Definition 5.1

- Let X be a random variable with certain probability distribution, $f_X(x)$.
- Let $X_1, X_2, \cdots, X_n$ be n independent random variables each having the same distribution as X,
- then $(X_1, X_2, \cdots, X_n)$ is called a random sample of size n from a population with distribution $f_X(x)$.
- The joint p.f. (or p.d.f.) of $(X_1, X_2, \cdots, X_n)$ is given by $f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$, where $f_X(x)$ is the p.f. (or p.d.f.) of the population.

✓ When doing the sampling, the population could be finite or infinite. We need to consider two scenarios, when we talk about the distribution of the sampled random variables:

★ The population is finite and the sample is drawn without replacement.

★ The population is infinite and the sample is drawn with/without replacement; or the population is finite, but the sample is drawn with replacement.

Read pages 5-3 to 5-20 and the corresponding lecture videos for details.

✓ A more convenient way to describe a sample is given on page 5-21. We say that $(X_1, X_2, \ldots, X_n)$ is a random sample from a distribution with probability function $f_X(x)$. Here "a random sample" is equivalent to state "$X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.)". Also pay attention to the joint pf given in this page of the lecture slide.

Pages 5-23

5.3.2 Statistic and Sampling Distribution

- A function of a random sample $(X_1, X_2, \cdots, X_n)$ is called a **statistic**. (e.g. $\bar{X}$ is a statistic as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$)
- Hence a statistic is a random variable. It is meaningful to consider the probability distribution of a statistic.
- The probability distribution of a statistic is called a **sampling distribution**.

We give some discussion about the the random sample, parameter, and statistic.

✓ The goal of drawing a random sample, $(X_1, X_2, \ldots, X_n)$, is to make inference for some unknown parameters of the population. For simplicity, you can think of the unknown parameter as the expected value (i.e., μ) of a population to start with.

✓ The population parameter, μ say, is **not observed**; but the values of $X_1, X_2, \ldots, X_n$ are observed as $x_1, x_2, \ldots, x_n$. So, we hope to establish an "estimation rule" to estimate the unknown μ using the observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$; such a rule $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called a statistic. In the literature, $\bar{X}$ is also called an "estimator" for μ. $\bar{x}$ is **computed/observed** based on the sample, and is called an estimate. Make sure that you can distinguish among the population mean μ, the sample mean $\bar{X}$, and the computed sample mean $\bar{x}$.

✓ The statistic is a function of the sample; the fundamental requirement is that it does not depend on the unknown parameters. Any function of the sample can be viewed as a statistic, for example $g_1(X_1,\ldots,X_n)=1$, $g_2(X_1,\ldots,X_n)=\min\{X_1,\ldots,X_n\}$, $g_3(X_1,\ldots,X_n)=X_1$, $g_4(X_1,\ldots,X_n)=(X_1+X_2)/2$ can all be called as a statistic, or an estimator for μ.

✓ The statistical performance of the statistic decides which one is better to use in practice. In other words, we would be interested in studying the distribution of a statistic, called "sampling distribution". The sampling distribution is also crucial for the subsequent inference for the corresponding parameter.

✓ The sampling distribution for a statistic, e.g., $\bar{X}$, is not how the sample $(X_1,\ldots,X_n)$ performs; instead, it is the distribution for the corresponding statistic, $\bar{X}$ say. So, constructing the histogram based on the observed $X_1=x_1, X_2=x_2,\ldots,X_n=x_n$ to have a view of the sampling distribution of $\bar{X}$ makes nonsense. Instead, we should do the following:

- ★ Get the first sample $\left(X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}\right)$ from the distribution $f_X(x)$, and compute the sample mean $\bar{X}^{(1)}$.
- ★ Get the second sample $\left(X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)}\right)$ from the distribution $f_X(x)$, and compute the sample mean $\bar{X}^{(2)}$.
- ★ Continue with this procedure $\ldots,\ldots,\ldots$
- ★ Get the Kth sample $\left(X_1^{(K)}, X_2^{(K)}, \ldots, X_n^{(K)}\right)$ from the distribution $f_X(x)$, and compute the sample mean $\bar{X}^{(K)}$.
- ★ Draw the histogram of $\bar{X}^{(1)}, \bar{X}^{(2)}, \ldots, \bar{X}^{(K)}$ to have a view of the distribution for $\bar{X}$.

