Practice questions

1. Let X be an Exponential(λ) random variable. Find the PDF of the random variables for

   (a) $Y = X^2$

   (b) $Z = e^{-\lambda X}$.

   Solution:

   (a) For $y \geq 0$, the CDF of Y is $F_Y(y) = \mathrm{P}(X^2 \leq y) = \mathrm{P}(X \leq \sqrt{y}) = 1 - e^{-\lambda\sqrt{y}}$. The PDF is the derivative of the CDF which is

   $$f_Y(y) = \begin{cases} \dfrac{\lambda}{2\sqrt{y}} e^{-\lambda\sqrt{y}} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

   (b) Since X only takes nonnegative values, Z will take values between 0 and 1. For $0 < z \leq 1$, the CDF of Z is

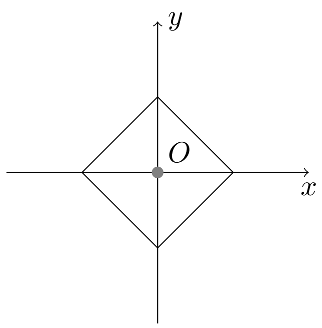
   $$F_Z(z) = \mathrm{P}(e^{-\lambda X} \leq z) = \mathrm{P}\Big(X \geq -\frac{\log z}{\lambda}\Big) = e^{-\lambda(-\log z/\lambda)} = z.$$

   Its derivative is

   $$f_Z(z) = \begin{cases} 1 & \text{if } 0 < z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   In words, Z is a Uniform$(0, 1)$ random variable.

2. We define the Manhattan distance between two points $A(x_A, y_A)$ and $B(x_B, y_B)$ in a two-dimensional space as $|x_A - x_B| + |y_A - y_B|$, denoted by $d_M(A, B)$. A point is chosen uniformly at random inside a square with each side of length $\sqrt{2}$. The vertices of the square are on the x or y axis and the centre of the square is at the origin (see the figure below). Let X be the Manhattan distance from the point to the centre of the square (the origin O).



   (a) What is the CDF of X

   (b) What is the PDF of X

   (c) What is the expected value of X

   (d) What is the variance of X

Solution:

(a) The PDF of the point is uniform in the square which has area 2, so it has value 1/2 inside the center and zero outside. The event $X \leq x$ consists of all the points in the square that are at distance less than or equal to x from the center, which is a square of side length $\sqrt{2}x$. Therefore the CDF is $\mathrm{P}(X \leq x) = 1/2 \times (\sqrt{2}x)^2 = x^2$,

(b) The PDF is $f_X(x) = d\,\mathrm{P}(X \leq x)/dx = 2x$ for $0 \leq x \leq 1$.

(c) The expected value of X is $\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty} 2x^2 = \frac{2}{3}x^3 \mid_0^1 = \frac{2}{3}$.

(d) The variance of X is $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$, where

$$\mathrm{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_{-\infty}^{\infty} 2x^3 = \tfrac{1}{2}x^4 \mid_0^1 = \frac{1}{2}.$$

Therefore, $\mathrm{Var}(X) = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$.

3. The arrival times of the 200 ENGG 2760A / ESTR 2018 students to class are normal random variables with a mean value of 2:25pm and a **standard deviation** of 5 minutes.

(**Note:** You can use the computer software or calculator to obtain the CDF of the normal random variable. (e.g. Python function `scipy.stats.norm.cdf`))

(a) What is the expected number of students that have arrived by 2:30pm?

(b) Assuming students' arrivals are independent, what is the probability that everyone has made it by 2:45pm?

Solution:

(a) Let N be a Normal$(0, 1)$ random variable. The delay of each student in minutes is modeled by $5N$. The probability that any given student has arrived by 2:30am is the probability that N is at most 1 which is about 84.13%. By linearity of expectation, the expected number of students that make it in time for class is about $84.13\% \cdot 200 \approx 168$.

(b) The probability that any given student hasn't made it by 2:45 is $\mathrm{P}(N > 4) = \mathrm{P}(N < -4) \approx 3.167 \cdot 10^{-5}$, so the probability that all the students have made is by 2:45 is $(1 - \mathrm{P}(N > 4))^{200}$, which is about 99.37%.

