Sampling Distribution of a Statistic

Engineering Statistics
Section 5.3

Josh Engwer

TTU

21 March 2016

PART I:

RANDOM SAMPLES

(A Priori) Samples vs. Samples-to-be-Collected

Recall from Chapter 1 the definition of a sample of a population:

Definition

A **sample** is a subset of a population.

Every sample encountered in Chapter 1 was an **a priori sample**.
Just saying "sample" by itself will always translate to "a priori sample."

TYPE OF SAMPLE	NOTATION	HAS SAMPLE BEEN ALREADY COLLECTED?
(a priori) Sample	$x: \quad x_1, x_2, \ldots, x_n$	**Yes**
Sample-to-be-Collected	$X_1, X_2, \ldots, X_n$	**No**

By contrast, a **sample-to-be-collected** has not been collected yet.
(as the name immediately suggests)

This means data points of a sample-to-be-collected have some uncertainty, and thus each data point is really a random variable!!

Random Samples

Statistical Inference methods to be encountered later in the course require that sample(s) to be collected must be of a very special kind:

Definition

(Random Sample)

A sample-to-be-collected $X_1, \ldots, X_n$ is called a **random sample** if:

1. the X_i's are all identical:
   - If the X_i's are all discrete, then the X_i's all have the exact same pmf $p_X(k)$.
   - If the X_i's are all continuous, then the X_i's all have the exact same pdf $f_X(x)$.
   - Regardless of random variable type, the X_i's have the exact same cdf $F_X(x)$.
2. the X_i's are all independent.

i.e. The rv's comprising the random sample are identical & independent.

Examples of Random Samples

Random Sample of size $n = 4$ from a discrete population with pmf $p_X(k)$:

$$X_1, X_2, X_3, X_4 \overset{iid}{\sim} \text{pmf } p_X(k)$$

Random Sample of size $n = 6$ from a continuous population with pdf $f_X(x)$:

$$X_1, X_2, X_3, X_4, X_5, X_6 \overset{iid}{\sim} \text{pdf } f_X(x)$$

Random Sample of size $n = 3$ from a population with cdf $F_X(x)$:

$$X_1, X_2, X_3 \overset{iid}{\sim} \text{cdf } F_X(x)$$

<u>NOTATION:</u> "iid" is shorthand for "**i**dentically and **i**ndependently **d**istributed"

Examples of Random Samples

Random Sample of size $n = 4$ from a Binomial$(5, 0.3)$ population:

$$X_1, X_2, X_3, X_4 \overset{iid}{\sim} \text{Binomial}(5, 0.3)$$

Random Sample of size $n = 2$ from a Normal(μ, σ^2) population:

$$X_1, X_2 \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$$

Random Sample of size $n = 3$ from an Exponential$(\lambda = 10)$ population:

$$X_1, X_2, X_3 \overset{iid}{\sim} \text{Exponential}(\lambda = 10)$$

NOTATION: "iid" is shorthand for "**i**dentically and **i**ndependently **d**istributed"

Careful Examination of a Random Sample

Random Sample of size $n = 4$ from a Binomial$(5, 0.3)$ population:

$$X_1, X_2, X_3, X_4 \stackrel{iid}{\sim} \text{Binomial}(5, 0.3)$$

What does this mean exactly??

- The X_i's are <u>identical</u>, meaning the X_i's have the exact same pmf:
  - $p_{X_1}(k) = p_{X_2}(k) = p_{X_3}(k) = p_{X_4}(k) = \binom{5}{k} 0.3^k \, 0.7^{5-k}$
  - $\text{Supp}(X_1) = \text{Supp}(X_2) = \text{Supp}(X_3) = \text{Supp}(X_4) = \{0, 1, 2, 3, 4, 5\}$
- The X_i's are <u>independent</u>:

$$\begin{aligned}
\mathbb{P}(X_1 = 3 \;\cap\; X_2 > 1) &= \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 > 1) \\
\mathbb{P}(X_2 \le 1 \;\cap\; X_4 \le 1) &= \mathbb{P}(X_2 \le 1) \cdot \mathbb{P}(X_4 \le 1) \\
\mathbb{P}(X_1 > 3 \;\cap\; X_2 \le 4 \;\cap\; X_3 = 0) &= \mathbb{P}(X_1 > 3) \cdot \mathbb{P}(X_2 \le 4) \cdot \mathbb{P}(X_3 = 0) \\
\mathbb{P}(X_1 > 2 \;\cap\; X_3 > 2 \;\cap\; X_4 > 2) &= \mathbb{P}(X_1 > 2) \cdot \mathbb{P}(X_3 > 2) \cdot \mathbb{P}(X_4 > 2) \\
&\vdots
\end{aligned}$$

Careful Examination of a Random Sample

Random Sample of size $n = 2$ from a Normal(μ, σ^2) population:

$$X_1, X_2 \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$$

What does this mean exactly??

