Quiz

When a fair coin is flipped 10 times, what is the chance of getting 7 or more heads? Answer in three ways.

(a) Exactly, using binomial distribution.

(b) Approximately, using the normal distribution.

(c) Approximately, using the normal approximation with continuity correction.

Note that the normal approximation with continuity correction can be an excellent approximation, even for n as small as 10.

Solution:

Let X be a number of heads after 10 throws of a fair coin. $X \sim \text{Bin}\left(10, \frac{1}{2}\right)$ with

$$\mathsf{E}(X) = 10 \cdot \frac{1}{2} = 5 \qquad \text{and} \qquad \mathsf{V}(X) = 10 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{2}.$$

(a) Exact probability:

$$\mathsf{P}(X \geq 7) = \sum_{k=7}^{10} \mathsf{P}(X = k) = \sum_{k=7}^{10} C_{10}^{k} \left(\frac{1}{2}\right)^{k} \left(\frac{1}{2}\right)^{10-k} = \left(\frac{1}{2}\right)^{10} \sum_{k=7}^{10} C_{10}^{k} = \boxed{\frac{11}{64} \approx 0.172}.$$

(b) According to CLT:

$$X \sim \text{Bin}\left(10, \frac{1}{2}\right) \overset{d}{\approx} X_{\text{CLT}} \sim \mathcal{N}\left(5, \frac{5}{2}\right).$$

Using standardization:

$$\mathsf{P}(X \geq 7) \approx \mathsf{P}(X_{\text{CLT}} \geq 7) = \mathsf{P}\left(Z \geq \frac{7-5}{\sqrt{5/2}}\right) \approx 1 - \Phi(1.265) \approx \boxed{0.103}.$$

(c) With continuity correction (point $X = 7$ is included):

$$\mathsf{P}(X \geq 7) \approx \mathsf{P}(X_{\text{CLT}} \geq 6.5) = \mathsf{P}\left(Z \geq \frac{6.5-5}{\sqrt{5/2}}\right) \approx 1 - \Phi(0.949) \approx \boxed{0.171}.$$

Problem 1

Let $X_1, \dots, X_8$ be i.i.d. from $\mathcal{N}(0, 16)$, $Y_1, \dots, Y_5$ be i.i.d. from $\mathcal{N}(1, 9)$, all X_i and Y_j are independent.

(a) Find $\mathsf{P}\left(\sum\limits_{i=1}^{8} X_i^2 < 20\right)$.

(b) Find $\mathsf{P}\left(\sum\limits_{j=1}^{5} (Y_j - 1)^2 > 12\right)$.

(c) Could you find so easily $\mathsf{P}\left(\sum\limits_{i=1}^{8} X_i^2 + \sum\limits_{j=1}^{5} (Y_j - 1)^2 > 32\right)$?

Solution:

Problems to find if the sum of squares of independent normal distributions is greater than some constant value can be solved via χ^2-distribution and its tables.

If $Z_i \sim \mathcal{N}(0, 1^2)$ are variables with independent standard normal distributions $\forall i \in \{1, k\}$, then:

$$Q = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$$

where k is degree of the χ^2-distribution with $\mathsf{E}(Q) = k$ and $\mathsf{V}(Q) = 2k$. Since Q is composed of sum of squares, its support and range are non-negative. From the definition of the χ^2-distribution we can derive its important property: if variables $Q_1, Q_2, \dots, Q_m$ are independent and distributed as $\chi_{k_1}^2, \chi_{k_2}^2, \dots \chi_{k_m}^2$ respectively, then:

$$Q_1 + Q_2 + \ldots + Q_m \sim \chi^2_{k_1 + k_2 + \ldots k_m}.$$

(a) To use the definition χ^2-distributions in the sum of X_i^2, we have to standardize them:

$$\mathsf{P}\left(\sum_{i=1}^{8} X_i^2 < 20\right) = \mathsf{P}\left(4^2 \cdot \sum_{i=1}^{8} \left(\frac{X_i - 0}{4}\right)^2 < 20\right) = \mathsf{P}\left(\sum_{i=1}^{8} Z_i^2 < \frac{20}{4^2}\right) =$$
$$= \mathsf{P}\left(\chi_8^2 < 1.25\right) = 1 - \mathsf{P}\left(\chi_8^2 > 1.25\right).$$

