ECS 315: Probability and Random Processes **2019/1**

HW Solution 12 — Due: Not Due

Lecturer: Prapun Suksompong, Ph.D.

Problem 1. A random variable X is a Gaussian random variable if its pdf is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2},$$

for some constant m and positive number σ. Furthermore, when a Gaussian random variable has $m = 0$ and $\sigma = 1$, we say that it is a standard Gaussian random variable. There is no closed-form expression for the cdf of the standard Gaussian random variable. The cdf itself is denoted by Φ and its values (or its complementary values $Q(\cdot) = 1-\Phi(\cdot)$) are traditionally provided by a table.

Suppose Z is a standard Gaussian random variable.

(a) Use the Φ table to find the following probabilities:

(i) $P[Z < 1.52]$

(ii) $P[Z < -1.52]$

(iii) $P[Z > 1.52]$

(iv) $P[Z > -1.52]$

(v) $P[-1.36 < Z < 1.52]$

(b) Use the Φ table to find the value of c that satisfies each of the following relation.

(i) $P[Z > c] = 0.14$

(ii) $P[-c < Z < c] = 0.95$

Solution:

(a)

(i) $P[Z < 1.52] = \Phi(1.52) = \boxed{0.9357.}$

(ii) $P[Z < -1.52] = \Phi(-1.52) = 1 - \Phi(1.52) = 1 - 0.9357 = \boxed{0.0643.}$

(iii) $P[Z > 1.52] = 1 - P[Z < 1.52] = 1 - \Phi(1.52) = 1 - 0.9357 = \boxed{0.0643.}$

(iv) It is straightforward to see that the area of $P[Z > -1.52]$ is the same as $P[Z < 1.52] = \Phi(1.52)$. So, $P[Z > -1.52] = \boxed{0.9357.}$
Alternatively, $P[Z > -1.52] = 1 - P[Z \le -1.52] = 1 - \Phi(-1.52) = 1 - (1 - \Phi(1.52)) = \Phi(1.52)$.

(v) $P[-1.36 < Z < 1.52] = \Phi(1.52) - \Phi(-1.36) = \Phi(1.52) - (1 - \Phi(1.36)) = \Phi(1.52) + \Phi(1.36) - 1 = 0.9357 + 0.9131 - 1 = \boxed{0.8488.}$

(b)

(i) $P[Z > c] = 1 - P[Z \leq c] = 1 - \Phi(c)$. So, we need $1 - \Phi(c) = 0.14$ or $\Phi(c) = 1 - 0.14 = 0.86$. In the Φ table, we do not have exactly 0.86, but we have 0.8599 and 0.8621. Because 0.86 is closer to 0.8599, we answer the value of c whose $\phi(c) = 0.8599$. Therefore, $c \approx \boxed{1.08.}$

(ii) $P[-c < Z < c] = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1$. So, we need $2\Phi(c) - 1 = 0.95$ or $\Phi(c) = 0.975$. From the Φ table, we have $c \approx \boxed{1.96.}$

Problem 2. The peak temperature T, as measured in degrees Fahrenheit, on a July day in New Jersey is a $\mathcal{N}(85, 100)$ random variable.

Remark: Do not forget that, for our class, the second parameter in $\mathcal{N}(\cdot, \cdot)$ is the variance (not the standard deviation).

(a) Express the cdf of T in terms of the Φ function.

(b) Express each of the following probabilities in terms of the Φ function(s). Make sure that the arguments of the Φ functions are positive. (Positivity is required so that we can directly use the Φ/Q tables to evaluate the probabilities.)

(i) $P[T > 100]$

(ii) $P[T < 60]$

(iii) $P[70 \leq T \leq 100]$

(c) Express each of the probabilities in part (b) in terms of the Q function(s). Again, make sure that the arguments of the Q functions are positive.

(d) Evaluate each of the probabilities in part (b) using the Φ/Q tables.

