Chapter 10

STAT 588

Chapter Summary

- ☐ Introduction to Point Estimation
- ☐ Unbiased Estimators
- ☐ Efficiency
- ☐ Sufficiency
- ☐ Method of Moments
- ☐ Method of Maximum Likelihood
- ☐ Bayesian Estimation (time permitting)

Introduction

- We are now going to start talking about inference
  - The process of making informed decisions regarding a population based on information from the sample
  - Two types:
    - Estimation (Chapters 10 & 11)
    - Hypothesis testing (Chapter 12)

Introduction

- There are two types of estimation
  - Point estimation (Chapter 10)
  - Interval estimation (Chapter 11)

DEFINITION 1. POINT ESTIMATION. *Using the value of a sample statistic to estimate the value of a population parameter is called* ***point estimation****. We refer to the value of the statistic as a* ***point estimate****.*

Introduction

- What makes an estimate good?
- How can we tell if one estimate is better than another?
- Some measures include:
  - Unbiased (has the right mean)
  - Efficiency (has the right variance)
    - Bias - Variance Trade-off
  - Sufficiency (has the right amount of information)

Unbiased Estimators

Unbiased Estimators

DEFINITION 2. UNBIASED ESTIMATOR. *A statistic $\hat{\Theta}$ is an* ***unbiased estimator*** *of the parameter θ of a given distribution if and only if $E(\hat{\Theta}) = \theta$ for all possible values of θ.*

- This result must be true for any/all sample sizes (value of n)
- There may be more than on unbiased estimator of a given parameter
  - i.e. unbiased estimators are NOT unique
- Not every parameter has an unbiased estimator

Example: Unbiased Estimator

- Consider a random sample $X_1, X_2, \ldots, X_n$ from a Bernoulli population with parameter θ
- Show that the sample proportion of successes is an unbiased estimator of θ

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

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{1}{n}(n\theta) = \theta$$

Example: Biased Estimator

- Consider a random sample $X_1, X_2, \ldots, X_n$ from a population with PDF given by

$$f(x) = \begin{cases} e^{-(x-\delta)}, & x > \delta \\ 0, & x \le \delta \end{cases}$$

- Show that the sample mean is a biased estimator of δ

$$E(X) = \int_\delta^\infty x e^{-(x-\delta)} dx = uv \Big|_\delta^\infty - \int_\delta^\infty v\, du$$

$u = x \quad dv = e^{-(x-\delta)} dx$

$du = dx \quad v = -e^{-(x-\delta)}$

$$= -xe^{-(x-\delta)} \Big|_\delta^\infty + \int_\delta^\infty e^{-(x-\delta)} dx$$

$$= 0 + \delta - \left[e^{-(x-\delta)}\right] \Big|_\delta^\infty = \delta + 1$$

Example: Biased Estimator

- Consider a random sample $X_1, X_2, \ldots, X_n$ from a population with PDF given by

$$f(x) = \begin{cases} e^{-(x-\delta)}, & x > \delta \\ 0, & x \leq \delta \end{cases}$$

$\mu = E(x) = \delta + 1$

- Show that the sample mean is a biased estimator of δ

From Theorem 8.1, $E(\bar{X}) = \mu = \delta + 1 \neq \delta$

Since $E(\bar{X}) \neq \delta$, the $\bar{X}$ is biased

NOTE: $\bar{X} - 1$ is unbiased since $E(\bar{X} - 1) = E(\bar{X}) - 1 = \delta + 1 - 1 = \delta$

Unbiased Estimators

DEFINITION 3. ASYMPTOTICALLY UNBIASED ESTIMATOR. *Letting* $\mathrm{b_n}(\theta) = \mathrm{E}(\hat{\Theta}) - \theta$ *express the* ***bias*** *of an estimator* $\hat{\Theta}$ *based on a random sample of size* n *from a given distribution, we say that* $\hat{\Theta}$ *is an* ***asymptotically unbiased estimator*** *of* θ *if and only if*

$$\lim_{n\to\infty} b_n(\theta) = 0$$

- ☐ This result only needs to be true as the sample size, n, gets larger

Example: Asymptotically Unbiased

- Consider a random sample from a normal population with mean μ and variance σ^2
- Show that $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$ is a biased estimator of σ^2