We are using probability $\mathsf{P}\left(\chi_8^2 > 1.25\right)$ since tables of χ^2-distributions use values of the distribution's right tail. Also, those tables have limited amount of different probabilities, so we will point out the interval of probabilities, to which the chosen value belongs to, with highlighting in **bold** the closest one.

$$\mathsf{P}\left(\sum_{i=1}^{8} X_i^2 < 20\right) \in 1 - (\mathbf{1}, 0.99) = \boxed{(\mathbf{0}, 0.01)}.$$

Simulation in Python gives approximate result $\boxed{0.0039}$, which fully corresponds to the previously obtained interval.

(b) Using standardization of Y_j and using tables of χ^2-distribution as in previous task:

$$\mathsf{P}\left(\sum_{j=1}^{5}(Y_j-1)^2>12\right)=\mathsf{P}\left(3^2\cdot\sum_{j=1}^{5}\left(\frac{Y_i-1}{3}\right)^2>12\right)=\mathsf{P}\left(\sum_{j=1}^{5}Z_j^2>\frac{12}{3^2}\right)=$$
$$=\mathsf{P}\left(\chi_5^2>1.33\right)\in\boxed{(0.9,\mathbf{0.95})}.$$

Simulation in Python gives approximate result $\boxed{0.9325}$.

(c) Firstly, let's convert probability to expression, containing χ^2-variables:

$$\mathsf{P}\left(\sum_{i=1}^{8}X_i^2+\sum_{j=1}^{5}(Y_j-1)^2>32\right)=\mathsf{P}\left(4^2\cdot\sum_{i=1}^{8}Z_i^2+3^2\sum_{j=1}^{5}Z_j^2>32\right)=$$
$$=\mathsf{P}\left(16\chi_8^2+9\chi_5^2>32\right).$$

There is no way to represent this expression in the form with single χ^2, since we can't add up terms $16\chi_8^2$ and $9\chi_5^2$ without violating independence of original variables. That's why we can only estimate the lower bound of the probability – probability that the sum of 2 variables is greater than some value is larger than the probability that either of these variables is greater than same value:

$$\mathsf{P}\left(16\chi_8^2+9\chi_5^2>32\right)>\mathsf{P}\left(16\chi_8^2>32\cup 9\chi_5^2>32\right)=1-\mathsf{P}\left(16\chi_8^2<32\cap 9\chi_5^2<32\right).$$

Since variables of $16\chi_8^2$ and $9\chi_5^2$ are independent:

$$\mathsf{P}\left(16\chi_8^2+9\chi_5^2>32\right)>1-\mathsf{P}\left(16\chi_8^2<32\right)\cdot\mathsf{P}\left(9\chi_5^2<32\right).$$

Using tables of χ^2-distributions:

$$\mathsf{P}\left(16\chi_8^2<32\right)=1-\mathsf{P}\left(\chi_8^2>2\right)\in 1-(0.99,\mathbf{0.975})=(0.01,\mathbf{0.025}).$$
$$\mathsf{P}\left(9\chi_5^2<32\right)=1-\mathsf{P}\left(\chi_5^2>3.56\right)\in 1-(0.75,\mathbf{0.5})=(0.25,\mathbf{0.5}).$$

Substituting and changing borders due to subtraction:

$$\mathsf{P}\left(16\chi_8^2+9\chi_5^2>32\right)>1-(0.025,0.01)\cdot(0.5,0.25)=(0.9875,0.9975).$$

Since we have inequality, we will take the lower bound to guarantee the correctness for all possible values:

$$\mathsf{P}\left(16\chi_8^2+9\chi_5^2>32\right)\in\boxed{(0.9875,1)}.$$

Simulation in Python gives approximate result $\boxed{0.9993}$.

Problem 2

Let $X_1, \dots, X_8$ be i.i.d. from $\mathcal{N}(0, 16)$, $Y_1, \dots, Y_5$ be i.i.d. from $\mathcal{N}(1, 9)$, all X_i and Y_j are independent.

(a) Find $\mathsf{P}\left(X_1 < \sqrt{\sum_{j=1}^{5}(Y_j - 1)^2}\right)$.

