i.i.d. sampling

$\overline{X} = \frac{1}{n} S$

Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with mean μ and variance σ^2. We call μ, σ^2 the **population mean/-variance**. We know

$\mathrm{Var}(\overline{X}) = \frac{1}{n^2} \mathrm{Var}(S)$

- The sum $S = X_1 + \cdots + X_n$ has mean $\mathrm{E}(S) = n\mu$ and variance $n\sigma^2$.
- The mean $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ has mean $\mathrm{E}(S) = \mu$ and variance σ^2/n.
- If $\mathbf{X} = \langle X_1, \ldots, X_n \rangle \overset{\text{i.i.d.}}{\sim} \mathsf{Norm}(\mu, \sigma)$, then S and $\overline{X}$ as defined above both have normal distributions.

$S \sim \mathrm{Norm}(n\mu, \sqrt{n}\,\sigma)$

Std. Error

In a finite population, an i.i.d. sample of size n is like drawing n times from a bag with replacement and (reshuffling/reshaking between draws).

$\overline{X} \sim \mathrm{Norm}(\mu, \sigma/\sqrt{n})$

In RStudio with the Mosaic package: `resample()` or `sample(..., replace=TRUE)`

$4/\sqrt{10}$

$SE_{\overline{X}_{10}}$

Simple random sampling

A sample is an SRS if every member of the population has equal chance to be in the sample.

Note: In a finite population, an SRS of size n is like drawing n times from a well-mixed bag without replacement.

In RStudio: `sample(..., replace=FALSE)` (`replace=FALSE` is the default)

Results: Let $\mathbf{X} = \langle X_1, \ldots, X_n \rangle$ be an SRS from a population of size N with mean and variance μ and σ^2, respectively.

- (Lemma 4.3.2, p. 240) Then $\mathrm{Cov}(X_i, X_j) = \begin{cases} \sigma^2, & i = j, \\ -\sigma^2/(N-1), & i \neq j. \end{cases}$
- (Corollary 4.3.3, p. 241) We have $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$.

~1 often

N is size of population
n = sample size.

$X_1, \ldots, X_n$ represents an SRS (sampling w/out replacement)

$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \longrightarrow \mathrm{E}(\overline{X}) = \mu$

Unbiased and consistent estimators

$\overline{X}$ is an unbiased estimator of μ, and for large n, $\overline{X}$ has a high prob. of being close to μ.

Definition 1: Call an estimator $\hat{\theta}$ for a parameter θ an **unbiased estimator** if $\mathrm{E}(\hat{\theta}) = \theta$.

A sequence $(\hat{\theta}_n)$ of estimators for a parameter θ is **consistent** if for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|\hat{\theta}_n - \theta| < \epsilon) = 1.$$

Alternatively, one says the estimators $\hat{\theta}_n$ **converge in probability** to θ.

Note that, in the case of both SRS and i.i.d. samples, $\overline{X}$ is an unbiased estimator of μ, the population mean. Moreover, we have:

Theorem 1 (Weak Law of Large Numbers, Coro. 4.3.10, p. 243): The sample mean $\overline{X}_n$, taken as a an estimator for the population mean μ, is consistent in both the cases of i.i.d. and simple random samples.

This result was one of the earliest major discoveries proved in the field of probability, by one of its early pioneers, Jacob Bernoulli. The modern proof, simpler than Bernoulli's, is greatly simplified by Chebyshev's inequality:

Theorem 2 (Chebyshev's Inequality): Let X be a r.v. with finite mean μ and variance σ^2. For each $\epsilon > 0$,

$$\Pr(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

When applied to the r.v. $\overline{X}_n$ whose mean and variance are

- μ and σ^2/n, respectively, for an i.i.d. random sample, and
- μ and $(\sigma^2/n)(N-n)/(N-1)$, respectively, for an SRS,

we see that $\Pr(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$, proving the Weak Law of Large Numbers.

In summary: Whether the random variables $X_1, \ldots X_n$ represent an i.i.d. random sample or an SRS from a population with mean μ and variance σ^2, the sample mean $\overline{X}$

1. is an unbiased estimator of μ—that is, $\mathrm{E}(\overline{X}) = \mu$.

