IEOR 151 – Lecture 2
Probability Review II

1 Motivating Example

Suppose that the amount of time required to perform a surgery is distributed according an exponential random variable (distribution function of $F_x(u) = 1 - \exp(-\lambda u)$) with rate $\lambda = 1/2$. And suppose that the price charged to a patient for a surgery is given by $C(u) = 2u + 5$. What is the probability that the amount of revenue for 100 surgeries will exceed 950, assuming that the amount of time required for each surgery is (mutually) independent?

Let X_i denote the duration of the i-th surgery. We are interested in computing $\mathbb{P}(\sum_{i=1}^{100} C(X_i) \geq 950)$, which can be rewritten as

$$\mathbb{P}(\textstyle\sum_{i=1}^{100} C(X_i) \geq 950) = \mathbb{P}(\sum_{i=1}^{100} 2X_i \geq 450).$$

Observe that $Y_i = 2X_i$ is distributed as an exponential random variable with rate $\lambda_y = 1/4$, and so $S = \sum_{i=1}^{100} Y_i$ is distributed as an Erlang distribution $F_S(u) = 1 - \sum_{n=0}^{99} \frac{1}{n!} \exp(-u/4) \cdot (u/4)^n$. Thus, we have that

$$\mathbb{P}(\textstyle\sum_{i=1}^{100} C(X_i) \geq 950) = \sum_{n=0}^{99} \frac{1}{n!} \exp(-450/4) \cdot (450/4)^n \approx 0.1085.$$

This computation is cumbersome, and there is in fact an easier way to approximately compute the probability. As will be shown at the end of the lecture, the distribution is approximately given by that of a Gaussian distribution and so

$$\mathbb{P}(\textstyle\sum_{i=1}^{100} C(X_i) \geq 950) \approx 1 - \frac{1}{2}[1 + \text{erf}(\frac{950-900}{\sqrt{2}\cdot\sqrt{100}\cdot\sqrt{16}})] \approx 0.1056,$$

where erf$(\cdot)$ is the Gauss error function.

2 Concentration About the Mean

Consider an infinite sequence of (mutually) independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$, and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average. There are a number of results that show that the sample average $\overline{X}_n$ is "close" to the mean of the distribution. The intuition is that the sample average can only deviate "far" from the mean if the random variables act in concert to pull the average in one direction, but the probability that the random variables pull in the same direction is small because of their independence.

2.1 Weak Law of Large Numbers

If the random variables X_i have a finite first moment $\mathbb{E}|X_i| < \infty$, then $\overline{X}_n \xrightarrow{p} \mu$ where $\mu = \mathbb{E}(X_i)$. In words — the weak law of large numbers states that if i.i.d. random variables have a finite first moment, then their sample average converges in probability to the mean of the distribution.

2.2 Central Limit Theorem

A more precise statement of the convergence of sample averages is given by the following theorem: If the random variables X_i with mean $\mathbb{E}(X_i) = \mu$ have a finite variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\mathcal{N}(0, \sigma^2)$ is the distribution of a Gaussian random variable with mean 0 and variance σ^2. This is a more precise statement because it describes the distribution of the sample average when it is appropriately scaled by $\sqrt{n}$. This scaling is important because otherwise $\overline{X}_n \xrightarrow{d} \mu$ by the weak law of large numbers (and the fact that convergence in probability implies convergence in distribution).

3 Extensions of Central Limit Theorem

There are a number of extensions of the Central Limit Theorem, but understanding and deriving these extensions requires several useful results.

3.1 Continuous Mapping Theorem

Suppose that $g(u)$ is a continuous function. The continuous mapping theorem states that convergence of a sequence of random variables $\{X_n\}$ to a limiting random variable X is preserved under continuous mappings. More specifically:

- If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.
- If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$.