Recall from Chapter 8 that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

we need to find $E(\hat{\sigma}^2)$ using $E(S^2) = n-1$

NOTE: $\frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)}{\sigma^2} \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}$

Example: Asymptotically Unbiased

- Consider a random sample from a normal population with mean μ and variance σ^2
- Show that $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$ is a biased estimator of σ^2

$$E(\hat{\sigma}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = \frac{1}{n}E\left[(n-1)S^2\right]$$

$$= \frac{1}{n}E\left[\sigma^2\cdot\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n}E\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n}\cdot(n-1) = \frac{(n-1)}{n}\sigma^2 \neq \sigma^2$$

Example: Asymptotically Unbiased

- Consider a random sample from a normal population with mean μ and variance σ^2

- Show that $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$ is an asymptotically unbiased estimator of σ^2

$$b_n(\sigma^2) = E(\hat{\sigma}^2) - \sigma^2 = \frac{(n-1)}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

$$\lim_{n\to\infty} b_n(\sigma^2) = \lim_{n\to\infty} -\frac{\sigma^2}{n} = 0$$

Unbiased Estimators

THEOREM 1. If S^2 is the variance of a random sample from an infinite population with the finite variance σ^2, then $E(S^2) = \sigma^2$.

- So, the sample variance is an unbiased estimate of the population variance!
  - This is one of the reasons why we divide by n-1 instead of n when computing the sample variance

THEOREM 1. If S^2 is the variance of a random sample from an infinite population with the finite variance σ^2, then $E(S^2) = \sigma^2$.

- **Proof:** From Theorem 8.1, $E(\bar{X}) = \mu$ and $var(\bar{X}) = \frac{\sigma^2}{n}$

We also know that $E(X_i) = \mu$ and $var(X_i) = \sigma^2$ for $i = 1, 2, \ldots, n$

NOTE: $var(X_i) = E[(X_i - \mu)^2]$ and $var(\bar{X}) = E[(\bar{X} - \mu)^2]$

$$E(S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2\right]$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2\right] \longrightarrow$$

We had something like this in Chapter 8

THEOREM 1. If S^2 is the variance of a random sample from an infinite population with the finite variance σ^2, then $E(S^2) = \sigma^2$.

- **Proof:** $E(S^2) = \frac{1}{n-1} E\left(\left[\sum_{i=1}^{n} (x_i - \mu)^2 \right] - n(\bar{x} - \mu)^2 \right)$

$$= \frac{1}{n-1} \left(\sum_{i=1}^{n} E\left[(x_i - \mu)^2\right] - n E\left[(\bar{x} - \mu)^2\right] \right)$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} \mathrm{var}(x_i) - n\,\mathrm{var}(\bar{x}) \right]$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^{n} \sigma^2 - n \cdot \frac{\sigma^2}{n} \right) = \frac{1}{n-1} \left(n\sigma^2 - \sigma^2 \right)$$

$$= \frac{(n-1)\sigma^2}{(n-1)} = \sigma^2$$

Statistics is fun and exciting!

- ☐ **Fun with Determining Bias of Estimators**

★ From the Fun Sheet, we had to unbiased estimators $Y = 2X_1$ and $Z = 2\bar{X}$. Which one do you think is the better estimator?

Z will be "right" more often ... this is the idea of efficiency!

Efficiency

Efficiency

Definition 4. minimum variance unbiased estimator. *The estimator for the parameter* θ *of a given distribution that has the smallest variance of all unbiased estimators for* θ *is called the* ***minimum variance unbiased estimator****, or the* ***best unbiased estimator*** *for* θ.

- This makes intuitive sense …
  - If we have two unbiased estimators, we want to pick the one that is closer to the right value MORE OFTEN (or with less variability/ variance)

Efficiency

- Cramer-Rao Inequality

$$\mathrm{var}(\hat{\Theta}) \geqq \boxed{\frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}}$$

This is called the Cramer-Rao lower bound (CRLB)

- Used to find the most efficient estimators …

Efficiency

Theorem 2. If $\hat{\Theta}$ is an unbiased estimator of θ and

Starts out assuming we already know our estimator is unbiased

$$\operatorname{var}(\hat{\Theta}) = \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

then $\hat{\Theta}$ is a minimum variance unbiased estimator of θ.

- A.K.A. the Best Unbiased Estimator

Efficiency

Theorem 2. If $\hat{\Theta}$ is an unbiased estimator of θ and

$$\operatorname{var}(\hat{\Theta}) = \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

then $\hat{\Theta}$ is a minimum variance unbiased estimator of θ.

