Distribution of a linear combination of independent random variables:
Let $X_1, X_2, \ldots, X_n$ be a collection of independent random variables with mgfs $m_{X_i}(t), i = 1,2,3,\ldots,n$.
Let $\mathrm{Y} = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n$, $a_0, a_1, \ldots, a_n$ are real numbers. Then the moment generating function for Y is given by:

$$m_Y(t) = e^{a_0 t} \prod_{i=1}^{n} m_{X_i}(a_i t)$$

Practice Example: Let X_1 $and X_2$ be independent normal random variables with means 2 and 5 and variances 9 and 1 respectively. Let $Y = 3X_1 + 6X_2 - 8$. What is the distribution of Y?

$$m_{X_1}(t) = e^{2t+\frac{9}{2}t^2}; m_{X_2}(t) = e^{5t+\frac{1}{2}t^2}$$

$$m_Y(t) = e^{-8t} m_{X_1}(3t) m_{X_2}(6t) = e^{-8t} e^{6t+\frac{81}{2}t^2} e^{30t+18t^2} = e^{28t+\frac{117}{2}t^2}$$

Hence, Y is a normal random variable with $\mu_Y = 28\ and\ \sigma_Y^2 = 117$, i.e., $Y \sim N(-1, 10)$

To verify the mean and variance of Y, we have,

$$E[Y\} = E[3X_1 + 6X_2 - 8] = 3E[X_1] + 6E[X_2] - 8 = 3 \times 2 + 6 \times 5 - 8 = 28$$

$$VarY = Var(3X_1 + 6X_2 - 8) = 9VarX_1 + 36VarX_2 = 9 \times 9 + 36 = 117$$

Theorem 7.3.4: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then $\bar{X}$ is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.

Proof:

Since $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, we have, $m_{X_i}(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, i = 1,2 \ \ldots, n$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1}{n} + \frac{X_2}{n} + \cdots + \frac{X_n}{n}$$

The moment generating function of $\bar{X}$ can be found as:

$$m_{\bar{X}}(t) = m_{X_1}\left(\frac{t}{n}\right) m_{X_2}\left(\frac{t}{n}\right) \ldots m_{X_n}\left(\frac{t}{n}\right)$$

$$= e^{\frac{\mu t}{n} + \frac{1}{2}\sigma^2 (\frac{t}{n})^2} e^{\frac{\mu t}{n} + \frac{1}{2}\sigma^2 (\frac{t}{n})^2} \cdots e^{\frac{\mu t}{n} + \frac{1}{2}\sigma^2 \left(\frac{t}{n}\right)^2} = e^{\mu t + \frac{1}{2}\frac{\sigma^2}{n} t^2}$$

Hence, we can claim that $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Theorem $Z \sim N(0,1)$. $Let\ X = Z^2$. $Then\ X \sim \mathrm{X}_1^2$ (chi-squared distribution with degree of 1)

Proof:

Recall a Gamma random variable X with parameters $\alpha\ and\ \beta$ has the density function:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}, \qquad x > 0, \alpha > 0, \beta > 0$$

When $\beta = 2\ and\ \alpha = \frac{\gamma}{2}$, we have a chi-squared distribution with γ degree of freedom, X_{γ}^2

$$f(x) = \frac{x^{\frac{\gamma}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{\gamma}{2}) 2^{\frac{\gamma}{2}}}$$

When $\gamma = 1$, the density function of X_1^2 is:

$$f(x) = \frac{x^{-1/2} e^{-\frac{x}{2}}}{\Gamma(\frac{1}{2}) 2^{1/2}} = \frac{x^{-1/2} e^{-\frac{x}{2}}}{\sqrt{2\pi}}, \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Now, we have to prove that $X = Z^2$ has the same density function of X_1^2

Proof (continued):
First, let's find the cumulative distribution function of X.

$$
\begin{aligned}
F_X(x) &= P[X \leq x] = P[Z^2 \leq x] = P[-\sqrt{x} \leq Z \leq \sqrt{x}] \\
&= P[Z \leq \sqrt{x}] - P[Z \leq -\sqrt{x}] = F_Z(\sqrt{x}) - F_Z(-\sqrt{x})
\end{aligned}
$$

Then, to find the density function $f_X(x)$, we take the derivative of $F_X(x)$ w.r.t. x

$$f_X(x) = \frac{1}{2} x^{-\frac{1}{2}} f_Z(\sqrt{x}) + \frac{1}{2} x^{-\frac{1}{2}} f_Z(-\sqrt{x})$$

