Lecture 3

Properties of Summary Statistics: Sampling Distribution
Main Theme

How can we use math to justify that our numerical summaries from the sample are good summaries of the population?
Today, we focus on two summary statistics of the sample and study its theoretical properties:

- Sample mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Sample variance: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

They are aimed to get an idea about the population mean and the population variance (i.e., parameters).

First, we’ll study, on average, how well our statistics do in estimating the parameters.

Second, we’ll study the distribution of the summary statistics, known as sampling distributions.
Setup

• Let $X_1, \ldots, X_n \in \mathbb{R}^p$ be i.i.d. samples from the population $F_\theta$

• $F_\theta$: distribution of the population (e.g. normal) with features/parameters $\theta$
  – Often the distribution AND the parameters are unknown.
  – That’s why we sample from it to get an idea of them!

• i.i.d.: independent and identically distributed
  – Every time we sample, we redraw $n$ members from the population and obtain $(X_1, \ldots, X_n)$. This provides a representative sample of the population.

• $\mathbb{R}^p$: dimension of $X_i$
  – For simplicity, we’ll consider univariate cases (i.e. $p = 1$)
Loss Function

• How “good” are our numerical summaries (i.e. statistics) in capturing features of the population (i.e. parameters)?

• **Loss Function**: Measures how good the statistic is in estimating the parameter
  – 0 loss: the statistic is the perfect estimator for the parameter
  – ∞ loss: the statistic is a terrible estimator for the parameter

• Example: \( l(T, \theta) = (T - \theta)^2 \) where \( T \) is the statistic and \( \Lambda \) is the parameter. Called square-error loss
It is **impossible** to compute the values for the loss function.

Why? We don’t know what the parameter is! (since it’s an **unknown feature** of the population and we’re trying to study it!)

More importantly, the statistic is **random**! It changes every time we take a different sample from the population. Thus, the **value** of our loss function **changes per sample**
A Remedy

• A more manageable question: On average, how good is our statistic in estimating the parameter?

• Risk (i.e. expected loss): The average loss incurred after repeated sampling

\[ R(\theta) = E[l(T, \theta)] \]

• Risk is a function of the parameter

• For the square-error loss, we have

\[ R(\theta) = E[(T - \theta)^2] \]
Bias-Variance Trade-Off: Square Error Risk

- After some algebra, we obtain another expression for square error Risk
  
  \[ R(\theta) = E[(T - \theta)^2] = (E[T] - \theta)^2 + E[(T - E[T])^2] \]

  \[ = Bias_\theta(T)^2 + Var(T) \]

- **Bias**: On average, how far is the statistic away from the parameter (i.e. accuracy)
  
  \[ Bias_\theta(T) = E[T] - \theta \]

- **Variance**: How variable is the statistic (i.e. precision)

- In estimation, there is always a bias-variance tradeoff!
Sample Mean, Bias, Variance, and Risk

- Let $T(X_1, ..., X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$. We want to see how well the sample mean estimates the population mean. We’ll use square error loss.

- $Bias_\mu(T) = 0$, i.e. the sample mean is unbiased for the population mean.
  - Interpretation: On average, the sample mean will be close to the population mean, $\mu$.

- $Var(T) = \frac{\sigma^2}{n}$
  - Interpretation: The sample mean is precise up to an order $\frac{1}{\sqrt{n}}$. That is, we decrease the variability of our estimate for the population mean by a factor of $\frac{1}{\sqrt{n}}$.

- Thus, $R(\mu) = \frac{\sigma^2}{n}$
  - Interpretation: On average, the sample mean will be close to the population mean by $\frac{\sigma^2}{n}$. This holds for all population mean $\mu$. 

Sample Variance and Bias

• Let $T(X_1, \ldots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. We want to see how well the sample variance estimates the population variance. We’ll use square error loss.

• $Bias_{\sigma^2}(T) = 0$, i.e. the sample variance is unbiased for the population mean
  – Interpretation: On average, the sample variance will be close to the population variance, $\sigma^2$

• $Var(T) = ?$, $R(\sigma^2) = ?$
  – Depends on assumption about fourth moments.
In Summary...

• We studied how good, on average, our statistic is in estimating the parameter

• Sample mean, $\bar{X}$, and sample variance, $\hat{\sigma}^2$, are both unbiased for the population mean, $\mu$, and the population variance, $\sigma^2$
But…

• But, what about the distribution of our summary statistics?

• So far, we only studied the statistics “average” behavior.

• Sampling distribution!
Sampling Distribution when $F$ is Normal

Case 1 (Sample Mean): Suppose $F$ is a normal distribution with mean $\mu$ and variance $\sigma^2$ (denoted as $N(\mu, \sigma^2)$). Then $\bar{X}$ is distributed as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$$

Proof: Use the fact that $X_i \sim N(\mu, \sigma^2)$. 
Sample mean for n=30 after 1000 experiments

(Theory-Based) Sampling Distribution
Mean (\mu)
Effect on the sampling distribution as $n$ changes after 5000 experiments
Sampling Distribution Example

• Suppose you want to know the probability that your sample mean, $\bar{X}$, is $\epsilon > 0$ away from the population mean.

