1 Random Samples, CB Chapter 5

Sample $X^n = X_1, \ldots, X_n \sim F$ with pdf/pmf $f_X$ means iid (independent, identically distributed). The joint pdf or pmf of $X_1, \ldots, X_n$ is given by

$$f(x^n) = f_{X^n}(x_1, \ldots, x_n) = f_X(x_1)f_X(x_2)\cdots f_X(x_n) = \prod_{i=1}^{n} f_X(x_i).$$

**Definition 1** Let $X_{(1)}, \ldots, X_{(n)}$ denote the ordered values:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

Then $X_{(1)}, \ldots, X_{(n)}$ are called the order statistics.

We’ll discuss order statistics in more details later as needed. More generally, a statistic is any function

$$T = g(X_1, \ldots, X_n)$$

which itself is a random variable. The probability distribution of $T$ is called the sampling distribution of $T$. The sample summary given by a statistic include many types of information.

**Examples of statistics:**

- order statistics, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
- sample mean: $\overline{X}_n = \frac{1}{n} \sum_i X_i$,
- sample variance: $S_n^2 = \frac{1}{n-1} \sum_i(X_i - \overline{X}_n)^2$,
- sample median: middle value of ordered statistics,
- sample minimum: $X_{(1)}$
- sample maximum: $X_{(n)}$
- sample range: $X_{(n)} - X_{(1)}$
Sample mean and variances

Theorem 2 Let \(X_1, \ldots, X_n\) be a random sample from a population with mean \(\mu\) and variance \(\sigma^2 < \infty\). That is, \(\mu = \mathbb{E}(X_i)\) and \(\sigma^2 = \text{Var}(X_i)\). Then

\[
\mathbb{E}(\overline{X}_n) = \mu, \quad \text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}, \quad \mathbb{E}(S^2_n) = \sigma^2.
\]

(1)

Proof. We show only the last one in (1).

\[
E(S^2_n) = E \left( \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \right)
= \frac{1}{n-1} \left( E \left( \sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2 \right) \right)
= \frac{1}{n-1} \left( nE(X_1^2) - nE(\overline{X}_n^2) \right)
= \frac{1}{n-1} \left( nE(\sigma^2 + \mu^2) - n \left( \frac{\sigma^2}{n} + \mu^2 \right) \right) = \sigma^2
\]

where in the last line we used the first two facts from (1).

Definition 3 We call \(\overline{X}_n\) an unbiased estimator of \(\mu\), and \(S^2_n\) an unbiased estimator of \(\sigma^2\) given that (1) holds.

Exercise: Check the following fact:

\[
(n-1)S^2_n = \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 = \sum_{i=1}^{n} X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2.
\]

2 Moment generating function: Review

Definition 4 Let \(X\) be a RV with cdf \(F_X\). The Moment generating function (mgf) of \(X\) (or \(F_X\)), denoted by \(M_X(t)\) is

\[
M_X(t) = \mathbb{E} \left( e^{tX} \right),
\]

provided that the expectation exists for \(t\) in some neighborhood of 0. That is, there is some \(h > 0\) such that for all \(t\) in \(-h < t < h\), \(E \left( e^{tX} \right)\) exists.

Theorem 5 (Theorem 4.2.10, CB) Let \(X\) and \(Y\) be independent random variables.

- (a) For any \(A \subset \mathbb{R}\) and \(B \subset \mathbb{R}\), \(P(X \in A, Y \in B) = P(X \in A)P(Y \in B)\); that is, the events \(\{X \in A\}\) and \(\{Y \in B\}\) are independent events.
• (b) Let \( g(x) \) be a function of only \( x \) and \( h(y) \) be a function of only \( y \). Then
\[
E(g(X)h(Y)) = Eg(X)Eh(Y)
\]

**Theorem 6** For any sum of independent random variables \( Y = X_1 + X_2 \),
\[
M_Y(t) = E(e^{tY}) = M_{X_1}(t)M_{X_2}(t).
\]

**Proof.** We apply Theorem 5 with \( g(X_1) = e^{tX_1} \) and \( h(X_2) = e^{tX_2} \) to obtain:
\[
M_Y(t) = E(e^{tY}) = E(e^{tX_1} \cdot e^{tX_2}) = E(e^{tX_1})E(e^{tX_2}).
\]

Thus the theorem holds by definition. ■

**Exercise.** Read the fact sheet, and work everything out!

In summary, we have

- \( n \)’th moment: \( E(X^n) \)
- Central moments: \( E((X - \mu)^n) \).
- Note: can use **Moment generating function (mgf)** to obtain the moments:
\[
M_X(t) = E(e^{tX}),
\]
provided that the expectation exists for \( t \) in some neighborhood of 0.