- Not all parameters have a Best Unbiased Estimator

Sometimes, NONE of the estimators attain the CRLB

Efficiency

THEOREM 2. If $\hat{\Theta}$ is an unbiased estimator of θ and

$$\operatorname{var}(\hat{\Theta}) = \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

then $\hat{\Theta}$ is a minimum variance unbiased estimator of θ.

- ☐ **If the Best Unbiased Estimator exists, then it is unique!**

Example: Best Unbiased Estimator

- Consider a random sample from a normal population with mean μ and variance σ^2
- Is the sample mean the best unbiased estimator of μ?

1. Check that the estimator is unbiased
2. Find the variance of the estimator
3. Find the CRLB
4. Check to see if 2 = 3

Example: Best Unbiased Estimator

- Consider a random sample from a normal population with mean μ and variance σ^2
- Is the sample mean the best unbiased estimator of μ?
  - STEP 1: Is the sample mean unbiased?

YES!

From Theorem 8.4, we know $E(\bar{X}) = \mu$ so $\bar{X}$ is unbiased.

Example: Best Unbiased Estimator

- Consider a random sample from a normal population with mean μ and variance σ^2
- Is the sample mean the best unbiased estimator of μ ?
  - STEP 2: Find the variance of the sample mean.

From Theorem 8.4, we know that $var(\overline{X}) = \frac{\sigma^2}{n}$

For a $N(\mu,\sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for $-\infty < x < \infty$

Example: Best Unbiased Estimator

$\ln f(x) = \ln\left(\sqrt{2\pi\sigma^2}\right) - \left[\frac{(x-\mu)^2}{2\sigma^2}\right]$

- Consider a random sample from a normal population with mean μ and variance σ^2
- Is the sample mean the best unbiased estimator of μ?
  - STEP 3: Find the CRLB

$\frac{\partial \ln f(x)}{\partial \mu} = -\frac{2(x-\mu)}{2\sigma^2} \cdot (-1)$

$= \frac{(x-\mu)}{\sigma^2} = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)$

$E\left[\left(\frac{\partial \ln f(x)}{\partial \mu}\right)^2\right] = E\left[\frac{1}{\sigma^2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$

$\left(\frac{x-\mu}{\sigma}\right) \sim N(0,1)$

$\left(\frac{x-\mu}{\sigma}\right)^2 \sim \chi^2_1$

Example: Best Unbiased Estimator

- Consider a random sample from a normal population with mean μ and variance σ^2
- Is the sample mean the best unbiased estimator of μ?
  - STEP 3: Find the CRLB

$$CRLB = \frac{1}{n\,E\left[\left(\frac{\partial \ln f(x)}{\partial \mu}\right)^2\right]}$$

$$E\left[\frac{1}{\sigma^2}\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_1 \qquad = \frac{1}{\sigma^2} \cdot 1 = \frac{1}{\sigma^2}$$

$$= \frac{1}{n\left(\frac{1}{\sigma^2}\right)} = \frac{\sigma^2}{n}$$

Example: Best Unbiased Estimator

- Consider a random sample from a normal population with mean μ and variance σ^2
- Is the sample mean the best unbiased estimator of μ?
  - STEP 3: Check to see if $var(\bar{X})$ is equal to the CRLB

$var(\bar{x}) = \frac{\sigma^2}{n}$

$CRLB = \frac{\sigma^2}{n}$

Yes $\bar{X}$ is the best unbiased estimator of μ!

Efficiency

- NOTE ABOUT MEAN SQUARED ERROR:
  - The efficiency results in the chapter assume we have an unbiased estimator
  - If we have a biased estimator, then the mean squared error (MSE) is used in the results in place of the variance of the estimator
    - MSE: $E[(\hat{\Theta} - \theta)^2] = var(\hat{\Theta}) + [b_n(\theta)]^2$

★ For fixed MSE, reducing variance leads to increased bias. So, to improve efficiency of biased estimators, we must reduce MSE!

Statistics is fun and exciting!

- Fun with Determining Efficiency of Estimators

Sufficiency

Sufficiency

- Let's start with an example
  - Suppose we are flipping a coin a large number of times
  - We don't know the true probability of getting heads, but we want to find it
  - The sequence of heads and tails flipped is H H T H T H H T T H T H T T T H T H H H T H H T H T H T T T
  - If I wanted to communicate this information to you, would I need to give you the entire string of heads and tails observed?

