Recap

The Central Limit Theorem (CLT)

Let X_1,X_2,...,X_n be i.i.d. random variables with expected value $EX_i = \mu < \infty$ and variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$\lim_{n \to \infty} P(Z_n \leq x) = \Phi(x), \qquad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.

The distribution of the X_i's does not matter.

How to Apply The Central Limit Theorem (CLT)

Here are the steps that we need in order to apply the CLT:

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variable X_i's:

$$Y = X_1 + X_2 + \ldots + X_n.$$

2. Find EY and $\mathrm{Var}(Y)$ by noting that

$$EY = n\mu, \qquad \mathrm{Var}(Y) = n\sigma^2,$$

where $\mu = EX_i$ and $\sigma^2 = \mathrm{Var}(X_i)$.

3. According to the CLT, conclude that $\frac{Y - EY}{\sqrt{\mathrm{Var}(Y)}} = \frac{Y - n\mu}{\sqrt{n}\sigma}$ is approximately standard normal; thus, to find $P(y_1 \leq Y \leq y_2)$, we can write

$$P(y_1 \leq Y \leq y_2) = P\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} \leq \frac{Y - n\mu}{\sqrt{n}\sigma} \leq \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right)$$
$$\approx \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right).$$

Example

A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $EX_i = 2$ (minutes) and $\mathrm{Var}(X_i) = 1$. We assume that service times for different bank customers are independent. Let Y be the total time the bank teller spends serving 50 customers. Find $P(90 < Y < 110)$.

$n = 50$, $EX_i = \mu = 2$, $Var(X_i) = 1$

$Y = X_1 + X_2 + X_3 + \dots + X_n$

$$P(90 \le Y \le 100) = P\left(\frac{90 - n\cdot\mu}{\sqrt{n}\cdot\sigma} \le \frac{Y - n\cdot\mu}{\sqrt{n}\cdot\sigma} \le \frac{110 - n\cdot\mu}{\sqrt{n}\cdot\sigma}\right)$$

$$= P\left(\frac{90 - 50\cdot 2}{\sqrt{50}\cdot 1} \le Z \le \frac{110 - 50\cdot 2}{\sqrt{50}\cdot 1}\right)$$

$$= P\left(\frac{-10}{\sqrt{50}} \le Z \le \frac{10}{\sqrt{50}}\right)$$

$$= P\left(-\frac{\sqrt{100}}{\sqrt{50}} \le Z \le \frac{\sqrt{100}}{\sqrt{50}}\right)$$

$$= P(-\sqrt{2} \le Z \le \sqrt{2}) = \Phi(\sqrt{2}) - \Phi(-\sqrt{2})$$

$= 0.8422$ (norm.cdf($\sqrt{2}$) − norm.cdf($-\sqrt{2}$))

```
from scipy import stats
import math
stats.norm.cdf(math.sqrt(2)) - stats.norm.cdf(-math.sqrt(2))
```

```
0.842700792949715
```

```
stats.norm.cdf(110, loc = 100, scale = math.sqrt(50)) - stats.norm.cdf(90, loc = 100, scale = math.sqrt(50))
```

```
0.8427007929497148
```

Example:

In a communication system each data packet consists of 1000 bits. Due to the noise, each bit may be received in error with probability 0.1. It is assumed bit errors occur independently. Find the probability that there are more than 120 errors in a certain data packet.

$X_i \sim \text{Bernoulli}(0.1)$

$EX_i = \mu = p = 0.1 \quad , Var(X_i) = \sigma^2 = p(1-p) = 0.09$

$$P(Y > 120) = 1 - P\left(Z < \frac{120 - 1000 \cdot 0.1}{\sqrt{1000 \cdot 0.09}}\right)$$

$$= 1 - P\left(\frac{20}{\sqrt{90}}\right) = 0.0175$$

Estimators and Bias:

$\theta \to$ parameters to be estimated

↳ Could b μ, σ, λ, etc.

