Introduction to Biostatistical Theory - Point Estimation

Robert Parker

Point Estimation

Given a sample $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$

- iid sample from a population with density $f(x|\theta)$
- θ is an unknown parameter

We want to find a good "estimator" of θ.

Point estimation uses the value of a statistic to estimate a population parameter. The value is the **point estimate** of the parameter.

Definition (2.1)

A statistic is a function of the data vector $(X_1, X_2, \ldots, X_n)$, which does not depend on unknown parameters.

Estimators

Consider a sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ which is iid $N(\mu, \sigma^2)$.

- μ and σ^2 are unknown population parameters
- $T(\mathbf{X}) = \sum_{i=1}^n X_i/n = \bar{X}$ is an estimator of μ
- $\bar{X} + \mu$ is not a statistics since it depends on the unknown parameter μ
- $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is an estimator of σ^2
- Note that statistics need not be univariate. $T(\mathbf{X}) = (\bar{X}, s^2)$ is a multivariate statistic.

Method of Moments

Idea: Equate the first few moments of a population to the corresponding moments of a sample to get as many equations as needed to solve fr the unknown parameters.

Setup: Given an iid sample $X_1, X_2, \ldots, X_n \sim f(x|\theta)$

- The kth population moment is $\mu_k = E[X_1^k]$
- The kth sample moment is $m_k = \sum_{i=1}^n X_i^k / n$

If you have 1 parameter: set $m_1 = \mu_1$ and solve for the parameter

If you have 2 parameters: set $m_1 = \mu_1$ and $m_2 = \mu_2$ and solve for the parameters

$\vdots$

Method of Moments

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(\alpha, 1)$, α unknown. Use the method of moments to estimate α.

Method of Moments

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Binomial(k, p), $k =\#$ of trials and $p =$probability of success are both unknown. Use the method of moments to estimate both k and p.

Method of Moments

Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Estimate μ and σ^2 using the method of moments.

Method of Moments

1. The MoM is generally very easy to find
2. If there is a unique solution to MoM equations, then the estimator is weakly consistent and under mild additional assumptions it is asymptotically normal, that is $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{D} N(0, \Sigma)$.
3. A drawback to the MoM is that the estimators need not make sense. For example, in the uniform example if we have $X_1 = 0.1, X_2 = 0.6, X_3 = 0.7, X_4 = 0.9$, then $\widehat{\alpha} = 0.2$, but $X_{(1)} = 0.1$ so we must have $\alpha < 0.1$. Similarly, for the Binomial example, it is possible to have data that result in $\widehat{p} < 0$ and $\widehat{k}$ having a non-integer value.

Maximum Likelihood Estimator (MLE)

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} f(x|\theta)$, θ an unknown parameter

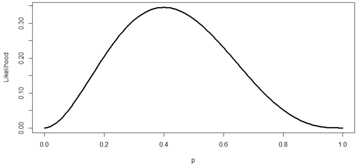
Idea: Based on the observed data, choose the value of θ such that the observed data is "most likely" to have occurred.

Suppose that $X \sim \text{Binomial}(5, 0.6)$, so the true value of n and p are $n = 5$ and $p = 0.6$. Suppose that n is known and p is unknown, but we observe $X = 2$. What is the most likely value of p?

- $p = 0$: $P(X = 2) = 0$
- $p = 0.2$: $P(X = 2) = \binom{5}{2}0.2^2(1 - 0.2)^3 = 0.2048$
- $p = 0.4$: $P(X = 2) = \binom{5}{2}0.4^2(1 - 0.4)^3 = 0.3456$
- $p = 0.6$: $P(X = 2) = \binom{5}{2}0.6^2(1 - 0.6)^3 = 0.23$
- $p = 0.8$: $P(X = 2) = 0.0512$
- $p = 1$: $P(X = 2 = 0)$

From this dataset the MLE is $\widehat{p} = 0.4$.