Sufficiency

- In the previous example, all we would need is the number of heads and the number of tails
- This is the idea of SUFFICIENCY
- It is a reduction/summary of the entire sample that preserves the amount of information necessary to investigate the parameter of interest

Sufficiency

DEFINITION 6. SUFFICIENT ESTIMATOR. *The statistic $\hat{\Theta}$ is a* ***sufficient estimator*** *of the parameter θ of a given distribution if and only if for each value of $\hat{\Theta}$ the conditional probability distribution or density of the random sample $X_1, X_2, \ldots, X_n$, given $\hat{\Theta} = \theta$, is independent of θ.*

- OK, great! But how do we find them in practice? …

Sufficiency

- Factorization Theorem:

THEOREM 4. The statistic $\hat{\Theta}$ is a sufficient estimator of the parameter θ if and only if the joint probability distribution or density of the random sample can be factored so that

$$f(x_1, x_2, \ldots, x_n; \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \ldots, x_n)$$

where $g(\hat{\theta}, \theta)$ depends only on $\hat{\theta}$ and θ, and $h(x_1, x_2, \ldots, x_n)$ does not depend on θ.

Example: Sufficient Statistic

- Consider a random sample from a Bernoulli population with parameter θ $f(x) = \theta^{x}(1-\theta)^{1-x}, \ x=0,1$
- Show that $\hat{\Theta} = \sum_{i=1}^{n} X_i$ is a sufficient estimator of θ
  - STEP 1: Find the joint PMF of the X's

$$f(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Example: Sufficient Statistic

- Consider a random sample from a Bernoulli population with parameter θ
- Show that $\hat{\Theta} = \sum_{i=1}^{n} X_i$ is a sufficient estimator of θ
  - STEP 2: Try to factor the joint PMF according to Theorem 4

$$f(x_1, x_2, \ldots, x_n; \theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} = \theta^{\hat{\theta}} (1-\theta)^{n-\hat{\theta}}$$

From Theorem 4, $g(\hat{\theta}, \theta) = \theta^{\hat{\theta}} (1-\theta)^{n-\hat{\theta}}$ and $h(x_1, x_2, \ldots, x_n) = 1$

Example: Sufficient Statistic

- Consider a random sample from a Bernoulli population with parameter θ
- Show that $\hat{\Theta} = \sum_{i=1}^{n} X_i$ is a sufficient estimator of θ

Since the joint PMF could be factored according to Theorem 4, then $\hat{\Theta}$ is a sufficient estimator for θ.

Example: Sufficient Statistic

- Consider a random sample from a Bernoulli population with parameter θ
- Show that $\hat{\Theta} = \sum_{i=1}^{n} X_i$ is a sufficient estimator of θ

★ In this example $h(x_1, x_2, \ldots, x_n) = 1$. This isn't always true!

★ In this example, you were given the statistic, and you just had to check Theorem 4. If you aren't given a statistic Theorem 4 can be used in the same way to find a sufficient estimator for a parameter!

Statistics is fun and exciting!

- Fun with Determining Sufficiency of Estimators

Now we’re going to transition …

- We have looked at different properties that make an estimator “good”
  - Unbiased
  - Efficient
  - Sufficient
  - etc.

Now we're going to transition …

- Now, we are going to look at ways to actually FIND the estimators in the first place
  - Method of Moments
  - Maximum Likelihood Estimates
  - These may or may not have "good" properties …
    we will find out!

Method of Moments

Method of Moments

- One of the ways we learned in Chapter 4 to find important parameters, like means and variances, was using MOMENTS
  - We learned about raw moments (moments about the origin)
    - Recall: The Kth raw moment was denoted by $\mu'_k = E(X^k)$
    - Recall: We can find raw moments using the MGF $\rightarrow$ if exists
  - We learned about central moments (moments about the mean)
- In an analogous manner, the Method of Moments (MOM) uses sample moments to find point estimates

