Ch 6. Point estimation

- Example 6.1

Suppose that we observe $n = 5$ numbers by independent sampling from a population having the same distribution: 1,3,4,6,7

- Different set of numbers could have been obtained due to the randomness as well as the nature of pupulation
- Postulate a family of plaussible pdf's, for example, $\mathcal{F} = \{N(\theta, 1) : -\infty < \theta < \infty\}$
- Make a guess for θ. How to make a guess?
  - How about the "arithmatic mean"? 4.2
  - How about the "median"? 4
- Which guess is better than the other? Why?
  - Comparing 4.2 ad 4 is meaningless
  - Need to compare the method yielding 4.2 ad 4!

Terminologies

- (parametric) model: $\mathcal{F} = \{f(x;\theta) : \theta \in \Omega\}$
- parameter space Ω: a set of plausible values of parameter θ
- Random sample: a set of iid random variables
- Statistic: Suppose that n random variables $X_1, \ldots, X_n$ constitute a sample from the distribution of a random variables X. Then any function $T = T(X_1, \ldots, X_n)$ of the sample is called a Statitic.
- A point estimator: A statistic which is used to make a guess for θ

- Tools to compare point estimators for θ

  - Bias of θ: $bias(\hat{\theta}) = E(\hat{\theta}) - \theta$
    If $bias(\hat{\theta}) = 0$ for all $\theta \in \Omega$, we say that $\hat{\theta}$ is an unbiased estimator for θ.
  - Mean square error of $\hat{\theta}$

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = V(\hat{\theta}) + (bias(\hat{\theta}))^2$$

- Remark

- Ideally $\hat{\theta}$ needs to be unbiased
- Unbiased estimator is not unique
- If an unbiased estimator $\hat{\theta}$ has the smallest variance among all unbiased estimators, we say that $\hat{\theta}$ is the Minimum Variance Unbiased Estimator (MVUE)
- What if $\hat{\theta}_1$ has a small bias but has a large variance, and $\hat{\theta}_1$ has a large bias but has a small variance?

Ch 6.1 Maximum Likelihood Estimation

- Example 6.2
  - Model: $\mathcal{F} = \{b(1,p) : p \in \{1/3, 2/3\}\}$
  - Realized random variables: 1,1,0,0,1,0,0,0,1,0
  - Only two possible point estimates: $\hat{p}_1 = 1/3$ and $\hat{p}_2 = 2/3$
  - Which one is better?

- Example 6.3
  - Model: $\mathcal{F} = \{b(1, p) : p \in (0, 1)\}$
  - Realized random variables: 1,1,0,0,1,0,0,0,1,0
  - What is the most likelihood estimate for unknown p?

More terminologies

Suppose that $X_1, \ldots, X_n$ is a random sample from pdf $f(x;\theta)$, $\theta \in \Omega$.

- The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

- The log-likelihood function is

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

- The maximum likelihood estimator (MLE) is the maximizer of $L(\theta)$ or $l(\theta)$.

Some remarks

- The maximizer of $L(\theta)$ and $l(\theta)$ are the same because $\log$ function is an increasing function
- In most cases, $l(\theta)$ is more convenient than $L(\theta)$.
- Under certain conditions, $\hat{\theta}^{MLE}$ is just a solution of $l'(\theta) = 0$ and $l'(\theta) = 0$ is often called "likelihood equation".

- Example 6.3 (revisited)
  - Model: $\mathcal{F} = \{b(1, p) : p \in (0, 1)\}$
  - Realized random variables: 1,1,0,0,1,0,0,0,1,0
  - What is the most likelihood estimate for unknown p?

- Example 6.4

  - $X_1, \ldots, X_n$ is a random sample from $N(\mu, 1)$, $-\infty < \mu < \infty$. Find the MLE of μ.

- Example 6.5
  - $X_1, \ldots, X_n$ is a random sample from $U[0, \theta]$, $\theta > 0$. Find the MLE of θ.

