Ch. 6 – Point Estimation

Parameters and Estimators

A **parameter** is a constant describing a distribution:

- In a normal distribution, the mean μ and variance σ^2 are parameters.
- In an exponential distribution, the rate λ is a parameter.
- We will often use the symbol θ to represent a parameter generically.

An **estimator** is a random variable which is used to estimate a parameter.

- Given a random sample $X_1, \ldots, X_n$, the sample mean $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is an estimator for the mean μ.
- The sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ is an estimator for the variance σ^2.
- We will often use the symbol $\hat{\theta}$ to represent an estimator generically.

Unbiased Estimators

An estimator $\hat{\theta}$ is an **unbiased** estimator for θ if $E(\hat{\theta}) = \theta$.

- If X is a binomial random variable, then the **sample proportion** $\hat{p} = \frac{X}{n}$ is an unbiased estimator of p:

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n} \cdot np = p$$

- If $X_1, \ldots, X_n$ is a random sample from a distribution with mean μ, the sample mean $\overline{X}$ is an unbiased estimator for μ:

$$\begin{aligned} E(\overline{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n\mu = \mu \end{aligned}$$

Example: Uniform Distribution

Given a random sample $X_1, \ldots, X_n$ from a uniform distribution on $[0, \theta]$, how do we estimate θ?

- The mean μ is the midpoint of the interval, $\mu = \theta/2$. Therefore, $\theta = 2\mu$. Since we can estimate μ with $\overline{X}$, we can estimate θ with $\hat{\theta} = 2\overline{X}$.
- Each of the observations $X_1, \ldots, X_n$ will be less than θ, and if n is large we expect one of them to be close to θ. So we may estimate θ using the maximum: $\hat{\theta} = \max\{X_1, \ldots, X_n\}$.

Example: Uniform Distribution

We gave two possible estimators for the parameter θ of a uniform distribution on $[0, \theta]$:

$$\hat{\theta}_1 = 2\overline{X}$$
$$\hat{\theta}_2 = \max\{X_1, \ldots, X_n\}$$

Question: Are these estimators unbiased?

We may calculate the expected value of $\hat{\theta}_1$:

$$E(\hat{\theta}_1) = E(2\overline{X}) = 2E(\overline{X}) = 2\mu = \theta$$

Since $E(\hat{\theta}_1) = \theta$, this means that $\hat{\theta}_1$ is an unbiased estimator for θ.

Example: Uniform Distribution

Since $\hat{\theta}_2 = \max\{X_1, \ldots, X_n\}$ is always less than θ, intuition suggests that $\hat{\theta}_2$ will underestimate θ and hence must be biased.

Now we will calculate $E(\hat{\theta}_2)$. For $0 \le x \le \theta$, the cdf $F(x)$ of $\hat{\theta}_2$ is

$$\begin{aligned} F(x) &= P(\hat{\theta}_2 \le x) = P(\max\{X_1, \ldots, X_n\} \le x) \\ &= P(X_1 \le x, X_2 \le x, \ldots, X_n \le x) \\ &= P(X_1 \le x)P(X_2 \le x) \cdots P(X_n \le x) = (x/\theta)^n \end{aligned}$$

Therefore the pdf of $\hat{\theta}_2$ is, for $0 \le x \le \theta$,

$$f(x) = F'(x) = \frac{d}{dx}(x/\theta)^n = nx^{n-1}/\theta^n$$

So the expected value of $\hat{\theta}_2$ is

$$E(\hat{\theta}_2) = \int_{-\infty}^{\infty} xf(x)\ dx = \int_0^{\theta} x \cdot \frac{nx^{n-1}}{\theta^n}\ dx = \frac{n}{n+1}\theta$$

Example: Uniform Distribution

We considered two estimators for the parameter θ of a uniform distribution on $[0, \theta]$,

$$\hat{\theta}_1 = 2\overline{X}$$
$$\hat{\theta}_2 = \max\{X_1, \ldots, X_n\}$$

We found that $\hat{\theta}_1$ was unbiased but that $\hat{\theta}_2$ was biased. However, it is easy to modify $\hat{\theta}_2$ to produce an unbiased estimator $\hat{\theta}_3$:

$$\hat{\theta}_3 = \frac{n+1}{n}\hat{\theta}_2 = \frac{n+1}{n}\max\{X_1, \ldots, X_n\}$$

Since $E(\hat{\theta}_2) = \frac{n}{n+1}\theta$, it follows that

$$E(\hat{\theta}_3) = E\left(\frac{n+1}{n}\hat{\theta}_2\right) = \frac{n+1}{n}E(\hat{\theta}_2) = \frac{n+1}{n} \cdot \frac{n}{n+1}\theta = \theta$$

so $\hat{\theta}_3$ is in fact an unbiased estimator for θ.