(b) Find $\mathsf{P}\left(X_1 + 2X_2 < \sqrt{\sum_{j=1}^{4}(Y_j - 1)^2}\right)$.

(c) Find $\mathsf{P}\left(Y_1 - 1 < \sqrt{\sum_{i=1}^{8} X_i^2}\right)$.

Solution:

In this type of problems where root mean squares of independent normal variables are compared with another normal variable, Student's t-distribution and its tables are used.
If $Z \sim \mathcal{N}\left(0, 1^2\right)$ and $Q \sim \chi^2_k$ are independent random variables then

$$T = \frac{Z}{\sqrt{Q/k}} \sim t_k$$

is a random variable with Student's distribution and degree of freedom k. It is symmetrical, having support over all real number line $\mathbb{R}$, and looks very similar to standard normal distribution, coinciding with the latter in the case of infinite degree of freedom:

$$t_k \xrightarrow[k\to\infty]{d} \mathcal{N}\left(0, 1^2\right).$$

Due to symmetry: $\mathsf{E}(T) = 0,\ k > 1$. t-distribution has heavier tails than Z, which is reflected in its variance: $\mathsf{V}(T) = \dfrac{k}{k-2},\ k > 2$.

(a) We have to standardize X_i and Y_j in order to use definition of χ^2 and t-distributions:

$$\mathsf{P}\left(X_1 < \sqrt{\sum_{j=1}^{5}(Y_j - 1)^2}\right) = \mathsf{P}\left(4 \cdot \frac{X_1 - 0}{4} < 3 \cdot \sqrt{\sum_{j=1}^{5}\left(\frac{Y_j - 1}{3}\right)^2}\right) =$$

$$= \mathsf{P}\left(\frac{Z}{\sqrt{\chi^2_5}} < 0.75\right) = \mathsf{P}\left(\frac{Z}{\sqrt{\chi^2_5/5}} < 0.75 \cdot \sqrt{5}\right) \approx \mathsf{P}(t_5 < 1.677) = 1 - \mathsf{P}(t_5 \geq 1.677).$$

As in the case with χ^2-distribution, tables use values of the right tail with limited amount of different probabilities.

$$\mathsf{P}\left(X_1 < \sqrt{\sum_{j=1}^{5}(Y_j-1)^2}\right) \in 1-(\mathbf{0.1}, 0.05) = \boxed{(\mathbf{0.9}, 0.95)}.$$

Simulation in Python gives approximate result $\boxed{0.9228}$.

(b) Let's find distribution of $X_1 + 2X_2$ first.

$$\mathsf{E}(X_1+2X_2) = \mathsf{E}(X_1) + 2\cdot\mathsf{E}(X_2) = 0 + 2\cdot 0 = 0.$$

$$\mathsf{V}(X_1+2X_2) \overset{\text{ind}}{=} \mathsf{V}(X_1) + \mathsf{V}(2X_2) = \mathsf{V}(X_1) + 2^2\cdot\mathsf{V}(X_2) = 16 + 4\cdot 16 = 80 = (4\sqrt{5})^2.$$

Using standardization of X_i, Y_j and their combination, and using tables of t-distribution as in previous task:

$$\mathsf{P}\left(X_1+2X_2 < \sqrt{\sum_{j=1}^{4}(Y_j-1)^2}\right) = \mathsf{P}\left(4\sqrt{5}\cdot\frac{X_1+2X_2-0}{4\sqrt{5}} < 3\cdot\sqrt{\sum_{j=1}^{4}\left(\frac{Y_j-1}{3}\right)^2}\right) =$$
$$= \mathsf{P}\left(\frac{Z}{\sqrt{\chi^2_4}} < \frac{3}{4\sqrt{5}}\right) = \mathsf{P}\left(\frac{Z}{\sqrt{\chi^2_4/4}} < \frac{3}{2\sqrt{5}}\right) \approx \mathsf{P}(t_4 < 0.671) = 1 - \mathsf{P}(t_4 \geq 0.671) \in$$
$$\in 1-(0.4, \mathbf{0.25}) = \boxed{(0.6, \mathbf{0.75})}.$$

Simulation in Python gives approximate result $\boxed{0.7306}$.