Let $\hat{\Theta} = h(X_1, X_2, \cdots, X_n)$ be a point estimator for θ. The **bias** of point estimator $\hat{\Theta}$ is defined by

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta.$$

Let $\hat{\Theta} = h(X_1, X_2, \cdots, X_n)$ be a point estimator for a parameter θ. We say that $\hat{\Theta}$ is an **unbiased** of estimator of θ if

$$B(\hat{\Theta}) = 0, \qquad \text{for all possible values of } \theta.$$

The **mean squared error** (MSE) of a point estimator $\hat{\Theta}$, shown by $MSE(\hat{\Theta})$, is defined as

$$MSE(\hat{\Theta}) = E\left[(\hat{\Theta} - \theta)^2\right].$$

Let X_1, X_2, X_3, ..., X_n be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1$, $X_2 = x_2$, $\cdots$, $X_n = x_n$.

1. If X_i's are discrete, then the **likelihood function** is defined as

$$L(x_1, x_2, \cdots, x_n; \theta) = P_{X_1 X_2 \cdots X_n}(x_1, x_2, \cdots, x_n; \theta).$$

2. If X_i's are jointly continuous, then the likelihood function is defined as

$$L(x_1, x_2, \cdots, x_n; \theta) = f_{X_1 X_2 \cdots X_n}(x_1, x_2, \cdots, x_n; \theta).$$

In some problems, it is easier to work with the **log likelihood function** given by

$$\ln L(x_1, x_2, \cdots, x_n; \theta).$$

Example:

For the following random samples, find the likelihood function:

1. $X_i \sim Binomial(3, \theta)$, and we have observed $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$.
2. $X_i \sim Exponential(\theta)$ and we have observed $(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12)$.

① If $X_i \sim Binomial(3, \theta)$

$$P_{X_i}(x; \theta) = \binom{3}{x} \theta^{x} (1-\theta)^{3-x}$$

Since we have I.I.D we just multiply all marginals to get joint!

$$L(x_i; \theta) = \prod_{1}^{4} \binom{3}{x_i} \theta^{x_i} (1-\theta)^{3-x_i}$$

$$= \binom{3}{x_1}\binom{3}{x_2}\binom{3}{x_3}\binom{3}{x_4} \cdot \theta^{x_1+x_2+x_3+x_4} (1-\theta)^{12-(x_1+x_2+x_3+x_4)}$$

$$= \binom{3}{1}\binom{3}{3}\binom{3}{2}\binom{3}{2} \theta^{8} \cdot (1-\theta)^{4}$$

$$= 27 \cdot \theta^{8} (1-\theta)^{4}$$ ← likelihood function!

② If $X_i \sim Exponential(\theta)$

$$f(x;\theta) = \theta e^{-\theta x}, \text{ so}$$

$$L(1.23, 3.32, 1.98, 2.12, \theta) = \theta^4 \cdot e^{-\theta(1.23+3.32+1.98+2.12)}$$

$$= \theta^4 \cdot e^{-8.65\theta}$$

likelihood function

Example:

Find the MLE (maximum likelihood estimator)

① $f'(x) = 0$

$$\left(27\theta^8 (1-\theta)^4\right)' = 0$$

```
import sympy as sp
```

```
[16] x = sp.symbols('x')
```

```
sp.solve(sp.diff(27*x**8*(1-x)**4, x), x)
[0, 2/3, 1]
```

, so

$\hat{\theta}_{ML} = 2/3$ (best estimator for p)

② $\left(\theta^4 \cdot e^{-8.65\theta}\right)' = 0$

Easier to work with log likelihood:

$$\ln(\theta^4 \cdot e^{-8.65\theta}) = 4 \cdot \ln\theta - 8.65\theta, \text{ so}$$

$$(4 \cdot \ln\theta - 8.65\theta)' = 0$$

$$\frac{4}{\theta} - 8.65 = 0 \Rightarrow \theta_{ML} = 0.46$$