Quiz

Random variables X and Y have a joint p.d.f.:

$$f(x, y) = (c + 0.1xy) \cdot I_{\{0 \le x, y \le 1\}}.$$

(a) Are X and Y independent?

(b) Find $\mathsf{P}(Y < X/2)$.

Solution:

(a) Let's find the parameter c. Joint p.d.f. $f(x, y)$ should be L^1-normalized to 1 (since p.d.f.-s are always non-negative, we omit absolute value sign):

$$\iint_{-\infty}^{\infty} f(x, y) dy dx = 1,$$

$$\int_0^1 \left(\int_0^1 (c + 0.1xy) dy \right) dx = \int_0^1 \left((cy + 0.05xy^2) \Big|_0^1 \right) dx = \int_0^1 (c + 0.05x)\, dx,$$

$$= (cx + 0.025x^2) \Big|_0^1 = c + 0.025.$$

Thus, $c = 1 - 0.025 = 0.975$.

Marginal distribution $f_X(x)$ is already calculated as a first integration in normalization line:

$$f_X(x) = \int_0^1 (0.975 + 0.1xy) dy = (0.975 + 0.05x) \cdot I_{\{0 \le x \le 1\}}.$$

Due to symmetry of $f(x, y)$ around x and y:

$$f_Y(y) = (0.975 + 0.05y) \cdot I_{\{0 \le y \le 1\}}.$$

X and Y are independent if $f(x, y) = f_X(x) \cdot f_Y(y)$. Clearly:

$$(0.975 + 0.1xy) \cdot I_{\{0 \le x, y \le 1\}} \ne (0.975 + 0.05x) \cdot (0.975 + 0.05y) \cdot I_{\{0 \le x, y \le 1\}}.$$

Which means that X and Y are $\boxed{\text{NOT independent}}$.

(b) Integrating by y internally first, and then by x externally:

$$
\begin{aligned}
\mathsf{P}(Y < X/2) &= \int_0^1 \left(\int_0^{x/2} (0.975 + 0.1xy)dy \right) dx = \\
&= \int_0^1 \left(\int_0^{x/2} (0.975y + 0.05xy^2)\Big|_0^{x/2} \right) dx = \\
&= \int_0^1 \left(\frac{0.975}{2}x + \frac{0.05}{4}x^3 \right) dx = \\
&= \left(\frac{0.975}{4}x^2 + \frac{0.05}{16}x^4 \right) \Big|_0^1 = \\
&= \frac{0.975}{4} + \frac{0.05}{16} = \boxed{\frac{79}{320} \approx 0.2469}.
\end{aligned}
$$

Problem 1

Cars independently pass a point on a busy road at an average rate of 150 per hour.

(a) Assuming a Poisson distribution, find the probability that none passes in a given minute.

(b) What is the expected number passing in two minutes?

(c) Find the probability that the expected number actually passes in a given two-minute period.

Other motor vehicles (vans, motorcycles etc.) pass the same point independently at the rate of 75 per hour. Assume a Poisson distribution for these vehicles too.

(d) What is the probability of one car and one other motor vehicle in a two-minute period?

Solution:

P.m.f. of Poisson random variable $X \sim \text{Poisson}(\lambda)$:

$$\mathsf{P}(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}.$$

Let $X \sim \text{Poisson}(150)$ be a number of cars' passes in 1 hour.

(a) Let Y be a number of cars' passes in 1 minute. Its distribution:

$$Y \sim \text{Poisson}\left(\frac{150 \text{ passes}}{\text{hour}} \cdot \frac{1}{60} \text{ hour}\right) = \text{Poisson}\left(\frac{150}{60}\right) = \text{Poisson}\,(2.5)\,.$$

Probability of none passes in a given minute:

$$\mathsf{P}(Y = 0) = \frac{e^{-2.5} \cdot 2.5^0}{0!} = e^{-2.5} \approx \boxed{0.082}.$$

(b) Let Z be a number of car passes in 2 minutes. Its distribution:

$$Z \sim \text{Poisson}\left(\frac{2.5 \text{ passes}}{\text{minute}} \cdot 2 \text{ minutes}\right) = \text{Poisson}\,(5)\,.$$

Expected number of Z then:

$$\mathsf{E}(Z) = \boxed{5}.$$

(c) Probability that 5 cars pass in given 2 minutes:

$$\mathsf{P}(Z = \mathsf{E}(Z)) = \mathsf{P}(Z = 5) = \frac{e^{-5} \cdot 5^5}{5!} \approx \boxed{0.175}.$$

(d) Let W be a number of motor vehicles' passes in 2 minutes. Its distribution:

$$W \sim \text{Poisson}\left(\frac{75 \text{ passes}}{\text{hour}} \cdot \frac{2}{60} \text{ hour}\right) = \text{Poisson}\left(\frac{150}{60}\right) = \text{Poisson}\,(2.5)\,.$$

We need to find probability $\mathsf{P}(Z = 1 \cap W = 1)$. Using independence:

$$\mathsf{P}(Z = 1 \cap W = 1) \overset{\text{ind}}{=} \mathsf{P}(Z = 1) \cdot \mathsf{P}(W = 1) = \frac{e^{-5} \cdot 5^1}{1!} \cdot \frac{e^{-2.5} \cdot 2.5^1}{1!} = 12.5 \cdot e^{-7.5} \approx \boxed{0.007}.$$

Problem 2

Telephone calls enter the hotel switchboard according to a Poisson process on the average of two every 3 minutes.