(c) Using standardization of X_i, Y_j and tables of t-distribution:

$$\mathsf{P}\left(Y_1 - 1 < \sqrt{\sum_{i=1}^{8}X_i^2}\right) = \mathsf{P}\left(3\cdot\frac{Y_1-1}{3} < 4\cdot\sqrt{\sum_{i=1}^{8}\left(\frac{X_i-0}{4}\right)^2}\right) =$$
$$= \mathsf{P}\left(\frac{Z}{\sqrt{\chi^2_8}} < \frac{4}{3}\right) = \mathsf{P}\left(\frac{Z}{\sqrt{\chi^2_8/8}} < \frac{8\sqrt{2}}{3}\right) \approx \mathsf{P}(t_8 < 3.771) = 1 - \mathsf{P}(t_8 \geq 3.771) \in$$
$$\in 1-(\mathbf{0.005}, 0.0005) = \boxed{(\mathbf{0.995}, 0.9995)}.$$

Simulation in Python gives approximate result $\boxed{0.9973}$.

Problem 3

Let $X_1, \ldots, X_8$ be i.i.d. from $\mathcal{N}(0, 16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $\mathcal{N}(1, 9)$, all X_i and Y_j are independent.

(a) Find $\mathsf{P}\left(\sum_{i=1}^{8} X_i^2 < \sum_{j=1}^{5} (Y_j - 1)^2\right)$.

(b) Find $\mathsf{P}\left(\sum_{i=1}^{4} X_i^2 < 7\sum_{j=1}^{3} (Y_j - 1)^2\right)$.

(c) Find $\mathsf{P}\left(\sum_{i=1}^{3} X_i^2 < \sum_{i=4}^{8} X_i^2\right)$.

(d) Find $\mathsf{P}\left(\sum_{i=1}^{7} X_i^2 < \sum_{i=3}^{8} X_i^2\right)$.

Solution:

When sums of squares of normal variables are compared, Fisher-Snedecor F-distribution and its tables are used.

If $Q \sim \chi^2_p$ and $R \sim \chi^2_k$ are independent random variables then

$$F = \frac{Q/p}{R/k} \sim F_{p,k}$$

is a random variable with F-distribution and degrees of freedom p and k. Since F is a combination of variables with χ^2-distribution, its support and range are also non-negative. By the composition of F it's clear that

$$\frac{1}{F} \sim F_{k,p} \qquad \text{and} \qquad T^2 \sim F_{1,k}$$

where $T \sim t_k$ is a Student's t-distributed variable. The latter identity sustains the fact that $\mathsf{E}(F) = \dfrac{k}{k-2}$, $k > 2$. The variance of F is not so neat: $\mathsf{V}(F) = \dfrac{2k^2(p+k-2)}{p(k-2)^2(k-4)}$, $k > 4$.

(a) We have to standardize X_i and Y_j in order to use definition of χ^2 and F-distributions:

$$\mathsf{P}\left(\sum_{i=1}^{8} X_i^2 < \sum_{j=1}^{5}(Y_j - 1)^2\right) = \mathsf{P}\left(4^2 \cdot \sum_{i=1}^{8}\left(\frac{X_i - 0}{4}\right)^2 < 3^2 \cdot \sum_{j=1}^{5}\left(\frac{Y_j - 1}{3}\right)^2\right) =$$
$$= \mathsf{P}\left(16\chi^2_8 < 9\chi^2_5\right).$$

Now, since a variable of F-distribution is a ratio of χ^2 variables, we have to decide, which one should be in numerator and which one in denominator.

Let's try χ^2_8 over χ^2_5:

$$\mathsf{P}\left(16\chi^2_8 < 9\chi^2_5\right) = \mathsf{P}\left(\frac{\chi^2_8}{\chi^2_5} < \frac{9}{16}\right) = \mathsf{P}\left(\frac{\chi^2_8/8}{\chi^2_5/5} < \frac{9\cdot 5}{16\cdot 8}\right) \approx \mathsf{P}(F_{8,5} < 0.352) = \\ = 1 - \mathsf{P}(F_{8,5} \geq 0.352).$$

As in the case with χ^2 and t-distributions, tables use values of the right tail with limited amount of different probabilities.

$$\mathsf{P}\left(\sum_{i=1}^{8} X_i^2 < \sum_{j=1}^{5}(Y_j - 1)^2\right) \in 1 - (1, 0.75) = (0, 0.25).$$

