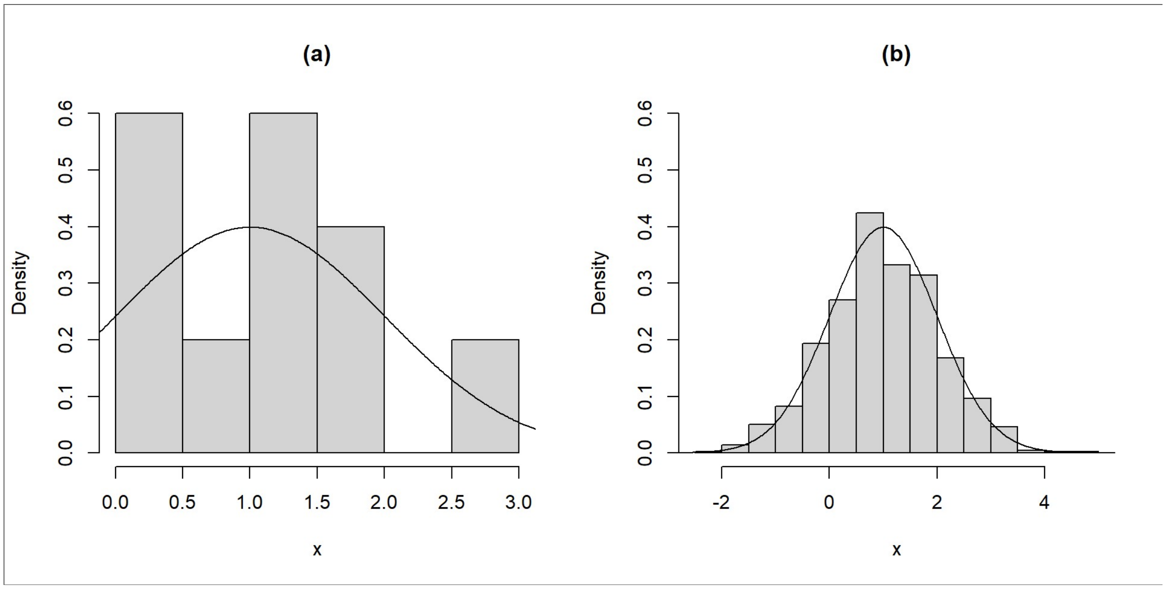
Chapter 2

Qualities of Esimators: Defining Good, Better, and Best

Estimation

Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, 1)$.
If n is large, then a histogram of the sampled values should look approximately like the distribution from which the sample came.

Estimation example



(a) is a histogram of a random sample of 10 values from a N(1, 1) distribution. (b) is a histogram of a random sample of 1,000 values from a N(1, 1) distribution

Estimation approach

A natural approach for estimating the mean is to use the **sample mean**

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- The notation $\hat{\mu}$ denotes an estimator of the parameter μ.
- We write $\hat{\mu} = \bar{X}$ to indicate that $\bar{X}$ is the estimator of μ.

Estimator vs estimate

An **estimator** is a random variable.

- It is a formula that tells us what to do when we get the data in the future.
- Uppercase $\bar{X}$ indicates that the estimator is a random variable.

An **estimate** is a single number obtained from the observed sample.

- It is not random since it is a function of the observed data.
- Lowercase $\bar{x}$ indicates that the estimate is not random.

Estimation ponderings

Estimating the mean μ with the sample mean $\bar{X}$ seems sensible, but leaves some questions unaddressed.

1. How “good” is the estimator?
2. What does it mean for an estimator to be “good”?
3. Can we find a better estimator?
4. How do we choose an estimator in other contexts, like the α parameter from a $\mathrm{Gamma}(\alpha, \beta)$ distribution?

2.1

Notation, Statistics, and Unbiasedness

Estimation context

Let θ denote a parameter or parameter vector.

- E.g., $\theta = \lambda$ for an exponential distribution.
- E.g., $\theta = (\alpha, \beta)$ for a Gamma distribution.

Estimation context

Distributions depends on parameters and we will emphasize that by explicitly including the relevant parameters in the description of the cdf, pmf, or pdf.

- E.g., The pdf might be denoted $f(x;\theta)$ or $f(x \mid \theta)$.
- E.g., If $X_1, X_2 \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, then the joint pdf is $f(x_1, x_2; \mu, \sigma^2) = f(\vec{x}; \mu, \sigma^2)$.

Estimation goal

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(x;\theta)$.
Our goal is to estimate a parameter θ or a function of θ, $\tau(\theta)$, using a **statistic**.
A **statistic** is a function of that only depends on the data and known values.