(a) Calculate the probability, that there will be exactly one call in 3 minutes.

(b) Calculate the probability, that there will be no more than one call in 3 minutes.

Let Y be the waiting time for the second call.

(c) Find the probability $\mathsf{P}(Y > 3)$.

Solution:

Let $X \sim \text{Poisson}(2)$ be a number of calls in 3 minutes.

(a) Probability of one call:

$$\mathsf{P}(X = 1) = \frac{e^{-2} \cdot 2^1}{1!} = 2e^{-2} \approx \boxed{0.2707}.$$

(b) Probability of no more than one call:

$$\mathsf{P}(X \leq 1) = \mathsf{P}(X = 0) + \mathsf{P}(X = 1) = \frac{e^{-2} \cdot 2^0}{0!} + 2e^{-2} = 3e^{-2} \approx \boxed{0.4060}.$$

(c) Probability that we have to wait more than 3 seconds for the second call is equivalent to probability of zero calls in 3 minutes (first call appears after given 3 minutes) plus probability of one call in 3 minutes (second call appears after given 3 minutes):

$$\mathsf{P}(Y > 3) \equiv \mathsf{P}(X \leq 1) = 3e^{-2} \approx \boxed{0.4060}.$$

Problem 3

(a) Let $X_1, X_2, \dots X_n$ be i.i.d. exponential random variables with parameter λ. Prove that $\min\{X_1, X_2, \dots X_n\}$ has exponential distribution.

(b) Prove the *memoryless property* of an exponential random variable X:

$$\mathsf{P}(X > s + t \mid X > s) = \mathsf{P}(X > t) \qquad \forall t, s > 0.$$

(c) When you enter the bank, you find that all three tellers are busy serving other customers, and there are no other customers in queue. Assume that the service times for you and for each of the customers being served are independent identically distributed exponential random variables. What is the probability that you will be the last to leave among the four customers?

Solution:

(a) C.d.f. of $X_i \sim \text{Exp}(\lambda)$ for $t \geq 0$ is:

$$\mathsf{P}(X_i \leq t) = 1 - e^{-\lambda t}.$$

Survival function of X_i for $t \geq 0$ is:

$$\mathsf{P}(X_i > t) = 1 - \mathsf{P}(X_i \leq t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

Let $L = \min\{X_1, X_2, \dots X_n\}$. If L is greater than some arbitrary value t, then all X_i must also be greater than t. Using independence of X_i:

$$\mathsf{P}(L > t) = \mathsf{P}\left(\bigcap_{i=1}^{n}(X_i > t)\right) \overset{\text{ind}}{=} \prod_{i=1}^{n}\mathsf{P}(X_i > t) = (e^{-\lambda t})^n = e^{-\lambda n t}.$$

For $t \geq 0$ c.d.f. of L is:

$$\mathsf{P}(L \leq t) = 1 - e^{-\lambda n t}.$$

It means that

$$\boxed{\min\{X_1, X_2, \dots X_n\} \sim \text{Exp}(\lambda n)}.$$

(b) Using definition of conditional probability (λ is arbitrary):

$$\mathsf{P}(X > s + t \mid X > s) = \frac{\mathsf{P}(X > s + t \cap X > s)}{\mathsf{P}(X > s)} = \frac{\mathsf{P}(X > s + t)}{\mathsf{P}(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} =$$
$$= e^{-\lambda(s+t) + \lambda s} = e^{-\lambda t} = \mathsf{P}(X > t).$$

$\boxed{\text{Q.E.D.}}$

(c) First of all, according to memoryless property of exponential distribution, we can choose any point in time to start Poisson process, probability to get an event won't change.

Let's choose as a reference point the moment, when the first customer finishes service at tellers.

In this situation there are 3 customers (including us), who expect tellers' service t finish. Service finish is a Poisson process, so we need to calculate probability, that one waiting time is greater than others'.

This problem is symmetrical in relation to 3 customers, which means that those probabilities for all of them are equal. Thus, it's $\boxed{\frac{1}{3}}$.

Problem 4

X_1 and X_2 are independent exponential random variables with rate λ. Let $Y = X_1 + X_2$. What is the p.d.f. of Y?

Use this result to compute probability from Problem 2(c).