4. The body temperatures of a healthy person and an infected person are Normal$(36.8, 0.5)$ and Normal$(37.8, 1.0)$ random variables, respectively. About 1% of the population is infected. What is the conditional PDF that I am infected given that my temperature is t? For which values of t am I more likely to be infected than not? (**Hint:** Bayes' rule)

Solution: Let A be the event that I am infected, and T be my body temperature. By the total probability theorem,

$$f_T(t) = \mathrm{P}(A)f_{T|A}(t) + \mathrm{P}(A^c)f_{T|A^c}(t),$$

where $T|A$ is a Normal$(37.8, 1.0)$ random variable and $T|A^c$ is a Normal$(36.8, 0.5)$ random variable. The (unconditional) PDF of X is

$$f_T(t) = \frac{0.01}{\sqrt{2\pi}}e^{-\frac{(t-37.8)^2}{2}} + \frac{0.99}{\sqrt{2\pi}(0.5)}e^{-\frac{(t-36.8)^2}{2(0.5)^2}} = \frac{0.01}{\sqrt{2\pi}}e^{-(t-37.8)^2/2} + \frac{1.98}{\sqrt{2\pi}}e^{-2(t-36.8)^2}.$$

By Bayes' rule, the conditional probability of A given T is

$$P(A|T = t) = \frac{P(A)f_{T|A}(t)}{f_T(t)} = \frac{0.01e^{-(t-37.8)^2/2}}{0.01e^{-(t-37.8)^2/2} + 1.98e^{-2(t-36.8)^2}}$$

I am more likely to be infected than not when $\mathrm{P}(A) > \mathrm{P}(A^c)$, namely when $0.01e^{-(t-37.8)^2/2} > 1.98e^{-2(t-36.8)^2}$. Taking logarithms of both sides this is equivalent to a quadratic inequality in t. Solving this inequality, we obtain that $\mathrm{P}(A) > \mathrm{P}(A^c)$ holds when $t < t_-$ or $t > t_+$, where $t_- \approx 34.4742$ and $t_+ \approx 38.4591$.

5. Bob's arrival time at a meeting with Alice is X hours past noon (noon is 12:00), where X is a random variable with PDF

$$f(x) = \begin{cases} cx + \frac{1}{2}, & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of the constant c.

(b) What is the probability that Bob arrives by 12.30?

(c) What is the expected hour of Bob's arrival?

(d) Given that Bob hasn't arrived by 12.30, what is the probability that he arrives by 12.45?

(e) Given that Bob hasn't arrived by 12.30, what is the expected hour of Bob's arrival?

Solution:

(a) By the axioms of probability, $\int_{-\infty}^{\infty} f(x)dx = 1$. Since

$$\int_{-\infty}^{\infty} f(x)dx = (\frac{1}{2}cx^2 + \frac{1}{2}x) \mid_0^1 = \frac{1}{2}c + \frac{1}{2},$$

c must be equal to 1.

(b) The CDF of X is $F(x) = \mathrm{P}(X \le x) = \int_{-\infty}^{\infty} f(x)dx = \frac{1}{2}x^2 + \frac{1}{2}x$. In particular, $\mathrm{P}(X \le 0.5) = \frac{3}{8}$.

(c) The expected value is $\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} x(x + \frac{1}{2})dx = \int_{-\infty}^{\infty} (x^2 + \frac{1}{2}x)dx = (\frac{1}{3}x^3 + \frac{1}{4}x^2) \mid_0^1 = \frac{7}{12}$. So Bob is expected to arrive at 12:35.

(d) The probability that Bob hasn't arrived by 12:30 is $P(X > 0.5) = 1 - P(X \le 0.5) = 1 - \frac{3}{8} = \frac{5}{8}$. The probability that Bob hasn't arrived by 12:30 but arrives by 12:45 is $P(0.75 \ge X > 0.5) = P(X \le 0.75) - P(X \le 0.5) = F(0.75) - F(0.5) = \frac{21}{32} - \frac{3}{8} = \frac{9}{32}$. Therefore, the conditional probability is

$$\mathrm{P}(X \le 0.75 \mid X > 0.5) = \frac{\mathrm{P}(0.75 \ge X > 0.5)}{\mathrm{P}(X > 0.5)} = \frac{9/32}{1 - 3/8} = \frac{9}{20}$$

.

(e) Given that Bob hasn't arrived by 12:30, the probability that Bob arrives by x hour past noon is

$$\mathrm{P}(X \le x \mid X > 0.5) = \frac{\mathrm{P}(x \ge X > 0.5)}{\mathrm{P}(X > 0.5)} = \begin{cases} \frac{\frac{1}{2}x^2 + \frac{1}{2}x - \frac{3}{8}}{\frac{5}{8}} & 0.5 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus, the conditional PDF is

$$f(x) = \begin{cases} \frac{8}{5}(x + \frac{1}{2}) & 0.5 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and the expected value is $\int_{-\infty}^{\infty} xf(x)dx = \int_{0.5}^{1} \frac{8}{5}(x^2 + \frac{1}{2}x)dx = \frac{8}{5}(\frac{1}{3}x^3 + \frac{1}{4}x^2)|_{0.5}^{1} = \frac{23}{30}$.