- Example 6.6 (Non-uniqueness of MLE)

  - $X_1, \ldots, X_n$ is a random sample from $U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, $-\infty < \mu < \infty$. Find the MLE of θ.

Theorem (Functional invariance of MLE, p.372)

$X_1, \ldots, X_n$ *is a random sample from* $f(x;\theta)$, $\theta \in \Omega$. *Let* $\eta = g(\theta)$ *be a parameter of interest. If* $\hat{\theta}$ *is the MLE of* θ, *then the MLE of* η *is* $\hat{\eta} = \widehat{g(\theta)} = g(\hat{\theta})$.

- Example 6.7

- $X_1, \ldots, X_n$ is a random sample from $b(1,p)$, $0 < p < 1$
  What is the MLE of $\eta = \frac{p}{1-p}$?

- $X_1, \ldots, X_n$ is a random sample from $Exp(\lambda)$, $0 < \lambda$
  What is the MLE of $\eta = P(X_1 > 1)$?

- If an estimator $\hat{\theta}$ converges to θ in probability, we say that $\hat{\theta}$ is a consistent estimator of θ.

- Example 6.4 (revisited)
$X_1, \ldots, X_n$ is a random sample from $N(\mu, 1)$. We know that the MLE of μ is $\bar{X}_n$, and $\bar{X}_n \xrightarrow{p} \mu$. That is, $\bar{X}_n$ is a consistent estimator of μ.

- Some comments

- To be a *good* estimator, we need "unbiasness"+"consistency".
- BUT, in most cases, "consistency" is enough to be a good estimator.
- Is the MLE unbiased?
- Is the MLE consistent?

Regularity conditions

(R0) The pdfs are identifiable.

i. e. $\theta_1 \neq \theta_2$ implies $f(x;\theta_1) \neq f(x;\theta_2)$

(R1) The pdfs have common support for all $\theta \in \Omega$

(R2) The true parameter θ_0 is an interior point of Ω

Theorem (Jensen's Inequality, p.95)

If $\phi(x)$ *is convex, then* $\phi(E(X)) \leq E(\phi(X))$.

Theorem (p. 370)

Let θ_0 be the true parameter. Under assumptions (R0) and (R1),

$$\lim_{n\to\infty} P_{\theta_0}\left[L(\theta_0) > L(\theta)\right] = 1, \quad \textit{for all } \theta \neq \theta_0$$

Theorem (p.373)

Suppose that $f(x;\theta)$ *is differentiable with respect to* θ *in* Ω*. Under (R0)*$\sim$*(R2), the MLE of* θ *is* *consistent*.

Exercises: 6.1.2, 6.1.4, 6.1.6, 6.1.9, 6.1.11, 6.1.12

- More regularity conditions

(R3) $f(x;\theta)$ is twice differentiable with respect to θ.

(R4) $\frac{d}{d\theta}\int f(x;\theta)dx = \int \frac{d}{d\theta} f(x;\theta)dx$

$\frac{d^2}{d\theta^2}\int f(x;\theta)dx = \int \frac{d^2}{d\theta^2} f(x;\theta)dx$

cf) score function:

$$s(\theta, x) = \frac{d}{d\theta}\log(f(x;\theta)) = \frac{\frac{d}{d\theta} f(x;\theta)}{f(x;\theta)}$$

Fact(p.376): Under (R0)$\sim$(R4), we have

$$(1)\ E\left[\frac{d}{d\theta}\log(f(X;\theta))\right] = 0$$

$$\begin{aligned}(2)\ V\left[\frac{d}{d\theta}\log(f(X;\theta))\right] &= E\left[\left(\frac{d}{d\theta}\log(f(X;\theta))\right)^2\right] \\ &= -E\left[\frac{d^2}{d\theta^2}\log(f(X;\theta))\right]\end{aligned}$$

Theorem (Rao-Cramer Lower bound, p.379)

Let $X_1, \ldots, X_n$ *be iid random variable with pdf* $f(x;\theta)$, $\theta \in \Omega$. *Let* $Y = u(X_1, \ldots, X_n)$ *be an unbiased estimator of* θ. *(i.e.* $E(Y) = \theta$*)*
Then, under (R0)∼(R4), $V(Y) \geq (nI(\theta))^{-1}$.

