CSE 312: Foundations of Computing II
Additional Problems #8: Normal Distribution, Central Limit Theorem (solutions)

1. Before putting any bets down on roulette, you watch 100 rounds, each of which results in an integer between 1 and 36. You count how many rounds have a result that is odd and, if the count exceeds 55, you decide the roulette wheel is unfair. Assuming the roulette wheel is fair, approximate the probability that you make the wrong decision.

   Let X be the number of rounds whose result is odd. If the roulette wheel is fair, then $X \sim \mathrm{Bin}(100, 0.5)$. $\mathbb{E}[X] = 100 \times 0.5 = 50$ and $\mathrm{Var}(X) = 100 \times 0.5 \times (1 - 0.5) = 25$. You will decide the roulette wheel is unfair if and only if $X > 55$.

$$\begin{aligned} \mathbb{P}(X > 55) &= \mathbb{P}(X > 55.5) = \mathbb{P}\left(\frac{X-50}{5} > \frac{55.5-50}{5}\right) \\ &\approx 1 - \Phi\left(\frac{55.5-50}{5}\right) = 1 - \Phi(1.1) \approx 1 - 0.8643 = 0.1357 \end{aligned}$$

2. Suppose 59 percent of voters favor Proposition 666. Use the Normal approximation to estimate the probability that a random sample of 100 voters will contain:

   Let X be the number of voters in the random sample that favor Proposition 666. Then $X \sim \mathrm{Bin}(100, 0.59)$, with $\mathbb{E}[X] = 100 \times 0.59 = 59$ and $\mathrm{Var}(X) = 100 \times 0.59 \times 0.41 = 24.19$.

   (a) at most 50 in favor.

$$\mathbb{P}(X \leq 50) = \mathbb{P}(X < 50.5) = \mathbb{P}\left(\frac{X-59}{\sqrt{24.19}} < \frac{50.5-59}{\sqrt{24.19}}\right) \approx \mathbb{P}\left(\frac{X-59}{\sqrt{24.19}} < -1.73\right) \approx 1-\Phi(1.73) \approx 0.0418$$

   (b) between 54 and 64 (inclusive) in favor.

$$\begin{aligned} \mathbb{P}(54 \leq X \leq 64) &= \mathbb{P}(53.5 < X < 64.5) = \mathbb{P}\left(\frac{53.5-59}{\sqrt{24.19}} < \frac{X-59}{\sqrt{24.19}} < \frac{64.5-59}{\sqrt{24.19}}\right) \\ &\approx \mathbb{P}\left(-1.12 < \frac{X-59}{\sqrt{24.19}} < 1.12\right) \approx 2\Phi(1.12) - 1 \approx 0.7372 \end{aligned}$$

   (c) fewer than 72 in favor.

$$\mathbb{P}(X < 72) = \mathbb{P}(X < 71.5) = \mathbb{P}\left(\frac{X-59}{\sqrt{24.19}} < \frac{71.5-59}{\sqrt{24.19}}\right) \approx \mathbb{P}\left(\frac{X-59}{\sqrt{24.19}} < 2.54\right) \approx \Phi(2.54) \approx 0.9945$$

3. Each day, the probability your computer crashes is 10%, independent of every other day. Approximate the probability of at least 87 crash-free days out of the next 100 days.

Let X be the number of crash-free days in the next 100 days. Then $X \sim \text{Bin}(100, 0.9)$, with $\mathbb{E}[X] = 100 \times 0.9 = 90$ and $\text{Var}(X) = 100 \times 0.9 \times 0.1 = 9$.

$$\begin{aligned}\mathbb{P}(X \geq 87) &= \mathbb{P}(86.5 < X < 100.5) = \mathbb{P}\left(\frac{86.5 - 90}{3} < \frac{X - 90}{3} < \frac{100.5 - 90}{3}\right) \\ &\approx \mathbb{P}\left(-1.17 < \frac{X - 90}{3} < 3.5\right) \approx \Phi(3.5) + \Phi(1.17) - 1 \approx 0.9998 + 0.8790 - 1 = 0.8788\end{aligned}$$

Notice that, if you had used $86.5 < X$ in place of $86.5 < X < 100.5$, your answer would have been nearly the same, because $\Phi(3.5)$ is so close to 1.