Recall that the density function of a standard normal random variable is $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Then,

$$f_X(x) = \frac{1}{2} x^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x} + \frac{1}{2} x^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x} = \frac{x^{-\frac{1}{2}} e^{-\frac{1}{2}x}}{\sqrt{2\pi}}$$

Therefore, $X \sim \mathrm{X}_1^2$, i.e., $Z^2 \sim \mathrm{X}_1^2$

Theorem *Distribution of a sum of independent chi-squared random variables*
Let $X_1, X_2, \ldots, X_n$ be independent chi-squared random variables with $\gamma_1, \gamma_2, \ldots, \gamma_n$ degree of freedom respectively. Let $Y = X_1 + X_2, + \cdots + X_n$. Then $Y \sim \mathrm{X}^2_\gamma, \ \ \gamma = \sum_{i=1}^n \gamma_i$

Proof: Recall the mgf of a Gamma R.V. with parameters $\alpha \ and \ \beta$ is $(1-\beta t)^{-\alpha}$. A chi-squared R.V. is a special case of Gamma R.V. with $\beta = 2 \ and \ \alpha = \frac{\gamma}{2}$. Hence the mgf of a chi-squared R.V. is $(1-2t)^{-\gamma/2}$
We have,

$$m_{X_i}(t) = (1-2t)^{-\gamma_i/2}, \qquad i = 1,2,\ldots,n$$

Then,

$$m_Y(t) = \prod_{i=1}^n m_{X_i}(t) = \prod_{i=1}^n (1-2t)^{-\gamma_i/2} = (1-2t)^{-\sum_{i=1}^n \gamma_i/2}$$

Therefore,

$$Y \sim \mathrm{X}^2_\gamma, \qquad \gamma = \sum_{i=1}^n \gamma_i$$

Theorem Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \mathrm{X}_n^2$$

Proof: Since $X_1, X_2, \dots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2. We have,

$$X_i \sim N(\mu, \sigma^2), \qquad \frac{X_i - \mu}{\sigma} \sim N(0,1), \qquad i = 1,2, \dots, n$$

