Functions of Normal Distribution Central Limit Theorem

5.5, 5.6

Linear Functions of Independent Normal Random Variables

$X_1,\quad X_2,\quad X_3$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

...

Theorem 5.5-1 If $X_1, X_2, \ldots, X_n$ are n mutually independent normal variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively, then the linear function

$\boxed{Y} = 2X_1 - X_2 + 3X_3$

$$Y = \sum_{i=1}^{n} c_i X_i$$

has the normal distribution

$$\star \quad Y \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

Proof By Theorem 5.4-1, we have, with $-\infty < c_i t < \infty$, or $-\infty < t < \infty$,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(c_i t) = \prod_{i=1}^{n} \exp\left(\mu_i c_i t + \sigma_i^2 c_i^2 t^2/2\right)$$

because $M_{X_i}(t) = \exp(\mu_i t + \sigma_i^2 t^2/2)$, $i = 1, 2, \ldots, n$. Thus,

$$M_Y(t) = \exp\left[\left(\sum_{i=1}^{n} c_i \mu_i\right) t + \left(\sum_{i=1}^{n} c_i^2 \sigma_i^2\right)\left(\frac{t^2}{2}\right)\right].$$

Notes

$X_1 \sim N(10, 5^2)$ $X_2 \sim N(100, 6^2)$

$$\boxed{Y = 3X_1 - X_2}$$

$\cancel{X_1 X_2}$

$225 + 36$

$Y \sim ?$

$$Y \sim N\left(3(10) + (-1)(100),\ 3^2 \cdot 5^2 + 6^2\right)$$

$$\boxed{Y \sim N(-70,\ 261)}$$

Notes

$$Var[X+Y] \quad (\text{if } X \perp Y) \quad = Var[X] + Var[Y]$$

$$Var[X-Y] = Var[X] + Var[Y]$$

$$Var[1X + (-1)Y]$$

$$Var[\underline{X+Y}] = Cov[\underline{X+Y}, \underline{X+Y}]$$

$[Y, Y]$

$$= Cov[X, X] + Cov[X, Y] + Cov[X, Y] + Cov[\ ]$$

$$= Var[X] + 0 + 0 + Var[Y]$$

Example

Example 5.5-1

Let X_1 and X_2 equal the number of pounds of butterfat produced by two Holstein cows (one selected at random from those on the Koopman farm and one selected at random from those on the Vliestra farm, respectively) during the 305-day lactation period following the births of calves. Assume that the distribution of X_1 is $N(693.2, 22820)$ and the distribution of X_2 is $N(631.7, 19205)$. Moreover, let X_1 and X_2 be independent. We shall find $P(X_1 > X_2)$. That is, we shall find the probability that the butterfat produced by the Koopman farm cow exceeds that produced by the Vliestra farm cow. (Sketch pdfs on the same graph for these two normal distributions.) If we let $Y = X_1 - X_2$, then the distribution of Y is $N(693.2 - 631.7, 22820 + 19205)$. Thus,

$= \mathcal{N}(\mu = 61.5, \sigma^2 = 42025)$

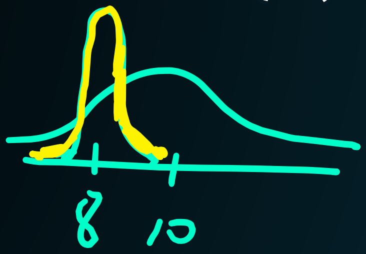
$$P(X_1 > X_2) = P(Y > 0) = P\left(\frac{Y - 61.5}{\sqrt{42025}} > \frac{0 - 61.5}{205}\right)$$

