University of Toronto
Department of Electrical and Computer Engineering

ECE286 Probability and Statistics

Final Exam
Winter 2020

Prof. Raviraj Adve April 18

Question 1
Given: Probability of any person in population not being infected is p.
Infections are independent.
We have n tests
X is a RV representing the number of tests required to test n subjects.
PMF of x?
Method 1: The probability of a person not being infected is p, hence probability of n people not being infected is p^n, hence we need only 1 test in this case.
When at least one person is infected, we need $n+1$ tests, where the probability is $1-p^n$. Hence, the PMF of X is

$$P(X=x) = \begin{cases} p^n, & \text{if } x=1 \\ 1-p^n, & \text{if } x=n+1 \\ 0, & \text{elsewhere} \end{cases} \tag{1}$$

Method 2: Let the RV Y represent the number of non-infected people the n Bernoulli trials (the n test subjects). The PMF of Y is

$$b(y;n,p) = \binom{n}{y} p^y (1-p)^{(n-y)}; \quad y = 0, 1 \ldots, n \tag{2}$$

For $n > 1$, the RV X is related to Y through the following

$$X = \begin{cases} 1, & \text{if } Y = n \\ n+1, & \text{else} \end{cases} \tag{3}$$

We have

$$P(Y=n) = b(n;n,p) = \binom{n}{n} p^n \tag{4}$$

and

$$P(Y \neq n) = 1 - P(Y=n) = 1 - p^n \tag{5}$$

Hence, the PMF of X is

$$P(X=x)=\begin{cases} p^n, & \text{if } x=1 \\ 1-p^n, & \text{if } x=n+1 \\ 0, & \text{elsewhere} \end{cases} \tag{6}$$

What is the average number of tests required to test n subjects ($n>1$)?

$$\begin{aligned} \mathbb{E}(X) &= \sum_x x f(x) = (1)p^n + (n+1)\left(1-p^n\right) = p^n + (n+1) - np^n - p^n \\ &= (n+1) - np^n \end{aligned} \tag{7}$$

Question 2

Given: X is non-zero in $(0,1]$ and the CDF is $F(x)=ax^2+b$.
We know that $f(x)=F(x)'=2ax$ and that $\int_x f(x)\,\mathrm{d}(x)=1$, hence $\int_0^1 2ax\,\mathrm{d}x=1$ which gives $\boxed{a=1}$.
Also, we know that $F(1)=1$ because $x\in(0,1]$, hence $a(1)+b=1$ which gives $\boxed{b=0}$. Therefore, the answer is True.

Question 3

Given: $f(x)$ is a valid PDF hence we should have:

- $\int_x f(x)=1$, and
- $f(x)\geq 0, \forall x$.

Is $[f(x)\star f(x)]$ a valid PDF also? where $\star$ is the convolution operator.

Method 1: Suppose that X_1 and X_2 are two iid RVs with PDF $f(x)$. Let $Y=X_1+X_2$ be another RV. For $X_1=\tau$, the variable $Y=y$ if and only if $X_2=y-\tau$. Then, the event $Y=y$ is the union of the disjoint events $X_1=\tau$ and $X_2=y-\tau$. Then

$$P(Y=y)=\int_{-\infty}^{\infty} f(\tau)f(y-\tau)\,\mathrm{d}\tau \tag{8}$$

which is the formula for the convolution. Hence, the PDF of Y is $g(y)=[f(x)\star f(x)]$, and $[f(x)\star f(x)]$ is a valid PDF.

Method 2: The convolution between two function $f_1(x)$ and $f_2(x)$ is defined as

$$[f_1(x)\star f_2(x)] = \left(f_1\star f_2\right)(x) \triangleq \int_{-\infty}^{\infty} f_1(\tau)f_2(x-\tau)\,\mathrm{d}\tau \tag{9}$$

In our case we need to prove that $\int_x [f(x) \star f(x)] = 1$. we have

$$\int_x [f(x) \star f(x)] = \int_x \int_{-\infty}^{\infty} f(\tau) f_2(x-\tau)\,\mathrm{d}\tau\,\mathrm{d}x = \int_{-\infty}^{\infty} f(\tau) \int_x f_2(x-\tau)\,\mathrm{d}x\,\mathrm{d}\tau$$
$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\tau)(1)\,\mathrm{d}\tau = 1 \tag{10}$$

where (a) follows from the fact that shifting the PDF of x to the left or to the right by τ won't change the area under it which is 1.

For the second condition $[f(x) \star f(x)] \geq 0, \forall x$, we already know that $f(x) \geq 0, \forall x$. Hence, if we flip $f(x)$ and multiply it with another $f(x)$ for all possible time shifts we won't get a negative number. Therefore, $[f(x) \star f(x)] \geq 0, \forall x$ directly follows from $f(x) \geq 0, \forall x$, which means $[f(x) \star f(x)]$ is a valid PDF.

Question 4

Using Bayes' rule we have

$$f(x|y) = \frac{f(x,y)}{h(y)} \tag{11}$$

where $h(y)$ is the marginal distribution of y. But $f(x,y) = f_1(x)f_2(y)$, hence

$$f(x|y) = \frac{f_1(x)f_2(y)}{h(y)} = \frac{f_1(x)f_2(y)}{\sum_x f(x,y)} = \frac{f_1(x)f_2(y)}{f_2(y)\sum_x f_1(x)} = \frac{f_1(x)}{\sum_x f_1(x)} \tag{12}$$

which is independent of Y, hence X and Y are independent because any knowledge about Y won't add any knowledge about X.

