MATH 3215 Assignment 7

Express answers in terms of the CDF Φ of $\mathcal{N}(0,1)$ when appropriate.

1. (a) Consider independent random variables $X_1 \sim \mathsf{Poi}(2)$ and $X_2 \sim \mathsf{Poi}(3)$. What is the probability that $X_1 + X_2 \le 1$?

(b) Consider independent random variables $X_1 \sim \mathsf{Bin}(2, 0.2)$ and $X_2 \sim \mathsf{Bin}(3, 0.2)$. What is the probability that $X_1 + X_2 \le 1$?

(c) Consider independent random variables $X_1 \sim \mathcal{N}(1, 2)$ and $X_2 \sim \mathcal{N}(-1, 3)$. What is the probability that $X_1 + X_2 \le 1$?

(a) Since $X_1 + X_2 \sim \mathsf{Poi}(5)$, we have $\mathbb{P}\{X_1 + X_2 \le 1\} = 6\,e^{-5}$.

(b) Since $X_1 + X_2 \sim \mathsf{Bin}(5, 0.2)$, we have $\mathbb{P}\{X_1 + X_2 \le 1\} = 0.8^5 + 5 \cdot 0.2 \cdot 0.8^4$.

(c) Since $X_1 + X_2 \sim \mathcal{N}(0, 5)$, we have $\mathbb{P}\{X_1 + X_2 \le 1\} = \mathbb{P}\left\{\frac{X_1+X_2}{\sqrt{5}} \le \frac{1}{\sqrt{5}}\right\} = \Phi\left(\frac{1}{\sqrt{5}}\right)$.

2. The score of a student in an exam is assumed to be a random variable with mean 75.

(a) Use Markov's inequality to give an upper bound on the probability that the score will exceed 85.

(b) Suppose in addition that the variance of the score is equal to 25. Use Chebyshev's inequality to give a lower bound on the probability that the score will be between 65 and 85.

(a) We have $\mathbb{P}\{X > 85\} \le \frac{75}{85} = \frac{15}{17}$.

(b) We have $\mathbb{P}\{65 \le X \le 86\} = 1 - \mathbb{P}\{|X - 75| > 10\} \ge 1 - \frac{25}{100} = \frac{3}{4}$.

3. Roll 10 dice. Let X be the sum of the 10 values observed. Compute the mean and the variance of X. Then, approximate the probability that X is between 30 and 40 (both inclusive) in terms of the CDF Φ of $\mathcal{N}(0,1)$. (Use the correction for continuity.)

Let X be the sum. We have $\mathbb{E}[X] = 35$ and $\mathsf{Var}(X) = 10\sum_{i=1}^{6}(i - 3.5)^2/6 = 175/6$. Therefore,

$$\mathbb{P}\{29.5 \le X \le 40.5\} = \mathbb{P}\left\{\frac{29.5 - 35}{\sqrt{175/6}} \le \frac{X - 35}{\sqrt{175/6}} \le \frac{40.5 - 35}{\sqrt{175/6}}\right\} = \Phi\left(\frac{5.5}{\sqrt{175/6}}\right) - \Phi\left(\frac{-5.5}{\sqrt{175/6}}\right).$$

The numerical value is about 0.6915.

4. The amount of nicotine in a cigarette produced by a tobacco company is a random variable with mean 2.2 mg and standard deviation 0.3 mg. Taking 100 randomly chosen cigarettes, let X_i denote the nicotine content of the ith cigarette for $i = 1, \dots, 100$, and let $\bar{X} := \frac{1}{100}\sum_{i=1}^{100} X_i$ be the sample mean of the nicotine content. Approximate the probability that $\bar{X}$ is higher than 2.3 in terms of the CDF Φ of $\mathcal{N}(0,1)$.

We have $\mathbb{E}[\bar{X}] = 2.2$ and $\mathsf{Var}(\bar{X}) = \frac{0.3^2}{100}$. Therefore,

$$\mathbb{P}\{\bar{X} > 2.3\} = \mathbb{P}\left\{\frac{\bar{X} - 2.2}{0.03} > \frac{2.3 - 2.2}{0.03}\right\} = 1 - \Phi(10/3) \approx 0.000429.$$

5. There are two classes, one of size 36 and the other of size 64. The exam scores of students in these classes are i.i.d. random variables with mean 77 pts and standard deviation 15 pts.

(a) What normal distribution approximates the average test score in the class of size 36?

(b) What normal distribution approximates the average test score in the class of size 64?

(c) What is the approximate probability that the average test score in the class of size 36 is higher than that in the class of size 64 by (at least) 5 pts? (No need to use the correction for continuity.)