The interval we've got turned out to be rather large, since tables of F-distribution are sparse in the area of left tail. We can improve our result, considering values at the right tail. This can be achieved via ratio of χ^2_5 on χ^2_8:

$$\mathsf{P}\left(16\chi^2_8 < 9\chi^2_5\right) = \mathsf{P}\left(\frac{\chi^2_5}{\chi^2_8} > \frac{16}{9}\right) = \mathsf{P}\left(\frac{\chi^2_5/5}{\chi^2_8/8} > \frac{16\cdot 8}{9\cdot 5}\right) \approx \mathsf{P}(F_{5,8} > 2.844).$$

Thus, we should bring an inequality to the form, where the side with number is greater than 1, since tables of F-distribution are usually more fine-grained on the support of right tails.

$$\mathsf{P}\left(\sum_{i=1}^{8} X_i^2 < \sum_{j=1}^{5}(Y_j - 1)^2\right) \in \boxed{(0.05, \mathbf{0.1})}.$$

Simulation in Python gives approximate result $\boxed{0.091}$.

(b) Using standardization of X_i, Y_j and tables of F-distribution:

$$\mathsf{P}\left(\sum_{i=1}^{4} X_i^2 < 7\sum_{j=1}^{3}(Y_j - 1)^2\right) = \mathsf{P}\left(4^2 \cdot \sum_{i=1}^{4}\left(\frac{X_i - 0}{4}\right)^2 < 7\cdot 3^2 \cdot \sum_{j=1}^{3}\left(\frac{Y_j - 1}{3}\right)^2\right) = \\ = \mathsf{P}\left(\frac{\chi^2_4}{\chi^2_3} < \frac{63}{16}\right) = \mathsf{P}\left(\frac{\chi^2_4/4}{\chi^2_3/3} < \frac{63\cdot 3}{16\cdot 4}\right) \approx \mathsf{P}(F_{4,3} < 2.953) = 1 - \mathsf{P}(F_{4,3} \geq 2.953) \in \\ \in 1 - (\mathbf{0.25}, 0.1) = \boxed{(\mathbf{0.75}, 0.9)}.$$

Simulation in Python gives approximate result $\boxed{0.800}$.

(c) Using standardization of X_i, and tables of F-distribution:

$$\mathsf{P}\left(\sum_{i=1}^{3} X_i^2 < \sum_{i=4}^{8} X_i^2\right) = \mathsf{P}\left(4^2 \cdot \sum_{i=1}^{3}\left(\frac{X_i - 0}{4}\right)^2 < 4^2 \cdot \sum_{i=4}^{8}\left(\frac{X_i - 0}{4}\right)^2\right) =$$
$$= \mathsf{P}\left(\frac{\chi_3^2}{\chi_5^2} < 1\right) = \mathsf{P}\left(\frac{\chi_3^2/3}{\chi_5^2/5} < \frac{5}{3}\right) \approx \mathsf{P}(F_{3,5} < 1.667) = 1 - \mathsf{P}(F_{3,5} \geq 1.667) \in$$
$$\in 1 - (0.5, \mathbf{0.25}) = \boxed{(0.5, \mathbf{0.75})}.$$

Simulation in Python gives approximate result $\boxed{0.712}$.

(d) Eliminating repeating elements:

$$\mathsf{P}\left(\sum_{i=1}^{7} X_i^2 < \sum_{i=3}^{8} X_i^2\right) = \mathsf{P}\left(X_1^2 + X_2^2 < X_8^2\right).$$

Using standardization of X_i, and tables of F-distribution:

$$\mathsf{P}\left(X_1^2 + X_2^2 < X_8^2\right) = \mathsf{P}\left(4^2 \cdot \sum_{i=1}^{2}\left(\frac{X_i - 0}{4}\right)^2 < 4^2 \cdot \left(\frac{X_8 - 0}{4}\right)^2\right) =$$
$$= \mathsf{P}\left(\frac{\chi_1^2}{\chi_2^2} > 1\right) = \mathsf{P}\left(\frac{\chi_1^2/1}{\chi_2^2/2} > 2\right) \approx \mathsf{P}(F_{1,2} > 2) \in \boxed{(\mathbf{0.25}, 0.5)}.$$

Simulation in Python gives approximate result $\boxed{0.293}$.

