Problem of Estimation

Problem: Given a sample, find (or guess) the distribution from which it was drawn.

Parametric approach: Simplify the problem by assuming the form of the distribution (e.g. Bernoulli, normal, etc.). Thus problem is narrowed to finding the true values of a few unknown parameters (e.g. \( p \), or \( \mu \) and \( \sigma^2 \), etc.).

Definition: An estimator is a formula, based on the data in a sample, that gives a value (an estimate) for the unknown parameter(s).

Remark: Since data are random (vary from sample to sample) an estimator is itself a random variable.

Definition: An estimate is a particular value taken by the estimator for a particular set of data.
Examples of estimators

Example 1: Suppose we have a random sample $X_1, \ldots, X_n$ which we believe is drawn from $N(\mu, \sigma^2)$ where $\mu$ and $\sigma^2$ are unknown. Some estimators of $\mu$ might be:

\[ \hat{\mu}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]

\[ \hat{\mu}_2 = \text{sample median}\{X_i\} \]

\[ \hat{\mu}_3 = \frac{1}{n+1} \sum_{i=1}^{n} X_i \]

\[ \hat{\mu}_4 = 47 \]
Examples of estimators

Example 2: Again suppose we have a random sample \(X_1, \ldots, X_n\) which we believe is drawn from \(N(\mu, \sigma^2)\) where \(\mu\) and \(\sigma^2\) are unknown. Some estimators of \(\sigma^2\) might be:

\[
\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^{n} \left( X_i - \overline{X} \right)^2
\]

\[
\hat{\sigma}_2^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left( X_i - \overline{X} \right)^2
\]
Properties of estimators

Suppose $\theta$ is an estimator for a certain distribution whose true but unknown probability density function is given by $f(x; \theta)$.

Recall: $\theta$ is itself a random variable, so in general it will have a mean, a variance, etc.

(1) **Bias:** $\text{Bias}(\theta) = \mathbb{E}(\theta) - \theta$.

If $\mathbb{E}(\theta) = \theta$, then $\theta$ is called *unbiased*. An unbiased estimator has a distribution centered around the true value of the parameter.

**Remark:** A biased estimator may still be useful if the bias is small.
Properties of estimators

Example 1 (cont'd): Easy to show that $\hat{\mu}_1$ is unbiased:

$$E(\hat{\mu}_1) = E \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} E \left( \sum_{i=1}^{n} X_i \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

It can be shown that $\hat{\mu}_2$ is also unbiased. However, $\hat{\mu}_3$ is biased since:

$$Bias(\hat{\mu}_3) = E(\hat{\mu}_3) - \mu = \frac{\mu n}{n+1} - \mu = \frac{-\mu}{n+1}$$

Also, $\hat{\mu}_4$ is biased since:

$$Bias(\hat{\mu}_4) = E(\hat{\mu}_4) - \mu = 47 - \mu$$

Example 2 (cont'd): Not very hard to show that $E(\hat{\sigma}_1^2) = \sigma^2 (n-1)/n \neq \sigma^2$, so $\hat{\sigma}_1^2$ is biased. However, $E(\hat{\sigma}_2^2) = \sigma^2$, so $\hat{\sigma}_2^2$ is unbiased.
Properties of estimators

(2) **Minimum variance**: An estimator $\theta$ is minimum-variance (MV) among a given set of estimators if it has the smallest variance for all possible values of $\theta$.

**Remark**: It is easy to find estimators with very low variance, but often they are not good estimators.
Properties of estimators

Example 1 (cont'd): Using the properties of the variance, one can easily show:

\[ Var(\hat{\mu}_1) = \left( \frac{1}{n} \right)^2 \sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n} \]

\[ Var(\hat{\mu}_3) = \left( \frac{1}{n+1} \right)^2 \sum_{i=1}^{n} Var(X_i) = \frac{n \sigma^2}{(n+1)^2} \]

\[ Var(\hat{\mu}_4) = 0 \]

It can be shown that \( Var(\hat{\mu}_2) \), though quite complicated, is approximately equal to \((\pi/2)(\sigma^2/n)\) for large \( n \). Clearly,

\[ Var(\hat{\mu}_4) < Var(\hat{\mu}_3) < Var(\hat{\mu}_1) < Var(\hat{\mu}_2) \]

So \( \hat{\mu}_4 \) is MV in this set, followed by \( \hat{\mu}_3 \).

Example 2 (cont'd): It can be shown that \( \hat{\sigma}_1^2 \) is MV among \( \{\hat{\sigma}_1^2, \hat{\sigma}_2^2\} \).
Properties of estimators

Remark: In ranking estimators, we need a way to trade off bias and variance.

(3) **Best unbiased**: An estimator $\hat{\theta}$ is best unbiased (BUE) if it is MV among all unbiased estimators of $\theta$.

