ (hours^{-1}), so the expected waiting time is 0.5 (hours). So the expected time of the next text is 8:30.