- The X_i's are <u>identical</u>, meaning the X_i's have the exact same cdf:
  - $F_{X_1}(x) = F_{X_2}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
  - $\text{Supp}(X_1) = \text{Supp}(X_2) = (-\infty, \infty)$
- The X_i's are <u>independent</u>:

$$\begin{aligned} \mathbb{P}(X_1 = 3 \ \cap \ X_2 > 1) &= \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 > 1) \\ \mathbb{P}(X_1 \leq 1 \ \cap \ X_2 \leq 1) &= \mathbb{P}(X_1 \leq 1) \cdot \mathbb{P}(X_2 \leq 1) \\ \vdots \qquad & \qquad \vdots \end{aligned}$$

Careful Examination of a Random Sample

Random Sample of size $n = 3$ from an Exponential($\lambda = 10$) population:

$$X_1, X_2, X_3 \overset{iid}{\sim} \text{Exponential}(\lambda = 10)$$

What does this mean exactly??

- The X_i's are identical, meaning the X_i's have the exact same pdf:
  - $f_{X_1}(x) = f_{X_2}(x) = f_{X_3}(x) = 10e^{-10x}$
  - Supp(X_1) = Supp(X_2) = Supp(X_3) = $[0, \infty)$
- The X_i's are independent:

$$\begin{aligned}
\mathbb{P}(X_1 = 3 \,\cap\, X_2 > 1) &= \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 > 1) \\
\mathbb{P}(X_2 \leq 1 \,\cap\, X_3 \leq 1) &= \mathbb{P}(X_2 \leq 1) \cdot \mathbb{P}(X_3 \leq 1) \\
\mathbb{P}(X_1 > 3 \,\cap\, X_2 \leq 4 \,\cap\, X_3 = 0) &= \mathbb{P}(X_1 > 3) \cdot \mathbb{P}(X_2 \leq 4) \cdot \mathbb{P}(X_3 = 0) \\
\mathbb{P}(X_1 > 2 \,\cap\, X_2 > 2 \,\cap\, X_3 > 2) &= \mathbb{P}(X_1 > 2) \cdot \mathbb{P}(X_2 > 2) \cdot \mathbb{P}(X_3 > 2) \\
&\vdots
\end{aligned}$$

PART II:

SAMPLING DISTRIBUTION OF A STATISTIC
OF A FINITE DISCRETE POPULATION

Statistic of a Random Sample

Recall from Chapter 1 the definition of a **sample statistic**:

Definition

A **statistic** of a sample is a meaningful characteristic of the sample.

More precisely, a statistic is a function of the data points of the sample.

	(A PRIORI) SAMPLE $x: \quad x_1, x_2, \ldots, x_n$	**RANDOM SAMPLE** $X_1, X_2, \ldots, X_n$
Sample Mean	$\overline{x} := \frac{x_1+x_2+\cdots+x_n}{n}$	$\overline{X} := \frac{X_1+X_2+\cdots+X_n}{n}$
Sample Minimum	$x_{(1)} := \min\{x_1, x_2, \ldots, x_n\}$	$X_{(1)} := \min\{X_1, X_2, \ldots, X_n\}$
Sample Maximum	$x_{(n)} := \max\{x_1, x_2, \ldots, x_n\}$	$X_{(n)} := \max\{X_1, X_2, \ldots, X_n\}$
Sample Range	$x_R := x_{(n)} - x_{(1)}$	$X_R := X_{(n)} - X_{(1)}$
Sample Variance	$s^2 := \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \overline{x})^2$	$S^2 := \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2$
Sample Total	$\sum x_k := x_1 + x_2 + \cdots x_n$	$\sum X_k := X_1 + X_2 + \cdots X_n$
Sample Proportion	x/n	X/n

Statistic of a Random Sample (Most Common)

	(A PRIORI) SAMPLE $x: \; x_1, x_2, \ldots, x_n$	RANDOM SAMPLE $X_1, X_2, \ldots, X_n$
Sample Mean	$\bar{x}$	$\bar{X}$
Sample Median	$\widetilde{x}$	$\widetilde{X}$
10% Trimmed Mean	$x_{tr(10\%)}$	$X_{tr(10\%)}$
Sample Range	x_R	X_R
Sample Variance	s^2	S^2
Sample Std Dev	s	S
Interhinge Range	x_{IHR}	X_{IHR}
Order Statistics	$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$	$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
Sample Minimum	$x_{(1)}$	$X_{(1)}$
Sample Maximum	$x_{(n)}$	$X_{(n)}$
Sample Total	$\sum x_k$	$\sum X_k$
Sample Proportion	x/n	X/n

Sampling Distribution of a Statistic (Definition)

Given a random sample, then since...

