July 25, 2014

Solutions of Math 302 HW assignment 4

1. (2 marks) The probability mass function of the POI(λ) distribution is

$$f(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Calculate the expected value and variance.

Solution:

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} k \cdot f(k) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k \cdot e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-1)!} =$$
$$e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = e^{-\lambda}\lambda e^{\lambda} = \lambda.$$

$$\mathbf{E}(X\cdot(X-1)) = \sum_{k=0}^{\infty} k\cdot(k-1)\cdot e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=2}^{\infty} k\cdot(k-1)\cdot e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-2)!} =$$
$$e^{-\lambda}\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = e^{-\lambda}\lambda^2\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2.$$

$$\lambda^2 = \mathbf{E}(X\cdot(X-1)) = \mathbf{E}(X^2 - X) = \mathbf{E}(X^2) - \mathbf{E}(X) = \mathbf{E}(X^2) - \lambda$$

Thus $\mathbf{E}(X^2) = \lambda^2 + \lambda$.

$$\mathrm{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

2. On a Pacific island beach there is at least one shark attack in a year with probability 1/3.

   (a) (2 marks) What is the probability that there will be at least four shark attacks next year?
   Solution: There are *many* days in a year. Every day of the year, there is a *tiny* chance of a shark attack, *independently* from other days. Thus the number X of shark attacks in a year has Poisson distribution, $X \sim \mathrm{Poi}(\lambda)$.
   We know $\mathbf{P}(X \geq 1) = 1/3$, thus $e^{-\lambda} = \mathbf{P}(X = 0) = 2/3$, thus $\lambda = \ln(3/2)$. Thus

$$\mathbf{P}(X \geq 4) = 1 - \mathbf{P}(X \leq 3) = 1 - \sum_{k=0}^{3} e^{-\lambda}\frac{\lambda^k}{k!} =$$
$$1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6}\right) = 1 - \frac{2}{3}\left(1 + \ln(3/2) + \frac{\ln(3/2)^2}{2} + \frac{\ln(3/2)^3}{6}\right)$$

   (b) (3 marks) Shark attacks are lethal 75% of the time. What is the probability that there will be no lethal shark attack next year?
   Solution: Conditioned on the event $\{X = k\}$, the probability of the event that no-one dies is a shark attack is 0.25^k, because each person attacked by a shark survives with 25% chance, independently from the others. Let's apply the total probability rule:

$$\mathbf{P}(\text{ no lethal shark attack }) = \sum_{k=0}^{\infty} 0.25^k f(k) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{(0.25\lambda)^k}{k!} =$$
$$e^{-\lambda}\sum_{k=0}^{\infty}\frac{(0.25\lambda)^k}{k!} = e^{-\lambda}e^{0.25\lambda} = e^{-0.75\ln(3/2)} = e^{\ln((3/2)^{-3/4})} = (2/3)^{3/4} = 0.738$$

3. Find the value of c such that the following functions become probability density functions. Calculate the corresponding cumulative distribution function. Sketch a picture of the graph of the p.d.f. as well as the c.d.f.

(a) (2 marks) $f(x) = c \cdot (2 - |x|)$ if $-2 \le x \le 2$ and $f(x) = 0$ otherwise.
Solution: The graph of the p.d.f. is triangle with area $4 \cdot 2/2 = 4$, thus $c = \frac{1}{4}$ if we want the area below the graph of $f(\cdot)$ to be equal to one. Let's find the c.d.f. $F(b)$. Note that $F(0) = \frac{1}{2}$ by symmetry of $f(\cdot)$. If $-2 \le b \le 0$ then

$$F(b) = \int_{-\infty}^{b} f(x)\mathrm{d}x = \int_{-2}^{b} \frac{1}{4}(2+x)\mathrm{d}x = \left[\frac{(2+x)^2}{8}\right]_{-2}^{b} = \frac{(2+b)^2}{8}.$$

If $0 \le b \le 2$ then

$$F(b) = \int_{-\infty}^{b} f(x)\mathrm{d}x = F(0) + \int_{0}^{b} \frac{1}{4}(2-x)\mathrm{d}x = \frac{1}{2} + \left[-\frac{(2-x)^2}{8}\right]_{0}^{b} = 1 - \frac{(2-b)^2}{8}.$$