MLE



Definition (2.2)

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$ (pdf or pmf). Given the data $\mathbf{X} = \mathbf{x}$, the **likelihood function** is defined as the joing pdf (or pmf) of the data viewed as a function of θ,

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

The maximum likelihood estimator (MLE) of θ is defined as

$$\widehat{\theta}_{MLE} = \text{argmax}_{\theta \in \Theta} L(\theta|\mathbf{x}).$$

MLE

If the likelihood function is differentiable in θ, then possible candidates for the MLE can be found by setting the set of first derivatives to 0 and solving for θ. To make this easer, we will often use the **log-likelihood** instead:

$$\ell(\theta|\mathbf{x}) = \log(L(\theta|\mathbf{x})) = \sum_{i=1}^{n} \log(f(x_i|\theta)).$$

We can do this because the log function is a monotonically increasing function, so

$$\text{argmax}_{\theta \in \Theta} L(\theta|\mathbf{x}) = \text{argmax}_{\theta \in \Theta} \ell(\theta|\mathbf{x})$$

MLE - Binomial

Let $X \sim \text{Binomial}(5, p)$. Find the MLE of p. Note that $p \in [0, 1]$.

MLE - Exponential

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\theta)$, $\theta > 0$. Find the MLE of θ.

MLE - Normal with known variance

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$, $\theta \in \mathbb{R}$. Find $\widehat{\theta}_{MLE}$.

MLE - Normal with known variance

Here we present an alternative least squares argument to fine the MLE of μ.

MLE - Normal with restricted parameter space

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$, $\theta \in [0, \infty)$. Find $\widehat{\theta}_{MLE}$.

MLE - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \tau)$, $\mu \in \mathbb{R}, \tau > 0$. Find the MLEs of μ and τ.

Bivariate Maximization

To ensure a local maximum, the following conditions need to be met:

a. At least one second partial derivative is negative: $\frac{\partial^2 l}{\partial \mu^2}\Big|_{\mu=\hat{\mu},\tau=\hat{\tau}} < 0$ or $\frac{\partial^2 l}{\partial \tau^2}\Big|_{\mu=\hat{\mu},\tau=\hat{\tau}} < 0$.

b. The determinant of the second derivatives is positive:

$$\left| \begin{array}{cc} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \tau} \\ \frac{\partial^2 l}{\partial \mu \partial \tau} & \frac{\partial^2 l}{\partial \tau^2} \end{array} \right|_{\mu=\hat{\mu},\tau=\hat{\tau}} > 0.$$

MLE - Normal

MLE - Uniform

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U[0, \theta]$, $\theta \in (0, \infty)$. Find $\widehat{\theta}_{MLE}$.

MLE - Uniform

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U[\theta, \theta + 1]$, $\theta \in (0, \infty)$. Find $\widehat{\theta}_{MLE}$.

MLE - Cauchy

Suppose $X_1, \ldots, X_n \overset{i.i.d.}{\sim} f(x|\theta) = \frac{1}{\pi\left[1+(x-\theta)^2\right]}$. The log-likelihood is $\ell(\theta; \boldsymbol{x}) = -n \log \pi - \sum_{i=1}^n \log\left[1 + (x_i - \theta)^2\right]$. Differentiating both sides with regard to θ and setting it to zero,

$$\ell'(\theta; \boldsymbol{x}) = 2\sum_{i=1}^{n} \frac{\theta - x_i}{1 + (x_i - \theta)^2} = 0.$$

When $\theta \to -\infty$, $l'(\theta; \boldsymbol{x}) \to 0-$ (from below); when $\theta \to \infty$, $l'(\theta; \boldsymbol{x}) \to 0+$ (from above). Hence, there are ≥ 1 and odd number of roots. Enforcing a common denominator for all $\frac{\theta - x_i}{1+(x_i-\theta)^2}$, we get a polynomial of degree $2n-1$. So the number of roots satisfies $R_n = 2K_n - 1$, $1 \leq K_n \leq n$. Reeds (1985) shows that

$$\mathbb{P}(K_n = k) \to \frac{\pi^{-k}}{k!} e^{-1/\pi}, \quad \text{as } n \to \infty.$$

Actually, $K_n \leq 4$ with probability close to 1 for all n.

