1. Sampling from the Normal Population

*Example: We wish to estimate the distribution of heights of adult US male. It is believed that the height of adult US male follows a normal distribution \( N(\mu, \sigma^2) \)

**Def. Simple** random sample: A sample in which every subject in the population has the same chance to be selected.

\( X \) : The Random Variable denote the height of a adult male we will choose randomly from the population

So \( X \sim N(\mu, \sigma^2) \): the distribution of a randomly selected subject is the population distribution.

**Theorem 1  Sampling from the normal population**

Let \( X_1, X_2, \ldots, X_n \) i.i.d. \( N(\mu, \sigma^2) \), where i.i.d stands for independent and identically distributed

1. \( \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \)

2. \( \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1} \) (Chi Square distribution with \((n-1)\) degrees of freedom),

*Reminder: The Sample variance \( S^2 \) is defined as: \( S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \)

*Def 1: The Chi-square distribution is a special gamma distribution (** Please find out which one it is.**)

*Def 2: Let \( Z_1, Z_2, \ldots, Z_k \) i.i.d. \( N(0,1) \),

Then \( W = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k \)

3. \( \bar{X} \) & \( S^2 \) are independent.

4. \( Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \)
5. \[ T = \frac{\bar{X} - \mu}{S \sqrt{n}} \sim t_{n-1} \] (t distribution with (n-1) degrees of freedom)

*Def. of t-distribution* (Student's t-distribution, first introduced by William Sealy Gosset)

Let \( Z \sim N(0,1) \), \( W \sim \chi^2_k \), where \( Z \& W \) are independent.

Then, \( T = \frac{Z}{\sqrt{W/k}} \sim t_k \)

Wiki: William Sealy Gosset (June 13, 1876–October 16, 1937) is best known by his pen name Student and for the Student's t-distribution.

*Proof of 5*

We know that \( Z = \frac{\bar{X} - \mu}{\sigma \sqrt{n}} \sim N(0,1) \) and \( W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \)
Furthermore, since \( \overline{X} \) & \( S^2 \) are independent, thus \( Z \) & \( W \) are independent.

Therefore by the definition of t-distribution, \[ \frac{Z}{\sqrt{\frac{W}{n-1}}} \sim t_{n-1} \]

*We will first prove #2 & #3 for the special case of n=2. In that case, we have:

Independent
\[
\begin{align*}
X_1 &\sim N(\mu, \sigma^2) \\
X_2 &\sim N(\mu, \sigma^2)
\end{align*}
\]

\[
S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2}{1} = \frac{1}{2}(X_1 - X_2)^2
\]

*#2

\[
W = \frac{(n-1)S^2}{\sigma^2} = \frac{S^2}{\sigma^2} = \frac{(X_1 - X_2)^2}{2\sigma^2}
\]

\[
Z = \frac{X_1 - X_2}{\sqrt{2\sigma}} \sim \chi^2_1 \quad \text{*You can prove this using the pivotal quantity method}
\]

\[
\therefore W = Z^2 \sim \chi^2_1 \quad \text{(using the 2nd Definition of Chi Square Distribution.)}
\]

*#3

If we can show \( X_1 + X_2 \) and \( X_1 - X_2 \) are independent then we have proven that \( \overline{X} \) and \( S^2 \) are independent.

**Approach 1:** p.d.f. \[ f_{X_1+X_2,X_1-X_2}(\mu, \nu) = f_{X_1+X_2}(\mu) f_{X_1-X_2}(\nu) \]

**Approach 2:** m.g.f. \[ M_{X_1+X_2,X_1-X_2}(t_1, t_2) = M_{X_1+X_2}(t_1) M_{X_1-X_2}(t_2) \]

The joint m.g.f. of X and Y is defined as:

\[ M_{X,Y}(t_1, t_2) = E(e^{t_1X-t_2Y}) \]

**Note:** We have done this already.

**Additional Questions and Solutions**

Q2. Prove \( E(S^2) = \sigma^2 \) for any distribution/population.

**Solution**
\[(n-1)S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2\]

\[= \sum_{i=1}^{n} (X_i - \mu + \mu - \bar{X})^2 = \sum_{i=1}^{n} [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2]\]

\[= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)\left(\sum_{i=1}^{n} X_i - n\mu\right)\]

\[= \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2\]

\[(n-1)E(S^2) = E\left[\sum_{i=1}^{n} (X_i - \mu)^2\right] - nE\left[(\bar{X} - \mu)^2\right]\]

\[= n\sigma^2 - n\text{Var}(\bar{X})\]

\[\rightarrow (n-1)\sigma^2\]

\[\therefore E(S^2) = \sigma^2\]

Q3.