Pages 5-31

Sampling Distribution of Sample Mean

Theorem 5.1

- For random samples of size n taken from an **infinite population** or from a **finite population with replacement** having population mean μ and population standard deviation σ,
- the **sampling distribution of the sample mean** $\bar{X}$ has its mean and variance given by

$$\mu_{\bar{X}} = \mu_X \quad \text{and} \quad \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}.$$

Deriving $\mu_{\bar{X}} = \mu_X$ is easy. Deriving $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$ relies on the formula $V(X_1 \pm X_2 \pm \ldots \pm X_n) = V(X_1) + V(X_2) + \ldots V(X_n)$ if $X_1, \ldots, X_n$ are independent. The condition "$X_1, X_2, \ldots, X_n$ are i.i.d. random variables" play an important role in this development.

Pages 5-33

Law of Large Number (LLN)

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and **finite** population variance σ^2. Then for any $\epsilon \in \mathbb{R}$,

$$P(|\bar{X} - \mu| > \epsilon) \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Law of large numbers and central limit theorem are fundamental theorems in statistics. That given on this slide is called the weak law of large numbers. "$P(|\bar{X} - \mu| > \epsilon) \to 0$ as $n \to \infty$" gives the definition of one type of statistical convergence: "convergence in probability". By convention, it is stated as "$\bar{X} \to \mu$ in probability, as $n \to \infty$. "

With Chebyshev's inequality, this theorem can be easily proved. Note that ϵ is an arbitrary, but fixed constant, we have

$$0 \leq P(|\bar{X} - \mu| > \epsilon) \leq \frac{V(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{as } n \to \infty.$$

Practically, Chebyshev's inequality can also give an upper bound of this probability. See pages 5-35 & 36 of the lecture slides for an example:

Law of Large Number (LLN) (Continued)

Example

Let $X_1, \cdots, X_n$ be a random sample from $U(0,1)$.

Then $E(X) = 1/2$ and $V(X) = 1/12$

Take $n = 3(10^6)$ and $\epsilon = 0.001$.

Let $\bar{X}_n$ be the sample mean from a sample of size $3(10)^6$.

Hence, $E(\bar{X}) = \frac{1}{2}$ and $V(\bar{X}) = \frac{1/12}{3(10^6)}$

Law of Large Number (LLN) (Continued)

Example (Continued)

Consider $\Pr(|\bar{X}_n - 0.5| > 0.001)$.

Find k such that $k\sqrt{V(\bar{X})} = 0.001$ or $k = 0.001\big(6(10^3)\big) = 6$

$$\Pr(|\bar{X}_n - 0.5| > 0.001) = \Pr\left(|\bar{X}_n - 0.5| > 6\sqrt{V(\bar{X})}\right)$$

By Chebyshev's Inequality,

$$\Pr\left(|\bar{X}_n - 0.5| > 6\sqrt{V(\bar{X})}\right) \leq \frac{1}{6^2} = 0.02778$$

Pages 5-37

5.4 Central Limit Theorem and its applications

Central Limit Theorem

- Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and finite population variance σ^2.
- The sampling distribution of the sample mean $\bar{X}$ is **approximately normal** with mean μ and variance σ^2/n if ***n* is sufficiently large**.
- Hence

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ follows approximately } N(0, 1)$$

Central limit theorem only gives the approximate distribution for $\bar{X}$, not the exact distribution. More rigorously, central limit theorem says, for any $z \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z),$$

where $\Phi(z)$ is the cdf for the standard normal distribution.

So, practically, when n is large, the distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is similar to that of a standard normal random variable.

Pages 5-39

Theorem 5.2

- If $X_i, i = 1, 2, \cdots, n$ are $N(\mu, \sigma^2)$, then $\bar{X}$ is $N(\mu, \frac{\sigma^2}{n})$ regardless of the sample size n.
- Similarly, if $X_i, i = 1, 2, \cdots, n$ are approximately $N(\mu, \sigma^2)$, then $\bar{X}$ is approximately $N(\mu, \frac{\sigma^2}{n})$ regardless of the sample size n.