Solution:

Both original variables are distributed as $X \sim \text{Exp}(\lambda)$ with p.d.f.:

$$f_X(x) = \lambda e^{-\lambda x} \cdot I_{\{x \geq 0\}}.$$

Using the convolution of probabilities for $y \geq 0$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1 = \int_0^y \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda(y - x_1)} \cdot I_{\{y \geq 0\}} dx_1.$$

We take limits of x_1 from 0 to y here, since for $y \geq 0$, p.d.f. $f_{X_2}(y - x_1)$ makes sense only for such x_1: $y - x_1 \geq 0 \Rightarrow x_1 \leq y$.

$$f_Y(y) = \lambda^2 e^{-\lambda y} \cdot I_{\{y \geq 0\}} \int_0^y dx_1 = \boxed{\lambda^2 y e^{-\lambda y} \cdot I_{\{y \geq 0\}}}.$$

In Problem 2(c) we had a Poisson process with rate of 2 events per 3 minutes. Declaring time unit as 1 minute, $\lambda = \frac{2}{3}$ for Y and:

$$f_Y(y) = \frac{4}{9} y e^{-\frac{2}{3}y} \cdot I_{\{y \geq 0\}}.$$

Finding requested probability:

$$\mathsf{P}(Y > 3) = \int_3^{\infty} \frac{4}{9} y e^{-\frac{2}{3}y} dy = \frac{4}{9} \left[-\frac{3}{2} y e^{-\frac{2}{3}y} \Big|_3^{\infty} + \int_3^{\infty} \frac{3}{2} e^{-\frac{2}{3}y} dy \right] = \frac{4}{9} \left[\frac{9}{2} e^{-2} - \frac{9}{4} e^{-\frac{2}{3}y} \Big|_3^{\infty} \right] = \boxed{3e^{-2}}.$$

The answer completely coincides with that in Problem 2(c).

Problem 5

Let X and Y be independent Poisson random variables with means λ and μ respectively. Find p.m.f. of $Z = X + Y$.

Solution:

Original variables are $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ with p.m.f.-s:

$$\mathsf{P}(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \qquad \text{and} \qquad \mathsf{P}(Y = y) = \frac{e^{-\mu} \cdot \mu^y}{y!}.$$

Using the convolution of probabilities for $z \geq 0$:

$$\mathsf{P}(Z = z) = \sum_{y=0}^{\infty} \mathsf{P}(Y = y) \cdot \mathsf{P}(X = z - y) = \sum_{y=0}^{z} \frac{e^{-y} \cdot \lambda^y}{y!} \cdot \frac{e^{-\mu} \cdot \mu^{z-y}}{(z-y)!}.$$

We take limits of y from 0 to z here, since for $z \geq 0$, probability $\mathsf{P}(X = z - y)$ makes sense only for such y: $z - y \geq 0 \Rightarrow y \leq z$.
Using binomial theorem:

$$\mathsf{P}(Z = z) = e^{-(\lambda+\mu)} \sum_{y=0}^{z} \frac{\lambda^y \mu^{z-y}}{y!(z-y)!} \cdot \frac{z!}{z!} = \frac{e^{-(\lambda+\mu)}}{z!} \sum_{y=0}^{z} C_z^y \lambda^y \mu^{z-y} = \boxed{\frac{e^{-(\lambda+\mu)} \cdot (\lambda+\mu)^z}{z!}}.$$

Turns out that the sum of independent Poisson variables is also Poisson:

$$Z \sim \text{Poisson}(\lambda + \mu).$$

Problem 6

Let X and Y be independent random variables with uniform distribution on $[0, 2]$ each. Find p.d.f. and c.d.f. of $Z = X + Y$.

Solution:

Both original variables are distributed as $X, Y \sim \mathcal{U}(0, 2)$ with p.d.f.:

$$f_X(x) = f_Y(x) = \frac{1}{2} \cdot I_{\{0 \le x \le 2\}}.$$

Using the convolution of probabilities for $z \in [0, 2]$:

$$f_Z^{(1)}(z) = \int\limits_{-\infty}^{\infty} f_X(y) f_Y(z-y) dy = \int\limits_{0}^{z} \frac{1}{2} \cdot \frac{1}{2} \cdot I_{\{0 \le z \le 2\}} dy.$$

We take limits of y from 0 to z here, since for $z \in [0, 2]$, p.d.f. $f_X(z - y)$ makes sense only for such y: $z - y \ge 0 \Rightarrow y \le z$.

$$f_Z^{(1)}(z) = \int\limits_{0}^{z} \frac{1}{4} \cdot I_{\{0 \le z \le 2\}} dy = \frac{z}{4} \cdot I_{\{0 \le z \le 2\}}.$$

Using the convolution of probabilities for $z \in (2, 4]$:

$$f_Z^{(2)}(z) = \int\limits_{-\infty}^{\infty} f_X(y) f_Y(z-y) dy = \int\limits_{z-2}^{2} \frac{1}{2} \cdot \frac{1}{2} \cdot I_{\{2 \le z \le 4\}} dy.$$

We take limits of y from 0 to z here, since for $z \in (2, 4]$, p.d.f. $f_X(z - y)$ makes sense only for such y: $z - y \le 2 \Rightarrow y \ge z - 2$.

$$f_Z^{(2)}(z) = \int\limits_{z-2}^{2} \frac{1}{4} \cdot I_{\{2 < z \le 4\}} dy = \left(1 - \frac{z}{4}\right) \cdot I_{\{2 < z \le 4\}}.$$

Unifying $f_Z^{(1)}(z)$ and $f_Z^{(2)}(z)$ into one p.d.f.:

$$f_Z(z) = f_Z^{(1)}(z) + f_Z^{(2)}(z) = \boxed{\frac{z}{4} \cdot I_{\{0 \le z \le 2\}} + \left(1 - \frac{z}{4}\right) \cdot I_{\{2 < z \le 4\}}}.$$