1. ...each data point is a random variable, and...
2. ...every statistic is a function of the data points, ...

...that implies that every statistic of a random sample is a random variable!!

This means that a statistic of a random sample follows a distribution:

Definition

Let $X_1, \ldots, X_n$ be a random sample of some population.
Let T be a statistic of the random sample.

Then the **sampling distribution** of statistic T is

- the pmf $p_T(k)$ if the population is discrete.
- the pdf $f_T(x)$ if the population is continuous.

Moreover, the statistic T has its own support, Supp(T).
Finally, the sampling distribution of T can be visualized as

- a density histogram if the population is discrete.
- a density curve if the population is continuous.

<u>WEX 5-3-1:</u> Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

Construct the sampling distribution for the following statistics:

Sample Mean, Sample Variance, Sample Total, Sample Min, Sample Max

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

Construct the sampling distribution for the following statistics:

Sample Mean, Sample Variance, Sample Total, Sample Min, Sample Max

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \ \cap \ X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	**0**						
0	**1**						
1	**0**						
1	**1**						

Enumerate all meaningful simultaneous values of X_1 & X_2. (**Order Matters!!**)

Supp$(X_1) = \{0, 1\} \implies$ The meaningful values for X_1 are 0 and 1.
Supp$(X_2) = \{0, 1\} \implies$ The meaningful values for X_2 are 0 and 1.

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	**0.36**					
0	1	**0.24**					
1	0	**0.24**					
1	1	**0.16**					

Compute the joint probabilities using the fact that X_1, X_2 are independent:

$$\mathbb{P}(X_1 = 0 \cap X_2 = 0) \overset{iid}{=} p_X(0) \cdot p_X(0) = (0.6)(0.6) = \mathbf{0.36}$$
$$\mathbb{P}(X_1 = 0 \cap X_2 = 1) \overset{iid}{=} p_X(0) \cdot p_X(1) = (0.6)(0.4) = \mathbf{0.24}$$
$$\mathbb{P}(X_1 = 1 \cap X_2 = 0) \overset{iid}{=} p_X(1) \cdot p_X(0) = (0.4)(0.6) = \mathbf{0.24}$$
$$\mathbb{P}(X_1 = 1 \cap X_2 = 1) \overset{iid}{=} p_X(1) \cdot p_X(1) = (0.4)(0.4) = \mathbf{0.16}$$

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24					
1	0	0.24					
1	1	0.16					

Compute the statistics for each meaningful simultaneous values of X_1 & X_2:

$$\overline{X} = \frac{X_1+X_2}{n} = \frac{0+0}{2} = 0$$

$$S^2 = \frac{1}{n-1}\left[(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2\right] = \frac{1}{2-1}\left[(0-0)^2 + (0-0)^2\right] = 0$$

$$X_1 + X_2 = 0 + 0 = 0$$
$$X_{(1)} = \min\{X_1, X_2\} = \min\{0, 0\} = 0$$
$$X_{(n)} = \max\{X_1, X_2\} = \max\{0, 0\} = 0$$

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24					
1	1	0.16					

Compute the statistics for each meaningful simultaneous values of X_1 & X_2:

$$\overline{X} = \frac{X_1+X_2}{n} = \frac{0+1}{2} = 0.5$$

$$S^2 = \frac{1}{n-1}\left[(X_1-\overline{X})^2 + (X_2-\overline{X})^2\right] = \frac{1}{2-1}\left[(0-0.5)^2 + (1-0.5)^2\right] = 0.5$$

$$X_1 + X_2 = 0 + 1 = 1$$

$$X_{(1)} = \min\{X_1, X_2\} = \min\{0, 1\} = 0$$

$$X_{(n)} = \max\{X_1, X_2\} = \max\{0, 1\} = 1$$

WEX 5-3-1: Let $X_1, X_2 \stackrel{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16					