(e) Observe that the Φ table ("Table 4" from the lecture) stops at $z = 2.99$ and the Q table ("Table 5" from the lecture) starts at $z = 3.00$. Why is it better to give a table for $Q(z)$ instead of $\Phi(z)$ when z is large?

Solution:

(a) Recall that when $X \sim \mathcal{N}(m, \sigma^2)$, $F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right)$. Here, $T \sim \mathcal{N}(85, 10^2)$. Therefore,

$$F_T(t) = \boxed{\Phi\left(\frac{t-85}{10}\right).}$$

(b)

(i) $P[T > 100] = 1 - P[T \le 100] = 1 - F_T(100) = 1 - \Phi\left(\frac{100-85}{10}\right) = 1 - \Phi(1.5)$

(ii) $P[T < 60] = P[T \le 60]$ because T is a continuous random variable and hence $P[T = 60] = 0$. Now, $P[T \le 60] = F_T(60) = \Phi\left(\frac{60-85}{10}\right) = \Phi(-2.5) = \boxed{1 - \Phi(2.5).}$ Note that, for the last equality, we use the fact that $\Phi(-z) = 1 - \Phi(z)$.

(iii)

$$P[70 \le T \le 100] = F_T(100) - F_T(70) = \Phi\left(\frac{100-85}{10}\right) - \Phi\left(\frac{70-85}{10}\right)$$
$$= \Phi(1.5) - \Phi(-1.5) = \Phi(1.5) - (1 - \Phi(1.5)) = \boxed{2\Phi(1.5) - 1.}$$

(c) In this question, we use the fact that $Q(x) = 1 - \Phi(x)$.

(i) $1 - \Phi(1.5) = \boxed{Q(1.5).}$

(ii) $1 - \Phi(2.5) = \boxed{Q(2.5).}$

(iii) $2\Phi(1.5) - 1 = 2(1 - Q(1.5)) - 1 = 2 - 2Q(1.5) - 1 = \boxed{1 - 2Q(1.5).}$

(d)

(i) $1 - \Phi(1.5) = 1 - 0.9332 = \boxed{0.0668.}$

(ii) $1 - \Phi(2.5) = 1 - 0.99379 = \boxed{0.0062.}$

(iii) $2\Phi(1.5) - 1 = 2(0.9332) - 1 = \boxed{0.8664.}$

(e) When z is large, $\Phi(z)$ will start with 0.999... The first few significant digits will all be the same and hence not quite useful to be there.

Problem 3. Suppose that the time to failure (in hours) of fans in a personal computer can be modeled by an exponential distribution with $\lambda = 0.0003$.

(a) What proportion of the fans will last at least 10,000 hours?

(b) What proportion of the fans will last at most 7000 hours?

[Montgomery and Runger, 2010, Q4-97]

Solution: Let T be the time to failure (in hours). We are given that $T \sim \mathcal{E}(\lambda)$ where $\lambda = 3 \times 10^{-4}$. Therefore,

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Here, we want to find $P\left[T > 10^4\right]$.

We shall first provide the general formula for the ccdf $P\left[T > t\right]$ when $t > 0$:

$$P\left[T > t\right] = \int_t^{\infty} f_T\left(\tau\right) d\tau = \int_t^{\infty} \lambda e^{-\lambda\tau} d\tau = -e^{-\lambda\tau}\Big|_t^{\infty} = e^{-\lambda t}. \tag{12.1}$$

Therefore,

$$P\left[T > 10^4\right] = e^{-3\times10^{-4}\times10^4} = \boxed{e^{-3} \approx 0.0498.}$$

(b) We start with $P\left[T \le 7000\right] = 1 - P\left[T > 7000\right]$. Next, we apply (12.1) to get

$$P\left[T \le 7000\right] = 1 - P\left[T > 7000\right] = 1 - e^{-3\times10^{-4}\times7000} = \boxed{1 - e^{-2.1} \approx 0.8775.}$$

Problem 4. Let a continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the probability density function of X is

$$f_X\left(x\right) = \begin{cases} 5, & 4.9 \le x \le 5.1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the probability that a current measurement is less than 5 milliamperes.