MLE

1. Under mild conditions, the MLE is both consistent and asymptotically normal.
2. Advantages over MoM estimator
   - Estimator is always in the parameter space
   - Always a function of a sufficient statistic (more on this later)
3. Disadvantages
   - Solution may be difficult to solve for or not available in closed form.
4. MLE's may not be unique.

MLE - Functions of the Parameter

If $\widehat{\theta}$ is the MLE of θ, then for any function $g(\theta)$, the MLE of $g(\theta)$ is $g(\widehat{\theta})$.

Ex. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U[0, \theta]$. What is the MLE of θ^2?

MLE - Consistency

Regularity Conditions:

R0 The pdfs are identifiable. That is, $\theta \neq \theta' \implies f(\cdot|\theta) \neq f(\cdot;\theta')$.

R1 The support of the pdfs does not depend on θ. That is, the support of $f(\cdot;\theta)$ is the same for all $\theta \in \Theta$.

R2 The true value of θ, θ_0, is an interior point of the parameters space Θ.

R3 The pdf $f(x|\theta)$ is differentiable in θ for all x.

Theorem

Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$ and suppose the regularity conditions R0-R3 hold. Then the MLE $\widehat{\theta}_n$ converges in probability to θ_0.

MLE - Asymptotic Normality

R4 The pdf $f(x|\theta)$ is three times differentiable in θ for all x, and we can exchange the order of integration and the first and second derivative with respect to θ.

R5 For all θ_0, there exists a c and a function $M(x)$ (both possible depending on θ_0) such that $E_{\theta_0}[M(X_1)] < \infty$ such that

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(x|\theta) \right| \leq M(x) \text{ for all x and for all } \theta \in [\theta_0 - c, \theta_0 + c].$$

Theorem

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$, and suppose the regularity conditions R0-R5 hold. Then the MLE, $\widehat{\theta}_n$, satisfies

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)),$$

where $I(\theta) = E\left[\left(\frac{\partial}{\partial \theta} \log f(X_1|\theta)\right)^2\right]$.

MLE - AN Proof

MLE - AN Proof

MLE - AN Proof

Unbiasedness

Definition (2.3)

The **bias** of an estimator T of a parameter θ is defined as

$$bias(T) = E_\theta[T] - \theta.$$

The estimator T is said to be **unbiased** if $bias(T) = 0$ for all θ. Let $b_n(\theta) = E_\theta[T_n] - \theta$. If $b_n(\theta) \to 0$ as $n \to \infty$ for all θ, then T_n is **asymptotically unbiased**.

Unbiased Estimator - Binomial

Let $X \sim \text{Binomial}(n, p)$ with n known and $p \in [0, 1]$ unknown. Find and unbiased estimator of p.

Unbiased Estimator - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

Unbiased Estimator - Uniform

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$. Is the $\widehat{\theta} = X_{(n)}$ unbiased?

Efficiency - UMVUE

We want the unbiased estimator with the lowest variance (least spread out sampling distribution).

Definition (2.4)

$\hat{\theta}$ is a uniformly minimum variance unbiased estimator (UMVUE) of a parameter θ if (1) $\hat{\theta}$ is unbiased, and (2) for any other unbiased estimator $\tilde{\theta}$, $\mathrm{Var}_\theta(\hat{\theta}) \leq \mathrm{Var}_\theta \tilde{\theta}$ uniformly for all θ.

There is no "best" estimator for all θ if we don't restrict the class of estimators. For example, the estimator $\hat{\theta} = 0$ is the best when $\theta = 0$ but a terrible estimator if $\theta \neq 0$. Even after restriction to unbiased estimators, it is often not easy to find the UMVUE. However, there exists a lower bound for the variances of all unbiased estimators.