- Note: For any distribution with a mgf, differentiate wrt \( t \).
\[
M_X^{(n)}(t)|_{t=0} = E(X^n)
\]
where \( M_X^{(n)}(t)|_{t=0} = \frac{d^n}{dt^n}M_X(t)|_{t=0} \) is the \( n \)th derivative of \( M_X(t) \) evaluated at \( t = 0 \).

**Theorem 7** If \( X \) has Moment generating function \( M_X(t) \), then
\[
E(X^n) = M_X^{(n)}(0)
\]
where we define
\[
M_X^{(n)}(0) = \frac{d^n}{dt^n}M_X(t)|_{t=0}.
\]
That is, the \( n \)th moment is equal to the \( n \)th derivative of \( M_X(t) \) evaluated at \( t = 0 \).
Example 8 Recall if $X \sim \Gamma(\alpha, \beta)$, then
\[
f_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \quad \alpha, \beta > 0
\]
where
\[
\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy
\]
Its moment-generating function (mgf) is
\[
M_X(t) = \left[ \frac{1}{1 - \beta t} \right]^\alpha \quad \text{for} \quad t < 1/\beta.
\]

Exercises: compute the mean and the variance for $X$ using the definitions as well as the mgf method as in Theorem 7.

3 Sample Mean: $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

Theorem 9 Let $X_1, \ldots, X_n$ be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is
\[
M_{\overline{X}_n} = [M_X(t/n)]^n = (E[e^{tX/n}])^n.
\]

Example 10 If $X_1, \ldots, X_n \sim \Gamma(\alpha, \beta)$, then $\overline{X}_n \sim \Gamma(n\alpha, \beta/n)$.

Proof. Given $X_1, \ldots, X_n$ are independent, we have by Theorem 5
\[
M_{\overline{X}_n} = E[e^{t\overline{X}_n}] = E[e^{\sum_{i=1}^n X_i/t}] = \prod_{i} E[e^{X_i(t/n)}]
\]
\[
= [M_X(t/n)]^n = \left[ \left( \frac{1}{1 - \beta t/n} \right)^\alpha \right]^n = \left[ \frac{1}{1 - \beta t/n} \right]^n \alpha.
\]
This is the mgf of $\Gamma(n\alpha, \beta/n)$.

Note:

1. Gamma $E[X_i] = \alpha \beta$ Var$[X_i] = \alpha \beta^2$.

\[
E[\overline{X}_n] = \alpha \beta
\]
\[
\text{Var}[\overline{X}_n] = (n\alpha)(\beta/n)^2 = \alpha \beta^2 / n \to 0 \text{ as } n \to \infty.
\]

2. Normal $E[X_i] = \mu$ Var$[X_i] = \sigma^2$

\[
E[\overline{X}_n] = \mu
\]
\[
\text{Var}[\overline{X}_n] = \sigma^2 / n \to 0 \text{ as } n \to \infty.
\]
3. Generally: If $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$ exist, then $E[\bar{X}_n] = \mu$ and $\text{Var}[\bar{X}_n] = \sigma^2/n \rightarrow 0$.

**Lemma 11** If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ then sample mean $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

**Proof.** Recall for $i = 1, \ldots, n$,

$$M_{X_i}(s) = \exp\{\mu s + \sigma^2 s^2/2\}.$$

Hence by independence of $X_1, \ldots, X_n$, we have

$$M_{\bar{X}_n}(t) = \mathbb{E}(e^{t\bar{X}_n}) = \mathbb{E}(e^{t \frac{1}{n} \sum_{i=1}^{n} X_i})$$

$$= \prod_{i=1}^{n} \mathbb{E}e^{t X_i/n} = (M_X(t/n))^n = \left(e^{(\mu t/n) + \sigma^2 t^2/(2n^2)}\right)^n$$

$$= \exp\left\{\mu t + \frac{\sigma^2 t^2}{2n}\right\}$$

which is the mgf of a $N(\mu, \sigma^2/n)$.