C.d.f. of Z for $z \in [0, 2]$ by definition:

$$F_Z^{(1)}(z) = \int\limits_{-\infty}^{z} f_Z^{(1)}(\zeta) d\zeta = \int\limits_{0}^{z} \frac{\zeta}{4} d\zeta = \frac{z^2}{8}.$$

C.d.f. of Z for $z \in (2, 4]$ by definition (with joint at previous interval):

$$F_Z^{(2)}(z) = F_Z^{(1)}(2) + \int\limits_{2}^{z} f_Z^{(2)}(\zeta) d\zeta = \frac{1}{2} + \int\limits_{2}^{z} \left(1 - \frac{\zeta}{4}\right) d\zeta = -\frac{z^2}{8} + z - 1.$$

The whole p.d.f. then:

$$F_Z(z) = \begin{cases} 0, & z < 0, \\ \dfrac{z^2}{8}, & 0 \le z \le 2, \\ -\dfrac{z^2}{8} + z - 1, & 2 < z \le 4, \\ 1, & z > 4. \end{cases}$$

Problem 7

Let $X_1, X_2, \dots X_n$ be independent random variables with X_i having an $\text{Exp}(\lambda_i)$ distribution. Find the probability that X_i is the minimum.

Solution:

Survival function ($1-$ c.d.f.) and p.d.f. of $X_i \sim \text{Exp}(\lambda_i)$:

$$\mathsf{P}(X_i > t) = \begin{cases} 1, & t < 0, \\ e^{-\lambda_i t}, & t \geq 0, \end{cases}$$

$$f_{X_i}(t) = \lambda_i e^{-\lambda_i t} \cdot I_{\{t \geq 0\}}.$$

Using marginalization:

$$\mathsf{P}(X_i \text{ is min}) = \mathsf{P}(X_i < X_j \quad \forall j \neq i) = \int\limits_{-\infty}^{\infty} \mathsf{P}(X_i < X_j \quad \forall j \neq i \mid X_i = t) f_{X_i}(t) dt =$$

$$= \int\limits_{0}^{\infty} \mathsf{P}\left(\bigcap_{j \neq i}^{n} (X_j > t)\right) \lambda_i e^{-\lambda_i t} dt \overset{\text{ind}}{=} \int\limits_{0}^{\infty} \prod_{j \neq i}^{n} \mathsf{P}\left(X_j > t\right) \lambda_i e^{-\lambda_i t} dt.$$

Since limits of the integration are $t = 0$ and $t = \infty$, survival function is $e^{-\lambda_j t}$:

$$\mathsf{P}(X_i \text{ is min}) = \int\limits_{0}^{\infty} \prod_{j \neq i}^{n} e^{-\lambda_j t} \cdot \lambda_i e^{-\lambda_i t} dt = \lambda_i \int\limits_{0}^{\infty} e^{-t \sum\limits_{j=1}^{n} \lambda_j} dt =$$

$$= \lambda_i \left(-\frac{e^{-t \sum\limits_{j=1}^{n} \lambda_j}}{\sum\limits_{j=1}^{n} \lambda_j} \Bigg|_0^{\infty} \right) = \boxed{\frac{\lambda_i}{\lambda_1 + \ldots + \lambda_n}}.$$

Problem 8

Solve Problem 3(c) using marginalization technique.

Solution:

Let $X_i \sim \text{Exp}(\lambda)$ be a waiting time for the end of service, starting from the leave of the first customer (we use memoryless property here). Let X_1 be your waiting time, and X_2 and X_3 be waiting times of other 2 customers.

It means that we need to find $\mathsf{P}(X_1 > X_2 \cap X_1 > X_3)$.

Using marginalization:

$$\mathsf{P}(X_1 > X_2 \cap X_1 > X_3) = \int\limits_{-\infty}^{\infty} \mathsf{P}(X_1 > X_2 \cap X_1 > X_3 \mid X_1 = t) f_{X_1}(t) dt =$$

$$= \int\limits_{0}^{\infty} \mathsf{P}\left(X_2 < t \cap X_3 < t\right) \lambda e^{-\lambda t} dt \overset{\text{ind}}{=} \int\limits_{0}^{\infty} \mathsf{P}(X_2 < t) \mathsf{P}(X_3 < t) \lambda e^{-\lambda t} dt =$$

$$= \int\limits_{0}^{\infty} \left(1 - e^{-\lambda t}\right)^2 \lambda e^{-\lambda t} dt = \lambda \int\limits_{0}^{\infty} \left(1 - 2e^{-\lambda t} + e^{-2\lambda t}\right) e^{-\lambda t} dt =$$

$$= \lambda \int\limits_{0}^{\infty} \left(e^{-\lambda t} - 2e^{-2\lambda t} + e^{-3\lambda t}\right) dt = -e^{-\lambda t}\Big|_0^{\infty} + e^{-2\lambda t}\Big|_0^{\infty} - \frac{1}{3} e^{-3\lambda t}\Big|_0^{\infty} = \boxed{\frac{1}{3}}.$$

Problem 9

A fair coin is flipped 1000 times. Let X be the number of heads, turned up in the experiment.

(a) Find $\mathsf{P}(X \geq 1)$.

(b) Find $\mathsf{P}(X \geq 10)$.

(c) Find $\mathsf{P}(467 < X \leq 513)$.