Fisher Information

In the case of unbiased estimators, we can establish a lower bound on the possible achievable variances under some conditions. To find this lower bound, we need to first introduce the concept of Fisher information.

Definition (2.5)

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$. Each X_i carries "some information about θ, so $X_1, X_2, \ldots, X_n$ carries "n pieces of information" about θ. The Fisher information is defined as

$$I(\theta) = E\left[\frac{\partial \log f(X|\theta)}{\partial \theta}\right]^2 = -E\left[\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2}\right].$$

Note that this is based on one observation. The final equality only holds under regularity conditions (such as R0-R4) which we will discuss later, but will hold for most distributions we use in this class. The total information is $I_n(\theta) = nI(\theta)$, in the iid case.

Fisher Information - Poisson

Let $X \sim \text{Poisson}(\lambda)$. Find the Fisher information.

Fisher Information - Binomial

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Binomial}(k, p)$, k known and $p \in [0, 1]$ unknown. Find the Fisher information for p.

Cramer-Rao Lower Bound

Theorem (3.2 Cramer-Rao lower bound)

Suppose $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f(\boldsymbol{x}|\theta)$*and* $f(\boldsymbol{x}|\theta)$ *satisfies the following regularity conditions:*

1. *the support of* $f(\boldsymbol{x}|\theta)$ *does not depend on* θ*; and*
2. *for any statistic* $T(\boldsymbol{X})$ *satisfying* $\mathrm{Var}(T) < \infty$*, the following exchangeability between integration and differentiation holds:*

$$\frac{d}{d\theta}\mathbb{E}T(\boldsymbol{X}) = \frac{d}{d\theta}\int T(\boldsymbol{x})f(\boldsymbol{x}|\theta)d\boldsymbol{x} = \int T(\boldsymbol{x})\frac{d}{d\theta}f(\boldsymbol{x}|\theta)d\boldsymbol{x}.$$

If T *is an unbiased estimator for* θ*, then* $\mathrm{Var}(T) \geq I_n^{-1}(\theta)$*, where* $I_n(\theta) = \mathbb{E}\left[\frac{d}{d\theta}\log f(\boldsymbol{X}|\theta)\right]^2$ *is the Fisher information.*
In addition, if T *is an unbiased estimator for* $g(\theta)$*,*

$$\mathrm{Var}(T) \geq [\frac{d}{d\theta}g(\theta)]^2 I_n^{-1}(\theta).$$

Cramer-Rao Lower Bound - Proof

Cramer-Rao Lower Bound - Comments

1. If $T(\boldsymbol{X})$ is an unbiased estimator of $g(\theta)$ and $\boldsymbol{X} = (X_1, \ldots, X_n)$ are iid, then

$$Var(T(\boldsymbol{X})) \geq \frac{[g'(\theta)]^2}{nI(\theta)}.$$

   If $\text{Var}(T(\boldsymbol{X})) = \dfrac{[g'(\theta)]^2}{nI(\theta)}$, then $T(\boldsymbol{X})$ has the smallest variance, and we call T efficient (the best unbiased estimator).

2. If the pdf $f(x|\theta)$ is a regular one parameter exponential family, then there exists an unbiased estimator $T(\boldsymbol{X})$ such that its variance achieves the Cramer-Rao lower bound. A one parameter exponential family has a density of the form

$$f(x|\theta) = h(\boldsymbol{x})c(\theta)e^{T(\boldsymbol{x})w(\theta)}.$$

   In particular, if $E[T(\boldsymbol{X})] = g(\theta)$, $\frac{d}{d\theta}w(\theta) \neq 0$ and is continuous, and the support does not depend on θ, then $T(\boldsymbol{X})$ is the UMVUE for $g(\theta)$ and achieves the C-R lower bound.