  $\hookrightarrow$ for the population moments

Method of Moments

DEFINITION 7. SAMPLE MOMENTS. *The* ***kth sample moment*** *of a set of observations* $x_1, x_2, \ldots, x_n$ *is the mean of their* kth *powers and it is denoted by* m'_k; *symbolically,*

$$m'_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$

Method of Moments

- The Kth sample moment is an estimate of the Kth raw moment
- So, the MOM sets them equal to each other

$$m_k' = \mu_k' = E(X^k)$$

  But, the key is to write each raw moment in terms of the parameters you are trying to estimate!
- Then, you solve the system of equations back for your parameter to find the estimate in terms of the sample moments
- Typically, you will need the same number of sample moments as you have parameters you are trying to estimate

Example: MOM Estimator

- Consider a random sample from a continuous uniform population with $\beta = 1$
- Find the MOM estimator for α

$$\mu_1' = E(X) = \frac{\alpha+\beta}{2} = \frac{\alpha+1}{2}$$

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\bar{x} = \frac{\hat{\alpha}+1}{2}$$

$$2\bar{x} = \hat{\alpha} + 1$$

$$\hat{\alpha} = 2\bar{x} - 1$$

Example: MOM Estimator

- Consider a random sample from a gamma population
- Find the MOM estimators for the parameters α, β

$$\mu_1' = E(X) = \alpha\beta$$

$$\mu_2' = E(X^2) = var(X) + [E(X)]^2$$

$$= \alpha\beta^2 + (\alpha\beta)^2 = \alpha\beta^2 + \alpha^2\beta^2$$

$$m_1' = \bar{X}$$

$$\bar{X} = \hat{\alpha}\hat{\beta}$$

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}}$$

$$m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\mu_2' = \left(\frac{\bar{X}}{\hat{\beta}}\right)\hat{\beta}^2 + \left(\frac{\bar{X}}{\hat{\beta}}\right)^2\hat{\beta}^2$$

$$= \bar{X}\hat{\beta} + \bar{X}^2$$

Example: MOM Estimator

- Consider a random sample from a gamma population
- Find the MOM estimators for the parameters α, β

$$m_2' = \bar{X}\hat{\beta} + \bar{X}^2$$

$$\bar{X}\hat{\beta} = m_2' - \bar{X}^2$$

$$\hat{\beta} = \frac{m_2' - \bar{X}^2}{\bar{X}}$$

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} = \frac{\bar{X}^2}{m_2' - \bar{X}^2}$$

$$\boxed{\hat{\alpha} = \frac{(m_1')^2}{m_2' - (m_1')^2}, \quad \hat{\beta} = \frac{m_2' - (m_1')^2}{m_1'}}$$

Statistics is fun and exciting!

- ☐ Fun with Method of Moments Estimators

Method of Maximum Likelihood

Method of Maximum Likelihood

- In previous chapters, we always started out with a parameter value, and we looked at properties of a random sample
  - So the parameter was fixed and the sample was unknown
- Let's start with an example:
  - Suppose you have 2 coins that look identical, but one coin has $P(H) = 0.999$ and the other has $P(H) = 0.001$
  - You lose one of the coins
  - You flip the remaining coin a bunch of times to see which one you lost
  - In this case, the sample is fixed and the parameter is unknown

Method of Maximum Likelihood

- In this case, the sample is fixed and the parameter is unknown
- This is the idea of the likelihood function!
- We know that for a random sample $X_1, X_2, \ldots, X_n$ the X's are IID
- Because of the independence, we know the joint PMF/PDF of the X's can be found by multiplying together all of the marginals
- Because they are identically distributed, we know all the marginals are the same

Method of Maximum Likelihood

DEFINITION 8. MAXIMUM LIKELIHOOD ESTIMATOR. *If* $x_1, x_2, \ldots, x_n$ *are the values of a random sample from a population with the parameter* θ, *the* ***likelihood function*** *of the sample is given by*

$$L(\theta) = f(x_1, x_2, \ldots, x_n; \theta)$$

for values of θ *within a given domain. Here,* $f(x_1, x_2, \ldots, x_n; \theta)$ *is the value of the joint probability distribution or the joint probability density of the random variables* $X_1, X_2, \ldots, X_n$ *at* $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. *We refer to the value of* θ *that maximizes* $L(\theta)$ *as the* ***maximum likelihood estimator*** *of* θ.