4. Suppose $Z = X + Y$, where $X \perp Y$. Z is called the convolution of two random variables. If X, Y, Z are discrete,

$$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x \mathbb{P}(X = x \cap Y = z - x) = \sum_x p_X(x)\, p_Y(z - x)$$

If X, Y, Z are continuous,

$$F_Z(z) = \mathbb{P}(X + Y \leq z) = \int_{-\infty}^{\infty} \mathbb{P}(Y \leq z - X \mid X = x) f_X(x) dx = \int_{-\infty}^{\infty} F_Y(z - x) f_X(x) dx$$

Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

(a) Find an expression for $\mathbb{P}(X_1 < 2X_2)$ using a similar idea to convolution, in terms of $F_{X_1}, F_{X_2}, f_{X_1}, f_{X_2}$. (Your answer will be in the form of a single integral, and requires no calculations – do not evaluate it).

We use the continuous version of the "Law of Total Probability" to integrate over all possible values of X_2. Take the probability that $X_1 < 2X_2$ given that value of X_2, times the density of X_2 at that value.

$$\mathbb{P}(X_1 < 2X_2) = \int_{-\infty}^{\infty} \mathbb{P}(X_1 < 2X_2 \mid X_2 = x_2) f_{X_2}(x_2)\, dx_2 = \int_{-\infty}^{\infty} F_{X_1}(2x_2)\, f_{X_2}(x_2)\, dx_2$$

(b) Find s, where $\Phi(s) = \mathbb{P}(X_1 < 2X_2)$ using the "reproductive" property of normal distributions.

Let $X_3 = X_1 - 2X_2$, so that $X_3 \sim \mathcal{N}(\mu_1 - 2\mu_2, \sigma_1^2 + 4\sigma_2^2)$ (by the reproductive property of normal distributions)

$$\mathbb{P}(X_1 < 2X_2) = \mathbb{P}(X_1 - 2X_2 < 0) = \mathbb{P}(X_3 < 0) = \mathbb{P}\left(\frac{X_3 - (\mu_1 - 2\mu_2)}{\sqrt{\sigma_1^2 + 4\sigma_2^2}} < \frac{0 - (\mu_1 - 2\mu_2)}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}\right)$$

$$= \mathbb{P}\left(Z < \frac{2\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}\right) = \Phi\left(\frac{2\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}\right) \to s = \frac{2\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + 4\sigma_2^2}}$$

5. Suppose $X_1, \ldots, X_n$ are iid $Poi(\lambda)$ random variables, and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, the sample mean. How large should we choose n to be such that $\mathbb{P}\left(\frac{\lambda}{2} \leq \overline{X}_n \leq \frac{3\lambda}{2}\right) \geq 0.99$? Use the CLT and give an answer involving $\Phi^{-1}(\cdot)$. Then evaluate it exactly when $\lambda = 1/10$ using the Φ table on the last page.

   We know $\mathbb{E}[X_i] = Var(X_i) = \lambda$. By the CLT, $\overline{X}_n \approx \mathcal{N}\left(\lambda, \frac{\lambda}{n}\right)$, so

$$\mathbb{P}\left(\frac{\lambda}{2} \leq \overline{X}_n \leq \frac{3\lambda}{2}\right) \approx \mathbb{P}\left(\frac{-\lambda/2}{\sqrt{\lambda/n}} \leq Z \leq \frac{\lambda/2}{\sqrt{\lambda/n}}\right) = \Phi\left(\frac{\lambda/2}{\sqrt{\lambda/n}}\right) - \Phi\left(\frac{-\lambda/2}{\sqrt{\lambda/n}}\right)$$

$$= \Phi\left(\frac{\lambda/2}{\sqrt{\lambda/n}}\right) - \left(1 - \Phi\left(\frac{\lambda/2}{\sqrt{\lambda/n}}\right)\right) = 2\Phi\left(\frac{\lambda/2}{\sqrt{\lambda/n}}\right) - 1 \geq 0.99 \to \Phi\left(\frac{\lambda/2}{\sqrt{\lambda/n}}\right) \geq 0.995$$

$$\to \frac{\sqrt{\lambda}}{2}\sqrt{n} \geq \Phi^{-1}(0.995) \to n \geq \frac{4}{\lambda}\left[\Phi^{-1}(0.995)\right]^2$$

   We have $\lambda = \frac{1}{10}$ and from the table, $\Phi^{-1}(0.995) \approx 2.575$ so that $n \geq \frac{4}{1/10} \cdot 2.575^2 = 265.225$. So $n = 266$ is the smallest value that will satisfy the condition.