Chapter 6 - Part 1

Central Limit Theorem

Expected Value and Variance of Sums and Averages of Random Variables

- Let
$$X_1, X_2, X_3, \ldots, X_n$$
be random random variables with
$$E(X_i) = \mu_i, \quad \mathrm{Var}(X_i) = \sigma_i^2, \quad \text{for } i = 1, 2, 3, \ldots, n.$$

- Call
$$\Sigma X = X_1 + X_2 + X_3 + \ldots + X_n$$
for their sum.
- Call
$$\overline{X} = \frac{1}{n}(\Sigma X) = \frac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n)$$
for their average.

Expected Value of Sum of Random Variables

- Then,

$$E\left(\Sigma X\right) = E\left(X_1 + X_2 + X_3 + \ldots + X_n\right)$$
$$= E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)$$
$$= \mu_1 + \mu_2 + \mu_3 + \ldots + \mu_n$$

- If $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n = \mu$, then

$$E\left(\Sigma X\right) = n\mu$$

Expected Value of Average of Random Variables

- Then,

$$
\begin{aligned}
E\left(\overline{X}\right) &= E\left(\frac{1}{n}\left(\Sigma X\right)\right) \\
&= \frac{1}{n}E\left(\Sigma X\right) \\
&= \frac{1}{n}\left(E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)\right) \\
&= \frac{\mu_1 + \mu_2 + \mu_3 + \ldots + \mu_n}{n}
\end{aligned}
$$

- If $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n = \mu$, then

$$E\left(\overline{X}\right) = \mu$$

Variance of Sum of Random Variables

- If $X_1, X_2, X_3, \ldots, X_n$ are independent,

$$\begin{aligned} \text{Var}\left(\Sigma X\right) &= \text{Var}\left(X_1 + X_2 + X_3 + \ldots + X_n\right) \\ &= \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \ldots + \text{Var}(X_n) \\ &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots + \sigma_n^2 \end{aligned}$$

  Therefore,

$$\text{StDev}\left(\Sigma X\right) = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots + \sigma_n^2}$$

- If $X_1, X_2, X_3, \ldots, X_n$ are independent and $\sigma_1 = \sigma_2 = \sigma_3 = \ldots = \sigma_n = \sigma \geq 0$, then

$$\text{Var}\left(\Sigma X\right) = n\sigma^2 \quad \Longrightarrow \quad \text{StDev}\left(\Sigma X\right) = \sigma\sqrt{n}$$

Variance of Average of Random Variables

- If $X_1, X_2, X_3, \ldots, X_n$ are independent,

$$\mathrm{Var}\left(\overline{X}\right) = \mathrm{Var}\left(\frac{1}{n}\left(\Sigma X\right)\right) = \frac{1}{n^2}\mathrm{Var}\left(\Sigma X\right)$$

$$= \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) + \ldots + \mathrm{Var}(X_n)\right)$$

$$= \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots + \sigma_n^2}{n^2}$$

  Therefore,

$$\mathrm{StDev}\left(\overline{X}\right) = \frac{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots + \sigma_n^2}}{n}$$

- If $X_1, X_2, X_3, \ldots, X_n$ are independent and $\sigma_1 = \sigma_2 = \sigma_3 = \ldots = \sigma_n = \sigma \geq 0$, then

$$\mathrm{Var}\left(\overline{X}\right) = \frac{\sigma^2}{n} \quad \Longrightarrow \quad \mathrm{StDev}\left(\overline{X}\right) = \frac{\sigma}{\sqrt{n}}$$

Sums and Averages of Normal Random Variables

If $X_1, X_2, X_3, \ldots, X_n$ are independent normal random variables with

$$X_i \sim N(\mu_i, \sigma_i), \quad \text{for } i = 1, 2, 3, \ldots, n$$

then ΣX is also normal with

$$\Sigma X \sim N\left(\mu_1 + \mu_2 + \mu_3 + \ldots + \mu_n, \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots + \sigma_n^2}\right)$$

and $\overline{X}$ is also normal with

$$\overline{X} \sim N\left(\frac{\mu_1 + \mu_2 + \mu_3 + \ldots + \mu_n}{n}, \frac{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots + \sigma_n^2}}{n}\right).$$