Problem 4

Suppose that we plan to take a random sample of size n from a normal distribution with mean μ and standard deviation $\sigma = 2$.

(a) Suppose $\mu = 4$ and $n = 20$.

i) What is the probability that the mean $\overline{X}$ of the sample is greater than 5?

ii) What is the probability that $\overline{X}$ is smaller than 3?

iii) What is $\mathsf{P}\left(\left|\overline{X} - \mu\right| \leq 1\right)$ in this case?

(b) How large should n be in order that $\mathsf{P}\left(\left|\overline{X} - \mu\right| \leq 0.5\right) \geq 0.95$ for every possible value of μ?

(c) It is claimed that the true value of μ is 5 in a population. A random sample of size $n = 100$ is collected from this population, and the mean for this sample is $\overline{x} = 5.8$. Based on the result in (b), what would you conclude from this value of $\overline{X}$?

Solution:

Since the population is distributed normally $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample mean (of sample size n) $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ is also distributed normally, but with parameters:

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

(a) We know that $\sigma = 2, \mu = 4$ and $n = 20$.

i) We need to find probability $\mathsf{P}\left(\overline{X} > 5\right)$. To use standard normal distribution table of c.d.f. $\Phi(z)$, we have to standardize the variable and limits:

$$\mathsf{P}\left(\overline{X} > 5\right) = \mathsf{P}\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} > \frac{5 - \mu}{\sigma/\sqrt{n}}\right) = \mathsf{P}\left(Z > \frac{5 - 4}{2/\sqrt{20}}\right) = \mathsf{P}\left(Z > \sqrt{5}\right)$$
$$= 1 - \mathsf{P}\left(Z < \sqrt{5}\right) \approx 1 - \Phi(2.236) = 1 - 0.9868 = \boxed{0.0132}.$$

ii) Since the mean of $\overline{X}$ is $\mu = 4$, and the distribution is symmetrical around the mean, the probabilities $\mathsf{P}\left(\overline{X} < 3\right)$ and $\mathsf{P}\left(\overline{X} > 5\right)$ are equal:

$$\mathsf{P}\left(\overline{X} < 3\right) = \mathsf{P}\left(\overline{X} > 5\right) \approx \boxed{0.0132}.$$

iii) Since $\mathsf{P}\left(\left|\overline{X}-4\right| \le 1\right) + \mathsf{P}\left(\overline{X} < 3\right) + \mathsf{P}\left(\overline{X} > 5\right) = 1$:

$$\mathsf{P}\left(\left|\overline{X}-4\right| \le 1\right) = 1 - \mathsf{P}\left(\overline{X} < 3\right) - \mathsf{P}\left(\overline{X} > 5\right) = 1 - 2 \cdot 0.0132 = \boxed{0.9736}.$$

(b) Let's rewrite the inequality $\mathsf{P}\left(\left|\overline{X}-\mu\right| \le 0.5\right) \ge 0.95$ in terms of $\Phi(z)$ to find the condition on n from standard distribution table:

$$\mathsf{P}\left(\left|\overline{X}-\mu\right| \le 0.5\right) = \mathsf{P}\left(-0.5 \le \overline{X}-\mu \le 0.5\right) = \mathsf{P}\left(-\frac{0.5}{\sigma/\sqrt{n}} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \frac{0.5}{\sigma/\sqrt{n}}\right) =$$
$$= \mathsf{P}\left(-\frac{0.5}{2/\sqrt{n}} \le Z \le \frac{0.5}{2/\sqrt{n}}\right) = \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right).$$

Using the symmetry of normal distribution $\Phi(-z) = 1 - \Phi(z)$:

$$\mathsf{P}\left(\left|\overline{X}-\mu\right| \le 0.5\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \ge 0.95.$$

$$\Phi\left(\frac{\sqrt{n}}{4}\right) \ge 0.975, \quad \Longrightarrow \quad \frac{\sqrt{n}}{4} \ge 1.96, \quad \Longrightarrow \quad n \ge 61.46.$$

Since the answer is the closest integer, which satisfies the last inequality:

$$n_{\min} = [61.46] + 1 = \boxed{62}.$$

(c) The probability $\mathsf{P}\left(\left|\overline{X}-\mu\right| \le 0.5\right)$ is greater than 0.95 already for $n = 62$, and for sample size $n = 100$ it will be even greater – approximately in the vicinity of 99%.