■

**Sample mean and variances: review**

**Example 12** Suppose we test a prediction method, a neural net for example, on a set of $n$ new test cases. Let $X_i = 1$ if the predictor is wrong and $X_i = 0$ if the predictor is right. Then

$$\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$$

is the observed error rate. Each $X_i$ may be regarded as a Bernoulli with unknown mean $p$. We would like to know the true — but unknown — error rate $p$. Intuitively, we expect that $\bar{X}_n$ should be close to $p$. How likely is $\bar{X}_n$ to not be within $\epsilon$ of $p$? We have that

$$\text{Var}(\bar{X}_n) = \text{Var}(X_1)/n = p(1-p)/n$$

and

$$\mathbb{P}(|\bar{X}_n - p| > \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{p(1-p)}{n \epsilon^2} \leq \frac{1}{4n \epsilon^2}$$

since $p(1-p) \leq \frac{1}{4}$ for all $p$. For $\epsilon = .2$ and $n = 100$ the bound is .0625. ■
4 Sampling from the Normal Distribution: I

**Theorem 13** The random variable $\bar{X}_n$ and the vector of random variables $(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$ are independent.

**Exercise.**

1. Prove Theorem 13 by showing that the covariance $\text{Cov}(\bar{X}_n, X_1 - \bar{X}_n) = 0$.

2. Question: Are $X_1 - \bar{X}_n$ and $X_2 - \bar{X}_n$ independent?

Sample $X^n = X_1, \ldots, X_n \sim F \equiv N(\mu, \sigma^2)$ with pdf/pmf $f_X$ means iid (independent, identically distributed). The joint pdf or pmf of $X_1, \ldots, X_n$ is given by

$$f(x^n) = f_{X^n}(x_1, \ldots, x_n) = f_X(x_1)f_X(x_2)\cdots f_X(x_n) = \prod_{i=1}^n f_X(x_i).$$

Recall

- sample mean: $\bar{X}_n = \frac{1}{n} \sum_i X_i$, where $\mathbb{E}(\bar{X}_n) = \mu$

- sample variance: $S_n^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X}_n)^2$, where $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$

and $\bar{X}_n \sim N(\mu, \sigma^2/n)$ and $\mathbb{E}(S_n^2) = \sigma^2$. In your HW 2, you will compute $\text{Var}(S_n^2)$.

We note that the following corollary follows immediately from Theorem 13 as $S_n^2$ is a function of the random vector $(X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$, which is independent of $\bar{X}_n$. We give a direct proof which shows a nice application of the Jacobian theorem we have seen. Here we actually show that $S_n^2$ is a function of the random vector $(X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$.

**Corollary 14** The random variable $\bar{X}_n$ and $S_n^2$ are independently distributed.

**Proof.** Throughout this proof, we let $S^2 = S_n^2$ and $X = \bar{X}_n$.

$$S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$$

$$= \frac{1}{n-1} \{(X_1 - \bar{X})^2 + \sum_{i=2}^n (X_i - \bar{X})^2\}$$

But $\sum_i (X_i - \bar{X}) = \sum_i X_i - n\bar{X} = 0$

which implies $(X_1 - \bar{X}) = -\sum_{i=2}^n (X_i - \bar{X})$

therefore $S^2 = \frac{1}{n-1} \{[\sum_{i=2}^n (X_i - \bar{X})]^2 + \sum_{i=2}^n (X_i - \bar{X})^2\}$

$$= \ell(X_2 - \bar{X}, X_3 - \bar{X}, \ldots, X_n - \bar{X})$$
where \( \ell(X_2 - \bar{X}, X_3 - \bar{X}, \ldots, X_n - \bar{X}) \) denotes a function of all random variables involved. Define

\[
\begin{align*}
Y_1 &= \bar{X} \\
Y_2 &= X_2 - \bar{X} \\
\vdots & \quad \vdots \\
Y_n &= X_n - \bar{X}
\end{align*}
\]

In order to prove independence of \( Y_1 = \bar{X} \) and random vector \( (Y_2, \ldots, Y_n) \), we want to show

\[
f_{Y^n}(y^n) = g(y_1)h(y_2, \ldots, y_n)
\]

where \( y^n = (y_1, y_2, \ldots, y_n) \). WLOG, assume \( X_1, \ldots, X_n \) iid \( N(0, 1) \).

1. \[
f_{X^n}(x^n) = \prod_i \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x_i^2\} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\frac{1}{2}\sum_i x_i^2\}
\]

2. \( X_i = Y_i + Y_1 \) for all \( i = 2, \ldots, n \) and

\[
X_1 = 2Y_1 - \sum_{i=1}^{n} Y_i = Y_1 - \sum_{i=2}^{n} Y_i
\]

following the fact that

\[
\sum_{i=1}^{n} Y_i = \bar{X} + \sum_{i=2}^{n} X_i - (n - 1)\bar{X} = \bar{X} + n\bar{X} - X_1 - (n - 1)\bar{X} = 2\bar{X} - X_1.
\]

We require the jacobian of the transformation from \( X \) to \( Y \)

\[
J = \begin{vmatrix}
1 & -1 & -1 & -1 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1
\end{vmatrix}
\]

