CSE 312: Foundations of Computing II

Section 8: Continuous RV's and Central Limit Theorem Solutions

0. Transformations

Suppose $X \sim$ Uniform$(0,1)$ has the continuous uniform distribution on $(0,1)$. Let $Y = -\frac{1}{\lambda}\log X$ for some $\lambda > 0$.

(a) What is Ω_Y?

Solution:

$\Omega_Y = (0,\infty)$ because $\log(x) \in (-\infty, 0)$ for $x \in (0,1)$.

(b) First write down $F_X(x)$ for $x \in (0,1)$. Then, find $F_Y(y)$ on Ω_Y.

Solution:

$F_X(x) = x$ for $x \in (0,1)$. Let $y \in \Omega_Y$.

$F_Y(y) = \Pr(Y \leq y) = \Pr\left(-\frac{1}{\lambda}\log X \leq y\right) = \Pr(\log X \geq -\lambda y) = \Pr\left(X \geq e^{-\lambda y}\right) = 1 - \Pr\left(X < e^{-\lambda y}\right)$

Then, because $e^{-\lambda y} \in (0,1)$

$$= 1 - F_X(e^{-\lambda y}) = 1 - e^{-\lambda y}$$

(c) Now find $f_Y(y)$ on Ω_Y. What distribution does Y have?

Solution:

$$f_Y(y) = F_Y'(y) = \lambda e^{-\lambda y}$$

Hence, $Y \sim$ Exponential(λ).

1. Poisson CLT

Suppose $X_1, \ldots, X_n$ are iid Poisson(λ) 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 $\Pr\left(\frac{\lambda}{2} \leq \overline{X}_n \leq \frac{3\lambda}{2}\right) \geq 0.99$? 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.

Solution:

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

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

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

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

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.

2. Convolutions

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

$$p_Z(z) = \Pr(X + Y = z) = \sum_x \Pr(X = x \cap Y = z - x) = \sum_x p_X(x)\, p_Y(z - x)$$

If X, Y, Z are continuous,

$$F_Z(z) = \Pr(X + Y \le z) = \int_{-\infty}^{\infty} \Pr(Y \le z - X \mid X = x)\, f_X(x) dx = \int_{-\infty}^{\infty} F_Y(z - x) f_X(x) dx$$

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 $\Pr(X_1 < 2X_2)$ using a similar idea to convolution, in terms of $F_{X_1}, F_{X_2}, f_{X_1}, f_{X2}$. (Your answer will be in the form of a single integral, and requires no calculations – do not evaluate it).

Solution:

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.

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

(b) Find s, where $\Phi(s) = \Pr(X_1 < 2X_2)$ using the fact that linear combinations of independent normal random variables are still normal.

Solution:

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)

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

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

3. Bad Computer

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next 100 days.

(a) Let X be the number of crash-free days in the next 100 days. What distribution does X have? Identify $\mathbb{E}[X]$ and $\text{Var}(X)$ as well. Write an exact (possibly unsimplified) expression for $\Pr(X \ge 87)$.

Solution:

$X \sim \text{Binomial}(100, 0.9)$. Hence, $\mathbb{E}[X] = np = 90$ and $\text{Var}(X) = np(1 - p) = 9$. Finally,

$$\Pr(X \ge 87) = \sum_{k=87}^{100} \binom{100}{k} (0.9)^k (1 - 0.9)^{100-k}$$

(b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem. Justify why we can use the CLT here.

Solution:

From the previous part, we know that $\mathbb{E}[X] = 90$ and Var$(X) = 9$.

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

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.