Useful website:
http://onlinestatbook.com/stat_sim/index.html

ST2334 Probability and Statistics CYM **Sampling and Sampling Distributions** 5-39

Note that this theorem has nothing to do with the central limit theorem. With the normality assumption, the distribution of $\bar{X}$ is exactly normal.

Compare this result with the central limit theorem; pay attention to the different conditions.

Some discussion on the chi-square distribution, t distribution and F distribution. **Note that all these distributions are founded on the standard normal distribution.**

✓ Let $Z_1, Z_2, \ldots, Z_n$ be i.i.d. $N(0,1)$ random variables, then $Y = Z_1^2 + Z_2^2 + \ldots + Z_n^2$ follows the $\chi^2(n)$ distribution. This can serve as the definition of the chi-square distribution.

This definition is helpful for us to understand/derive the properties of the chi-square distribution given on page 5-61 of the lecture slides:

Some properties of Chi-square distributions

1. If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$.
2. For large n, $\chi^2(n)$ approx $\sim N(n, 2n)$.
3. If $Y_1, Y_2, \cdots, Y_k$ are independent chi-square random variables with $n_1, n_2, \cdots, n_k$ degrees of freedom respectively, then $Y_1 + Y_2 + \cdots + Y_k$ has a chi-square distribution with $n_1 + n_2 + \cdots + n_k$ degrees of freedom. That is,

$$\sum_{i=1}^k Y_i \sim \chi^2\left(\sum_{i=1}^k n_i\right)$$

ST2334 Probability and Statistics CYM Sampling and Sampling Distributions 5-61

★ For property 1, since $Z_i \sim N(0,1)$, $E(Z_i^2) = V(Z_i) + (E(Z_i))^2 = 1$. Therefore $E(Y) = nE(Z_i^2) = n$. $V(Y) = nV(Z_i^2)$. The evaluation of $V(Z_i^2)$ is omitted.

★ Property 2 is the outcome of the expression $Y = Z_1^2 + Z_2^2 + \ldots + Z_n^2$, property 1, and the central limit theorem.

★ For property 3, we can write

$$
\begin{aligned}
Y_1 &= Z_{1,1}^2 + Z_{1,2}^2 + \ldots + Z_{1,n_1}^2 \\
Y_2 &= Z_{2,1}^2 + Z_{2,2}^2 + \ldots + Z_{2,n_2}^2 \\
\ldots \quad & \ldots \quad \ldots\ldots\ldots \\
Y_k &= Z_{k,1}^2 + Z_{k,2}^2 + \ldots + Z_{k,n_2}^2,
\end{aligned}
$$

where all the $Z_{i,j}$ are i.i.d. $N(0,1)$ random variables. Therefore $Y_1 + Y_2 + \ldots + Y_k$ is the summation of the square of $n_1 + n_2 + \ldots + n_k$ i.i.d. $N(0,1)$ random variables, which follows the $\chi^2(n_1 + n_2 + \ldots + n_k)$ distribution.

Note that this definition immediately verifies Theorem 5.5. given on page 5-62 of the lecture slides:

Theorem 5.5

1. If $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$.
2. Let $X \sim N(\mu, \sigma^2)$, then $[(X-\mu)/\sigma]^2 \sim \chi^2(1)$.
3. Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal population with mean μ, and variance σ^2. Define
$$Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$
Then $Y \sim \chi^2(n)$

ST2334 Probability and Statistics CYM Sampling and Sampling Distributions 5-62

✓ The t distribution defined on page 5-69 is also based on the standard normal distribution (and its resulted chi-square distribution):

5.8 The t-distribution

Definition 5.4

- Suppose $Z \sim N(0,1)$ and $U \sim \chi^2(n)$.
- If Z and U are independent, and let

$$T = \frac{Z}{\sqrt{U/n}}$$

- then the random variable T follows the t-distribution with n degrees of freedom. That is,

$$\frac{Z}{\sqrt{U/n}} \sim t(n)$$

Besides the distribution of Z and U, please also take note that Z in the numerator and U in the denominator need to be independent.