Compute the statistics for each meaningful simultaneous values of X_1 & X_2:

$$\overline{X} = \frac{X_1+X_2}{n} = \frac{1+0}{2} = 0.5$$

$$S^2 = \frac{1}{n-1}\left[(X_1-\overline{X})^2 + (X_2-\overline{X})^2\right] = \frac{1}{2-1}\left[(1-0.5)^2 + (0-0.5)^2\right] = 0.5$$

$$X_1 + X_2 = 1 + 0 = 1$$

$$X_{(1)} = \min\{X_1, X_2\} = \min\{1, 0\} = 0$$

$$X_{(n)} = \max\{X_1, X_2\} = \max\{1, 0\} = 1$$

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

Compute the statistics for each meaningful simultaneous values of X_1 & X_2:

$$\overline{X} = \frac{X_1+X_2}{n} = \frac{1+1}{2} = 1$$

$$S^2 = \frac{1}{n-1}\left[(X_1-\overline{X})^2 + (X_2-\overline{X})^2\right] = \frac{1}{2-1}\left[(1-1)^2 + (1-1)^2\right] = 0$$

$$X_1 + X_2 = 1 + 1 = 2$$

$$X_{(1)} = \min\{X_1, X_2\} = \min\{1, 1\} = 1$$

$$X_{(n)} = \max\{X_1, X_2\} = \max\{1, 1\} = 1$$

Sampling Distribution of a Statistic (Example)

<u>WEX 5-3-1:</u> Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf:

k			
$p_{\overline{X}}(k)$			

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	**0.36**	**0**	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf:

k	**0**		
$p_{\overline{X}}(k)$	**0.36**		

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	**0.24**	**0.5**	0.5	1	0	1
1	0	**0.24**	**0.5**	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf:

k	0	**0.5**	
$p_{\overline{X}}(k)$	0.36	**0.24** $+$ **0.24**	

Sampling Distribution of a Statistic (Example)

<u>WEX 5-3-1:</u> Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	**0.16**	**1**	0	2	1	1

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf:

k	0	0.5	**1**
$p_{\overline{X}}(k)$	0.36	0.48	**0.16**

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf:

k	0	0.5	1
$p_{\overline{X}}(k)$	0.36	0.48	0.16

$\implies$ Supp$(\overline{X}) = \{0, 0.5, 1\}$

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of S^2 is the pmf:

k		
$p_{S^2}(k)$		

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	**0.36**	0	**0**	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	**0.16**	1	**0**	2	1	1

$\therefore$ Sampling Dist. of S^2 is the pmf:

k	**0**	
$p_{S^2}(k)$	**0.36** + **0.16**	

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	**0.24**	0.5	**0.5**	1	0	1
1	0	**0.24**	0.5	**0.5**	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of S^2 is the pmf:

k	0	**0.5**
$p_{S^2}(k)$	0.52	**0.24** $+$ **0.24**

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of S^2 is the pmf:

k	0	0.5
$p_{S^2}(k)$	0.52	0.48

$\Longrightarrow$ Supp$(S^2) = \{0, 0.5\}$

Sampling Distribution of a Statistic (Example)

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf

k	0	0.5	1
$p_{\overline{X}}(k)$	0.36	0.48	0.16

$\therefore$ Sampling Dist. of S^2 is the pmf

k	0	0.5
$p_{S^2}(k)$	0.52	0.48

$\therefore$ Sampling Dist. of $X_1 + X_2$ is the pmf

k	0	1	2
$p_{X_1+X_2}(k)$	0.36	0.48	0.16

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$X_1 = j_1$	$X_2 = j_2$	$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2)$	$\overline{X}$	S^2	$X_1 + X_2$	$X_{(1)}$	$X_{(n)}$
0	0	0.36	0	0	0	0	0
0	1	0.24	0.5	0.5	1	0	1
1	0	0.24	0.5	0.5	1	0	1
1	1	0.16	1	0	2	1	1

$\therefore$ Sampling Dist. of $X_{(1)}$ is the pmf

k	0	1
$p_{X_{(1)}}(k)$	0.84	0.16

$\therefore$ Sampling Dist. of $X_{(n)}$ is the pmf

k	0	1
$p_{X_{(n)}}(k)$	0.36	0.64

Expected Value & Variance of a Statistic

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$$\implies \mu_X = \mathbb{E}[X] = \sum_{k \in \text{Supp}(X)} k \cdot p_X(k) = (0)(0.6) + (1)(0.4) = 0.4$$

$$\mathbb{E}[X^2] = \sum_{k \in \text{Supp}(X)} k^2 \cdot p_X(k) = (0^2)(0.6) + (1^2)(0.4) = 0.4$$

$$\implies \sigma_X^2 = \mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 0.4 - (0.4)^2 = 0.24$$