Method of Maximum Likelihood

- Notes about the likelihood function
  - Now that we have observed values of our sample, we use lower case letters instead of uppercase letters (they are fixed, no longer random variables)
  - This is a function of the unknown parameter(s)

Example: Likelihood Function

- ☐ Consider a random sample from a Bernoulli population with parameter θ
- ☐ Find the likelihood function

$$f(x) = \theta^{x}(1-\theta)^{1-x}$$

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{x_1+x_2+\cdots+x_n}(1-\theta)^{(1-x_1)+(1-x_2)+\cdots+(1-x_n)}$$

$$L(\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Method of Maximum Likelihood

- Notes about maximizing the likelihood function
  - From calculus, we learned that we can find a point of maximum of a function by finding the first derivative and setting it equal to zero (find where the function switches from increasing to decreasing) (it is a good idea to check that you aren't getting a minimum instead, where the function switches from decreasing to increasing)
  - Fun Fact: the log-likelihood function has the same maximum point as the regular likelihood function, but is often easier to take derivatives of it
  - In statistics, we are going to first find the log-likelihood function and then find the maximum $l(\theta) = \ln L(\theta)$

Example: Log-Likelihood Function

- Consider a random sample from a Bernoulli population with parameter θ
- Find the log-likelihood function

$$L(\theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$l(\theta) = \ln L(\theta)$$

$$= \left(\sum_{i=1}^{n} x_i\right) \ln(\theta) + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-\theta)$$

Example: MLE

- Consider a random sample from a Bernoulli population with parameter θ
- Find the MLE of θ

$$l(\theta) = \left(\sum_{i=1}^{n} x_i\right) \ln(\theta) + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-\theta)$$

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{\left(\sum_{i=1}^{n} x_i\right)}{\theta} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-\theta} = 0$$

$$\frac{\left(\sum_{i=1}^{n} x_i\right)}{\hat{\theta}} = \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-\hat{\theta}} \longrightarrow (1-\hat{\theta})(n\bar{x}) = \hat{\theta}(n - n\bar{x})$$

$$\hat{\theta} n - \hat{\theta} n\bar{x} + \hat{\theta} n\bar{x} = n\bar{x}$$

$$\frac{\hat{\theta} n}{n} = \frac{n\bar{x}}{n}$$

$$\boxed{\hat{\theta} = \bar{x}}$$

Example: MLE

- Some issues with finding MLEs
  - Setting derivative equal to 0 finds both maxima and minima
    - Need to check the second derivative for concavity
    - Negative means concave down —> indicates a maximum
  - Setting derivative equal to zero may not result in an equation involving the parameter of interest that you can solve
    - Need to reason it out from the (log-) likelihood function …

Example: MLE

- Consider a random sample from a continuous uniform population with $\beta = 1$
- Find the MLE for α

$$f(x) = \frac{1}{\beta - \alpha} = \frac{1}{1-\alpha}, \quad \alpha < x < 1$$

$$L(\alpha) = \prod_{i=1}^{n} \frac{1}{1-\alpha} = \left(\frac{1}{1-\alpha}\right)^n$$

$$l(\alpha) = -n \ln(1-\alpha)$$

$$\frac{\partial l(\alpha)}{\partial \alpha} = \frac{n}{1-\alpha} = 0$$

$$\Rightarrow n = 0$$

☆ This is a nonsensical equation that gives no information on MLE $\hat{\alpha}$

Example: MLE

- Consider a random sample from a continuous uniform population with $\beta = 1$
- Find the MLE for α

★ Need to look at $L(\alpha)$ to find MLE for α...

$L(\alpha) = \left(\frac{1}{1-\alpha}\right)^n$ is largest when $(1-\alpha)$ is smallest

OR when α is largest

From $\alpha < x < \beta$, then α must be less than all the X's

So, the largest α can be is the smallest X

Therefore, $\hat{\alpha} = \min(X_1, X_2, \ldots, X_n)$

Statistics is fun and exciting!