Solution:

X has a binomial distribution with number of experiments $n = 1000$ and probability of success $p = \dfrac{1}{2}$: $X \sim \text{Bin}\left(1000, \dfrac{1}{2}\right)$. P.m.f. of X is following:

$$\mathsf{P}\left(X = x\right) = C_{1000}^{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{1000-x} = C_{1000}^{x}\left(\frac{1}{2}\right)^{1000}.$$

(a) Since $X \in \{0, 1000\}$:

$$\mathsf{P}(X \geq 1) = 1 - \mathsf{P}(X = 0) = \boxed{1 - \left(\frac{1}{2}\right)^{1000} \approx 1 - 9.3 \cdot 10^{-302}}.$$

(b) Calculating 10 values manually is already frustrating:

$$\mathsf{P}(X \geq 10) = 1 - \sum_{x=0}^{9} \mathsf{P}(X = x) = \boxed{1 - \left(\frac{1}{2}\right)^{1000} \cdot \sum_{x=0}^{9} C_{1000}^{x} \approx 1 - 2.5 \cdot 10^{-280}}.$$

(c) Calculating 46 values seems to be impossible to do manually:

$$\mathsf{P}(467 < X \leq 513) = \sum_{x=468}^{513} \mathsf{P}(X = x).$$

Using De Moivre–Laplace theorem, let's approximate binomial distribution with normal. Expected value of X is $\mathsf{E}(X) = 1000 \cdot \dfrac{1}{2} = 500$ and variance of X is $\mathsf{V}(X) = 1000 \cdot \dfrac{1}{2} \cdot \dfrac{1}{2} = 250$, which means that:

$$X \sim \text{Bin}\left(1000, \frac{1}{2}\right) \overset{d}{\approx} X_{\text{CLT}} \sim \mathcal{N}(500, 250).$$

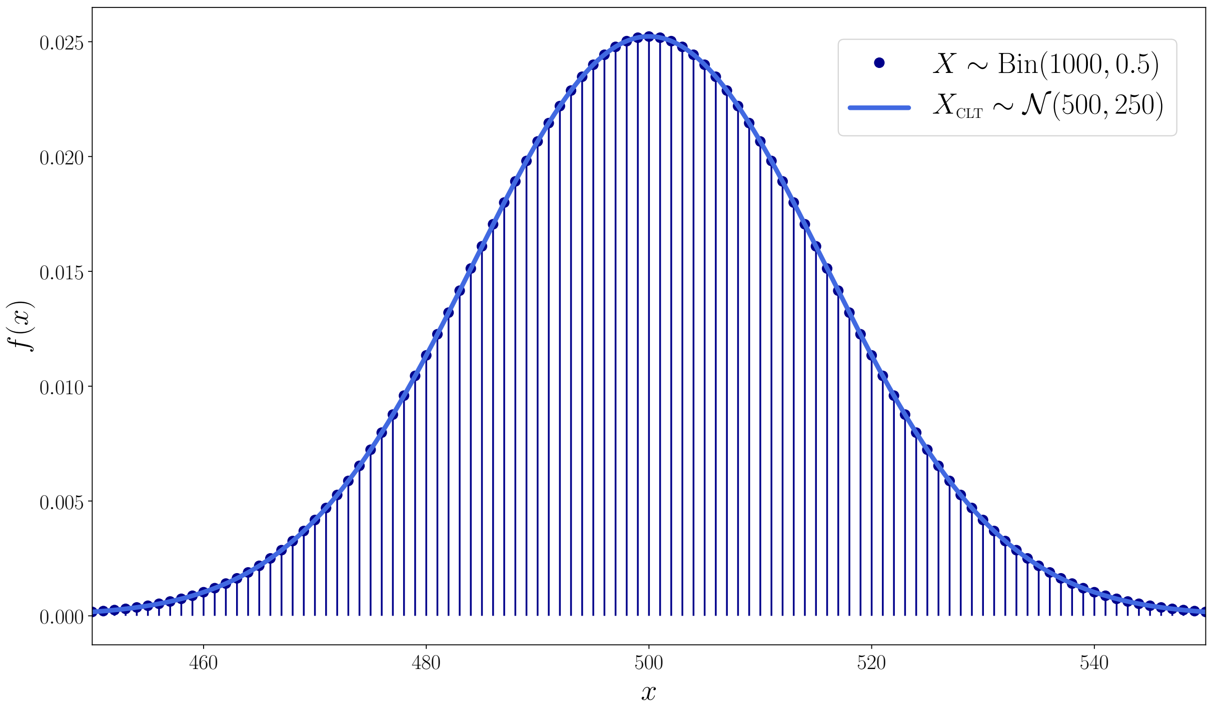

Figure 1: Approximation of $X \sim \mathrm{Bin}\left(1000, \frac{1}{2}\right)$ with $X_{\text{CLT}} \sim \mathcal{N}(500, 250)$.

This approximation is good since $np = 500 > 5$ and $n(1-p) = 500 > 5$, and it is shown in fig. 1.

Let's find probability $\mathsf{P}(467 < X_{\text{CLT}} \leq 513)$ as shown in fig. 2. Probabilities of X are expanded from points into rectangles with width 1 and respective height.