Cramer-Rao Lower Bound - Comments

The regularity conditions are critical. Let $\boldsymbol{X} \stackrel{iid}{\sim} U(0, \theta)$. Recall the MLE is $\widehat{\theta} = X_{(n)}$ and $E[X_{(n)}] = \frac{n}{n+1}\theta$.

Cramer-Rao Lower Bound - Bernoulli

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p) = \text{Binomial}(1, p)$.

Cramer-Rao Lower Bound - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find the C-R lower bound for the variance of unbiased estimators of μ and σ^2.

Cramer-Rao Lower Bound - Normal

Relative Efficiency

Definition (2.6)

If $T_1(\boldsymbol{X})$ and $T_2(\boldsymbol{X})$ are two unbiased estimators of $g(\theta)$, the relative efficiency of T_2 relative to T_1 is given by

$$\frac{\text{Var}(T_1(\boldsymbol{X}))}{\text{Var}(T_2(\boldsymbol{X}))}$$

Definition (2.7)

If $T_1(\boldsymbol{X})$ is an unbiased estimator of θ and

$$\frac{\text{Var}(T_1(\boldsymbol{X}))}{\text{C-R lower bound}} \to 1 \text{ as } n \to \infty,$$

then $T_1(\boldsymbol{X})$ is said to be **asymptotically efficient**.

Relative Efficiency - Uniform

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$. Consider the following two unbiased estimators of θ

1. $T_1(\boldsymbol{X}) = \frac{n+1}{n} X_{(n)}$
2. $T_2(\boldsymbol{X}) = 2\bar{X}$

$\text{Var}(T_1(\boldsymbol{X})) = \dfrac{\theta^2}{n(n+2)}$ and $\text{Var}(T_2(\boldsymbol{X})) = \dfrac{\theta^2}{3n}$, so

$$RE = \frac{\text{Var}(T_1(\boldsymbol{X}))}{\text{Var}(T_2(\boldsymbol{X}))} = \frac{\dfrac{\theta^2}{n(n+2)}}{\dfrac{\theta^2}{3n}} = \frac{3}{n+2} \to 0 \text{ as } n \to \infty.$$

Asymptotically Efficient

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Consider the sample variance S_n^2.

Mean Squared Error

We need not always use unbiased estimators. In such a case, we can compare the MSE's instead.

Definition (2.8)

Let $T_1(\boldsymbol{X})$ be an estimator of θ. The **mean squared error (MSE)** is defined as

$$MSE = E_\theta(T_1(\boldsymbol{X}) - \theta)^2.$$

Note: It is easy to see that

$$\mathbb{E}(T-\theta)^2 = \mathbb{E}(T - \mathbb{E}T)^2 + (\mathbb{E}T - \theta)^2 = \mathrm{Var}\,T + bias^2(T).$$

An estimator with good MSE usually has both small variance and small bias. For unbiased estimators, we have $MSE = \mathrm{Var}(T)$.

MSE

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. We want to estimate σ^2. Consider the two estimators:

1. $T_1(\boldsymbol{X}) = \dfrac{1}{n-1} \sum_{i=1}^{n}(X_i - \bar{X})^2$ (sample variance/unbiased)
2. $T_2(\boldsymbol{X}) = \dfrac{1}{n} \sum_{i=1}^{n}(X_i - \bar{X})^2$ (MLE/biased)

Which one is better?

Let $X_1, X_2, \ldots, X_n \sim f(x|\theta)$, θ is unknown. We want to estimate θ. We may not be able to achieve the C-R lower bound, but we still want the best estimator. We want an estimator that still contains all the information about the parameter, but how do we justify this? This is the concept of sufficiency.