$$= P(Z > -0.30) = 0.6179.$$

Let $X \sim N(10,4)$ $\quad Y \sim N(8, 1)$

Find $P[X > Y]$

$P[X - Y > 0]$ let $W = X - Y$



1 Create new random variable

2 Determine its distribution

3 Standardize

$P[W > 0]$

$N(2, \sqrt{5}^2)$

$W \sim N(\underbrace{10 + (-1)8}, \underbrace{4+1}) = N(2, 5)$

$P\left[Z > \frac{0-2}{\sqrt{5}}\right] = 1\text{-pnorm}(2/\sqrt{5})$

$= P[Z > -.894]$

$= 0.814$

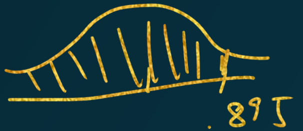
Find $P[2X > 3Y]$

$P[2x - 3y > 0]$ let $W = \underline{2x - 3y}$

$W \sim N(2(10) - 3(8), 2^2 \cdot 4 + (-3)^2 1)$

$\boxed{\sim N(-4, 25)}$

$P[W > 0]$



Notes

$P[2x-3y>0]$ let $w = 2x-3y$

$W \sim N\left(2(10)-3(8),\ 2^2\cdot 4+(-3)^2 1\right)$

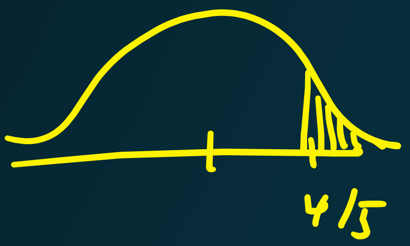
$W \boxed{\sim N(-4, 25)}$ $P[w>0]$

$\sigma_w^2 = 25$

$\sigma_w = 5$

$P\left[\frac{w-(-4)}{5} > \frac{0-(-4)}{5}\right]$

$= P\left[Z > \frac{4}{5}\right] = 0.211$



$$\boxed{X_i \sim N(\mu, \sigma^2)} \qquad \boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

Sample mean Normal random variables

Corollary 5.5-1 If $X_1, X_2, \ldots, X_n$ are observations of a random sample of size n from the normal distribution $N(\mu, \sigma^2)$, then the distribution of the sample mean $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ is $N(\mu, \sigma^2/n)$.

Proof Let $c_i = 1/n$, $\mu_i = \mu$, and $\sigma_i^2 = \sigma^2$ in Theorem 5.5-1. ◀

$$X_1, X_2, \ldots X_n \sim N(\mu, \sigma^2)$$

$$* \underbrace{\sum_{i=1}^{n} X_i}_{\text{sum}} \sim N(n\mu, n\sigma^2)$$

$$\boxed{\sum_{i=1}^{3} X_i \neq 3X_i} \quad \text{caution}$$

$$Var[X_1 + X_2 + X_3] = 3\sigma^2$$

$$Var[3X_1] = 9\sigma^2$$

$N(\mu, \sigma^2)$

Sample mean example

$N(50, 100)$

Suppose a team's points per game is normally distributed with mean = 50, and standard deviation = 10. $X \sim N(50, 10^2)$

Let $\underbrace{X_1, X_2, \ldots X_n}$ be a random sample from this distribution.

- A) What is the distribution of $\bar{X}$? $\bar{X} \sim N(50, \frac{10^2}{n})$

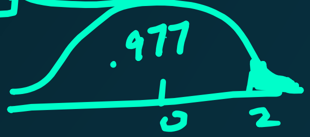
  $\bar{X}$

- B) Suppose Chloe takes a sample of size *n=16.* Find the probability that the average of this sample is greater than 55.