Using standardization, difference of c.d.f.s, symmetry of $Z \sim \mathcal{N}(0,1)$ and standard normal distribution table:

$$\begin{aligned}
\mathsf{P}(467 < X_{\text{CLT}} \leq 513) &= \mathsf{P}\left(\frac{467-500}{\sqrt{250}} < \frac{X_{\text{CLT}}-500}{\sqrt{250}} \leq \frac{513-500}{\sqrt{250}}\right) = \\
&= \mathsf{P}(-2.09 < Z \leq 0.82) = \Phi(0.82) - \Phi(-2.09) = \\
&= \Phi(0.82) - 1 + \Phi(2.09) \approx 0.794 - 1 + 0.982 = 0.776.
\end{aligned}$$

But we forgot about one important moment – variable X is discrete and we should take into account that limit 467 is excluded and limit 513 is included. Thus, in our previous calculation we did not distinguish between inequalities $467 < X \leq 513$, $467 \leq X \leq 513$, $467 \leq X < 513$ and $467 < X < 513$.

To include 513 and exclude 467, we conduct a continuity correction, shown in fig. 3. The whole rectangle, responsible for $X = 467$ is erased, and rectangle, responsible for $X = 513$ is marked. Thus, limits are shifted from $(467, 513)$ to $(467.5, 513.5)$.

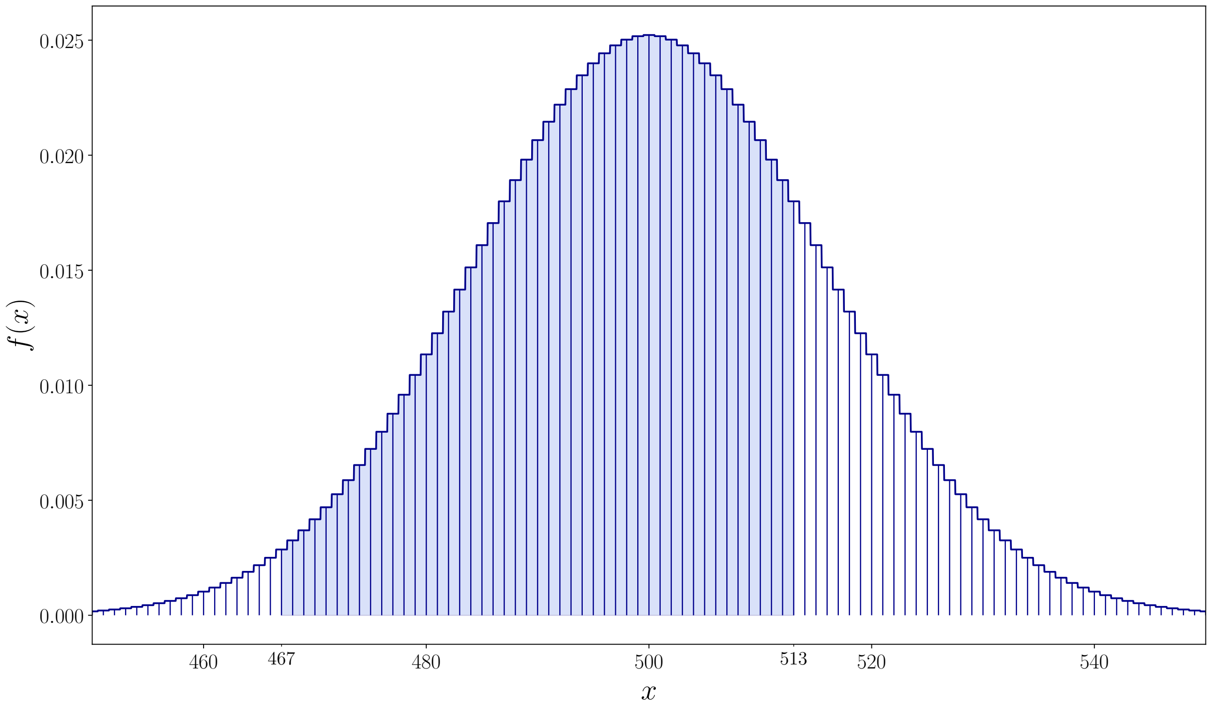

Figure 2: Pseudo-binomial p.d.f. to be approximated with $\mathsf{P}(467 < X_{\text{CLT}} \leq 513)$.