Note that $F(-b) = 1 - F(b)$ just like $\Phi(-b) = 1 - \Phi(b)$, where Φ is the c.d.f. of $\mathcal{N}(0,1)$. In both cases this is because the p.d.f. is an even function. But to conclude the solution:

$$F(b) = \begin{cases} 0 & \text{if } b \le -2, \\ \frac{(2+b)^2}{8} & \text{if } -2 \le b \le 0, \\ 1 - \frac{(2-b)^2}{8} & \text{if } 0 \le b \le 2, \\ 1 & \text{if } b \ge 2. \end{cases}$$

(b) (3 marks) $f(x) = c \cdot x^2 e^x$ if $x \le 0$ and $f(x) = 0$ if $x > 0$.
Solution: We calculated in class that $\int_0^\infty x^2 e^{-x}\mathrm{d}x = 2$, thus $c = \frac{1}{2}$. Let's calculate the c.d.f. Note that $F(0) = 1$ because $f(x) = 0$ if $x > 0$. The indefinite integral of $f(x)$ is $\int f(x)\mathrm{d}x = e^x\left(\frac{x^2}{2} - x + 1\right) + C$ according to Wolfram Alpha, and if we choose $C = 0$ and let $F(x) = e^x\left(\frac{x^2}{2} - x + 1\right)$ then we indeed have $F(0) = 1$ and $\lim_{x\to-\infty} F(x) = 0$ as we should. Thus

$$F(x) = \begin{cases} e^x\left(\frac{x^2}{2} - x + 1\right) & \text{if } x \le 0, \\ 1 & \text{if } x \ge 0. \end{cases}$$

(c) (2 marks) $f(x) = c \cdot \frac{1}{1+x^2}$ for any x.
Solution: This is the density function of the famous Cauchy distribution. Now $\arctan(x)$ is an antiderivative of $\frac{1}{1+x^2}$ and $\lim_{x\to\infty} \arctan(x) = \frac{\pi}{2}$, $\lim_{x\to-\infty} \arctan(x) = -\frac{\pi}{2}$, thus $\int_{-\infty}^{\infty} \frac{1}{1+x^2}\mathrm{d}x = \pi$ and $c = \frac{1}{\pi}$.
Therefore $\lim_{x\to\infty} \frac{1}{\pi}\arctan(x) = \frac{1}{2}$, $\lim_{x\to-\infty} \frac{1}{\pi}\arctan(x) = -\frac{1}{2}$, thus if we define the c.d.f. as $F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$ then we indeed have $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$ as we should. Also note that

$$F(-x) = \frac{1}{\pi}\arctan(-x) + \frac{1}{2} = \frac{-1}{\pi}\arctan(x) + \frac{1}{2} = 1 - \left(\frac{1}{\pi}\arctan(x) + \frac{1}{2}\right) = 1 - F(x),$$

just like $\Phi(-x) = 1 - \Phi(x)$, where Φ is the c.d.f. of $\mathcal{N}(0,1)$. In both cases this is because the p.d.f. is an even function.

4. There are two coffee shops on a street 300 meters from each other. You stroll around the street, and on a uniformly distributed point between the two coffee shops you decide to drink a coffee. You like one coffee shop twice as much as the other one, i.e., you are willing to walk twice as much (but not more) to get your coffee there. You decide between the two coffee shops and walk to one of them to get your coffee. You burn 30 calories per kilometre.

(a) (2 marks) Calculate the expectation and standard deviation of the calories burnt using the formula $\mathbf{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x$ rather than calculating the c.d.f. and the the p.d.f.!