- Remark

- In the textbook, thre is a more general result than this theorem. That is, when $E(Y) = k(\theta)$, we have

$$V(Y) \geq \frac{(k'(\theta))^2}{nI(\theta)}$$

- Rao-Cramer lower bound gives the theoretical lower bound of any unbiased estimator for θ.
- If your estimator is unbiased and its variance is $V(Y) = (nI(\theta))^{-1}$, then you can say that your estimator is the MVUE.

Definition (Efficient estimator)

Let $\hat{\theta}$ be an unbiased estimator of a parameter θ. $\hat{\theta}$ is called an efficient estimator of θ if the variance of $\hat{\theta}$ attains the Rao-Cramer lower bound.

Definition (Efficiency)

$$\text{Efficiency of } \hat{\theta} : \quad \frac{(nI(\theta))^{-1}}{Var(\hat{\theta})}$$

- Example 6.8(p.381): $X_1, \ldots, X_n \stackrel{iid}{\sim} Poisson(\theta)$

- Example 6.9(p. 381): $X_1, \ldots, X_n \overset{iid}{\sim} Beta(\theta, 1)$

- Remark

- Although the MLE is not unbiased in general, as we have seen, the MLE is asymptotically unbiased under some regularity conditions.
- The MLE is not efficient in general, but it is asymptotically efficient.

- Question: What is the distribution of the MLE?

$\rightarrow$ With finite sample, the distribution of the MLE is different over different statistical models. But, the MLE asymptotically follows a normal distribution. For this asymptotical normality, we need the following additional regularity condition:

- (R5) $f(x;\theta)$ is three times differentiable and

$$E\left(\left|\frac{d^3}{d\theta^3}\log f(X;\theta)\right|\right) < \infty$$

See p.382 Assumption 6.2.2 for more concrete condition.

Theorem (Asymptotical normality of MLE, p383)

Under (R0)∼(R5), if $0 < I(\theta) < \infty$*, then*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I^{-1}(\theta))$$

- Remark

- Although unbiasness is a desirable property of an estimator, there is an aymptotically equivalent class of estimators.
- When we extend our focus to the class of asymptotically unbiased estimators, we may need asymptotic version of "efficiency" to measure the quality of estimators.
- The previous theorem shows that the asymptotic variance of MLE attains the theoretical lower bound. So, the MLE is asymptotically efficient. Moreover, the limiting distribution of the MLE is normal.

Definition

Asymptotical Efficiency

1. If $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$, the asymptotical efficiency of $\hat{\theta}$ is

$$e(\hat{\theta}) = \frac{I^{-1}(\theta)}{\sigma^2}$$

.

2. If $e(\hat{\theta}) = 1$, we say that $\hat{\theta}$ is <u>asymptotically efficient</u> estimator.
3. For two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, if $\sqrt{n}(\hat{\theta}_1 - \theta) \xrightarrow{d} N(0, \sigma_1^2)$ and $\sqrt{n}(\hat{\theta}_2 - \theta) \xrightarrow{d} N(0, \sigma_2^2)$, the asymptotical relative efficiency (ARE) of $\hat{\theta}_1$ to $\hat{\theta}_2$ is

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sigma_2^2}{\sigma_1^2}$$

.

- Method of Moments (MOM) estimator
  - $X_1, \ldots, X_n \overset{iid}{\sim} f(x;\theta)$.
  - If $E(X) = g(\theta)$, we can expect $\bar{X} \approx E(X) = g(\theta)$
  - A method of moments estimator of θ can be defined as

$$\hat{\theta}^{MoM} = g^{-1}(\bar{X})$$

  - Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

- Example 6.9(revisited): $X_1, \ldots, X_n \overset{iid}{\sim} Beta(\theta, 1)$