MTH 201: Probability and Statistics

Quiz 3

21/04/2025

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen and eraser. Any exchange of information related to the quiz with a human or machine will be deemed as cheating. Institute rules will apply. Explain your answers. Show your steps. Approximate calculations are fine as long as the approximations are reasonable. Don't bother simplifying products and sums that are too time consuming. Leave your answer as products and sums, if you don't have the time. You have 45 minutes.

Question 1. 30 **marks** A bin has 12 balls of which 6 are red, 3 are blue, and 3 are black. You draw three balls randomly and without replacement. Let X_i, $i = 1, 2, 3$, be the random variable corresponding to outcome i. Each X_i maps red to 0, blue to 1, and black to 2. Answer the following questions.

1) Calculate $E[X_1 + X_2 + X_3]$.
2) Calculate $E[(X_1 + X_2 + X_3)^2]$

Question 2. 35 **marks** A bin consists of many balls of colours red, blue and black. You would like to estimate the fraction of blue balls. You draw 10 balls randomly and with replacement and obtain the sequence red, black, red, blue, blue, red, black, red, black, blue. Propose a suitable estimator and provide a confidence interval such that the probability the estimator is within ± 0.2 of the true fraction of blue balls in the bin is 0.75. Show all your steps, which would include clearly defining your estimator and stating any properties you use.

Question 3. 35 **marks** A temperature sensor outputs $X = t + N$, where t is the true unknown temperature and N is a Gaussian random variable with 0 mean and variance 16. Since the output is noisy, you choose to calculate the sample mean over $n \geq 1$ measurements of the temperature. Calculate the minimum value of n to achieve a confidence of 0.95 over an interval of ± 1. If using the Gaussian distribution, leave your answer in terms of the CDF $\Phi(x)$ of the standard normal. In case you believe that the Chebyshev's inequality is more likely to give you the minimum required n, see the section on useful facts at the end of this paper.

I. Useful Facts

An exponential RV X with expected value (mean) $1/\lambda$ has a PDF $f_X(x) = \lambda e^{-\lambda x}, x \geq 0$.
A Poisson random variable K with parameter α has PMF $P_K(k) = \alpha^k e^{-\alpha}/k!$. Its mean is α.
The Chebyshev's inequality states that for all $c > 0$, $P[|X - E[X]| \geq c] \leq \frac{\text{Var[X]}}{c^2}$.

Question 1. 30 marks A bin has 12 balls of which 6 are red, 3 are blue, and 3 are black. You draw three balls randomly and without replacement. Let X_i, $i = 1, 2, 3$, be the random variable corresponding to outcome i. Each X_i maps red to 0, blue to 1, and black to 2. Answer the following questions.

1) Calculate $E[X_1 + X_2 + X_3]$.
2) Calculate $E[(X_1 + X_2 + X_3)^2]$

Note that we need the marginal PMFs of X_1, X_2, and X_3 to calculate $E[X_1 + X_2 + X_3]$ and we need the joint PMFs $P_{X_i, X_j}(x, y)$ for all $i \neq j$ to calculate $E[(X_1 + X_2 + X_3)^2]$.

I mentioned in the lecture that the joint PMFs $P_{X_1,X_2}(x,y)$, $P_{X_2,X_3}(x,y)$, and $P_{X_1,X_3}(x,y)$ are all the same. That is it is sufficient to calculate just one of them. [Why are they the same? Note that it is easy to verify that they are the same.]

Consider the joint PMF $P_{X_1,X_2}(x_1, x_2)$. The table summarizes it.

$X_1 \backslash X_2$	0	1	2
0	$\frac{6}{12}\frac{5}{11}$	$\frac{6}{12}\frac{3}{11}$	$\frac{6}{12}\,{}^{3}/_{11}$
1	$\frac{3}{12}\frac{6}{11}$	$\frac{3}{12}\frac{2}{11}$	$\frac{3}{12}\frac{3}{11}$
2	$\frac{3}{12}\frac{6}{11}$	$\frac{3}{12}\frac{3}{11}$	$\frac{3}{12}\frac{2}{11}$

Calculating the joint PMFs. 10/30

The marginal PMFs are now easy to obtain.

$$P[X_1 = 0] = \frac{6}{12} = \frac{1}{2}$$
$$P[X_1 = 1] = \frac{3}{12} = \frac{1}{4}$$
$$P[X_1 = 2] = \frac{3}{12} = \frac{1}{4}.$$

Note that (easy to verify)

$$P_{X_1}(x) = P_{X_2}(x) = P_{X_3}(x) \text{ for every } x \in \mathbb{R}.$$

That is the marginal PMFs are the same. X_1, X_2, X_3 are identically distributed.

Calculating the marginal PMFs. 5/30

(1) $E[X_1 + X_2 + X_3]$

$$= E[X_1] + E[X_2] + E[X_3]$$
$$= 3E[X_1] = 3\left(\frac{1}{4} + 2\left(\frac{1}{4}\right)\right)$$
$$= 3\left(\frac{3}{4}\right) = 9/4.$$