Example: Uniform Distribution

We now have two unbiased estimators for the parameter θ of a uniform distribution on $[0, \theta]$:

$$\hat{\theta}_1 = 2\overline{X}$$
$$\hat{\theta}_3 = \frac{n+1}{n}\max\{X_1, \ldots, X_n\}$$

Question: Which of these estimators is better?

To answer this, we need a measure of how good an estimator is. One commonly used such measure is the *variance* of the estimator.

We can calculate the variance of $\hat{\theta}_1$:

$$V(\hat{\theta}_1) = V(2\overline{X}) = 4V(\overline{X}) = \frac{4}{n}V(X_1) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

Example: Uniform Distribution

Recall the pdf of $\hat{\theta}_2$ is $f(x) = nx^{n-1}/\theta^n$. Therefore,

$$E(\hat{\theta}_2^2) = \int_0^\infty x^2 f(x)\ dx = \int_0^\infty nx^{n+1}/\theta^n\ dx = \frac{n}{n+2}\theta^2$$

$$V(\hat{\theta}_2) = E(\hat{\theta}_2^2) - [E(\hat{\theta}_2)]^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2$$

$$= \left(\frac{n(n+1)^2 - n^2(n+2)}{(n+1)^2(n+2)}\right)\theta^2 = \frac{n}{(n+1)^2(n+2)}\theta^2$$

$$V(\hat{\theta}_3) = V\left(\frac{n+1}{n}\hat{\theta}_2\right) = \left(\frac{n+1}{n}\right)^2 V(\hat{\theta}_2)$$

$$= \left(\frac{n+1}{n}\right)^2 \cdot \frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{\theta^2}{n(n+2)}$$

Example: Uniform Distribution

We considered two unbiased estimators for the parameter θ of a uniform distribution on $[0, \theta]$:

$$\hat{\theta}_1 = 2\overline{X}$$
$$\hat{\theta}_3 = \frac{n+1}{n} \max\{X_1, \ldots, X_n\}$$

We calculated that their variances were $V(\hat{\theta}_1) = \frac{\theta^2}{3n}$ and $V(\hat{\theta}_3) = \frac{\theta^2}{n(n+2)}$. Therefore, for $n > 1$, $\hat{\theta}_3$ has a smaller variance.

More advanced statistical theory can be used to show that in fact $\hat{\theta}_3$ is a **minimum variance unbiased estimator**: it has a smaller variance than any other unbiased estimator.

Unbiasedness of Sample Variance

Given a distribution with variance σ^2, the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an unbiased estimator for σ^2.

$$
\begin{aligned}
E(S^2) &= E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(V(X_i) + [E(X_i)]^2) - n(V(\overline{X}) + [E(\overline{X})]^2)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)\right] = \sigma^2
\end{aligned}
$$

Maximum Likelihood Estimation

Suppose a Poisson process has unknown rate λ. We observe the process for 2 hours, and the number of events which occur is $X = 40$. What value of λ gives the largest probability $P(X = 40)$?

X is a Poisson random variable with mean 2λ, so

$$P(X = 40) = \frac{e^{-2\lambda}(2\lambda)^{40}}{40!}$$

The function $L(\lambda) = \frac{e^{-2\lambda}(2\lambda)^{40}}{40!}$ here is called the **likelihood function**. To maximize $L(\lambda)$, we take the derivative:

$$L'(\lambda) = \frac{1}{40!}(e^{-2\lambda} \cdot 40(2\lambda)^{39} \cdot 2 - 2e^{-2\lambda}(2\lambda)^{40})$$

Setting this equal to zero gives $40 - 2\lambda = 0$, so $\lambda = 20$. This is called the **maximum likelihood estimate** of λ.

Maximum Likelihood Estimation

In the previous problem, the likelihood function was given by

$$L(\lambda) = \frac{e^{-2\lambda}(2\lambda)^{40}}{40!}$$

The **log-likelihood function** is

$$\ell(\lambda) = \ln(L(\lambda)) = -2\lambda + 40\ln(2\lambda) - \ln(40!)$$

The value of λ which maximizes $\ell(\lambda)$ will also maximize $L(\lambda)$. The log-likelihood function is often easier to differentiate than the likelihood function. In this case,

$$\ell'(\lambda) = -2 + \frac{40}{\lambda}$$

Setting this equal to zero gives $\lambda = 20$ as before.