(b) Find and plot the cumulative distribution function of the random variable X.

(c) Find the expected value of X.

(d) Find the variance and the standard deviation of X.

(e) Find the expected value of power when the resistance is 100 ohms?

Solution:

(a) $P\left[X < 5\right] = \int_{-\infty}^{5} f_X\left(x\right) dx = \int_{-\infty}^{0} \underbrace{f_X\left(x\right)}_{0} dx + \int_{0}^{5} \underbrace{f_X\left(x\right)}_{5} dx = 0 + 5x|_{x=4.9}^{5} = \boxed{0.5}.$

(b) By definition, $F_X(x) = P\left[X \le x\right] = \int_{-\infty}^{x} f_X(t)dt$.

- For $x < 4.9$, $f_X(t) = 0$ for all t inside $(-\infty, 4.9)$. Therefore, $F_X(x) = \int_{-\infty}^{x} 0dt = 0$.
- For $4.9 \le x \le 5.1$,
  $F_X\left(x\right) = \int_{-\infty}^{x} f_X\left(t\right) dt = \int_{-\infty}^{4.9} \underbrace{f_X\left(t\right)}_{0} dt + \int_{4.9}^{x} \underbrace{f_X\left(t\right)}_{5} dt = 0 + 5t|_{t=4.9}^{x} = 5x - 24.5.$

- For $x > 5.1$,
  $F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-\infty}^{4.9} \underbrace{f_X(t)}_{0}dt + \int_{4.9}^{5.1} \underbrace{f_X(t)}_{5}dt + \int_{5.1}^{x} \underbrace{f_X(t)}_{0}dt = 0 + 5t|_{t=4.9}^{5.1} + 0 = 1.$

Combining the three cases above, we have the complete description of the cdf:

$$F_X(x) = \begin{cases} 0, & x < 4.9, \\ 5x - 24.5, & 4.9 \le x \le 5.1, \\ 1, & x > 5.1. \end{cases}$$

The corresponding plot is shown in Figure 12.1. Note that $F_X(x)$ is a continuous function; this is expected because X is a continuous RV.

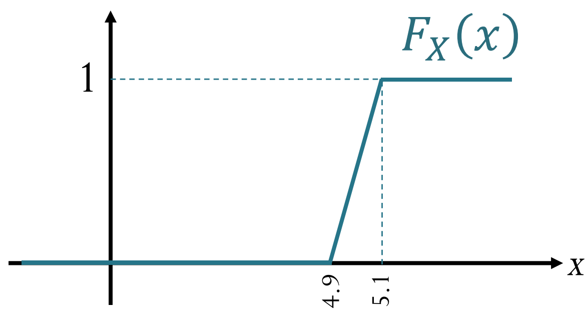

Figure 12.1: Plot of cdf for Problem 4.

(c) $\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{4.9} x \underbrace{f_X(x)}_{0}dx + \int_{4.9}^{5.1} x \underbrace{f_X(x)}_{5}dx + \int_{5.1}^{x} x \underbrace{f_X(x)}_{0}dx = 0 + 5\frac{x^2}{2}\Big|_{x=4.9}^{5.1} + 0 = \boxed{5}$ mA.

Alternatively, for $X \sim \mathcal{U}(a,b)$, we have $\mathbb{E}X = \frac{a+b}{2} = \frac{4.9+5.1}{2} = 5$.

(d) $\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2$. From the previous part, we know that $\mathbb{E}X = 5$. SO, to find $\operatorname{Var} X$, we need to find $\mathbb{E}[X^2]$: $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_{-\infty}^{4.9} x^2 \underbrace{f_X(x)}_{0}dx + \int_{4.9}^{5.1} x^2 \underbrace{f_X(x)}_{5}dx + \int_{5.1}^{x} x^2 \underbrace{f_X(x)}_{0}dx = 0 + 5\frac{x^3}{3}\Big|_{x=4.9}^{5.1} + 0 = 25 + \frac{1}{300}$.