Definition (2.9)

Let $(X_1, X_2, \ldots, X_n) \sim f(\boldsymbol{x}|\theta)$. A statistic $T(\boldsymbol{X})$ is a sufficient estimator of θ iff for each value of θ the conditional distribution of $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on θ, i.e.

$$f(\boldsymbol{x}|T(\boldsymbol{x})) \text{ is free of } \theta.$$

Sufficient - Bernoulli

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(θ). Show that $T(\boldsymbol{X}) = \bar{X}$ is a sufficient estimator of θ.

Sufficient - Bernoulli

Show that $Y = \frac{1}{6}(X_1 + 2X_2 + 3X_3)$ is not a sufficient estimator of the Bernoulli parameter θ.

Factorization Theorem

Theorem (Fisher-Neyman Factorization Theorem)

Let $f(\boldsymbol{x}|\theta)$ denote the joint pdf of a sample $\boldsymbol{X}$. A statistic $T(\boldsymbol{X})$ is a sufficient statistic of θ iff there exists functions $g(t, \theta)$ and $h(\boldsymbol{x})$ such that for all $\boldsymbol{x}$ and θ,

$$f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x}), \theta)h(\boldsymbol{x}).$$

Sufficient - Bernoulli

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$. Find a sufficient statistic for θ.

Sufficient - Normal with known variance

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where σ^2 is known. Find a sufficient statistic for μ.

Sufficient - Uniform

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$. Find a sufficient statistic for θ.

Sufficient - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where μ and σ^2 are unknown. Find a sufficient statistic for $\theta = (\mu, \sigma^2)$.

Sufficient - Exponential Family

Theorem (Exponential family sufficient statistics)

Let $X_1, X_2, \ldots, X_n$ be a random sample (iid) that belongs to an exponential family. That is the pdf (or pmf) can be written in the form

$$f(x|\theta) = h(x)c(\theta)\exp\left\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\right\}.$$

Then $T(\boldsymbol{X}) = (\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j))$ is a sufficient statistic for $\theta = (\theta_1, \ldots, \theta_d)$, $d \leq k$.

Note: When $d < k$, it is called a curved exponential family, and when $k = 1$ it is called a one parameter exponential family.

Sufficient - Bernoulli

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim}$ Bernoulli(θ). Find a sufficient statistic for θ.

Sufficient - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where μ and σ^2 are unknown. Find a sufficient statistic for $\theta = (\mu, \sigma^2)$.

Sufficient - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta, \theta^2)$. Find a sufficient statistic for θ.

Rao-Blackwell

Theorem (Rao-Blackwell)

Let $T(\boldsymbol{X})$ be any unbiased estimator of θ and let $W(\boldsymbol{X})$ be a sufficient statistic for θ. Define $\phi(W) = E[T|W]$. Then $E\phi(W) = \theta$ and $Var(\phi(W)) \leq Var(T)$ for all θ.

Note: The previous theorem says that $\phi(W)$ is unbiased

$$E\phi(W) = E[E[T|W]] = E[T] = \theta$$

and is a uniformly better unbiased estimator than T since

$$\text{Var}(T) = \underbrace{\text{Var}(E[T|W])}_{=\text{Var}(\phi(W))} + \underbrace{E[\text{Var}(T|W)]}_{\geq 0} \geq \text{Var}(\phi(W)).$$

Rao-Blackwell - Bernoulli

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta)$.

Uniqueness of UMVUE

Theorem

If T is a UMVUE for $g(\theta)$, then T is unique.

Uniqueness of UMVUE

Complete

Definition (2.10)

Let $f(t|\theta)$ be a family of pdfs for a statistic $T(\boldsymbol{X})$. The family of pdfs is called **complete** if $E_\theta[g(T)] = 0$ for all θ implies $P(g(T) = 0) = 1$ for all θ. Equivalently, $T(\boldsymbol{X})$ is called a **complete statistic**.

Complete - Bernoulli

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim}$ Bernoulli(p) $0 < p < 1$. Recall that $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for p. Is T complete?