Sums and Averages of Normal Random Variables

If $X_1, X_2, X_3, \ldots, X_n$ are independent normal random variables with

$$X_i \sim N(\mu, \sigma), \quad \text{for } i = 1, 2, 3, \ldots, n$$

then ΣX is also normal with

$$\Sigma X \sim N\left(n\mu, \sigma\sqrt{n}\right)$$

and $\overline{X}$ is also normal with

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Example

Suppose individual exam scores are normally distributed:

$$X_i \sim N(70, 12)$$

(mean 70, standard deviation 12).
Consider a sample of $n = 16$ students and let

$$\overline{X} = \text{average score of the sample.}$$

Then

$$\overline{X} \sim N\left(70, \frac{12}{\sqrt{16}}\right) = N(70, 3).$$

Interpretation: Individual scores vary a lot (sd = 12), but the sample average is much more stable (sd = 3).

Question: What is the probability that the sample average exceeds 73?

Example (cont'd)

Question: What is the probability that the sample average exceeds 73?

Solution: We have

$$\overline{X} \sim N(70, 3).$$

We want

$$P(\overline{X} > 73).$$

Standardize:

$$Z = \frac{\overline{X} - 70}{3}$$

Thus

$$P(\overline{X} > 73) = P\left(Z > \frac{73 - 70}{3}\right) = P(Z > 1) = 0.5 - 0.3413 = 0.1587.$$

Central Limit Theorem

If $X_1, X_2, X_3, \ldots, X_n$ are independent and identically distributed random variables with

$$E(X_i) = \mu, \quad \text{StDev}(X_i) = \sigma, \quad \text{for } i = 1, 2, 3, \ldots, n$$

then ΣX is approximately normal for 'large enough' n, with

$$\Sigma X \sim N\left(n\mu, \sigma\sqrt{n}\right)$$

and $\overline{X}$ is approximately normal for 'large enough' n, with

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Central Limit Theorem

If $X_1, X_2, X_3, \ldots, X_n$ are independent and identically distributed random variables with

$$E(X_i) = \mu, \quad \text{StDev}(X_i) = \sigma, \quad \text{for } i = 1, 2, 3, \ldots, n$$

then ΣX is approximately normal for 'large enough' n, with

$$\Sigma X \sim N\left(n\mu, \sigma\sqrt{n}\right)$$

and $\overline{X}$ is approximately normal for 'large enough' n, with

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- **Note:** Here it is not required that the each X_i have normal distributions. They can have any distributions, even discrete ones.
- In our course, 'large enough' means $n \geq 30$ (except for the 'normal approximation to binomial').

Example

You use the random number generator button on your calculator 40 times and add the results. The generator has a uniform pdf between 0 and 1, namely $U(0,1)$. What is the probabilty that the sum is between 21 and 23?

Solution: Since $n = 40 > 30$, we can apply CLT. Let $X_1, \ldots, X_{40} \sim U(0,1)$. Recall that

$$\mu = E(X_i) = \tfrac{1}{2}, \quad \sigma = \mathrm{StDev}(X_i) = \frac{1}{\sqrt{12}}.$$

Let $\Sigma X = X_1 + X_2 + X_3 + \ldots + X_{40}$. Then by CLT,

$$\Sigma X \sim N\left(\underbrace{n\mu}_{40\cdot\frac{1}{2}}, \underbrace{\sigma\sqrt{n}}_{\frac{1}{\sqrt{12}}\cdot\sqrt{40}}\right) = N\left(20, \sqrt{\frac{10}{3}}\right)$$

Example (cont'd)

$$\Sigma X \sim N\left(20, \sqrt{\frac{10}{3}}\right)$$

Thus

$$\begin{aligned} P(21 < \Sigma X < 23) &= P\left(\frac{21-20}{\sqrt{10/3}} < Z < \frac{23-20}{\sqrt{10/3}}\right) \\ &= P(0.55 < Z < 1.64) \\ &= 0.4495 - 0.2088 \\ &= 0.2407 \end{aligned}$$

Example

Babir Khan likes baklava. Due to his doctors' advice and his wife's threats, the number of plates of baklava he eats each night has the following distribution:

x	0	1	2
$p(x)$	0.10	0.65	0.25

What is the probability that his average number of plates per day over the next 80 days is between 1.125 and 1.25?