2. has a high probability of being close to μ for large n.
3. has a sampling distribution whose standard deviation is approximately $\sigma/\sqrt{n}$. To highlight that we are talking about the standard deviation of values of $\overline{X}$, as opposed to the population standard deviation, σ, we use the name **standard error**, or **standard error of the mean**, for this quantity $\sigma/\sqrt{n}$. A symbol used is $\mathrm{SE}_{\overline{X}}$.

Aside. There is, as you would expect, a *strong* law of large numbers. When applied to $\overline{X}_n$ from i.i.d. random samples, it says:

Theorem 3 (Strong Law of Large Numbers): Let $X_1, X_2, \ldots$ be an i.i.d. sequence of r.v.s with finite mean μ. Setting $S_n = X_1 + \cdots + X_n$ and $\overline{X}_n = S_n/n$ for each n, we have

$$\Pr(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$

The differences between the two versions of LLN, weak and strong, are subtle, and not elicited here in these notes. I only mention the strong law because it seems natural that there would be a counterpart to a weak law. The convergence of $\overline{X}_n$ to μ found in the strong law is of a different sort, called **almost sure convergence**.

Example 1:

(a) Suppose the random vector $\mathbf{X} \overset{\text{i.i.d.}}{\sim}= \langle X_1, \ldots, X_{50}\rangle \sim \mathsf{Norm}(15, 2)$. Illustrate

- the exact population distribution.
  What are μ and σ?
- a simulated sample distribution.
- a simulated sampling distribution for $\overline{X}_{50}$.
  Explain why the simulated sampling distribution looks the way it does. What are $\mathrm{E}(\overline{X}_{50})$ and $\mathrm{Var}(\overline{X}_{50})$?
- the exact sampling distribution for $\overline{X}_{50}$..

(b) Given $\sigma = 2$, if we wished to test the hypothesis

$$\mathbf{H}_0 : \ : \mu = 15 \qquad \text{vs. the alternative} \qquad \mathbf{H}_a : \ : \mu \neq 15,$$

how strong is the evidence (i.e., what is the P-value) associated with the sample $\bar{x} = 11.7$.

■

Example 2:

Repeat part (a) from the above example with $n = 200$, instead of $n = 50$.

■

Example 3:

Repeat again, but this time using i.i.d. random samples (various sizes n) from the 'PTSG' column of data file miaa05.csv. Compare the population distribution with the sampling distribution of the sample mean. Note that the population mean and standard deviation for this 'PTSG' variable are $\mu = 5.593$ and $\sigma = 4.919$ respectively.

■

Example 4:

Recall that a binomial r.v. $X \sim \mathsf{Binom}(n, \pi)$ is the sum $X = X_1 + X_2 + \cdots + X_n$ of n independent Bernoulli r.v.s, each $X_i \sim \mathsf{Binom}(1, \pi)$. That makes $X = n\overline{X} = n\hat{\pi}$, a rescaling of the mean/sample proportion. Whereas $\hat{\pi}$ has mean π and variance $\pi(1 - \pi)/n$, X has mean and variance $n\pi$ and $n\pi(1 - \pi)$. Look at the distributions of both $\hat{\pi}$ and X, and compare them with normal distributions, in the case where

(a) $n = 40$, $\pi = .3$

(b) $n = 40$, $\pi = .1$

(c) $n = 200$, $\pi = .1$

■

$X \sim \text{Binom}(n, \pi)$ i.i.d.

$X = X_1 + X_2 + \cdots + X_n$ = sum

each $X_i \sim \text{Binom}(1, \pi)$ Bernoulli trials

quite nonnormal

$X = n \bar{X}$

Each X_i has expected value π

variance $\pi(1-\pi)$

$\Longrightarrow$ X has expected value $n\pi$

variance $n\pi(1-\pi)$

$X_i \sim \text{Binom}(1, \pi)$

For $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

$= \hat{\pi}$ proportion of successes in n trials (sampled)