5/30

(2) $E[(X_1 + X_2 + X_3)^2]$

$$= E[X_1^2] + E[X_2^2] + E[X_3^2]$$
$$+ 2E[X_1 X_2]$$
$$+ 2E[X_1 X_3]$$
$$+ 2E[X_2 X_3]$$

$$E[X_1^2] = \frac{1}{4} + 2^2\left(\frac{1}{4}\right) = 1 + \frac{1}{4} = \frac{5}{4}$$
$$= E[X_2^2] = E[X_3^2].$$

$$E[X_1 X_2] = (1)(1)\frac{3}{12}\frac{2}{11}$$
$$+ (1)(2)\frac{3}{12}\frac{3}{11}$$
$$+ (2)(1)\frac{3}{12}\frac{3}{11} + (2)(2)\frac{3}{12}\frac{2}{11}$$
$$= \frac{3}{12}\frac{2}{11} + 4\frac{3}{12}\frac{3}{11} + 4\frac{3}{12}\frac{2}{11}$$
$$= \frac{3}{12}\frac{2}{11} + 4\left(\frac{3}{12}\right)\left(\frac{5}{11}\right)$$
$$= \frac{3}{66} + \frac{5}{11} = \frac{33}{66} = \frac{1}{2}.$$
$$= E[X_2 X_3] = E[X_1 X_3].$$

$$\therefore E[(X_1 + X_2 + X_3)^2] = \left(\frac{5}{4}\right)3$$
$$+ 2(3)\frac{1}{2}$$
$$= \frac{15}{4} + 3 = \frac{27}{4}.$$

10/30

Question 2. 35 **marks** A bin consists of many balls of colours red, blue and black. You would like to estimate the fraction of blue balls. You draw 10 balls randomly and with replacement and obtain the sequence red, black, red, blue, blue, red, black, red, black, blue. Propose a suitable estimator and provide a confidence interval such that the probability the estimator is within ± 0.2 of the true fraction of blue balls in the bin is 0.75. Show all your steps, which would include clearly defining your estimator and stating any properties you use.

Let rv X describe whether a randomly chosen ball is blue or not. X is Bernoulli(p) 5/35

The sample mean $M_n(X) = \frac{X_1 + X_2 + \cdots + X_n}{n}$

is an estimator of the frequency of blue balls, which is the unknown p.

Note that one can't fix the interval, the confidence, and also the n arbitrarily. 5/35

You could do one of the following, given the data. (Either one is okay)

(1) Hold the interval for ± 0.2, stick for $n = 10$, and calculate the confidence.

The confidence is

$$P\left[\left|M_n(X) - E[X]\right| \le 0.2\right]$$

$$= P\left[\left|M_n(X) - E[M_n(X)]\right| \le 0.2\right]$$ Since $M_n(X)$ is an unbiased estimator of $E[X]$.

$$\ge 1 - \frac{Var[M_n(X)]}{(0.2)^2}$$ We need the Chebyshev's inequality.

$$= 1 - \frac{Var[X]}{10\,(0.2)^2}$$

10/35

Our confidence is the minimum value of the above probability, for any $Var[X]$, when X is Bernoulli.

Confidence is $1 - \frac{0.25}{(10)(0.2)^2}$

$$= 1 - \frac{0.25}{(10)(0.04)}$$

$$= 1 - \frac{(0.25)10}{4}$$

$$= 1 - \frac{2.5}{4} = \frac{1.5}{4} = \frac{3}{2(4)}$$

$$= 3/8.$$

10/35

The confidence interval is $\left(\frac{3}{10} - 0.2, \frac{3}{10} + 0.2\right)$ with confidence $\frac{3}{8}$.

5/35

(2) $n = 10$, confidence $= 0.75$.
Calculate the interval.

We have:

$$P\left[\left|M_n(X) - E[X]\right| \le c\right]$$

$$= P\left[-c \le M_n(X) - E[M_n(X)] \le c\right]$$

$$\ge 1 - \frac{Var[X]}{(10)\,c^2}$$

Reasons as above.

10/35

We want a c that satisfies

$1 - \frac{Var[X]}{(10)\,c^2} = 0.75$ for any Bernoulli X.

Pick maximum possible $Var[X]$.

$$1 - \frac{0.25}{(10)\,c^2} = 0.75$$

$$\Rightarrow 0.25 = \frac{0.25}{10\,c^2}$$

$$10c^2 = 1$$

$$c^2 = \frac{1}{10} = 0.1$$

$$c = \sqrt{0.1}.$$

10/35

The confidence interval is $\left(\frac{3}{10} - \sqrt{0.1}, \frac{3}{10} + \sqrt{0.1}\right)$ with confidence 0.75.

5/35

Question 3. 35 **marks** A temperature sensor outputs $X = t + N$, where t is the true unknown temperature and N is a Gaussian random variable with 0 mean and variance 16. Since the output is noisy, you choose to calculate the sample mean over $n \geq 1$ measurements of the temperature. Calculate the minimum value of n to achieve a confidence of 0.95 over an interval of ± 1. If using the Gaussian distribution, leave your answer in terms of the CDF $\Phi(x)$ of the standard normal. In case you believe that the Chebyshev's inequality is more likely to give you the minimum required n, see the section on useful facts at the end of this paper.

Sample mean $M_n(X) = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$. 5/35

Note: X is Gaussian with mean t and variance 16. (Since $X = t + N$) 5/35

Further note that $M_n(X)$ is Gaussian with mean t and variance $\dfrac{16}{n}$. 10/35

We require to choose the smallest integer n that satisfies

$$P\left[\,|M_n(X) - t| \leq 1\right] = 0.95.$$

$$P\left[\,|M_n(X) - t| \leq 1\right]$$

$$= P\left[-1 \leq M_n(X) - t \leq 1\right]$$

$$= P\left[\frac{-1}{(4/\sqrt{n})} \leq \frac{M_n(X) - t}{(4/\sqrt{n})} \leq \frac{1}{(4/\sqrt{n})}\right]$$

$$= \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right)$$

10/35

Solve for n^* that satisfies

$$\Phi\left(\frac{\sqrt{n^*}}{4}\right) - \Phi\left(-\frac{\sqrt{n^*}}{4}\right) = 0.95.$$

The size of sample $n = \text{ceil}(n^*)$. 5/35