This expression can also help us understand why the t distribution approaches the standard normal distribution when its degrees of freedom n approaches infinity: its denominator

$$\frac{U}{n} = \frac{Z_1^2 + Z_2^2 + \ldots + Z_n^2}{n} \to E(Z_i^2) = 1,$$

based on Law of Large Numbers.

Another view of this result is given on page 5-72 of the lecture slide:

Properties of a *t*-distribution (Continued)

2. It can be shown that the p.d.f. of t-distribution with n d.f. is approaching to the p.d.f. of standard normal distribution when $n \to \infty$.

   That is

$$\lim_{n\to\infty} f_T(t) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

   as $n \to \infty$.

✓ F distribution is defined as the ratio of two independent chi-square distributions; note that it has two degrees of freedoms, one is from the numerator and the other is from the denominator.

The *F*-distribution

Definition 5.5

- Let U and V be independent random variables having $\chi^2(n_1)$ and $\chi^2(n_2)$, respectively,
- then the distribution of the random variable,

$$F = \frac{U/n_1}{V/n_2},$$

  is called a F distribution with (n_1, n_2) degrees of freedom.

This definition immediately leads to Theorem 5.7 given on page 5-84 of the lecture slides:

NUS National University of Singapore

Theorem 5.7

- If $F \sim F(n, m)$, then $1/F \sim F(m, n)$.
- This theorem follows immediately from the definition of F-distribution.
- Values of the F-distribution can be found in the statistical tables.
- The table gives the values of $F(n_1, n_2; \alpha)$ such that $\Pr\big(F > F(n_1, n_2; \alpha)\big) = \alpha$.

Adobe Acrobat Document

which further leads to Theorem 5.8 given on page 5-86:

NUS National University of Singapore

Theorem 5.8

$$F(n_1, n_2; 1-\alpha) = 1/\,F(n_2, n_1; \alpha).$$

For example,

- $F(10, 5; 0.95) = 1/F(5, 10; 0.05) = 1/3.33 = 0.30$ which means $\Pr(F > 0.30) = 0.95$, where $F \sim F(10, 5)$.

Here is a derivation: Let $F \sim F(n_1, n_2)$, then $1/F \sim F(n_2, n_1)$ based on Theorem 5.7. Based on the definition of $F(n_1, n_2, 1-\alpha)$,

$$Pr\left(F > F(n_1, n_2, 1-\alpha)\right) = 1-\alpha,$$

which leads to

$$Pr(F < F(n_1, n_2, 1-\alpha)) = 1 - Pr\left(F > F(n_1, n_2, 1-\alpha)\right) = \alpha.$$

That is

$$Pr(1/F < 1/F(n_1, n_2, 1-\alpha)) = \alpha,$$

which together with the fact $1/F \sim F(n_2, n_1)$ implies

$$1/F(n_1, n_2, 1-\alpha) = F(n_2, n_1, \alpha).$$

✓ Checking the definition of the t and F distributions, we can find one connection between them: if $Y \sim t(n)$, then $Y^2 \sim F(1, n)$.

We give a brief summary of the sampling distribution of some statistic established on the random sample $X_1, X_2, \ldots, X_n$. Recall that the observations in a random sample are all independent. Take notes of the conditions for the corresponding results to be correct.

- ✓ If $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$, $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ follows $N(0, 1)$ regardless of the sample size n; see page 5-39. If $X_1, X_2, \ldots, X_n$ have mean μ and finite variance σ^2, and n is sufficiently large, then $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ approximately follows the $N(0, 1)$ distribution; see page 5-37.

- ✓ If $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$, then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

where S^2 is the sample variance of the random sample; see pages 5-67 & 68.

- ✓ If $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$, then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1);$$

see pages 5-76 & 77.

- ✓ Let $X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}$ be $N(\mu_1, \sigma_1^2)$ and let $X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}$ be $N(\mu_2, \sigma_2^2)$. Denote by S_1^2 and S_2^2 be the sample variances of $X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}$ and $X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}$ respectively. Then we have

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1);$$

see page 5-82 & 83.