Solution: Since $n = 80 > 30$, we can use CLT. Let X_i be the number of plates eaten in the i-th day. Then

$$\mu = E(X_i) = 0 \times 0.10 + 1 \times 0.65 + 2 \times 0.25 = 1.15$$

and

$$E(X_i^2) = 0^2 \times 0.10 + 1^2 \times 0.65 + 2^2 \times 0.25 = 1.65.$$

So

$$\sigma^2 = E(X_i^2) - \mu^2 = 1.65 - (1.15)^2 = 0.3275,$$

hence

$$\sigma = \sqrt{0.3275} \approx 0.5723.$$

Example (cont'd)

By CLT,

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(1.15, \frac{\sqrt{0.3275}}{\sqrt{80}}\right).$$

Therefore,

$$P(1.125 < \overline{X} < 1.25) = P\left(\frac{1.125 - 1.15}{\sqrt{0.3275/80}} < Z < \frac{1.25 - 1.15}{\sqrt{0.3275/80}}\right)$$

$$= P(-0.39 < Z < 1.56) = 0.1517 + 0.4406$$

$$= 0.5923$$

The Normal Approximation to Binomial

Recall that if $X \sim \mathrm{Binomial}(n, p)$, then $X = X_1 + X_2 + X_3 + \ldots X_n$ where X_i are independent and each X_i has the following pdf

x_i	0	1
$p(x_i)$	1-p	p

Also recall that

$$E(X) = np, \quad \mathrm{StDev}(X) = \sqrt{np(1-p)}$$

For $\overline{X} = \frac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n) = \frac{1}{n} \cdot X$, we have

$$E\left(\overline{X}\right) = p, \quad \mathrm{StDev}\left(\overline{X}\right) = \sqrt{\frac{p(1-p)}{n}}$$

We can apply the central limit theorem to X_i's.

The Normal Approximation to Binomial

If X is a random variable with $X \sim \mathrm{Binomial}(n, p)$, then for 'large enough' n,

$$X \sim N\left(np, \sqrt{np(1-p)}\right)$$

Also, by letting $\overline{X} = \frac{X}{n}$, for 'large enough' n,

$$\overline{X} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

- Here 'large enough' means $np \geq 15$ and $n(1-p) \geq 15$.

Example

We have a coin with $P(H) = 0.44$ is tossed 55 times. What is the probability of getting $19, 20, 21, 22$ or 23 heads?

Solution using binomial formula:

$$P(19 \leq X \leq 23) = \sum_{i=19}^{23} \binom{55}{i} (0.44)^i (0.56)^{55-i} \approx 0.3674$$

Example (cont'd)

We have a coin with $P(H) = 0.44$ is tossed 55 times. What is the probability of getting $19, 20, 21, 22$ or 23 heads?

Solution using normal approximation:

$$X \sim N\left(\underbrace{np}_{55\cdot 0.44}, \underbrace{\sqrt{np(1-p)}}_{\sqrt{55\cdot 0.44\cdot 0.56}}\right) = N\left(24.2, \sqrt{13.552}\right)$$

We will look for $P(18.5 < X < 23.5)$, this is called a continuity correction.

$$P(18.5 < X < 23.5) = P\left(\frac{18.5 - 24.2}{\sqrt{13.552}} < Z < \frac{23.5 - 24.2}{\sqrt{13.552}}\right) = P\left(-1.55 < Z < -0.19\right)$$

$$= 0.4394 - 0.0753 = 0.3641$$