$P[\bar{X} > 55]$ $\bar{X} \sim N(50, \frac{100}{16})$

$P[Z > \frac{55-50}{10/4}] = P[Z > 2]$

$P[\bar{X} > 55] = 0.02275$



Effect of sample size on distribution

σ^2 / n

$X_i \sim$

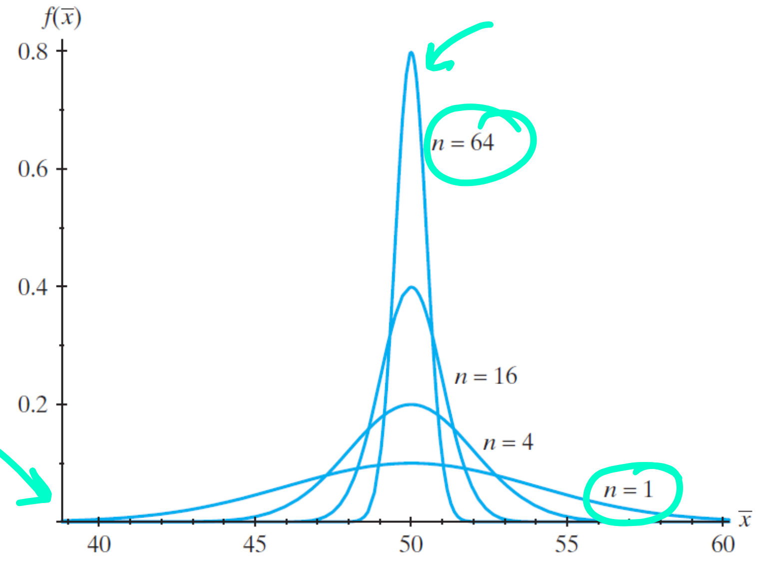


Figure 5.5-1 pdfs of means of samples from $N(50, 16)$

Central Limit Theorem

Theorem 5.6-1 **(Central Limit Theorem)** If $\overline{X}$ is the mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution with a finite mean μ and a finite positive variance σ^2, then the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$$

is $N(0, 1)$ in the limit as $n \to \infty$.

In other words: if $X_1, X_2, \ldots X_n$ all have mean μ, and variance σ^2, then

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$ as $n \to \infty$

Using the normal distribution:

If the population is already normally distributed, the sampling distribution of the sample mean is normal for **any** sample size n.

$X_1, X_2, \dots \sim N(\mu, \sigma^2)$

Using the CLT:

If the population is **not** normally distributed, **but** the sample size, n, is *large enough*, the sampling distribution of the sample mean, *X*, of *n* independent samples is **approximately** normal with mean μ and standard deviation $\sigma/\sqrt{n}$.

$n \to \infty$

Generally, $\boxed{n>30}$ or n>25 is considered large enough. Not a strict cutoff.

Central Limit Theorem (5.6)

Examples

$n = 100$

1 Let $\bar{X}$ be the mean of a random sample of size 100 from an exponential distribution with mean 5. Approximate the probability that the sample mean is between 4 and 6.

$X \sim Exp(\theta = 5)$ $\mu_x = 5$ $\sigma^2_x = 25$

$\bar{X} \overset{CLT}{\sim} N\left(5, \frac{25}{100}\right)$ $N\left(5, \frac{1}{4}\right)$ $\frac{1}{2}^2$ $\sigma_{\bar{x}} = 1/2$

$P\left[\frac{4-5}{1/2} < \frac{\bar{x}-5}{1/2} < \frac{6-5}{1/2}\right] = P\left[-2 < Z < 2\right]$

$P[4<\bar{X}<6] =$ $= .954$

2 The tensile strength of paper, X, has μ=30 and σ=3 pounds per square inch. A random sample of size n=64 is taken from this distribution. Compute the probability that the sample mean is greater than 29.5 pounds per square inch.

SD

$$P[\bar{X} > 29.5]$$

$$\bar{X} \sim N\left(30, \frac{9}{64}\right)$$

$$\sigma_{\bar{x}} = \frac{3}{8}$$

$$= P\Bigg[\underbrace{\frac{\bar{X} - 30}{3/8}}_{Z} > \frac{29.5 - 30}{3/8}\Bigg]$$

$$=$$

$P[\bar{X}>29.5] = P[Z > -1.333] = 0.9087$

3

Example 5.6-1 Let $\overline{X}$ be the mean of a random sample of $n = 25$ currents (in milliamperes) in a strip of wire in which each measurement has a mean of 15 and a variance of 4. Then $\overline{X}$ has an approximate $N(15, 4/25)$ distribution. As an illustration,

$$P(14.4 < \overline{X} < 15.6) = P\left(\frac{14.4 - 15}{0.4} < \frac{\overline{X} - 15}{0.4} < \frac{15.6 - 15}{0.4}\right)$$

$$\approx \Phi(1.5) - \Phi(-1.5) = 0.9332 - 0.0668 = 0.8664.$$