The determinant of a matrix is unchanged under linear transformation. Replace the first row by the sum of all rows to obtain:

\[
J = \begin{vmatrix}
n & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1
\end{vmatrix} = n
\]
Now the matrix is lower triangular and the determinant is the product of the diagonal terms.

\[
f_{Y^n}(y^n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp \left\{ -\frac{1}{2} \left[ \left( y_1 - \sum_{i=2}^{n} y_i \right)^2 + \sum_{i=2}^{n} (y_i + y_1)^2 \right] \right\} \cdot n
\]

\[
= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot n \cdot \exp \left\{ -\frac{ny_1^2}{2} \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \sum_{i=2}^{n} y_i^2 + \left( \sum_{i=2}^{n} y_i \right)^2 \right] \right\}
\]

We now apply Theorem 15 to conclude independence of \(\bar{X}\) and \(S_n^2\), which is a function of random vector \((Y_2, \ldots, Y_n)\). ■

**Theorem 15** (CB Lemma 4.2.7) Let \((X, Y)\) be a bivariate random vector with \(f_{XY}(x, y)\). \(X\) and \(Y\) are independent iif there exists functions \(g, h\) such that

\[
f_{XY}(x, y) = g(x)h(y).
\]

### 5 Sampling from the Normal Distribution: II

**Sample Variance, \(S_n^2\)**

**Definition 16** \(\chi_p^2 = \Gamma(p/2, 2)\) is the chi squared pdf with \(p\) degrees of freedom, denoted by \(\chi^2_p\), which is the distribution of

\[
V = \sum_{i=1}^{p} X_i^2
\]

where \(X_i \sim N(0, 1)\) independently.

Clearly for \(V \sim \chi_p^2\),

\[
f(v) = \frac{1}{\Gamma(p/2)2^{p/2}} v^{(p/2)-1} e^{-v/2}, \quad v > 0.
\]

- Mgf is \(M(t) = (1 - 2t)^{-p/2}\).
- If \(U\) and \(V\) are independent and \(U \sim \chi_m^2\) and \(V \sim \chi_n^2\), then \(U + V \sim \chi_{m+n}^2\).

**Theorem 17** If \(X_1, \ldots, X_n \sim N(\mu, \sigma^2)\) then

\[
\frac{(n-1)}{\sigma^2} S_n^2 \sim \chi_{n-1}^2,
\]

which is the chi-square distribution with \(n - 1\) degrees of freedom.
Proof. Define $U = \frac{(n-1)S^2}{\sigma^2}$ and $T = \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$, we have

$$(n - 1)S^2 = \sum_i (X_i - \bar{X})^2 = \sum_i (X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

$$\frac{(n - 1)S^2}{\sigma^2} = \sum \left( \frac{X_i - \mu}{\sigma} \right)^2 - \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$$

$$U = V - T$$

where $U$ and $T$ are independent by Corollary 14.

It follows that that the mgfs $M_{U+T}$ and $M_V$ are equal. Furthermore by Theorem 6, we have

$$M_{U+T} = M_UM_T = M_V$$

where both $V$ and $T$ follow $\chi^2_p$ distributions, with $p = n$ and 1 respectively. Thus

$$M_U = \frac{M_V}{M_T} = \frac{(1 - 2t)^{-n/2}}{(1 - 2t)^{-1/2}} = (1 - 2t)^{-(n-1)/2}$$

which happens to be the mgf of a $\chi^2_{n-1}$ distribution. ■

Definition 18 If $Z \sim N(0,1)$ and $U \sim \chi^2_n$ and $Z$ and $U$ are independent, then the distribution of $Z/\sqrt{U/n}$ is called the t distribution with $n$ degrees of freedom.

Corollary 19 If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ then

$$T_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1} \approx N(0,1).$$

Proposition 20 The density function of t distribution with $n$ degrees of freedom is

$$f(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left( 1 + \frac{t^2}{n} \right)^{-(n+1)/2}.$$

Proof. Homework 2. ■

Definition 21 If $U \sim \chi^2_m$ and $V \sim \chi^2_n$ are independent, then the distribution of $W = \frac{U/m}{V/n}$ is called the F distribution with $m$ and $n$ degrees of freedom and is denoted by $F_{m,n}$.

Proposition 22 The density function of F distribution with $m$ and $n$ degrees of freedom is

$$f(w) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left( \frac{m}{n} \right)^{m/2} w^{(m/2)-1} \left( 1 + \frac{m}{n} w \right)^{-(m+n)/2}, \quad w > 0.$$