Practice questions

1. In an exam question half of the students scored 5 points, a quarter scored 3 points, and the rest scored no points. You are trying to figure out the average score by sampling two random students (with repetition) and asking for their score.

   (a) What is the PMF of the average score of the two sampled students?

   Solution: The PMF of a student's score X is

x	0	3	5
$\mathrm{P}(X = x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

   The PMF of the sum of the scores of the two sampled students can be found using the convolution formula: $\mathrm{P}(X_1 + X_2 = x) = \sum_{x_1} \mathrm{P}(X_1 = x_1)\,\mathrm{P}(X_2 = x - x_1)$, which yields the following PMF:

x	0	3	5	6	8	10
$\mathrm{P}(X_1 + X_2 = x)$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{4}$

   The PMF of the sample mean $\overline{X} = (X_1 + X_2)/2$ is then obtained by scaling down the value by a factor of two:

x	0	1.5	2.5	3	4	5
$\mathrm{P}(\overline{X} = x)$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{4}$

   (b) What is the probability that the sample mean is equal to the actual mean?

   Solution: The actual mean is $\mathrm{E}[X] = 0 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 5 \cdot \frac{1}{2} = \frac{13}{4}$. This is not a possible value for $\overline{X}$, so the probability of $\overline{X} = \mu$ is zero.

   (c) What is the probability that the sample mean is within one point of the actual mean?

   Solution: The desired probability is

$$\mathrm{P}(\mu - 1 \leq \overline{X} \leq \mu + 1) = \mathrm{P}(9/4 \leq \overline{X} \leq 17/4) = \mathrm{P}(\overline{X} \in \{2.5, 3, 4\}) = \frac{9}{16}.$$

2. Let X_1, X_2, X_3 be independent samples of an Indicator(1/4) random variable. Calculate the PMF of the (a) sample mean (b) sample variance (c) sample standard deviation and (d) sample maximum.

   Solution:

   (a) The sum $X_1 + X_2 + X_3$ is a Binomial$(3, \frac{1}{4})$ random variable, so the sample mean $\overline{X}$ is a Binomial$(3, \frac{1}{4})$ scaled down by a factor of 3:

x	0	$\frac{1}{3}$	$\frac{2}{3}$	1
$\mathrm{P}(\overline{X} = x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$.

   (b) The sample variance V is the derived random variable

$$\begin{aligned} V &= \frac{X_1^2 + X_2^2 + X_3^2}{3} - \Big(\frac{X_1 + X_2 + X_3}{3}\Big)^2 \\ &= \frac{X_1 + X_2 + X_3}{3} - \Big(\frac{X_1 + X_2 + X_3}{3}\Big)^2 \\ &= \overline{X} - \overline{X}^2. \end{aligned}$$

   Here, we used the fact that $X_i^2 = X_i$ for indicator random variables. Its PMF is therefore

v	0	$\frac{2}{9}$
$P(V = v)$	$\frac{7}{16}$	$\frac{9}{16}$

(c) The sample standard deviation is the square root $\sqrt{V}$ of the sample variance with PMF:

s	0	$\frac{\sqrt{2}}{3}$
$P(\sqrt{V} = s)$	$\frac{7}{16}$	$\frac{9}{16}$

(d) The only possible values of the sample max MAX are 0 and 1. The value 0 is taken when all three of the samples are zero, so $P(MAX = 0) = (3/4)^3 = 27/64$ and so $P(MAX = 1) = 1 - P(MAX = 0)$ and the PDF is

m	0	1
$P(MAX = m)$	$\frac{27}{64}$	$\frac{37}{64}$

3. A food processing company packages honey in glass jars. The volume of honey (in millilitres) in a random jar is a Normal$(\mu, 10)$ random variable for some unknown μ.

(a) What is the PDF of the sample mean volume of six random jars?

Solution: Let X_1, X_2, $\cdots$, X_6 be the random variables denoting the volume in the six sampled jars. As they are independent their sum is a Normal$(6\mu, 10\sqrt{6})$ random variable, and so their sample mean is a Normal$(\mu, 10/\sqrt{6})$ random variable. Its PDF is

$$f_{\overline{X}}(x) = \frac{\sqrt{3}}{10\sqrt{\pi}} e^{\frac{-(x-\mu)^2}{100/3}}$$

(b) What is the probability that the sample mean is within 3 millilitres of the true mean μ?

Solution: The random variable $N = (\overline{X} - \mu)/(10/\sqrt{6})$ is Normal$(0, 1)$, so

$$P(-3 \leq \overline{X} - \mu \leq 3) = P\Big(\frac{-3\sqrt{6}}{10} \leq Z \leq \frac{3\sqrt{6}}{10}\Big) \approx 0.5346.$$

4. Take $n = 100$ samples of an Indicator(0.01) random variable. Let $\overline{X}$ be the sample mean.

(a) What is the probability that the sample mean $\overline{X}$ is within 0.005 of the true mean μ?

Solution: The sample mean $\overline{X}$ is 0.01 times a Binomial$(100, 0.01)$ random variable S, so $\overline{X}$ takes value between 0.005 and 0.015 exactly when S takes value 1:

$$P(0.005 \leq \overline{X} \leq 0.015) = P(0.5 \leq S \leq 1.5) = P(S = 1) = 100 \cdot 0.01 \cdot (1 - 0.01)^{99} \approx 0.3697.$$

(b) The Central Limit Theorem says that the event $\mu - \epsilon \leq \overline{X} \leq \mu + \epsilon$ should have similar probability to $-t \leq N \leq t$ for large n, a Normal$(0, 1)$ random variable N, and a suitable choice of t. What is the probability predicted for the event in part (a)?

Solution: The standard deviation σ of an Indicator(0.01) r.v. is $\sqrt{0.01 \cdot (1 - 0.01)} \approx 0.099$, so the standard deviation of $\overline{X}$ is $\sigma/\sqrt{100} \approx 0.0099$. The event $0.005 \leq \overline{X} \leq 0.015$ is that of $\overline{X}$ being within about half a standard deviation from its mean, so the Central Limit Theorem predicts a probability of

$$P(0.005 \leq \overline{X} \leq 0.015) \approx P(-0.5 \leq N \leq 0.5) \approx 0.383.$$

(c) How do the answers change if we lower the sampling error from 0.005 to 0.001? **Solution:**

The answer in part (a) does not change. The answer in part (b) becomes $P(-0.001 \leq N \leq 0.001) \approx 0.079$, a much worse estimate.