(d) Suppose the average scores in the two classes are 76 and 83. Which class do you think was more likely to have averaged 83? (It suffices to explain the answer intuitively.)

(a) $\mathcal{N}(77, \frac{15^2}{36})$

(b) $\mathcal{N}(77, \frac{15^2}{64})$

(c) Let the two average scores be denoted by X and Y respectively. Then

$$\mathbb{P}\{X - Y \geq 5\} = \mathbb{P}\left\{\frac{X-Y}{\sqrt{\frac{15^2}{36} + \frac{15^2}{64}}} \geq \frac{5}{\sqrt{\frac{15^2}{36} + \frac{15^2}{64}}}\right\} \approx 1 - \Phi(8/5).$$

(d) The class with 36 students, because the variance of the sample average is larger.

6. Let $X \sim \mathsf{Bin}(150, 0.6)$. Using R or any software of your choice, compute the probability that $X \leq 80$ in the following ways:

(a) the exact probability;

(b) the normal approximation using the correction for continuity;

(c) the normal approximation without using the correction for continuity.

For each value, show 5 digits after the decimal point. (No need to show the code.)

We have $\mathbb{E}[X] = 90$ and $\mathsf{Var}(X) = 36$. The following R code computes the three values:

```
pbinom(80,150,0.6)
pnorm(80.5,90,6)
pnorm(80,90,6)
```

The results are (a) 0.05746; (b) 0.05667; (c) 0.04779. Clearly, the correction for continuity is important.

7. Using R or any software of your choice, plot the following two curves in one figure: the PDF of a t-distribution with 10 degrees of freedom, and the PDF of a standard normal distribution, both in the range $[-4, 4]$. (Distinguish the two curves using different colors or line styles. No need to show the code.) Then explain why they are close using the law of large numbers.

The following code works in R:

```
x = c(-40:40)/10
y = dt(x, 10)
z = dnorm(x)
plot(x, y, type="p", col="blue", pch=4)
lines(x, z, type="p", col="red")
```

An explanation of why they are close: The denominator of a t-random variable in the canonical form is close to 1 by the law of large numbers.

8. Using R or any software of your choice, plot the following two curves in one figure: the PMF of $\mathsf{Poi}(20)$ and the PDF of $\mathcal{N}(20, 20)$, both in the range $[10, 30]$. (Distinguish the two curves using different colors or line styles. No need to show the code.) Then explain why they are close using the central limit theorem.

The following code works in R:

```
x = c(10:30)
y = dpois(x, 20)
z = dnorm(x, 20, sqrt(20))
plot(x, y, type="p", col="blue", pch=4)
lines(x, z, type="p", col="red")
```

An explanation of why they are close: A random variable $X \sim \mathsf{Poi}(20)$ can be written as $X = \sum_{i=1}^{20} X_i$ for i.i.d. $X_i \sim \mathsf{Poi}(1)$. Moreover, $\mathsf{Poi}(20)$ has the same mean and variance as $\mathcal{N}(20, 20)$.

9. (a) Let X be a continuous random variable taking values in (a, b) where $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R} \cup \{\infty\}$. (This simply means that we allow the sample space to be (a, ∞), $(-\infty, b)$, or $\mathbb{R}$.) Suppose that the CDF $F(x)$ of X is strictly increasing on (a, b). Show that the random variable $Y := F(X)$ has the uniform distribution over $(0, 1)$.

(b) Let Y be a uniform random variable over $(0, 1)$. Let $F : \mathbb{R} \to [0, 1]$ be a continuous and increasing function that is strictly increasing on (a, b) and satisfies $F(a) = 0$ and $F(b) = 1$. (We allow $a = -\infty$ or $b = \infty$. If $a = -\infty$, this means that $F(x) \to 0$ as $x \to -\infty$; similarly for $b = \infty$.) As a result, the inverse $F^{-1} : (0, 1) \to (a, b)$ is well-defined. Show that the CDF of the random variable $X := F^{-1}(Y)$ is equal to $F(x)$.

(Hint: This problem is not as difficult as it might seem. Both parts have very short proofs. All you need is using the definition of the CDF and inverting functions when necessary. Note that by part (b), we can generate a random variable with any distribution as long as the CDF of that distribution is given.)

(a) We have

$$\mathbb{P}\{Y \le y\} = \mathbb{P}\{F(X) \le y\} = \mathbb{P}\{X \le F^{-1}(y)\} = F(F^{-1}(y)) = y$$

for $y \in (0, 1)$, so the conclusion follows.

(b) We have

$$\mathbb{P}\{X \le x\} = \mathbb{P}\{F^{-1}(Y) \le x\} = \mathbb{P}\{Y \le F(x)\} = F(x).$$