Expected Value & Variance of a Statistic

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$$\mu_X = \mathbb{E}[X] = 0.4 \qquad \sigma_X^2 = \mathbb{V}[X] = 0.24$$

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf

k	0	0.5	1
$p_{\overline{X}}(k)$	0.36	0.48	0.16

$$\implies \mu_{\overline{X}} = \mathbb{E}\left[\overline{X}\right] = \sum_{k \in \text{Supp}(\overline{X})} k \cdot p_{\overline{X}}(k) = (0)(0.36) + (0.5)(0.48) + (1)(0.16) = 0.4$$

$$\mathbb{E}\left[\overline{X}^2\right] = \sum_{k \in \text{Supp}(\overline{X})} k^2 \cdot p_{\overline{X}}(k) = (0^2)(0.36) + (0.5^2)(0.48) + (1^2)(0.16) = 0.28$$

$$\implies \sigma_{\overline{X}}^2 = \mathbb{V}\left[\overline{X}\right] = \mathbb{E}\left[\overline{X}^2\right] - \left(\mathbb{E}\left[\overline{X}\right]\right)^2 = 0.28 - (0.4)^2 = 0.12$$

Expected Value & Variance of a Statistic

WEX 5-3-1: Let $X_1, X_2 \overset{iid}{\sim}$ Bernoulli(0.4).

Then, X_1 & X_2 follow the pmf:

k	0	1
$p_X(k)$	0.6	0.4

$$\mu_X = \mathbb{E}[X] = 0.4 \qquad \sigma_X^2 = \mathbb{V}[X] = 0.24$$

$\therefore$ Sampling Dist. of $\overline{X}$ is the pmf

k	0	0.5	1
$p_{\overline{X}}(k)$	0.36	0.48	0.16

$$\mu_{\overline{X}} = \mathbb{E}\left[\overline{X}\right] = 0.4 \qquad \sigma_{\overline{X}}^2 = \mathbb{V}\left[\overline{X}\right] = 0.12$$

Proposition

(Construction of the Sampling Distribution of a Statistic)

GIVEN: Random sample $X_1, \dots, X_n$ of finite discrete population w/ pmf $p_X(k)$.
TASK: Find the sampling distribution $p_T(k)$ of statistic T of random sample.

(1) Enumerate all meaningful simultaneous values of the X_i's.
*Use the support of X_1, Supp(X_1), as guidance. (**Order Matters!!**)*

(2) For each enumeration of meaningful simultaneous values of the X_i's, compute the statistic T & the joint probability using iid & pmf $p_X(k)$:

$$\mathbb{P}(X_1 = j_1 \cap X_2 = j_2 \cap \cdots \cap X_n = j_n) \overset{iid}{=} p_X(j_1) \cdot p_X(j_2) \cdots p_X(j_n)$$

(3) The support of statistic T, Supp(T), is the set of all values of T attained.

(4) The probability of statistic T being a value in its support is the sum of the joint probabilities corresponding to that value of T.

Textbook Logistics for Section 5.3

- Difference(s) in Notation:

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Probability of Event	$P(A)$	$\mathbb{P}(A)$
Support of a r.v.	"All possible values of X"	Supp(X)
pmf of a r.v.	$p_X(x)$	$p_X(k)$
Expected Value of r.v.	$E(X)$	$\mathbb{E}[X]$
Variance of r.v.	$V(X)$	$\mathbb{V}[X]$
Sample Total	T_o	$\sum X_k$
pmf of Sample Mean	$p_{\overline{X}}(\overline{x})$	$p_{\overline{X}}(k)$
pmf of Sample Variance	$p_{S^2}(s^2)$	$p_{S^2}(k)$

Textbook Logistics for Section 5.3

- Ignore EXAMPLE 5.22 (pg 225)
    - The statistic of sample of a continuous population involves multiple integrals!
    - Multivariable Calculus (CalcIII) will never show up on homework and exams.

- Skip "Simulation Experiments" section (pgs 225-229)
    - Simulations were briefly encountered in Ch2 when developing the "deep" interpretation of Probability.
    - Simulations will be briefly encountered again in section 5.4

Fin

Fin.