- Please point out a chi-square random variable with k degrees of freedom corresponds to which particular gamma distribution.
- Please write down the pdf, mgf, mean, and variance of a general gamma distribution and of a chi-square random variable with k degrees of freedom.

Solution

Let \( W \sim \chi^2_k \). \( W \) is indeed a special random variable.

**Gamma distribution**

\( X \sim \text{gamma}(\alpha, \beta) \)  
(Some books use \( r=\alpha, \quad \lambda = \frac{1}{\beta} \))

where

\[f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad 0 < x < \infty\]

or
\[
f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad 0 < x < \infty
\]

if \( r \) is a non-negative integer, then

\[
\Gamma(r) = (r-1)!
\]

\[
1 = \int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}dx
\]

\[
\Gamma(r) = \lambda^r \int_{0}^{\infty} x^{r-1} e^{-\lambda x}dx
\]

\[
M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r} = \left(\frac{\lambda - t}{\lambda}\right)^{-r}
\]

\[
\Rightarrow M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r
\]

\[
E(X) = \frac{r}{\lambda}
\]

\[
Var(X) = \frac{r}{\lambda^2}
\]

**Special case**: when

\[
r = \frac{k}{2}, \quad \lambda = \frac{1}{2} \quad \Rightarrow \quad X \sim \chi_k^2
\]

**Special case**: when \( r = 1 \) \( \Rightarrow \) \( X \sim \exp(\lambda) \)

**Review**

\( X \sim \exp(\lambda) \)

p.d.f.

\[
f(x) = \lambda e^{-\lambda x}, \quad x > 0
\]

m.g.f.

\[
M_X(t) = \frac{\lambda}{\lambda - t}
\]

**Q4.** Let \( X_i \overset{i.i.d.}{\sim} \exp(\lambda), \quad i = 1, \cdots, n \). What is the distribution of \( \sum_{i=1}^{n} X_i \) ?
Solution

\[ M_{\sum X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left( \frac{\lambda}{\lambda - t} \right)^n \]

\[ \therefore \sum X_i \sim \text{gamma}(r = n, \lambda) \]

Q5. Let \( W \sim \chi^2_k \). What is the mgf of \( W \)?

Solution

\[ M_W(t) = \left( \frac{1/2}{1/2 - t} \right)^{k/2} \]

\[ \Rightarrow M_W(t) = \left( \frac{1}{1 - 2t} \right)^{k/2} \]

Q6. Let \( W_1 \sim \chi^2_{k_1} \), \( W_2 \sim \chi^2_{k_2} \), and \( W_1 \) and \( W_2 \) are independent. What is the distribution of \( W_1 + W_2 \)?

Solution

\[ M_{W_1 + W_2}(t) = M_{W_1}(t) \cdot M_{W_2}(t) = \left( \frac{1}{1 - 2t} \right)^{k_1/2} \left( \frac{1}{1 - 2t} \right)^{k_2/2} = \left( \frac{1}{1 - 2t} \right)^{(k_1 + k_2)/2} \]

\[ \Rightarrow W_1 + W_2 \sim \chi^2_{k_1 + k_2} \]
2. More general understanding of sampling and the sample mean distribution.

**Definition:** Sampling error is the error resulting from using a sample to infer a population characteristic.

**Example:** We want to estimate the mean amount of Pepsi-Cola in 12-oz. cans coming off an assembly line by choosing a random sample of 16 cans, and using the sample mean as an estimate of the mean for the population of cans. Suppose that we choose 100 random samples of size 16 and compute the sample mean for each of these samples. These 100 values of $\bar{X}$ will differ from each other somewhat due to sampling error; but the values should all be close to 12-oz.

**Definition:** For a random variable $X$, and a given sample size $n$, the distribution of the variable $\bar{X}$, i.e., of all possible values of $\bar{X}$, is called the sampling distribution of the mean. This probability distribution is a set of pairs of numbers. In each pair, the first number is a possible value of the sample mean, and the second number is the probability of obtaining that value of the mean occur when we select a random sample from the population.

**Properties of the Sampling Distribution of the Mean:**

1) For samples of size $n$, the expectation (mean) of $\bar{X}$, equals the expectation (mean) of $X$. In other words, $\mu_{\bar{X}} = \mu_X$.

2) The possible values of $\bar{X}$ cluster closer around the population mean for larger samples than for smaller samples. In other words, the larger the sample size, the smaller the sampling error. In particular, the standard deviation of the sampling distribution of the means, $\sigma_{\bar{X}}$, will be smaller than the population standard deviation, $\sigma_X$. In particular, we have $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$, where $n$ is the sample size.