Solution: Let $X \sim \mathrm{Unif}[0, 300]$ denote the location on the street where you decide to drink coffee. Let's say your favourite café is at 0 and the other one is at 300. If $0 \le X \le 200$ then you walk to the left and walk X meters, if $200 < X \le 300$ then you walk to the right and you walk $300 - X$ meters. You burn $\frac{3}{100}$ calories per meter, thus if we define $g(x)$ as the amount of calories that you burn if $\{X = x\}$, then

$$g(x) = \begin{cases} \frac{3}{100}x & \text{if } 0 \le x \le 200, \\ \frac{3}{100}(300 - x) & \text{if } 200 < x \le 300. \end{cases}$$

The amount of calories you actually burn is $g(X)$.
The density of X is $f(x) = \frac{1}{300}$ if $0 \le x \le 300$, thus

$$\mathbf{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x = \int_0^{300} g(x)\frac{1}{300}\,\mathrm{d}x = \frac{1}{300}\left(\frac{200 \cdot 6}{2} + \frac{100 \cdot 3}{2}\right) = 2.5$$

$$\mathbf{E}(g(X)^2) = \int_0^{300} g(x)^2 \frac{1}{300}\,\mathrm{d}x = \frac{1}{300}\left(\int_0^{200} \frac{9}{10000}x^2\,\mathrm{d}x + \int_{200}^{300} \frac{9}{10000}(300 - x)^2\,\mathrm{d}x\right) = \frac{1}{300}\left(\frac{9}{10000}\frac{200^3}{3} + \frac{9}{10000}\frac{100^3}{3}\right) = 9.$$

Thus $\mathrm{Var}(g(X)) = 9 - (2.5)^2 = 2.75$.

(b) (2 marks) Calculate the c.d.f. and the p.d.f. of the calories burnt.
Solution: Recall from class that for any $[a, b] \subseteq [0, 300]$ we have $\mathbf{P}(X \in [a, b]) = \frac{b-a}{300}$.
Let $Y = g(X)$. Then $\mathbf{P}(0 \le Y \le 6) = 1$. Denote by $F(y) = \mathbf{P}(Y \le y)$ the c.d.f. of Y.
For any $0 \le y \le 3$ we have

$$F(y) = \mathbf{P}(g(X) \le y) = \mathbf{P}(\frac{3}{100}X \le y) + \mathbf{P}(\frac{3}{100}(300 - X) \le y) = \mathbf{P}(X \in [0, \frac{100}{3}y]) + \mathbf{P}(X \in [300 - \frac{100}{3}y, 300]) = \frac{\frac{100}{3}y}{300} + \frac{\frac{100}{3}y}{300} = \frac{2}{9}y$$

For any $3 \le y \le 6$ we have

$$F(y) = \mathbf{P}(g(X) \le y) = \mathbf{P}(\frac{3}{100}X \le y) + \mathbf{P}(X \in [200, 300]) = \mathbf{P}(X \in [0, \frac{100}{3}y]) + \frac{100}{300} = \frac{\frac{100}{3}y}{300} + \frac{1}{3} = \frac{1}{9}y + \frac{1}{3}$$

Thus

$$F(y) = \begin{cases} 0 & \text{if } y \le 0, \\ \frac{2}{9}y & \text{if } 0 \le y \le 3, \\ \frac{1}{9}y + \frac{1}{3} & \text{if } 3 \le y \le 6, \\ 1 & \text{if } y \ge 6. \end{cases}$$

Now we obtain the p.d.f. of Y as $f(y) = \frac{\mathrm{d}}{\mathrm{d}y}F(y)$:

$$f(y) = \begin{cases} \frac{2}{9} & \text{if } 0 \le y \le 3, \\ \frac{1}{9} & \text{if } 3 < y \le 6, \\ 0 & \text{otherwise }. \end{cases}$$

(c) (2 marks) Your coffee has 4.5 calories in it. What is the probability of the event that you burn less calories than what you consume, conditioned on the event that you walk more than 50 meters?
Solution: You burn less than 4.5 calories if you walk less than 150 meters. You walk more than 50 meters if $A = \{50 \le X \le 250\}$ occurs. You walk less than 150 meters if $B = \{0 \le X \le 150\} \cup \{200 \le X \le 300\}$ occurs.

$$\mathbf{P}(B \mid A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(\{50 \le X \le 150\} \cup \{200 \le X \le 250\})}{\mathbf{P}(50 \le X \le 250)} = \frac{(100 + 50)/300}{200/300} = \frac{3}{4}.$$

5. How to generate any random variable using a uniform random generator?

(a) (2 marks) Let $X \sim \text{Unif}[0,1]$ and let $Y = -2\ln(X)$. Calculate the c.d.f. of Y. Now Y is a famous random variable: name it and identify its parameter(s).