3.2 Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} y_0$, where y_0 is a constant, then

- $X_n + Y_n \xrightarrow{d} X + y_0$;
- $Y_n X_n \xrightarrow{d} y_0 X$;

There is a technical point to note about this theorem: Convergence of Y_n to a constant is a subtle but important feature for this result, because the theorem will not generally hold when Y_n converges to a non-constant random variable. Consider the example where $X_n \sim \mathcal{N}(0, 1)$ and $Y_n = X_n$, then $X_n + Y_n = \mathcal{N}(0, 4)$ which obviously does not converge in distribution to $\mathcal{N}(0, 1) + \mathcal{N}(0, 1) = \mathcal{N}(0, 2)$.

3.3 Delta Method

Using the Continuous Mapping Theorem and Slutsky's Theorem, we can now state and derive an extension of the Central Limit Theorem. Consider an infinite sequence of random variables $\{X_n\}$ that satisfies

$$\sqrt{n}[X_n - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where θ and σ^2 are finite constants. If $g(u)$ is continuously differentiable at θ and $g'(\theta) \neq 0$, then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} g'(\theta)\mathcal{N}(0, \sigma^2).$$

To derive this result, which is known as the Delta Method, we will need to use the "zeroth-order" Lagrange form of Taylor's Theorem. This result states that there exists some $\xi(u, \theta)$ in between u and θ such that

$$g(u) = g(\theta) + g'(\xi(u, \theta))(u - \theta).$$

This theorem is also known as the Mean Value Theorem.

First, note that because $1/\sqrt{n} \to 0$ as $n \to \infty$, then Slutsky's Theorem implies

$$1/\sqrt{n} \cdot \sqrt{n}[X_n - \theta] = X_n - \theta \xrightarrow{p} 0 \cdot \mathcal{N}(0, \sigma^2) = 0.$$

Thus, $X_n - \theta \xrightarrow{p} 0$. Now using the form of Taylor's Theorem stated above, we have

$$\sqrt{n}[g(X_n) - g(\theta)] = \sqrt{n}[g(\theta) + g'(\xi(X_n, \theta))(X_n - \theta) - g(\theta)] = \sqrt{n}[g'(\xi(X_n, \theta))(X_n - \theta)].$$

But because $|\xi(X_n, \theta) - \theta| \leq |X_n - \theta|$ and $X_n - \theta \xrightarrow{p} 0$ (as shown above), we have that $|\xi(X_n, \theta) - \theta| \xrightarrow{p} 0$ and so $\xi(X_n, \theta) \xrightarrow{p} \theta$. Because $g'(u)$ is continuous at θ by assumption, we can apply the Continuous Mapping Theorem to infer that $g'(\xi(X_n, \theta)) \xrightarrow{p} g'(\theta)$. Next, we can apply Slutsky's Theorem to the two facts (i) $\sqrt{n}[X_n - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and (ii) $g'(\xi(X_n, \theta)) \xrightarrow{p} g'(\theta)$, to conclude

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} g'(\theta)\mathcal{N}(0, \sigma^2).$$

This is the result that was to be shown.

4 Motivating Example, cont.

Returning to the motivating example presented above, we can apply the Delta Method to approximate the probability to be computed. By the Central Limit Theorem and for an exponential distribution with rate of $\lambda = 1/2$, we have $\sqrt{n}[\overline{X}_n - 2] \xrightarrow{d} \mathcal{N}(0, 4)$. Applying the Delta Method to $C(u) = 2u + 5$, we have $\sqrt{n}[C(\overline{X}_n) - 9] \xrightarrow{d} \mathcal{N}(0, 16)$. But

$$\textstyle\sum_{i=1}^{100} C(X_i) = nC(\overline{X}_n) = \sqrt{n} \cdot \sqrt{n}[C(\overline{X}_n) - 9] + 9n.$$

Hence, we can approximate the probability $\mathbb{P}(\sum_{i=1}^{100} C(X_i) \geq 950)$ by the distribution $\sqrt{100}\mathcal{N}(0, 16) + 9 \cdot 100$. This gives an approximate probability of

$$\mathbb{P}(\textstyle\sum_{i=1}^{100} C(X_i) \geq 950) \approx 1 - \frac{1}{2}[1 + \text{erf}(\frac{950-900}{\sqrt{2}\cdot\sqrt{100}\cdot\sqrt{16}})] \approx 0.1056.$$