Therefore, $\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \left(25 + \frac{1}{300}\right) = \boxed{\frac{1}{300}} \approx 0.0033\ (\text{mA})^2$

and

$$\sigma_X = \sqrt{\operatorname{Var} X} = \boxed{\frac{1}{10\sqrt{3}}} \approx 0.0577 \text{ mA}.$$

Alternatively, for $X \sim \mathcal{U}(a,b)$, we have $\operatorname{Var} X = \frac{(b-a)^2}{12} = \frac{(5.1-4.9)^2}{12} = \frac{1}{300}$.

(e) Recall that $P = I \times V = I^2 r$. Here, $I = X$. Therefore, $P = X^2 r$ and $\mathbb{E}P = \mathbb{E}\left[X^2 r\right] = r\mathbb{E}\left[X^2\right] = 100 \times \left(25 + \frac{1}{300}\right) = 2500 + \frac{1}{3} \approx 2.50033 \times 10^3$ [(mA)$^2\Omega$]. Factoring out m^2, we have $\mathbb{E}P \approx 2.50033$ mW. ([A$^2\Omega$] = [W].)

Problem 5. Let X be a uniform random variable on the interval $[0,1]$. Set

$$A = \left[0, \frac{1}{2}\right), \quad B = \left[0, \frac{1}{4}\right) \cup \left[\frac{1}{2}, \frac{3}{4}\right), \quad \text{and } C = \left[0, \frac{1}{8}\right) \cup \left[\frac{1}{4}, \frac{3}{8}\right) \cup \left[\frac{1}{2}, \frac{5}{8}\right) \cup \left[\frac{3}{4}, \frac{7}{8}\right).$$

Are the events $[X \in A], [X \in B]$, and $[X \in C]$ independent?

Solution: Note that

$$P\left[X \in A\right] = \int_0^{\frac{1}{2}} dx = \frac{1}{2},$$

$$P\left[X \in B\right] = \int_0^{\frac{1}{4}} dx + \int_{\frac{1}{2}}^{\frac{3}{4}} dx = \frac{1}{2}, \text{ and}$$

$$P\left[X \in C\right] = \int_0^{\frac{1}{8}} dx + \int_{\frac{1}{4}}^{\frac{3}{8}} dx + \int_{\frac{1}{2}}^{\frac{5}{8}} dx + \int_{\frac{3}{4}}^{\frac{7}{8}} dx = \frac{1}{2}.$$

Now, for pairs of events, we have

$$P\left([X \in A] \cap [X \in B]\right) = \int_0^{\frac{1}{4}} dx = \frac{1}{4} = P\left[X \in A\right] \times P\left[X \in B\right], \tag{12.2}$$

$$P\left([X \in A] \cap [X \in C]\right) = \int_0^{\frac{1}{8}} dx + \int_{\frac{1}{4}}^{\frac{3}{8}} dx = \frac{1}{4} = P\left[X \in A\right] \times P\left[X \in C\right], \text{ and} \tag{12.3}$$

$$P\left([X \in B] \cap [X \in C]\right) = \int_0^{\frac{1}{8}} dx + \int_{\frac{1}{2}}^{\frac{5}{8}} dx = \frac{1}{4} = P\left[X \in B\right] \times P\left[X \in C\right]. \tag{12.4}$$

Finally,

$$P([X \in A] \cap [X \in B] \cap [X \in C]) = \int_0^{\frac{1}{8}} dx = \frac{1}{8} = P[X \in A] P[X \in B] P[X \in C]. \quad (12.5)$$

From (12.2), (12.3), (12.4) and (12.5), we can conclude that the events $[X \in A], [X \in B]$, and $[X \in C]$ are $\boxed{\text{independent}}$.

Problem 6. Cholesterol is a fatty substance that is an important part of the outer lining (membrane) of cells in the body of animals. Its normal range for an adult is 120–240 mg/dl. The Food and Nutrition Institute of the Philippines found that the total cholesterol level for Filipino adults has a mean of 159.2 mg/dl and 84.1% of adults have a cholesterol level below 200 mg/dl. Suppose that the cholesterol level in the population is normally distributed.