The log-likelihood function was found to be

$$l(\lambda) = \ln(L(\lambda)) = -2\lambda + 40\ln(2\lambda) - \ln(40!)$$

The maximum of log-likelihood function may be found by setting the derivative equal to zero:

$$l'(\lambda) = -2 + \frac{40}{2\lambda} \cdot 2 = -2 + \frac{40}{\lambda} = 0$$

Solving for λ gives $\lambda = 20$. This is the **maximum likelihood estimate** for λ.

Maximum Likelihood Estimation

In general, given a random variable X with pmf or pdf $f(x;\theta)$ depending on a parameter θ, the **maximum likelihood estimator** (mle) is the value $\hat{\theta}$ of θ that maximizes $f(X;\theta)$.

More generally, given a random sample $X_1, \ldots, X_n$, the mle is the value $\hat{\theta}$ that maximizes the joint pmf or joint pdf of $X_1, \ldots, X_n$: $f(X_1, \ldots, X_n;\theta) = f(X_1;\theta)f(X_2;\theta)\cdots f(X_n;\theta)$.

Under fairly general conditions, the maximum likelihood estimator $\hat{\theta}$ satisfies some desirable statistical properties:

- $\hat{\theta}$ exists and is unique.
- $\hat{\theta}$ is *asymptotically unbiased*: for large sample sizes, it is practically an unbiased estimator.
- $\hat{\theta}$ is *asymptotically efficient*: for large sample sizes, it approximately achieves the smallest possible variance for an unbiased estimator.
- $\hat{\theta}$ is *asymptotically normal*: for large sample sizes, it has approximately a normal distribution.

Example

Given a random sample $X_1, \ldots, X_n$ from an exponential distribution with rate λ, find the mle for λ.

The pdf of each X_i is $f(x; \lambda) = \lambda e^{-\lambda x}$. Since $X_1, \ldots, X_n$ are independent, the likelihood function is

$$L(\lambda) = f(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

The log-likelihood is therefore

$$\ell(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

Setting the derivative equal to zero gives $\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$. Solving for λ, we find the mle:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = 1/\overline{X}$$

Example

Given an observation of an exponential random variable X with unknown rate λ, find the mle for λ.

The pdf of each X is $f(x) = \lambda e^{-\lambda x}$. The log-likelihood is therefore

$$\ell(\lambda) = \ln \lambda - \lambda x$$

Setting the derivative equal to zero gives $\frac{1}{\lambda} - x = 0$. Solving for λ, we find the mle:

$$\hat{\lambda} = \frac{1}{X}$$

Invariance Principle for MLE

Let $h(\theta)$ be a function of θ. If $\hat{\theta}$ is the mle for θ, then $h(\hat{\theta})$ is the mle for $h(\theta)$.

Example: Given an observation of an exponential random variable X with unknown rate λ, we found that the mle for λ was

$$\hat{\lambda} = \frac{1}{X}$$

Therefore, the mle for $\mu = 1/\lambda$ is

$$\hat{\mu} = 1/\hat{\lambda} = \frac{1}{1/X} = X$$

Thus, the mle for μ is simply the observed value X. It can be shown that this in fact is the minimum variance unbiased estimator for μ.

MLE of Several Parameters

There are many families of distributions where we would want to estimate two parameters at once. For example,

- For a normal distribution, we want to estimate μ and σ.
- For a gamma distribution, we want to estimate k and λ.
- For a Weibull distribution, we want to estimate α and β.

The definition of mle extends in an natural way to distributions with several parameters:

Given a random sample $X_1, \ldots, X_n$ from a distribution with several parameters $\theta_1, \ldots, \theta_m$, the mle of $(\theta_1, \ldots, \theta_m)$ is the combination of values $(\hat{\theta}_1, \ldots, \hat{\theta}_m)$ maximizing the joint pmf or joint pdf $f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$ of $X_1, \ldots, X_n$.

Estimating Parameters of a Normal Distribution

Given a random sample $X_1, \ldots, X_n$ from a normal distribution:

The MLE for μ is the sample mean $\overline{X}$, and this is also the minimum variance unbiased estimator.

This means that alternative estimators for μ, such as the sample median $\tilde{X}$, must have a greater variance. (In fact, $V(\overline{X}) = \frac{\sigma^2}{n}$, while $V(\tilde{X}) \approx \frac{\pi}{2} \cdot \frac{\sigma^2}{n}$ for large n.)

The MLE for σ^2 is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. However, this estimator is biased; the minimum variance unbiased estimator is the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.