Problem set for Probability and Statistics 10 — 20 April 2026

Recall that the cumulative distribution function F_X is defined by

$$F_X(x) = \mathbb{P}[X \leq x].$$

If X is continuous, then

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

for a non-negative function f_X (the density of X). Then

$$\mathbb{P}[X \in A] = \int_A f_X(t)dt, \qquad \text{thus} \qquad \mathbb{P}[a \leq X \leq b] = \int_a^b f_X(t)dt$$

Also, $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \; f_X(x)dx$ and in general

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(t) \; f_X(t)dt.$$

Just as for discrete random variables, here also holds that $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

- Relationship between joint density and joint cumulative distribution function

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)dtds$$
$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

- Marginal density from joint density

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

- **Independence:** $X \perp Y \iff F_{X,Y}(x,y) = F_X(x)F_Y(y) \iff f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- **Law of total expectation:** $\mathbb{E}[g(X)|B] = \int_{-\infty}^{\infty} g(x) f_{X|B}(x)dx$
- **Convolution formula:** For continuous independent random variables X, Y, the variable $Z = X + Y$ has the density

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx.$$

- **Markov's inequality:** $\mathbb{P}[X \geq a\mathbb{E}[X]] \leq \frac{1}{a}$ for $X \geq 0$.
- **Chebyshev's inequality:** $\mathbb{P}[|X - \mathbb{E}[X]| \geq t\sigma_X] \leq \frac{1}{t^2}$.
- **Central limit theorem:** Let $X_1, X_2, \dots$ a sequence of i.i.d. $L^2(\Omega, F, P)$ random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 > 0$. Then for the sequence $(S_n^*)_{n=1}^{\infty}$, where

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

we have

$$S_n^* \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Continuous distributions

1. Let X be a random variable with density $f_X(t) = \frac{1}{t^2}$ for $t \geq 1$ and $f_X(t) = 0$ otherwise.
(a) Verify that this is a probability density function.
(b) Determine $\mathbb{E}[X]$.
(c) Compute the cumulative distribution function F_X.
(d) Determine $\mathbb{P}[2 \leq X \leq 3]$.
(e) Let $Y = 1/X$. What is the cumulative distribution function of the random variable Y?
(f) Determine the probability density function of the random variable Y. Name its distribution.

2. Mr. Chen visited Prague and at a uniformly random time (0:00-24:00), he appears in the Old Town Square. Every hour from 9:00 to 23:00, 12 apostle figures appear on the astronomical clock.
(a) What is the probability that Mr. Chen will see the apostles without waiting for more than 15 minutes?
(b) What if Mr. Chen arrives at the Old Town Square at a uniformly random time after noon, i.e., 12:00–24:00?

3. We break a one-metre stick into two pieces, at a uniformly random point. Let X be the length of the longer piece.
(a) What is the distribution of X?
(b) Determine $\mathbb{E}[X]$.
(c) * If we choose two points on a one-metre stick independently and uniformly at random, then break the stick into three pieces at these, what is the probability that the lengths of the three pieces are the side lengths of a triangle? (See the hints at the end for help.)

4. Let Y be the maximum of n uniformly distributed numbers in the interval $[0, 1]$.
(a) Find the cumulative distribution function F_Y.
(b) From there, determine the density function f_Y.
(c) Calculate $\mathbb{E}[Y]$.
(d) How about for the minimum of those numbers?
(e) * And for the k^{th} smallest number?

Named continuous distributions

distribution	pdf	cdf	expectation	variance
$\mathsf{Unif}(a, b)$	$\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$	$\min\{\frac{x-a}{b-a}, 1\}\mathbb{1}_{[a,\infty)}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathsf{Exp}(\lambda)$	$\lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$	$(1 - e^{-\lambda x})\mathbb{1}_{[0,\infty)}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathsf{Cauchy}(x_0, \gamma)$	$\frac{1}{\pi\gamma(1+(\frac{x-x_0}{\gamma})^2)}$	$\frac{1}{2} + \frac{1}{\pi} \arctan \frac{x-x_0}{\gamma}$	—	—
$\mathcal{N}(\mu, \sigma^2)$	(*) $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$	* $\Phi(\frac{x-\mu}{\sigma})$	μ	σ^2
$\mathsf{Gamma}(r, \alpha)$	$\frac{\alpha^r x^{r-1} e^{-\alpha x}}{\Gamma(r)}\mathbb{1}_{[0,\infty)}(x)$	* $\frac{\gamma(r,\alpha x)}{\Gamma(r)}\mathbb{1}_{[0,\infty)}(x)$	$\frac{r}{\alpha}$	$\frac{r}{\alpha^2}$

z	-4	-3	-2	-1	0	1	2	3	4
$\Phi(z)$	0.00003	0.00135	0.02275	0.15866	0.500000	0.84135	0.97725	0.99865	0.99997

5. We say that a random variable X (or its distribution) is *memoryless* if

$$\mathbb{P}[X > s + t | X \geq s] = \mathbb{P}[X > t]$$

for $s, t \geq 0$. In other words, the time we have already waited does not affect the time we will still wait. Show that the exponential distribution is memoryless.

6. Plutonium-238 has a half-life of 87.7 years. We will model its decay using an exponential distribution: for each atom, we consider the time until decay as an independent random variable with the distribution $\mathsf{Exp}(\lambda)$.