• We assume $\sigma$ is known and $X_i \sim N(\mu, \sigma^2)$, i.i.d.

$$ P(|\bar{X} - \mu| \leq \epsilon) = P \left( \frac{|\bar{X} - \mu|}{\frac{\epsilon}{\sigma}} \leq \frac{\epsilon}{\sigma} \frac{\sqrt{n}}{\sqrt{n}} \right) = P \left( |Z| \leq \frac{\epsilon \sqrt{n}}{\sigma} \right) $$

where $Z \sim N(0,1)$.
Prob. that sample mean is within 1 of the pop. mean

Probability

Sample size

Sigma=1
Sigma=2
Sigma=5
Sampling Distribution when $F$ is Normal

**Case 1 (Sample Variance):** Suppose $F$ is a normal distribution with mean $\mu$ and variance $\sigma^2$ (denoted as $N(\mu, \sigma^2)$). Then $(n - 1)\frac{\hat{\sigma}^2}{\sigma^2}$ is distributed as

$$(n - 1)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2_{n-1}$$

where $\chi^2_{n-1}$ is the Chi-square distribution with $n - 1$ degrees of freedom.
Sampling Distribution of \((n-1)S^2/\sigma^2\)

(Theory-Based) Sampling Distribution
Some Preliminaries from Stat 430

• **Fact 0**: \((\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})\) is jointly normal
  Proof: Because \(ar{X}\) and \(X_i - \bar{X}\) are linear combinations of normal random variables, they must be jointly normal.

• **Fact 1**: For any \(i = 1, \ldots, n\), we have
  \[\text{Cov}(\bar{X}, X_i - \bar{X}) = 0\]
  Proof:
  \[
  E(\bar{X}(X_i - \bar{X})) = \frac{n - 1}{n} \mu^2 + \frac{1}{n}(\mu^2 + \sigma^2) - (\mu^2 + \frac{\sigma^2}{n}) = 0
  
  E(\bar{X})E(X_i - \bar{X}) = \mu(\mu - \mu) = 0.
  
  Thus, \text{Cov}(\bar{X}, X_i - \bar{X}) = E(\bar{X}(X_i - \bar{X})) - E(\bar{X})E(X_i - \bar{X}) = 0\]
• Since $\bar{X}$ and $X_i - \bar{X}$ are jointly normal, the zero covariance between them implies that $\bar{X}$ and $X_i - \bar{X}$ are independent.

• Furthermore, because
  \[
  \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2
  \]
  is a function of $X_i - \bar{X}$, $\hat{\sigma}^2$ is independent of $\bar{X}$. 

• Fact 2: If $W = U + V$ and $W \sim \chi_{a+b}^2$, $V \sim \chi_b^2$, and $U$ and $V$ are independent, then $U \sim \chi_a^2$

Proof: Use moment generating functions

• Now, we can prove this fact. (see blackboard)

$$W = \sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} \left( \frac{X_i - \bar{X} + \bar{X} - \mu}{\sigma} \right)^2 = U + V \sim \chi_n^2$$

$$U = \frac{(n-1)S^2}{\sigma^2}$$

$$V = \left( \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \right)^2 \sim \chi_1^2$$

Thus, $U \sim \chi_{n-1}^2$
Sampling Distribution when $F$ is not Normal

Case 2: Suppose $F$ is an arbitrary distribution with mean $\mu$ and variance $\sigma^2$ (denoted as $F(\mu, \sigma^2)$). Then as $n \to \infty$,

$$\lim_{n \to \infty} \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \to N(0,1)$$

Proof: Use the fact that $X_i \sim F(\mu, \sigma^2)$ and use the central limit theorem
Properties

• As sample size increases, $\bar{X}$ is unbiased
  – Asymptotically unbiased

• As sample size increases, $\bar{X}$ approaches the normal distribution at a rate $\frac{1}{\sqrt{n}}$
  – Even if we don’t have infinite sample size, using $N\left(\mu, \frac{\sigma^2}{n}\right)$ as an approximation to $\bar{X}$ is meaningful for large samples
  – How large? General rule of thumb: $n \geq 30$
Example

• Suppose $X_i \sim Exp(\lambda)$ in i.i.d.
  
  – Remember, $E(X_i) = \lambda$
  and $Var(X_i) = \frac{1}{\lambda^2}$

• Then, for large enough $n$, $\bar{X} \approx N\left(\frac{1}{\lambda}, \frac{1}{\lambda^2 n}\right)$
Effect on the sampling distribution as $n$ changes after 5000 experiments from $\text{Exp}(4)$.
An Experiment
Lecture Summary

• **Risk**: Average loss/mistakes that the statistics make in estimating the population parameters
  – Bias-variance tradeoff for squared error loss.
  – $\bar{X}$ and $\hat{\sigma}^2$ are **unbiased** for the population mean and the population variance

• **Sampling distribution**: the distribution of the statistics used for estimating the population parameters.
  – If the population is normally distributed:
    • $\bar{X} \sim N \left( \mu, \frac{\sigma^2}{n} \right)$ and $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$
  – If the population is not normally distributed
    • $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \rightarrow N(0,1)$ or $\bar{X} \approx N \left( \mu, \frac{\sigma^2}{n} \right)$