- ☐ Fun with Maximum Likelihood Estimators

Bayesian Estimation

Bayesian Estimators

- In Bayesian estimation problems, the unknown parameter value is treated like a random variable
- Random variables have distributions associated with them
- Let's go back to our example with the 2 identical coins, one with P(H) = 0.999 and the other P(H) = 0.001
  - Suppose we just realized we lost one of the coins
  - Before we start flipping the remaining coin, we have some distribution associated with the unknown parameter value

Bayesian Estimators

- Prior distribution
  - This distribution is called a prior distribution because it has to be set prior to observing any data
  - Denoted $h(\theta)$ in our book
  - Denoted $\pi(\theta)$ in many other places
- Now, let's go ahead and flip the remaining coin to get some data
  - This sample will have an associated likelihood function

$$L(\theta) = f(x_1, x_2, \ldots, x_n; \theta)$$

Bayesian Estimators

- Posterior distribution
  - An updated version of the prior distribution after some data has been observed
  - Denoted in our book by $\varphi(\theta \mid x_1, x_2, \ldots, x_n) = \dfrac{h(\theta)f(x_1, x_2, \ldots, x_n \mid \theta)}{g(x_1, x_2, \ldots, x_n)}$
  - Denoted in many other places by $\pi(\theta \mid x)$
  - Can think of it like a conditional probability density for the parameter given the data/observations

Bayesian Estimators

- A note about the marginal distribution
  - In the posterior distribution $\varphi(\theta \mid x) = \dfrac{h(\theta)f(x \mid \theta)}{g(x)}$

    the marginal distribution of the data is given by g
  - You find the marginal distribution in the normal way

$$g(x) = \int_\theta h(\theta)f(x \mid \theta)\ d\theta$$

Bayesian Estimators

- Bayes Estimate
  - Since the posterior distribution for the parameter given the data is the most up-to-date information about the value of the parameter, then the Bayes Estimate is the expected value of the posterior distribution
  - Denoted by $E(\Theta \mid x)$
  - This is often called the "posterior mean"

Bayesian Estimators

THEOREM 5. If X is a binomial random variable and the prior distribution of Θ is a beta distribution with the parameters α and β, then the posterior distribution of Θ given $X = x$ is a beta distribution with the parameters $x + \alpha$ and $n - x + \beta$.

☆ The beta distribution is called the conjugate prior for binomial random variables because the posterior distribution is also a beta (but different parameters)

THEOREM 5. If X is a binomial random variable and the prior distribution of Θ is a beta distribution with the parameters α and β, then the posterior distribution of Θ given $X = x$ is a beta distribution with the parameters $x+\alpha$ and $n-x+\beta$.

- Proof:

$$f(x|\theta) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}, \quad h(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$x = 1, 2, \ldots, n \qquad 0 < \theta < 1$

$$g(x) = \int_0^1 h(\theta) f(x|\theta)\, d\theta = \frac{\binom{n}{x}}{B(\alpha,\beta)} \int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\, d\theta$$

$$= \frac{\binom{n}{x} B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)} \int_0^1 \frac{1}{B(x+\alpha, n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\, d\theta$$

PDF of Beta$(x+\alpha, n-x+\beta)$

THEOREM 5. If X is a binomial random variable and the prior distribution of Θ is a beta distribution with the parameters α and β, then the posterior distribution of Θ given $X = x$ is a beta distribution with the parameters $x+\alpha$ and $n-x+\beta$.

- Proof:

$$\varphi(\theta|x) = \frac{h(\theta) f(x|\theta)}{g(x)} = \frac{\dfrac{\binom{n}{x}}{B(\alpha,\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{\dfrac{\binom{n}{x}B(x+\alpha,n-x+\beta)}{B(\alpha,\beta)}}$$

$$= \frac{1}{B(x+\alpha,n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

☆ This is the PDF of a beta with parameters $x+\alpha$ and $n-x+\beta$!

Example: Bayesian Estimators

- ☐ Consider a single observation x from a Binomial population with parameters n (known) and θ (unknown)
- ☐ Let the prior distribution for θ be a Beta with parameters α, β
- ☐ Find the Bayesian estimator of θ

From Theorem 5, the posterior distribution is beta with parameters $x+\alpha$ and $n-x+\beta$. From the table, then

$$E(\theta|x) = \frac{(x+\alpha)}{(x+\alpha)+(n-x+\beta)} = \frac{x+\alpha}{n+\alpha+\beta}$$

Bayesian Estimators

Theorem 6. If $\overline{X}$ is the mean of a random sample of size n from a normal population with the known variance σ^2 and the prior distribution of M (capital Greek *mu*) is a normal distribution with the mean μ_0 and the variance σ_0^2, then the posterior distribution of M given $\overline{X} = \bar{x}$ is a normal distribution with the mean μ_1 and the variance σ_1^2, where

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{and} \quad \frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

Statistics is fun and exciting!

- Fun with Bayesian Estimators