(a) What is λ?
(b) What is the average lifespan of a plutonium-238 atom?
(c) After how much time will 90% of the atoms decay?
(d) What percentage of atoms will decay after 50 years? (Some cardiac pacemakers use plutonium-238 as an energy source. `https://en.wikipedia.org/wiki/Plutonium-238#Nuclear_powered_pacemakers`)

7. Let $Z \sim \mathcal{N}(0, 1)$. Use the Φ function table to calculate:
(a) $\mathbb{P}[|Z| \leq 1]$
(b) $\mathbb{P}[|Z| \leq 2]$
(c) $\mathbb{P}[|Z| \leq 3]$
(d) Rewrite what this means for a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$.

8. We will model the amount of snow that will lie on the ground in a Krkonoše ski resort, on New Year's eve. We will use normal distribution with a mean of 40 (centimetres) and a standard deviation of 10.
(a) What is the probability that the model will give us a negative value for the snow cover?
(b) What is the probability that the snow cover will be between 50 and 70 cm?

Joint density, convolution

9. Let X, Y have a joint density $f_{X,Y}(x, y) = e^{-x-y}$ for $x, y > 0$ (and 0 otherwise).
(a) Determine the marginal densities f_X, f_Y.
(b) Also find the cumulative distribution functions F_X, F_Y, $F_{X,Y}$.
(c) Are X, Y independent?
(d) Find $\mathbb{P}[X + Y \leq 1]$ and $\mathbb{P}[X > Y]$.

10. For independent continuous random variables $X \sim \mathsf{Unif}(0, 2)$ and $Y \sim \mathsf{Unif}(0, 1)$, we examine $\mathbb{P}[X < Y]$. Compute it:
(a) with a geometrical argument;
(b) using $\mathbb{P}[X < Y] = \mathbb{E}[\mathbb{1}_{\{X<Y\}}]$, applying LOTUS to the function $g(x, y) = \mathbb{1}_{\{x<y\}}$, and computing the integral. Do it for both orders of integration.

11. Let $X, Y, Z \sim \mathsf{Unif}(0, 1)$ be independent random variables.
(a) What is the distribution of $X + Y$? Determine the density (in two ways) – using the convolution formula and by a geometrical argument.
(b) * What is the distribution of $X + Y + Z$?

12. Let $X, Y, Z \sim \mathsf{Exp}(\lambda)$ be independent random variables.
(a) What is the distribution of $X + Y$?
(b) What is the distribution of $X + Y + Z$?

13. Choose a point uniformly randomly from the semicircle with radius 1, centered at the origin, in the upper half-plane. Let X, Y be the coordinates of the chosen point.
(a) Find the joint density $f_{X,Y}$.
(b) Find the marginal density f_Y and compute $\mathbb{E}[Y]$ using it.
(c) For verification, calculate $\mathbb{E}[Y]$ directly (using the law of total expectation).

Inequalities, central limit theorem

14. We know that the average number of points on a test was 40 (out of 100). Give an upper bound for the proportion of students with at least 80 points. Improve the upper bound if you know that the standard deviation of the number of points is 10.

15. Estimate $\binom{100}{40}$ by estimating $\mathbb{P}[39.5 < X < 40.5]$ for a suitable random variable X using the central limit theorem.

16. Let $S = \sum_{k=56}^{100} \binom{100}{k}$. Also, let $X = \sum_{i=1}^{100} X_i$, where X_i is 0 or 1, both with probability $\frac{1}{2}$ and the variables $X_1, \ldots, X_n$ are independent. Thus, $X \sim \mathsf{Bin}(100, 1/2)$.

(a) Express S using the cumulative distribution function F_X.

(b) Use CLT to estimate this probability.

(c) We want to estimate whether our coin (and the way we flip it) is fair. If we get more than 55 heads out of a hundred flips, we will say it is not fair. What is the probability of making a mistake? That is, if we have a fair coin, what is the probability that we get more than 55 heads out of a hundred flips? Also compute upper bounds for that probability using Markov's inequality, then Chebyshev's inequality.

Bonus problems

17. (**Buffon's needle**) We throw a needle of length l onto an infinite floor. The floor consists of boards, with edges forming parallel lines at a distance $d \geq l$. Determine the probability that the needle will cross the edge of any board. (See the hints at the end for help.)

18. Suppose that each of $m \geq 1$ pigeons independently and at random enter one of $n \geq 1$ pigeonholes. If $m \geq 1.2\sqrt{n} + 1$, then show that the probability that two pigeons go into the same pigeonhole is greater than $\frac{1}{2}$.

For part c) of problem 3, by the triangle inequality the three pieces can form a triangle if and only if each of them has length below 0.5 metres. If we plot the point whose x-coordinate is the location of the first breaking point and the y-coordinate is the location of the second breaking point, it is uniformly distributed on $[0, 1] \times [0, 1]$. Determine the regions that correspond to neither of the breaking points occurring within the first 0.5 m, neither of the breaking points occurring within the last 0.5 m and the two breaking points being at least 0.5 m apart from each other; the area of the remaining region gives the required probability.

For problem 17, let θ be the acute angle the needle forms with one of the edges and let r be the distance of the centre of the needle from the closest edge. Determine the distribution of r and θ and express the event of the needle crossing the edge of the nearest board in terms of d, l, r, θ. Use these to evaluate the probability.