(a) Determine the standard deviation of this distribution.

(b) What is the value of the cholesterol level that exceeds 90% of the population?

(c) An adult is at moderate risk if cholesterol level is more than one but less than two standard deviations above the mean. What percentage of the population is at moderate risk according to this criterion?

(d) An adult is thought to be at high risk if his cholesterol level is more than two standard deviations above the mean. What percentage of the population is at high risk?

Solution: Let X be the cholesterol level of a randomly chosen adult. It is given that $X \sim \mathcal{N}(m, \sigma^2)$ where $m = 159.2$ mg/dl. We also know that $P[X < 200] = 0.841$.

(a) For any Gaussian random variable, $P[X < 200] = \Phi\left(\frac{200-m}{\sigma}\right)$. It is given that this probability should be 0.841. Our plan is then to first find the number z whose $\Phi(z) = 0.841$. Then, solve for σ from $z = \frac{200-m}{\sigma}$.

From the Φ table, $\Phi(0.99) \approx 0.8389$ and $\Phi(1) \approx 0.8413$. Because 0.841 is closer to 0.8413 than 0.8389, we conclude that the value of z that makes $\Phi(z) = 0.841$ is $z \approx 1$.

From $z = \frac{200-m}{\sigma}$, plugging-in $z \approx 1$ gives $\sigma \approx 200 - m = 200 - 159.2 = \boxed{40.8 \text{ mg/dl.}}$

(b) Here, we want to find the value of x such that $P[X \le x] = 0.9$. (90% of the population has cholesterol level lower than this x.)

For any Gaussian random variable, $P[X \le x] = \Phi\left(\frac{x-m}{\sigma}\right)$. Our plan is then to first find the number z whose $\Phi(z) = 0.9$. Then, solve for x from $z = \frac{x-m}{\sigma}$.

From the Φ table, $\Phi(1.28) \approx 0.8997$ and $\Phi(1.29) \approx 0.9015$. Because 0.9 is closer to 0.8997 than 0.9015, we conclude that the value of z that makes $\Phi(z) = 0.9$ is $z \approx 1.28$.

From $z = \frac{x-m}{\sigma}$, plugging-in $z \approx 1.28$ gives $x \approx 1.28\sigma + m \approx \boxed{211.424 \text{ mg/dl.}}$

(c) Here we want to find the probability that $m+\sigma < X < m+2\sigma$:

$$P[m+\sigma < X < m+2\sigma] = F_X(m+2\sigma) - F_X(m+\sigma).$$

For any Gaussian random variable, $F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right)$. Therefore,

$$\begin{aligned} P[m+\sigma < X < m+2\sigma] &= \Phi\left(\frac{(m+2\sigma)-m}{\sigma}\right) - \Phi\left(\frac{(m+\sigma)-m}{\sigma}\right) \\ &= \Phi(2) - \Phi(1) \approx 0.97725 - 0.8413 \approx \boxed{0.1359 = 13.59\%} \end{aligned}$$

(d) $P[X > m+2\sigma] = 1 - P[X \le m+2\sigma] = 1 - F_X(m+2\sigma) = 1 - \Phi\left(\frac{(m+2\sigma)-m}{\sigma}\right) = 1 - \Phi(2) \approx 1 - 0.97725 \approx \boxed{0.0228 = 2.28\%}$.

Problem 7 (Q3.5.6)**.** Solve this question using the Φ/Q table.

A professor pays 25 cents for each blackboard error made in lecture to the student who points out the error. In a career of n years filled with blackboard errors, the total amount in dollars paid can be approximated by a Gaussian random variable Y_n with expected value $40n$ and variance $100n$.

(a) What is the probability that Y_{20} exceeds 1000?

(b) How many years n must the professor teach in order that $P[Y_n > 1000] > 0.99$?