Synonyms: "efficient," and "uniformly minimum variance unbiased estimator (UMVUE)."

Example 1 (cont'd): It can be shown that $\hat{\mu}_1$ is BUE for this problem.

Example 2 (cont'd): It can be shown that $\hat{\sigma}_2^2$ is BUE for this problem.
Properties of estimators

(4) **Mean squared error:** \( \text{MSE}(\theta) = \mathbb{E} (\theta - \theta)^2 \).

Remark 1: If \( \theta \) unbiased, then \( \text{MSE}(\theta) = \text{Var}(\theta) \).

Remark 2: Not too hard to show that \( \text{MSE}(\theta) = \text{Var}(\theta) + [\text{Bias}(\theta)]^2 \).

Example 1 (cont'd): Remark 1 implies the MSEs of \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) are equal to their variance, so \( \text{MSE}(\hat{\mu}_1) < \text{MSE}(\hat{\mu}_2) \).

Remark 2 implies:

\[
\text{MSE}(\hat{\mu}_4) = (47 - \mu)^2
\]

which does not depend on \( n \), and also:

\[
\text{MSE}(\hat{\mu}_3) = \frac{n \sigma^2}{(n+1)^2} + \left( \frac{-\mu}{n+1} \right)^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}
\]

This last result shows that \( \text{MSE}(\hat{\mu}_1) < \text{MSE}(\hat{\mu}_3) \) if \( \mu \) is large relative to \( \sigma^2 \) and \( n \).
Example 2 (cont'd): It can be shown that $\text{MSE}(\hat{\sigma}_1^2) < \text{MSE}(\hat{\sigma}_2^2)$, even though $\hat{\sigma}_1^2$ is biased.
Properties of estimators

(5) Linearity: An estimator $\theta$ is linear if it is a linear function of the sample data.

Example 1 (cont'd): Clearly $\hat{\mu}_1$ and $\hat{\mu}_3$ are linear. $\hat{\mu}_4$ is also linear (trivially). $\hat{\mu}_2$ is not linear.

Example 2 (cont'd): Neither $\hat{\sigma}_1^2$ nor $\hat{\sigma}_2^2$ are linear.

Remark: It is easier to analyze the theoretical properties of linear estimators.
Asymptotic properties of estimators

Definition: Asymptotic properties of estimators describe their behavior as the sample size grows without bound.

Purposes:

- Indicators of reasonableness.
- Handy approximations for computation.

Synonym: "Large-sample properties."

(1) Asymptotic bias: Definition:

\[ \lim_{n \to \infty} Bias(\hat{\theta}) = \lim_{n \to \infty} E(\hat{\theta}) - \theta \]

Example 1 (cont'd): Since \( Bias(\hat{\mu}_3) = -\mu/(n+1) \), \( \hat{\mu}_3 \) is biased but asymptotically unbiased.

Example 2 (cont'd): It is not too hard to show that \( \hat{\sigma}_1^2 \) is biased but asymptotically unbiased.
Asymptotic properties of estimators

(2) Consistency: \( \theta \) is consistent if the probability that \( \theta \) is more than any finite distance from the true value converges to zero as the sample size grows without bound. Formally, \( \theta \) is consistent if for any number \( \delta > 0 \):

\[
\lim_{n \to \infty} \text{Prob} \{ |\hat{\theta} - \theta| > \delta \} = 0
\]

Remark: An inconsistent estimator is suspect, because it does not take advantage of each new observation to improve its accuracy.

Remark: It can be shown that an estimator is consistent if:

\[
\lim_{n \to \infty} \text{MSE}(\hat{\theta}) = 0
\]

Example 1 (cont'd): The formulas for their MSE's show that \( \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\mu}_3 \) are consistent but \( \hat{\mu}_4 \) is inconsistent.

Example 2 (cont'd): It can be shown that \( \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \) are
both consistent.
Asymptotic properties of estimators

(3) **Asymptotic normality:** An asymptotically normal estimator has a distribution which approaches the normal as the sample size grows without bound.

**Abbreviation:**

\[ \hat{\theta} \overset{\mathcal{A}}{\sim} N\left(\theta, \text{Var}(\hat{\theta})\right) \]

**Remark:** Often the exact distribution of an estimator is hopelessly complicated, but the asymptotic distribution provides a good approximation for large samples.

**Remark:** The mean of a sample drawn from any distribution is asymptotically normal, according to the central limit theorem. Hence, \( \hat{\mu}_i \) would be asymptotically normal even if the underlying data were not normally distributed.

**Remark:** Almost all estimators encountered in practice can be shown to be asymptotically normal. In particular, all the estimators in examples 1 and 2
are asymptotically normal except $\hat{\mu}_4$. 