Complete - Uniform

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} U(0, \theta)$, $\theta \in (0, \infty)$. We have shown that $X_{(n)}$ is a sufficient statistic and is the MLE for θ. Is it complete?

Complete - Exponential Family

Theorem (Complete statistic in an exponential family)

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$ *(*θ *may be a vector). If* $f(x|\theta)$ *forms an exponential family given by*

$$f(x|\theta) = h(x)c(\theta)\exp\{\sum_{j=1}^{k} w_j(\theta)t_j(x)\}$$

where $\theta = (\theta_1, \ldots, \theta_d), d \leq k$*, then*
$T(\boldsymbol{X}) = (\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i))$ *is complete if*
$\{(w_1(\theta), \ldots, w_k(\theta)) : \theta \in \Theta)\}$ *contains an open set in* $\mathbb{R}^k$.

Complete - Bernoulli

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p) $0 < p < 1$. Find a complete statistic for θ.

Lehmann-Scheffe

Theorem

Let T be a complete and sufficient statistic for the parameter θ and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator (UMVUE) of its expected value. That is, if $E[\phi(T)] = \tau(\theta)$, then $\phi(T)$ is the UMVUE of $\tau(\theta)$.

UMVUE - Bernoulli

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ $0 < p < 1$. Find the UMVUE of p.

UMVUE - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find the UMVUE of (μ, σ^2).

UMVUE - Uniform

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$. Find the UMVUE of θ.

UMVUE - Binomial

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim}$ Binomial(k, θ), k known and $\theta \in (0,1)$, and let

$$\tau(\theta) = P(X_1 = 1) = k\theta(1-\theta)^{k-1}, k > 1.$$

Find the UMVUE of $\tau(\theta)$.

UMVUE - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$, $\theta \in \mathbb{R}$. $\bar{X}$ is the UMVUE of θ. What is the UMVUE of θ^2?

Unbiased Estimator of Zero and UMVUEs

Definition (2.11)

$U(\boldsymbol{X})$ is said to be an **unbiased estimator of 0** if $E_\theta[U(\boldsymbol{X})] = 0$ for all $\theta \in \Theta$.

Theorem

$W(\boldsymbol{X})$ is the UMVUE for $\tau(\theta)$ if and only if $W(\boldsymbol{X})$ is unbiased for $\tau(\theta)$ and for every unbiased estimator of 0, $U(\boldsymbol{X})$, we have

$$Cov_\theta(W(\boldsymbol{X}), U(\boldsymbol{X})) = E_\theta[W(\boldsymbol{X})U(\boldsymbol{X})] = 0, \quad \textit{for all } \theta \in \Theta.$$

That is $W(\boldsymbol{X})$ is uncorrelated with all unbiased estimators of 0, $U(\boldsymbol{X})$ for all $\theta \in \Theta$.

Proof

Proof

Proof

Non-existance of a UMVUE

Let $X \sim p(x|\theta) = p_\theta(x)$, $\theta \in \mathbb{Z}$, where

$$P_\theta(X = \theta) = P_\theta(X = \theta - 1) = P_\theta(X = \theta + 1) = \frac{1}{3}.$$

Show that there does not exists a UMVUE for θ.

Non-existance of a UMVUE

Non-existance of a UMVUE

Not a UMVUE - Uniform MoM Estimator

Let $X \sim U(\theta, \theta + 1)$, $\theta \in \mathbb{R}$. Is the method of moments estimator of θ

$$\widehat{\theta}_{MoM} = X - \frac{1}{2}$$

the UMVUE of θ?.

Minimal Sufficiency

We are looking for a sufficient statistic achieving the most data reduction but keeping all of the information about the parameter θ.

Definition

A sufficient statistic $T(\boldsymbol{X})$ is called a **minimal sufficient statistic** if and only if for any other sufficient statistic, $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$ or more explicitly if $T'(\boldsymbol{X}) = T'(\boldsymbol{Y})$ then $T(\boldsymbol{X}) = T(\boldsymbol{Y})$.