Solution: Denote by $F(y)$ the c.d.f. of Y. First note that $-2\ln(x) \geq 0$ if $0 \leq x \leq 1$, thus $F(y) = 0$ for $y \leq 0$. For $y \geq 0$ we have

$$F(y) = \mathbf{P}(-2\ln(X) \leq y) = \mathbf{P}(\ln(X) \geq -\frac{1}{2}y) =$$

$$\mathbf{P}(X \geq e^{-\frac{1}{2}y}) = \mathbf{P}(X \in [e^{-\frac{1}{2}y}, 1]) = \frac{1 - e^{-\frac{1}{2}y}}{1-0} = 1 - e^{-\frac{1}{2}y}.$$

Thus $Y \sim \text{EXP}(\frac{1}{2})$.

(b) (2 marks) Let $X \sim \text{Unif}[0,1]$ and let $Y = g(X)$. How to choose $g(\cdot)$ if I want $Y \sim \mathcal{N}(0,1)$?

Solution: Let us assume in advance that $g : [0,1] \to \mathbb{R}$ is strictly increasing and continuous, so it has an inverse function $g^{-1} : \mathbb{R} \to [0,1]$. Let's work backwards. We want $\mathbf{P}(Y \leq y) = \Phi(y)$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution, thus we want

$$\Phi(y) = \mathbf{P}(g(X) \leq y) = \mathbf{P}(X \leq g^{-1}(y)) = \frac{g^{-1}(y) - 0}{1-0} = g^{-1}(y).$$

Thus we want $\Phi(\cdot)$ to be the inverse function of $g(\cdot)$. Otherwise said, we want $g(\cdot)$ to be the inverse function of $\Phi(\cdot)$. Note that this is OK, because the function $\Phi(\cdot)$ is indeed strictly increasing and continuous. Thus $g(x) = \Phi^{-1}(x)$, $0 \leq x \leq 1$.

6. (4 marks) Let us define two probability density functions:

$$f_1(x) = \begin{cases} c_1 \cdot x^{-5/2} & \text{if } 1 \leq x \leq 10^{100} \\ 0 & \text{otherwise} \end{cases} \qquad f_2(x) = \begin{cases} c_2 \cdot x^{-7/2} & \text{if } 1 \leq x \leq 10^{100} \\ 0 & \text{otherwise} \end{cases}$$

Find c_1 and c_2 so that $f_1(\cdot)$ and $f_2(\cdot)$ are indeed probability density functions (you can omit small errors). Denote by X_1 and X_2 the corresponding random variables. Calculate the corresponding expectations μ_1, μ_2 and variances σ_1, σ_2 numerically (you can omit small errors).

Given this data calculate the lower bound that Chebychev's inequality gives on the probabilities

$$\mathbf{P}(|X_1 - \mu_1| < 1000) \qquad \text{and} \qquad \mathbf{P}(|X_2 - \mu_2| < 1000).$$

Solution:

$$c_1 = \int_1^{10^{100}} x^{-5/2}\mathrm{d}x = \left[-\frac{2}{3}x^{-3/2}\right]_1^{10^{100}} \approx \frac{2}{3}.$$

$$c_2 = \int_1^{10^{100}} x^{-7/2}\mathrm{d}x = \left[-\frac{2}{5}x^{-5/2}\right]_1^{10^{100}} \approx \frac{2}{5}.$$

$$\mu_1 = \int_1^{10^{100}} x \cdot \frac{2}{3}x^{-5/2}\mathrm{d}x = \frac{2}{3}\int_1^{10^{100}} x^{-3/2}\mathrm{d}x = \frac{2}{3}\left[-2x^{-1/2}\right]_1^{10^{100}} \approx \frac{4}{3}.$$