Then,

$$\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \mathrm{X}_1^2$$

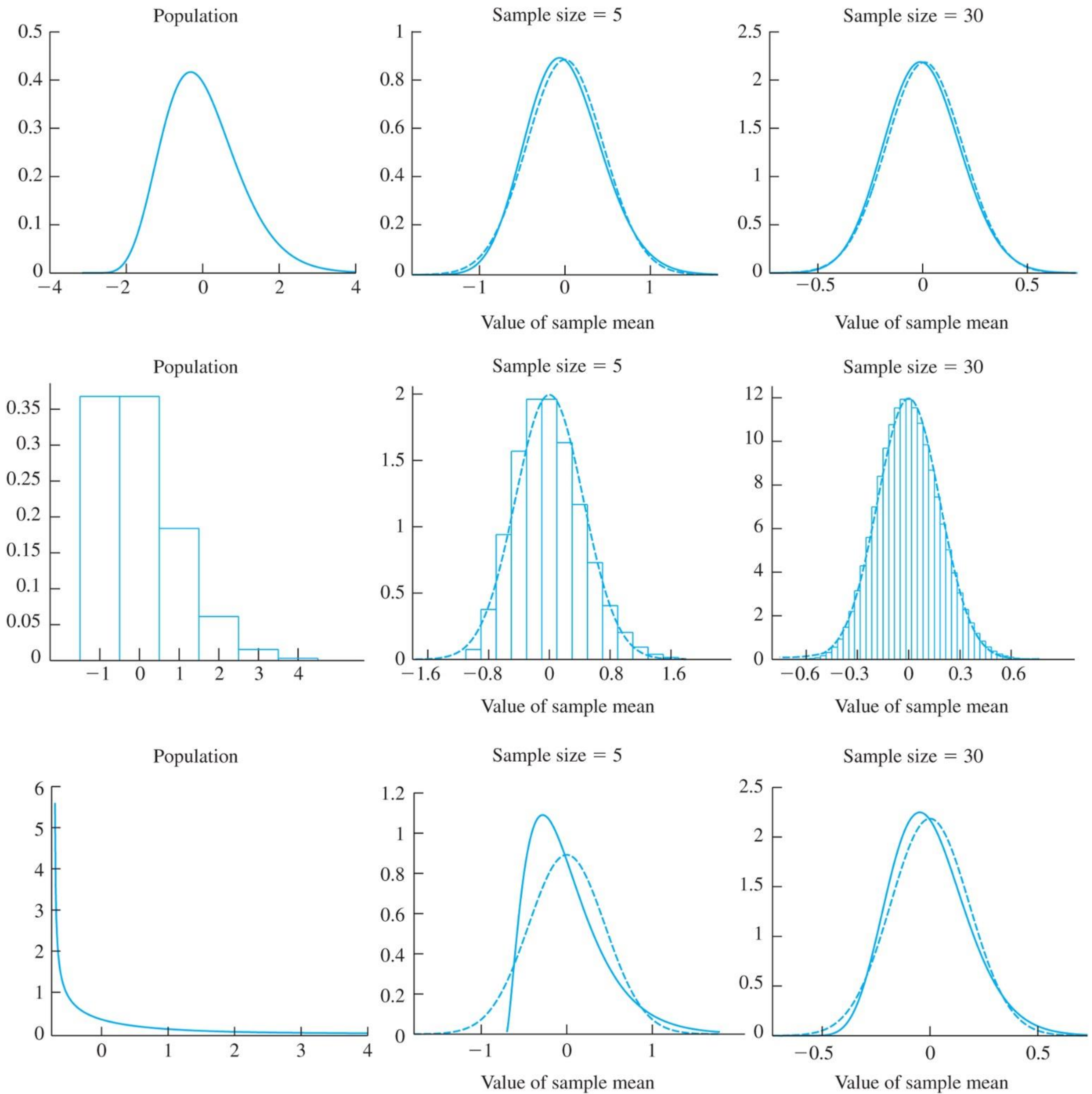
Therefore,

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} \sim \mathrm{X}_n^2$$

The Central Limit Theorem

Theorem 7.4.2: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, $\bar{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$. The random variable $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is approximately standard normal.

- ***This applies to any population (not only normal population!!!)***
- For most population, if the sample size is greater than 30, the central limit theorem approximation is good.
- If the sample is drawn from a nearly symmetric distribution, the normal approximation can be good for a fairly small value of n;
- If the population if heavily skewed, a fairly large n maybe necessary.


Population
Sample size = 5
Sample size = 30
Value of sample mean
Value of sample mean
Population
Sample size = 5
Sample size = 30
Value of sample mean
Value of sample mean
Population
Sample size = 5
Sample size = 30
Value of sample mean
Value of sample mean

The Central Limit Theorem

Example: Let $X_1, X_2, \ldots, X_{100}$ be a random sample of size 100 from a Gamma distribution with $\alpha = 5\ and\ \beta = 3$. Find the distribution of $\bar{X}$ using the moment generating function method and the central limit theorem.

Solution: (the moment generating function method)

Since $X_i, i = 1,2, \ldots, 100$ are the random samples from a Gamma distribution with $\alpha = 5\ and\ \beta = 3$, The moment generating function of X_i is $(1-3t)^{-5}$. Then the mgf of $\bar{X} = \frac{1}{100}\sum_{i=1}^{100} X_i$ is:

$$m_{\bar{X}}(t) = \prod_{i=1}^{100} m_{X_i}\left(\frac{t}{100}\right) = (1-0.03t)^{-500}$$

Hence, $\bar{X}$ is a Gamma R.V. with $\alpha = 500\ and\ \beta = 0.03$. i.e.,

$$f_{\bar{X}}(x) = \frac{1}{\Gamma(500) \times 0.03^{500}} x^{499} e^{-\frac{x}{0.03}}$$

This is very hard to evaluate!

The Central Limit Theorem

Example: Let $X_1, X_2, \dots, X_{100}$ be a random sample of size 100 from a Gamma distribution with $\alpha = 5\ and\ \beta = 3$. Find the distribution of $\bar{X}$ using the moment generating function method and the central limit theorem. Find the probability that $\bar{X}$ is at most 14.

Solution: (the central limit theorem method)

According to the central limit theorem, $\bar{X}$ is approximately following a normal distribution with mean μ and variance $\frac{\sigma^2}{100}$, where $\mu\ and\ \sigma^2$ are the mean and variance of the population variable.

Since the population variable follows a Gamma distribution with $\alpha = 5\ and\ \beta = 3$, we have,

$$\mu = \alpha\beta = 15;\ \sigma^2 = \alpha\beta^2 = 45$$

$$\bar{X} \sim N(15, 0.45)$$

$$P[\bar{X} \le 14] = P[\bar{X} - 15 \le -1] = P\left[\frac{\bar{X} - 15}{\sqrt{0.45}} \le -\frac{1}{\sqrt{0.45}}\right] = P[Z \le 1.49] = 0.0681$$