What is a statistic?

A statistic will be denoted by $T = t(X_1, X_2, \ldots, X_n) = t(\vec{(}X))$.

- $T = t(\vec{X}) = \bar{X}$ is a one-dimensional statistic.
- $T = t(\vec{X}) = (\bar{X}, \sum_{i=1}^{n} X_i^2, X_{(1)})$ is a multi-dimensional statistic.

Definition 2.1.1 (Unbiased)

An estimator T is **unbiased** for $\tau(\theta)$ if
$E[T] = \tau(\theta)$.

Mean of sample mean

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(x;\theta)$, with $E(X) = \mu < \infty$.
Prove that
$E(\bar{X}) = \mu$.
$\bar{X}$ is an unbiased estimator of μ.

Proof

Variance of sample mean

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(x;\theta)$, with $\mathrm{var}(X) = \sigma^2 < \infty$.
Prove that
$\mathrm{var}(\bar{X}) = \sigma^2/n$.

Proof

2.2

Mean Squared Error and Bias

Definition 2.2.1 (Bias)

The **bias** of an estimator $\hat{\theta}$ of θ is
$B(\hat{\theta}) = E(\hat{\theta}) - \theta.$

Definition 2.2.2 (Mean Squared Error)

The mean squared error (MSE) of an estimator $\hat{\theta}$ of θ is
$\mathrm{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$.

Bias of an unbiased estimator

The bias of an unbiased estimator $\hat{\theta}$ of θ is zero.

Proof

MSE of an unbiased estimator

If $\hat{\theta}$ is an unbiased estimator of θ, then
$\mathrm{MSE}(\hat{\theta}) = \mathrm{var}(\hat{\theta})$.

Proof

MSE, variance, and bias relationship

If $\hat{\theta}$ is an estimator of θ, then
$\mathrm{MSE}(\hat{\theta}) = \mathrm{var}(\hat{\theta}) + [B(\hat{\theta})]^2.$

Proof

Connecting the dots

- If multiple estimators are unbiased, we prefer the estimator with the smaller variance since it tends to be closer to to the truth, on average.
- If we are comparing two estimators, one biased and one unbiased, then the MSE is a fairer measure of each estimators quality.
- In the next section, we will consider other ways of assessing the quality of an estimator.

Example 2.2.1 (Long Estimation Example)

Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \text{Exponential}(\lambda)$ with $n \geq 3$.
We will consider estimating $\tau(\lambda) = 1/\lambda$ and λ.

Example 2.2.1 (cont)

Example 2.2.1 (cont)

Example 2.2.1 (cont)

Example 2.2.1 (cont)

Example 2.2.1 (cont)

Example 2.2.1 (cont)

2.3

Convergence in Probability

Large sample properties

The quality of an estimator often depends on the sample size used to compute the estimator.
We hope that an estimator will tend to be closer to θ as the sample size increases.
To highlight the dependence of an estimator on the sample size, we might include the subscript n in the estimator.

- $\hat{\theta} \equiv \hat{\theta}_n$.
- $\bar{X} \equiv \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Definition 2.3.1 (Convergence in Probability)

A sequence of random variables $\{X_n\}$ **converges in probability** to a random variable X if, for any $\epsilon > 0$,
$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0$.
(Equivalently, if $\lim_{n \to \infty} P(|X_n - X| \leq \epsilon) = 1$.)
We write $X_n \overset{P}{\to} X$.

Some Convergence in Probability interpretations

- As n gets large, X_n is almost always close to X.
- As the sample size increases, the probability that X_n is arbitrarily close to X is very high.
- As n gets large, it is very unlikely that X_n differs much from X.

Convergence in probability notes

1. Probabilities are numbers, so the limits are for a sequence of numbers.
2. A constant is a (boring) random variable, so you can have convergence in probability a number.
3. In practice, we want to know if $\hat{\theta}_n \xrightarrow{P} \theta$.
4. The inequalities can include equals without problems, i.e., $>$ or $\geq$, $<$ or $\leq$.

2.3.1

Markov's Inequality

Generalized Markov Inequality

If X is a random variable and $g(x)$ is a non-negative, real-valued function, then for any $c > 0$,

$$P[g(X) \geq c] \leq \frac{E[g(X)]}{c}.$$

Proof (Generalized Markov Inequality)

Proof (cont)

Proof (cont)

Markov's Inequality

For a random variable X and number $r, c > 0$,

$P(|X| \geq c) \leq \dfrac{E(|X|^r)}{c^r}$.

Proof (Markov's Inequality)