$$\mu_2 = \int_1^{10^{100}} x \cdot \frac{2}{5}x^{-7/2}\mathrm{d}x = \frac{2}{5}\int_1^{10^{100}} x^{-5/2}\mathrm{d}x = \frac{2}{5}\left[-\frac{2}{3}x^{-3/2}\right]_1^{10^{100}} \approx \frac{4}{15}.$$

$$\mathbf{E}(X_1^2) = \int_1^{10^{100}} x^2 \cdot \frac{2}{3}x^{-5/2}\mathrm{d}x = \frac{2}{3}\int_1^{10^{100}} x^{-1/2}\mathrm{d}x = \frac{2}{3}\left[2x^{1/2}\right]_1^{10^{100}} \approx \frac{4}{3}10^{50}.$$

$$\mathbf{E}(X_2^2) = \int_1^{10^{100}} x^2 \cdot \frac{2}{5}x^{-7/2}\mathrm{d}x = \frac{2}{5}\int_1^{10^{100}} x^{-3/2}\mathrm{d}x = \frac{2}{5}\left[-2x^{-1/2}\right]_1^{10^{100}} \approx \frac{4}{5}.$$

$$\sigma_1^2 = \text{Var}(X_1) \approx \frac{4}{3}10^{50} - \left(\frac{4}{3}\right)^2 \approx \frac{4}{3}10^{50}, \qquad \sigma_2^2 = \text{Var}(X_2) = \frac{4}{5} - \left(\frac{4}{15}\right)^2 \approx 0.73$$

Let's apply Chebyshev's inequality:

$$\mathbf{P}(|X_1 - \mu_1| \geq 1000) \leq \frac{\sigma_1^2}{10^6} \approx \frac{4}{3}10^{44}, \qquad \mathbf{P}(|X_2 - \mu_2| \geq 1000) \leq \frac{\sigma_2^2}{10^6} \approx 5.3 \times 10^{-7}.$$

Now the bound on $\mathbf{P}(|X_1 - \mu_1| \geq 1000)$ is useless, since $\frac{4}{3}10^{44} > 1$ and by the axioms of probability we know that the probability of any event is less than or equal to 1. Chebyshev's inequality $\mathbf{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$ only gives useful information if $\sigma < \varepsilon$. This is exactly what happens in the case of X_2. We also have $\mathbf{P}(|X_2 - \mu_2| < 1000) \geq 1 - 5.3 \times 10^{-7}$.

7. (2 marks) The IQ distribution of students at UBC can be well approximated by a normal distribution with mean 130 and standard deviation 20. In order to design an appropriate final exam, the professor must find the shortest interval $[a, b]$ that contains the IQ of 90% of UBC students. Help the professor with this question.

   Solution: By looking at the shape of the density function f_X of $X \sim \mathcal{N}(\mu, \sigma)$ (a bell curve), it is clear that if we are aiming for the shortest interval $[a, b]$ such that $\int_a^b f_X(x)\mathrm{d}x = 0.9$, then we'd better choose an interval whose midpoint is μ. In our case $X \sim \mathcal{N}(130, 20)$, thus we want to find $x > 0$ such that $0.9 = \mathbf{P}(130 - x \leq X \leq 130 + x)$. Let us introduce $Z = \frac{X-130}{20}$, then $Z \sim \mathcal{N}(0, 1)$ and

$$0.9 = \mathbf{P}(130 - x \leq X \leq 130 + x) = \mathbf{P}(-\frac{x}{20} \leq \frac{X - 130}{20} \leq \frac{x}{20}) = \mathbf{P}(-\frac{x}{20} \leq Z \leq \frac{x}{20}) =$$
$$\Phi(x/20) - \Phi(-x/20) = \Phi(x/20) - (1 - \Phi(x/20)) = 2\Phi(x/20) - 1.$$

   Thus $\Phi(x/20) = 0.95$. We can find $\Phi^{-1}(0.9505) = 1.65$, thus $x = 20 \cdot 1.65 = 33$, thus the IQ of 90% of UBC students lies in the interval $[130 - 33, 130 + 33]$, that is $[97, 163]$.