Theorem

Let $f(x|\theta)$ be the joint pdf of $\boldsymbol{X}$. Suppose there exists a function $T(\boldsymbol{X})$ such that for any two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$ the ratio $f(\boldsymbol{x}|\theta)/f(\boldsymbol{y}|\theta)$ is constant as a function of θ if and only if $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Then $T(\boldsymbol{X})$ is a minimal sufficient statistic for θ.

Minimal Sufficiency - Normal

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find a minimal suffcient statistic for (μ, σ^2).

Minimal Sufficiency

Theorem

(Bahadur's Theorem) If a minimal sufficient statistic exists, then any sufficient statistic that is complete is minimal sufficient.

Note: If T is a finite dimensional complete sufficient statistic, then it is minimal sufficient.

Theorem

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ denote an iid sample from a distribution with pdf $f(x|\theta)$, $\theta \in \Theta$. If a sufficient statistic $T(\boldsymbol{X})$ exists for θ and if the MLE, $\widehat{\theta}$, of θ exists uniquely, then $\widehat{\theta}$ is a function of $T(\boldsymbol{X})$. If $\widehat{\theta}_{MLE}$ exists uniquely and is sufficient for θ, then it must be minimal sufficient.

Example: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known and $\mu \in \mathbb{R}$. We know that $\bar{X}$ is the unique MLE for θ and $\bar{X}$ is sufficient. Therefore, $\bar{X}$ is minimal sufficient.

Ancillary

Definition

A statistic $S(\boldsymbol{X})$ is said to be **ancillary** for θ if the distribution of $S(\boldsymbol{X})$ does not depend on θ.

Note: An ancillary statistic contains no information about θ, but could be informative about θ in junction with other statistics.

Ancillary

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known. Show that

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is ancillary for μ.

Ancillary - Basu's Theorem

Theorem (Basu's Theorem)

A complete and minimal sufficient statistic is independent of any ancillary statistic.

Note: This theorem allows us to show that statistics are independent without needing to find their joint distribution.

Example: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known. We have already shown that $\bar{X}$ is complete and minimal sufficient for μ and S^2 is ancillary for μ, so by Basu's theorem $\bar{X}$ and S^2 are independent.

Ancillary - Basu's Theorem

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$, $\theta > 0$. Show that $X_{(n)}$ and $X_1 / \sum_{i=1}^{n} X_i$ are independent.

Let $\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$, $\theta \in \Theta$. In the Bayesian paradigm, θ is treated as a random variable and assigned a prior distribution $\theta \sim \pi(\theta)$. Inference is then based on the posterior distribution of $\theta|\boldsymbol{X}$,

$$\pi(\theta|\boldsymbol{X}) = \frac{f(x,\theta)}{\int_{-\infty}^{\infty} f(x,\theta)\, d\theta} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x,\theta)\, d\theta}.$$

If we observe $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, we update the distribution of θ base on the observed data $\boldsymbol{X}$. We call the mean of the posterior distribution of θ, $E[\theta|\boldsymbol{X}]$, the **Bayes estimate** of θ.

Bayes Estimator - Normal-Normal

Let $X_1, X_2, \ldots, X_n | \theta \overset{iid}{\sim} N(\theta, \sigma^2)$, σ^2 known. Using the prior $\theta \sim N(\mu, \tau^2)$, find the Bayes estimator of θ.

Bayes Estimator - Normal-Normal

Bayes Estimator - Beta-Binomial

Let $X|p \sim \text{Binomial}(k, p)$ k known. Let $p \sim \text{Beta}(a, b)$. Find the Bayes estimator of p.

Definition

Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$. A class Π of prior distributions is a conjugate family of $\mathcal{F}$ if the posterior distributions are in the class Π for all $f \in \mathcal{F}$, all priors in Π and all $x \in \mathcal{X}$.

Example: Show that the Gamma family is conjugate for the Poisson family.