RANDOM SAMPLES [DEVORE 5.3]

- **(A PRIORI) SAMPLE VS. SAMPLE-TO-BE-COLLECTED:**

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

  the X_i's are all identical and independent.

  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 $F_X(x)$.

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

- **PARTICULAR EXAMPLES OF RANDOM SAMPLES:**

  Random Sample of size $n = 4$ from a Binomial$(5, 0.3)$ population: $X_1, X_2, X_3, X_4 \overset{iid}{\sim} \text{Binomial}(5, 0.3)$

  The X_i's are identical, meaning they have the 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 independent:

  $$\mathbb{P}(X_1 = 3 \ \cap \ X_2 > 1) = \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 > 1)$$
  $$\mathbb{P}(X_1 > 3 \ \cap \ X_2 \le 4 \ \cap \ X_3 = 0) = \mathbb{P}(X_1 > 3) \cdot \mathbb{P}(X_2 \le 4) \cdot \mathbb{P}(X_3 = 0)$$

  Random Sample of size $n = 2$ from a Normal(μ, σ^2) population: $X_1, X_2 \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$

  The X_i's are identical, meaning they have the 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 independent:

  $$\mathbb{P}(X_1 = 3 \ \cap \ X_2 > 1) = \mathbb{P}(X_1 = 3) \cdot \mathbb{P}(X_2 > 1)$$
  $$\mathbb{P}(X_1 \le 1 \ \cap \ X_2 \le 1) = \mathbb{P}(X_1 \le 1) \cdot \mathbb{P}(X_2 \le 1)$$

  Random Sample of size $n = 3$ from an Exponential$(\lambda = 10)$ population: $X_1, X_2, X_3 \overset{iid}{\sim} \text{Exponential}(\lambda = 10)$

  The X_i's are identical, meaning they have the same pdf:

  $$f_{X_1}(x) = f_{X_2}(x) = f_{X_3}(x) = 10e^{-10x}$$
  $$\text{Supp}(X_1) = \text{Supp}(X_2) = \text{Supp}(X_3) = [0, \infty)$$

  The X_i's are independent:

  $$\mathbb{P}(X_2 > 1 \ \cap \ X_3 \le 1) = \mathbb{P}(X_2 > 1) \cdot \mathbb{P}(X_3 \le 1)$$
  $$\mathbb{P}(X_1 > 2 \ \cap \ X_2 > 2 \ \cap \ X_3 > 2) = \mathbb{P}(X_1 > 2) \cdot \mathbb{P}(X_2 > 2) \cdot \mathbb{P}(X_3 > 2)$$

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

SAMPLING DISTRIBUTION OF A STATISTIC [DEVORE 5.3]

- **STATISTIC OF A RANDOM SAMPLE:** Recall from Chapter 1 the definition of a **sample statistic**:

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 $x: \ x_1, x_2, \ldots, x_n$	**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

- **SAMPLING DISTRIBUTION OF A STATISTIC (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.

- **SAMPLING DISTRIBUTION OF A STATISTIC (PROCEDURE):**

GIVEN: Random sample $X_1, \ldots, X_n$ 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.

EX 5.3.1: Let size ($n = 2$) random sample $X_1, X_2 \overset{iid}{\sim}$ pmf $p_X(k)$ such that:

k	0	1	2
$p_X(k)$	0.65	0.25	0.10

(a) What is the population mean μ and population variance σ^2?

(b) Construct the sampling distributions for the sample mean $\overline{X}$ and the sample variance S^2.

(c) What is the expected value & variance of the sample mean, $\mu_{\overline{X}}$ & $\sigma^2_{\overline{X}}$?

(d) What is the expected value & variance of the sample variance, μ_{S^2} & $\sigma^2_{S^2}$?

EX 5.3.2: Let size $(n = 3)$ random sample $Y_1, Y_2, Y_3 \overset{iid}{\sim}$ pmf $p_Y(k)$ such that:

k	2	3
$p_Y(k)$	0.2	0.8

(a) What is the population mean μ and population variance σ^2?

(b) Construct the sampling distributions for the sample mean $\overline{Y}$ and the sample total $Y_1 + Y_2 + Y_3$.

(c) What is the expected value & variance of the sample mean, $\mu_{\overline{Y}}$ & $\sigma^2_{\overline{Y}}$?

(d) What is the expected value & variance of the sample total, $\mu_{Y_1+Y_2+Y_3}$ & $\sigma^2_{Y_1+Y_2+Y_3}$?

EX 5.3.3: Let size ($n = 2$) random sample $W_1, W_2 \overset{iid}{\sim}$ Binomial$(2, 0.35)$.

(a) What is the population mean μ and population variance σ^2?

(b) Construct the sampling distributions for the sample mean $\overline{W}$ and the sample minimum $W_{(1)}$.

(c) What is the expected value & variance of the sample mean, $\mu_{\overline{W}}$ & $\sigma^2_{\overline{W}}$?

(d) What is the expected value & variance of the sample total, $\mu_{W_{(1)}}$ & $\sigma^2_{W_{(1)}}$?