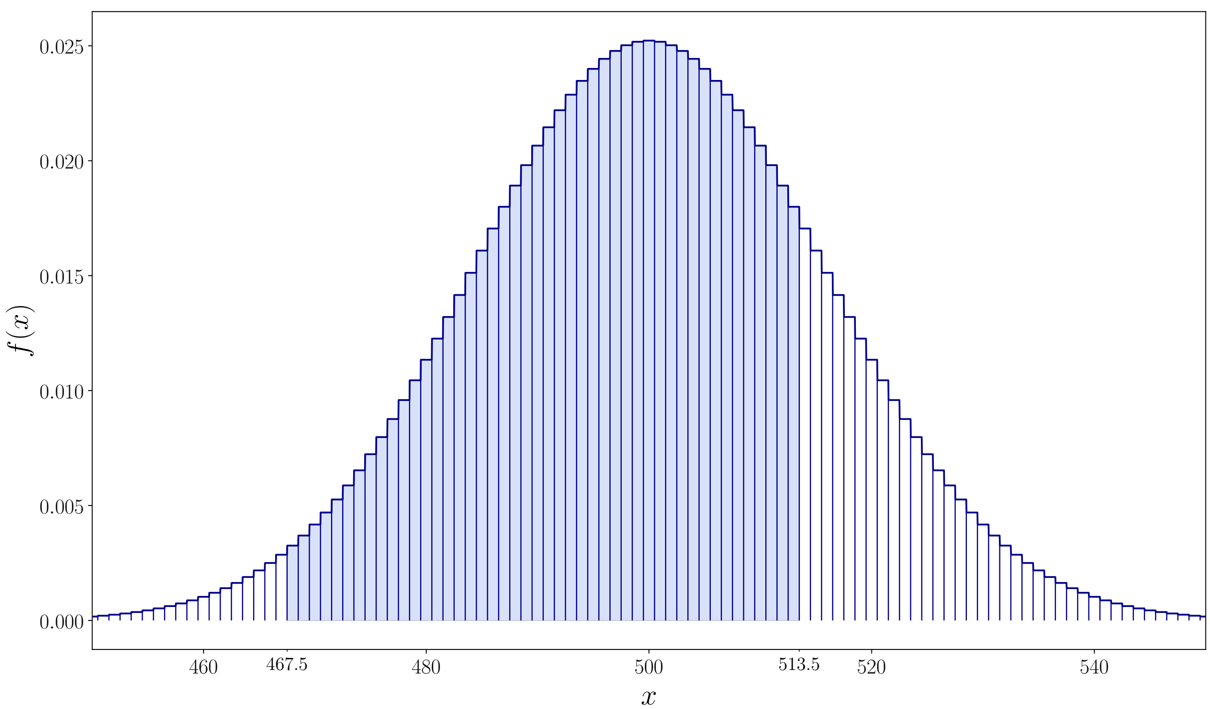

Figure 3: Continuity corrected area is given by $\mathsf{P}(467.5 < X_{\text{CLT}} \leq 513.5)$.

Using standardization, difference of c.d.f.s, symmetry of $Z \sim \mathcal{N}(0,1)$ and standard normal distribution table:

$$\begin{aligned}\mathsf{P}(467.5 < X_{\text{CLT}} \leq 513.5) &= \mathsf{P}\left(\frac{467.5-500}{\sqrt{250}} < \frac{X_{\text{CLT}}-500}{\sqrt{250}} \leq \frac{513.5-500}{\sqrt{250}}\right) = \\ &= \mathsf{P}(-2.06 < Z \leq 0.85) = \Phi(0.85) - \Phi(-2.06) = \\ &= \Phi(0.85) - 1 + \Phi(2.06) \approx 0.802 - 1 + 0.980 = \boxed{0.782}.\end{aligned}$$

Let's compare our results with direct calculation of $\mathsf{P}(467 < X \leq 513)$ in Python:

$$\mathsf{P}(467 < X \leq 513) \approx 0.783.$$

Continuity corrected result is only 0.001 away from the true one (and we don't take into account rounding and table usage inaccuracies). This will always be the case for De Moivre–Laplace approximations – continuity corrected results are always better.

Problem 10

An emergency towing service receives on average seventy calls per day for assistance. For any given day, what is the probability that fewer than fifty calls will be received?

Solution:

Let $X \sim \text{Poisson}(70)$ be a number of calls per day for assistance with $\mathsf{E}(X) = 70$ and $\mathsf{V}(X) = 70$.

We need to find $\mathsf{P}(X < 50)$.

According to CLT:

$$X \sim \text{Poisson}(70) \overset{d}{\approx} X_{\text{CLT}} \sim \mathcal{N}(70, 70),$$

and since we approximate discrete distribution with continuous, we should conduct continuity correction. Point $X = 50$ is excluded:

$$\mathsf{P}(X < 50) \approx \mathsf{P}(X_{\text{CLT}} < 49.5).$$

Using standardization and symmetry of standard normal Z:

$$\mathsf{P}(X_{\text{CLT}} < 49.5) = \mathsf{P}\left(Z < \frac{49.5 - 70}{\sqrt{70}}\right) \approx \mathsf{P}(Z < -2.45) = \Phi(-2.45) =$$
$$= 1 - \Phi(2.45) \approx 1 - 0.993 = 0.007.$$

Thus,

$$\mathsf{P}(X < 50) \approx \boxed{0.007}.$$

Python simulation gives $\mathsf{P}(X < 50) \approx 0.005$.

Problem 11

Suppose that the step length is uniformly distributed in the interval $[0.9, 1.1]$ meters. A person walks 100 steps. Find the probability that the walked distance will be more than 105 meters.

Solution:

Let $X_i \sim \mathcal{U}(0.9, 1.1)$ be an i^{th} step length in meters. Moments of X_i:

$$\mathsf{E}(X_i) = \frac{0.9 + 1.1}{2} = 1, \qquad \mathsf{V}(X_i) = \frac{(1.1 - 0.9)^2}{12} = \frac{1}{300}.$$

Distance after 100 independent steps is $S = \sum_{i=1}^{100} X_i$. We need to find $\mathsf{P}(S > 105)$.