Solution: We are given[1] that $Y_n \sim \mathcal{N}(40n, 100n)$. Recall that when $X \sim \mathcal{N}(m, \sigma^2)$,

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right). \tag{12.6}$$

(a) Here $n = 20$. So, we have $Y_n \sim \mathcal{N}(40 \times 20, 100 \times 20) = \mathcal{N}(800, 2000)$. For this random variable $m = 800$ and $\sigma = \sqrt{2000}$.

We want to find $P[Y_{20} > 1000]$ which is the same as $1 - P[Y_20 \le 1000]$. Expressing this quantity using cdf, we have

$$P[Y_{20} > 1000] = 1 - F_{Y_{20}}(1000).$$

Apply (12.6) to get

$$P[Y_{20} > 1000] = 1 - \Phi\left(\frac{1000-800}{\sqrt{2000}}\right) = 1 - \Phi(4.472) \approx Q(4.47) \approx \boxed{3.91 \times 10^{-6}}.$$

[1]Note that the expected value and the variance in this question are proportional to n. This naturally occurs when we consider the sum of i.i.d. random variables. The approximation by Gaussian random variable is a result of the central limit theorem (CLT).

(b) Here, the value of n is what we want. So, we will need to keep the formula in the general form. Again, from (12.6), for $Y_n \sim \mathcal{N}(40n, 100n)$, we have

$$P[Y_n > 1000] = 1 - F_{Y_n}(1000) = 1 - \Phi\left(\frac{1000 - 40n}{10\sqrt{n}}\right) = 1 - \Phi\left(\frac{100 - 4n}{\sqrt{n}}\right).$$

To find the value of n such that $P[Y_n > 1000] > 0.99$, we will first find the value of z which make

$$1 - \Phi(z) > 0.99. \tag{12.7}$$

At this point, we may try to solve for the value of Z by noting that (12.7) is the same as

$$\Phi(z) < 0.01. \tag{12.8}$$

Unfortunately, the tables that we have start with $\Phi(0) = 0.5$ and increase to something close to 1 when the argument of the Φ function is large. This means we can't directly find 0.01 in the table. Of course, 0.99 is in there and therefore we will need to solve (12.7) via another approach.

To do this, we use another property of the Φ function. Recall that $1 - \Phi(z) = \Phi(-z)$. Therefore, (12.7) is the same as

$$\Phi(-z) > 0.99. \tag{12.9}$$

From our table, we can then conclude that (12.8) (which is the same as (12.9)) will happen when $-z > 2.33$. (If you have `MATLAB`, then you can get a more accurate answer of 2.3263.)

Now, plugging in $z = \frac{100-4n}{\sqrt{n}}$, we have $\frac{4n-100}{\sqrt{n}} > 2.33$. To solve for n, we first let $x = \sqrt{n}$. In which case, we have $\frac{4x^2-100}{x} > 2.33$ or, equivalently, $4x^2 - 2.33x - 100 > 0$. The two roots are $x = -4.717$ and $x > 5.3$. So, We need $x < -4.717$ or $x > 5.3$. Note that $x = \sqrt{n}$ and therefore can not be negative. So, we only have one case; that is, we need $x > 5.3$. Because $n = x^2$, we then conclude that we need $\boxed{n > 28.1 \text{ years}}$.

Problem 8. The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function

$$F_X(x) = \begin{cases} 1 - e^{-0.01x}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine the probability density function of X.

(b) What proportion of reactions is complete within 200 milliseconds?

Solution: Note that the cdf $F_X(x)$ is a continuous function. Therefore, X is a continuous RV.

(a) $f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 0.01 e^{-0.01x}, & x > 0, \\ 0, & x < 0. \end{cases}$

At $x = 0$, the derivative does not exist. However, X is a continuous RV. Therefore, we can assign $f_X(0)$ to be any arbitrary value. Here, we set $f_X(0) = 0$:

$$f_X(x) = \boxed{\begin{cases} 0.01 e^{-0.01x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}}$$

(b) $P[X < 200] = P[X \le 200] = F_X(200) = 1 - e^{-0.01 \times 200} = \boxed{1 - e^{-2} \approx 0.8647.}$