Question 5

Give: $C = 100$ passengers, $\lambda = 10$ passengers/minute.

Let Y be the inter-arrival time of the train which is given as

$$Y = \begin{cases} 9 \text{ minutes}, & \text{with probability } 0.5 \\ 11 \text{ minutes}, & \text{with probability } 0.5 \end{cases} \tag{13}$$

(a) Let X be a RV representing the number of passengers arriving at the train station during an interval of time. From our course, we know that the number of outcomes occurring during a given time interval results in a Poisson experiment, hence X follows a Poisson distribution with average number of outcomes per unit time λ, and it has a PDF

$$p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \tag{14}$$

where $\mu_p = \lambda t$ can be called the rate.

Hence, the probability that the station reaches its capacity before a train arrives is

$$P(X \geq 100) = P(X \geq 100; \lambda y | Y = 9)\,P(Y = 9) + P(X \geq 100; \lambda y | Y = 11)\,P(Y = 11) \tag{15}$$

Both the Poisson and the Gaussian distribution can be used to approximate the Binomial distribution under specific conditions.
When the number of Bernoulli trials n_b is large and the probability of success p_b of the trials is small we can approximate the Binomial distribution by a Poisson distribution $p(x; \mu_p)$, where $\mu_p = \lambda t = n_b p_b$.
When the number of trials n_b is large, we can approximate the Binomial distribution by a Gaussian distribution $n(x; \mu, \sigma)$, where the mean $\mu = n_b p_b$ and the standard deviation $\sigma = \sqrt{n_b p_b (1 - p_b)}$. The approximation will be good if $n_b p_b$ and $n_b(1 - p_b)$ are greater than or equal to 5.
In our question we are required to use the Q-function, and we already have $\mu_p = \lambda t = n_b p_b = 10 > 5$, hence we need to approximate the Poisson distribution by a Gaussian one. Hence, using the mean and variance of the Poisson distribution found on page 2 of the Exam sheet, we approximate $X \sim n(x; \mu_p, \sqrt{\mu_p})$ with $\mu_p = \lambda t$.

$$
\begin{aligned}
P(X \geq 100) &= P(X \geq 100; \lambda y | Y = 9) P(Y = 9) + P(X \geq 100; \lambda y | Y = 11) P(Y = 11) \\
&\overset{(a)}{\simeq} P\left(Z \geq \frac{100 + 0.5 - 9\lambda}{\sqrt{9\lambda}}\right)(0.5) + P\left(Z \geq \frac{100 + 0.5 - 11\lambda}{\sqrt{11\lambda}}\right)(0.5) \\
&= P\left(Z \geq \frac{10.5}{\sqrt{90}}\right)(0.5) + P\left(Z \geq \frac{-9.5}{\sqrt{110}}\right)(0.5) \\
&= 0.5Q\left(\frac{10.5}{\sqrt{90}}\right) + 0.5Q\left(\frac{-9.5}{\sqrt{110}}\right) \\
&= 0.0671 + 0.4088 = 0.4759
\end{aligned} \tag{16}
$$

where $Z = (X - \mu)/\sigma$ is the standard Normal distribution, the $+0.5$ term is a continuity correction term that is introduced to be more accurate when we seek the area under the normal curve, and $Q(\cdot)$ is the Q-function. The step in (a) follows from approximating the Poisson distribution by a Gaussian distribution and then transforming X into the standard Gaussian distribution.
(b)

$$
\begin{aligned}
P(Y = 9 | X = 100) &= \frac{P(Y = 9, X = 100)}{P(X = 100)} \\
&= \frac{P(X = 100 | Y = 9) P(Y = 9)}{P(X = 100 | Y = 9) P(Y = 9) + P(X = 100 | Y = 11) P(Y = 11)} \\
&= \frac{0.5n(x; \mu_1, \sigma_1)}{0.5n(x; \mu_1, \sigma_1) + 0.5n(x; \mu_2, \sigma_2)}
\end{aligned} \tag{17}
$$

with $x = 100$, $\mu_1 = 9\lambda$, $\sigma_1 = \sqrt{9\lambda}$, $\mu_2 = 11\lambda$, and $\sigma_2 = \sqrt{11\lambda}$.

Question 11

The 95% confidence interval for the mean corresponds to $\alpha = 0.05$. A 95% confidence interval for μ for a population with unknown variance is given by

$$
\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \tag{18}
$$

where $t_{\alpha/2}$ is the t-value with $\nu = n - 1 = 4$ degrees of freedom leaving an area of $\alpha/2$ to the right. The average time it takes to reach the office recovered from the sample is $\bar{x} = (1/5)(31+27+31+32+29) = 30$ and the variance is $\sigma^2 = (\frac{1}{5-1})((31-30)^2+(27-30)^2+(31-30)^2+(32-30)^2+(29-30)^2) = 4$, hence the standard deviation is $s = 2$.
From Table A.4, we have $t_{0.025} = 2.776$ for 4 degrees of freedom. The confidence interval is

$$\begin{aligned} &30 - (2.776)\left(\frac{2}{\sqrt{5}}\right) < \mu < 30 + (2.776)\left(\frac{2}{\sqrt{5}}\right) \\ &\rightarrow 27.5171 < \mu < 32.4829 \end{aligned} \tag{19}$$