It means that with probability 99% expression $\left|\overline{X}-\mu\right|$ is less than 0.5. We are told that $\mu = 5$, but the outcome of $\overline{X}$ turned out to be $\overline{x} = 5.8$:

$$|\overline{x}-\mu| = |5.8-5| = 0.8 \not\le 0.5.$$

This result can be interpreted as the following:

1. Either with $< 1\%$ chance it's true and we are just really unlucky.
2. Or the claim on $\mu = 5$ is wittingly false.

The second assumption is much more probable. Thus, we make a conclusion that we were probably lied to.

Problem 5

Suppose $X_1, X_2, \dots, X_{40}$ are i.i.d. random variables with c.d.f.

$$F(x) = \begin{cases} 0, & x < 0, \\ x^3, & 0 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

(a) Find $\mathsf{P}(X_1 > 0.8 \mid X_1 > 0.5)$ and $\mathsf{P}(X_1 > 0.8 \mid X_2 > 0.5)$.

(b) Let the mean be $\overline{X} = \frac{1}{40}\sum_{i=1}^{40} X_i$. Estimate probability $\mathsf{P}\left(\overline{X} > 0.7\right)$.

Solution:

(a) By definition of conditional probability:

$$\mathsf{P}(X_1 > 0.8 \mid X_1 > 0.5) = \frac{\mathsf{P}(X_1 > 0.8 \cap X_1 > 0.5)}{\mathsf{P}(X_1 > 0.5)} = \frac{\mathsf{P}(X_1 > 0.8)}{\mathsf{P}(X_1 > 0.5)} =$$

$$= \frac{1 - \mathsf{P}(X_1 \le 0.8)}{1 - \mathsf{P}(X_1 \le 0.5)} = \frac{1 - F(0.8)}{1 - F(0.5)} = \frac{1 - 0.8^3}{1 - 0.5^3} = \boxed{\frac{488}{875} \approx 0.558}.$$

Since X_1 and X_2 are independent:

$$\mathsf{P}(X_1 > 0.8 \mid X_2 > 0.5) = \mathsf{P}(X_1 > 0.8) = 1 - F(0.8) = 1 - 0.8^3 = \boxed{\frac{61}{125} = 0.488}.$$

(b) According to CLT:

$$\overline{X} \overset{d}{\approx} \overline{X}_{\text{CLT}} \sim \mathcal{N}\left(\mathsf{E}(X_i), \frac{\mathsf{V}(X_i)}{40}\right).$$

In order to find $\mathsf{E}(X_i)$ and $\mathsf{V}(X_i)$ we should derive a p.d.f. of X_i:

$$f(x) = \frac{d}{dx}F(x) = 3x^2 \cdot I_{\{0 \le x \le 1\}}.$$

Moments of X_i then:

$$\mathsf{E}(X_i) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \cdot 3x^2 dx = \left.\frac{3x^4}{4}\right|_0^1 = \frac{3}{4},$$

$$\mathsf{E}\left(X_i^2\right) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 \cdot 3x^2 dx = \left.\frac{3x^5}{5}\right|_0^1 = \frac{3}{5},$$

$$\mathsf{V}(X_i) = \mathsf{E}\left(X_i^2\right) - \mathsf{E}(X_i)^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{80}.$$

Using standardization to standard normal variable and symmetry of $\Phi(z)$:

$$\mathsf{P}\left(\overline{X} > 0.7\right) \approx \mathsf{P}\left(\overline{X}_{\text{CLT}} > 0.7\right) = \mathsf{P}\left(Z > \frac{0.7 - \mathsf{E}(X_i)}{\sqrt{\mathsf{V}(X_i)/40}}\right) = \mathsf{P}\left(Z > \frac{0.7 - 0.75}{\sqrt{3/3200}}\right) \approx$$
$$= \mathsf{P}(Z > -1.633) = 1 - \Phi(-1.633) = \Phi(1.633) \approx \boxed{0.949}.$$

Python simulation gives $\mathsf{P}\left(\overline{X} > 0.7\right) \approx 0.945$.