According to CLT:

$$S \overset{d}{\approx} S_{\text{CLT}} \sim \mathcal{N}\left(100 \cdot \mathsf{E}(X_i) = 100, 100 \cdot \mathsf{V}(X_i) = \frac{1}{3}\right),$$

Using standardization and symmetry of standard normal Z:

$$\mathsf{P}(S_{\text{CLT}} > 105) = \mathsf{P}\left(Z > \frac{105 - 100}{1/\sqrt{3}}\right) \approx \mathsf{P}(Z > 3.46) = 1 - \Phi(3.46) \approx 1 - 0.9997 = 0.0003.$$

Thus,

$$\mathsf{P}(S > 105) \approx \boxed{0.0003}.$$

Python simulation gives $\mathsf{P}(S > 105) \approx 0.0001$.

Problem 12

Your gain in one round of a game with zero mean is 1 (with probability 0.5), 0 (with probability 0.4), and -5 (with probability 0.1). Estimate the probability that your total gain after 100 independent rounds of the game will be greater than 20.

Solution:

Let X_i be your gain in the round i. All X_i are i.i.d. with

$$\mathsf{E}(X_i) = 1 \cdot 0.5 + 0 \cdot 0.4 + (-5) \cdot 0.1 = 0,$$

$$\mathsf{E}\left(X_i^2\right) = 1^2 \cdot 0.5 + 0^2 \cdot 0.4 + (-5)^2 \cdot 0.1 = 3,$$
$$\mathsf{V}(X_i) = \mathsf{E}\left(X_i^2\right) - \mathsf{E}(X_i)^2 = 3 - 0^2 = 3,$$

Total gain after 100 rounds is $S = \sum\limits_{i=1}^{100} X_i$. We need to find probability $\mathsf{P}\left(S > 20\right)$.

According to CLT:

$$S \overset{d}{\approx} S_{\text{CLT}} \sim \mathcal{N}\left(100 \cdot \mathsf{E}(X_i) = 0, 100 \cdot \mathsf{V}(X_i) = 300\right),$$

and since we approximate discrete distribution with continuous, we should conduct continuity correction. Point $S = 20$ is excluded:

$$\mathsf{P}(S > 20) \approx \mathsf{P}(S_{\text{CLT}} > 20.5).$$

Using standardization and symmetry of standard normal Z:

$$\mathsf{P}(S_{\text{CLT}} > 20.5) = \mathsf{P}\left(Z > \frac{20.5 - 0}{10\sqrt{3}}\right) \approx \mathsf{P}(Z > 1.18) = 1 - \Phi(1.18) \approx 1 - 0.882 = 0.118.$$

Thus,

$$\mathsf{P}(S > 20) \approx \boxed{0.118}.$$

Python simulation gives $\mathsf{P}(S > 20) \approx 0.116$.

Problem 13

Let $X_1, \ldots, X_{48}$ be a random sample of size 48 from the distribution with p.d.f.:

$$f(x) = \frac{1}{x^2} \cdot I_{\{x>1\}}.$$

Approximate probability, that at most 10 of these random variables have values greater than 4.

Solution:

Let's introduce following random variables:

$$Y_i = \begin{cases} 1, & \text{if } X_i > 4, \\ 0, & \text{if } X_i \leq 4, \end{cases}$$

Those are Bernoulli random variables with the probability of success:

$$p = \mathsf{P}(X_i > 4) = \int\limits_4^\infty \frac{1}{x^2} dx = -\left.\frac{1}{x}\right|_4^\infty = \frac{1}{4}.$$

Now we have i.i.d. $Y_i \sim \text{Bernoulli}\left(\frac{1}{4}\right)$. Let $S = \sum\limits_{i=1}^{48} Y_i$. Sum of i.i.d. Bernoulli variables is a random variable with binomial distribution:

$$S \sim \text{Bin}\left(48, \frac{1}{4}\right) \text{ with } \mathsf{E}(S) = 48 \cdot \frac{1}{4} = 12 \text{ and } \mathsf{V}(S) = 48 \cdot \frac{1}{4} \cdot \frac{3}{4} = 9.$$

We need to find $\mathsf{P}(S \leq 10)$.
According to De Moivre–Laplace theorem:

$$S \sim \text{Bin}\left(48, \frac{1}{4}\right) \overset{d}{\approx} S_{\text{CLT}} \sim \mathcal{N}(12, 9),$$

and since we approximate discrete distribution with continuous, we should conduct continuity correction. Point $S = 10$ is included:

$$\mathsf{P}(S \leq 10) \approx \mathsf{P}(S_{\text{CLT}} < 10.5).$$

Using standardization and symmetry of standard normal Z:

$$\mathsf{P}(S_{\text{CLT}} < 10.5) = \mathsf{P}\left(Z < \frac{10.5 - 12}{\sqrt{9}}\right) \approx \mathsf{P}(Z < -0.5) = \Phi(-0.5) =$$
$$= 1 - \Phi(0.5) \approx 1 - 0.691 = 0.309.$$

Thus,

$$\mathsf{P}(S \leq 10) \approx \boxed{0.309}.$$

Python simulation gives $\mathsf{P}(S \leq 10) \approx 0.315$.