**Example:** We can easily list the sampling distribution of the mean only when both the population size and the sample size are very small. Suppose that a professor gives an eight-point quiz to a class of four students. Let the class be the population, of size $N = 4$. Suppose the scores are 2, 4, 6, and 8. We can easily calculate the population mean and the population standard deviation, using the formulae given in chapter 3. We have

$$\mu = \frac{2 + 4 + 6 + 8}{4} = 5, \quad \text{and} \quad \sigma = \sqrt{\frac{(2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2}{4}} = 2.236.$$ 

The population distribution is discrete uniform; i.e., if we randomly select one member from the population, we are equally likely to find $x = 2, x = 4, x = 6$ or $x = 8$, where the random
variable $X$ is the score of our randomly selected student.

The probability distribution for $X$ is therefore

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>0.25</td>
</tr>
<tr>
<td>8</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Suppose, now, that the professor does not want to calculate the population mean using the above formula, but wants to estimate the population mean using a sample of 2 students. There are several such samples which could be selected. How many? If we sample with replacement; i.e., if we allow the possibility that the same student may be selected more than once for a sample, then the number of possible samples is (order of selection is important here) 16. These are {2, 2}, {2, 4}, {2, 6}, {2, 8}, {4, 2}, {4, 4}, {4, 6}, {4, 8}, {6, 2}, {6, 4}, {6, 6}, {6, 8}, {8, 2}, {8, 4}, {8, 6}, {8, 8}. If we then compute the sample mean for each one of these samples, we find the sampling distribution of the mean listed in the following table.

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0.0625</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.125</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.1875</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.1875</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0.125</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

Thus, if the professor randomly selects a sample of size 2 from the class, she will have a 0.0625 probability of obtaining a sample mean $\bar{x} = 2$, a 0.125 probability of obtaining a sample mean $\bar{x} = 3$, and so forth. The most likely value of the sample mean is 5, with a probability of 0.25.

The expectation of $\bar{x}$ may be easily calculated from the above table. We get:

$$\mu_{\bar{x}} = \frac{2 + 3 + 3 + 4 + 4 + 4 + 5 + 5 + 5 + 5 + 5 + 6 + 6 + 6 + 7 + 7 + 8}{16} = 5.$$  

For the standard deviation, we obtain

$$\sigma_{\bar{x}} = \sqrt{\frac{(2 - 5)^2 + (3 - 5)^2 + (3 - 5)^2 + \ldots + (7 - 5)^2 + (7 - 5)^2 + (8 - 5)^2}{16}} = 1.581$$

We find that $\sigma_{\bar{x}} = 1.581 = \frac{2.236}{\sqrt{2}} = \frac{\sigma}{\sqrt{2}}$, consistent with property 2 given above.

**Definition**: The standard deviation of the sampling distribution of the mean is called the **standard error of the mean**.
Property 2 says that for a given population, and a given random variable defined for the members of that population, the standard error of the mean is smaller for larger sample sizes. For example, assume that our population standard deviation is 1. If our sample size is 4, then the standard error of the mean is 0.5. If our sample size is 16, then the standard error of the mean is 0.25. If our sample size is 100, then the standard error of the mean is 0.10.

The following theoretical result from probability theory is fundamental for our work in statistical inference.

**The Central Limit Theorem:** For large \( n \geq 30 \) sample sizes, the random variable \( X \) has an approximate normal distribution, with mean \( \mu_X = \mu_x \) and standard deviation \( \sigma_X = \frac{\sigma_x}{\sqrt{n}} \). In other words, the random variable \( Z = \frac{X - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} \) has an approximate standard normal distribution.

This theorem holds regardless of the type of population distribution. The population distribution could be normal; it could be uniform (equally likely outcomes); it could be strongly positively skewed; it could be strongly negatively skewed. Regardless of the shape of the population distribution, the sampling distribution of the mean will be approximately normal for large sample sizes.

3. **The Sampling Distribution of the Sample Proportion**

Assume that we have a (large) population, which is divided into two subpopulations. In one subpopulation, each member possesses a certain characteristic; in the other subpopulation, each member does not possess this characteristic. Assume that the proportion of members of the entire population who possess the characteristic is \( p \).

We select a simple random sample of size \( n \) from the population. We are interested in the proportion of the members of the sample who possess the characteristic of interest. This proportion is called the sample proportion, denoted by \( \hat{p} \).

The Central Limit Theorem tells us that, if the sample size is “large,” then

a) The shape of the sampling distribution of \( \hat{p} \) is approximately normal, having mean \( p \) and standard deviation \( \sqrt{\frac{p(1-p)}{n}} \), provided \( np(1-p) \geq 10 \).

b) Under the same conditions, the distribution of

\[
Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}
\]
is approximately standard normal.