Economics 403A

Likelihood Inference

Dr. Randall R. Rojas

Today's Class

- Introduction
- The Likelihood Function
- Maximum Likelihood Estimation
- Inferences Based on the MLE
  - Confidence Intervals
  - Tests of Hypotheses
- Distribution-Free Methods
  - Method of Moments
  - Bootstrapping

Introduction

- A point estimate is a reasonable value of a population parameter.
- $X_1, X_2, \ldots, X_n$ are random variables.
- Functions of these random variables (e.g., s^2), are also random variables called statistics.
- Statistics have their unique distributions which are called sampling distributions.
- A point estimate of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic. The statistic $\hat{\Theta}$ is called the point estimator.

Introduction

Wish List for a Good Estimate

- Unbiased
- Efficient
- Consistent
- Small Variance

The Likelihood Function

Introduction

- Example: Estimation by guessing: Suppose an urn contains 1 million marbles, a fraction of which are blue.
  - Denote the unknown fraction of blue marbles by π.
  - We draw 3 marbles at random from the urn and obtain: green(G), blue(B), green(G)
- Q: What is the probability of this (G B G) sequence?

The Likelihood Function

Introduction

- A: We can guess how likely the sequence G B G is for different values of π. Let L denote the probability of observing the sequence G B G, then
- If $\pi = 0.2 \rightarrow L = 0.8 \times 0.2 \times 0.8 = 0.128$
- If $\pi = 0.3 \rightarrow L = 0.7 \times 0.3 \times 0.7 = 0.147$
- If $\pi = 0.4 \rightarrow L = 0.6 \times 0.4 \times 0.6 = 0.144$
- If $\pi = 0.5 \rightarrow L = 0.5 \times 0.5 \times 0.5 = 0.125$
- If $\pi = 0.6 \rightarrow L = 0.4 \times 0.6 \times 0.4 = 0.096$
- If $\pi = 0.7 \rightarrow L = 0.3 \times 0.7 \times 0.3 = 0.063$
- If $\pi = 1/3 \rightarrow L = 2/3 \times 1/3 \times 2/3 =$ **0.148**

Which one would you choose for π?

Note: Likelihood inference about π is based on ordering

The Likelihood Function

Introduction

- Example: Analytical Estimate: Could we have derived π = 1/3 analytically instead of trial and error?

- Solution: $L(\pi) = (1-\pi)\pi(1-\pi) = \pi(1-\pi)^2$, using calculus, the max can be obtained by taking

$$d/d\pi\ (L) = 0$$

$$d/d\pi\ (L) = (3\pi - 1)(\pi - 1) = 0$$

$$\rightarrow \pi = 1,\ 1/3$$

This example illustrates that MLE is simply the best guess for π, and the procedure for obtaining it, the foundation of Maximum Likelihood Estimation.

The Likelihood Function

- Def: The likelihood function $\mathrm{L}(\bullet|x)$ is determined by the model and the data (x), where $\mathrm{L}(\theta|x)$ represents the probability of observing the data (x) given that θ is true. $\mathrm{L}(\theta|x)$ is referred to as the likelihood of θ.

The Likelihood Function

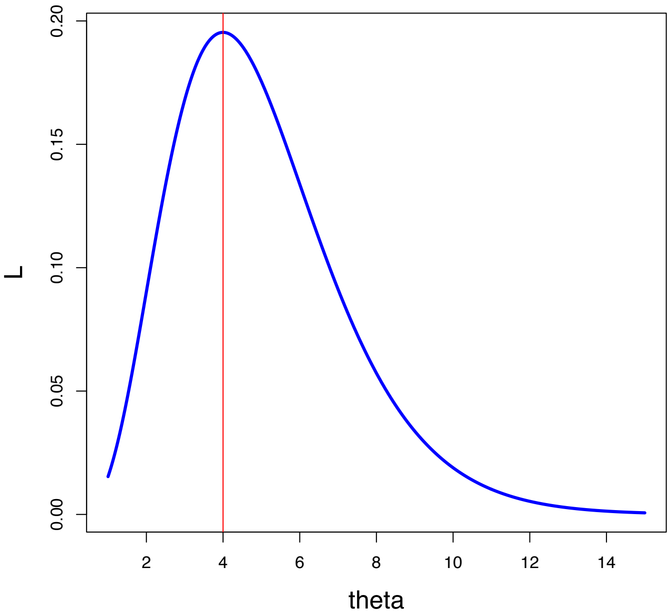
- Example (1 obs): Lets revisit the original insurance claims problem →Suppose that you work for an insurance company and historically they found that the number of claims they receive can be described by a Poisson distribution but the parameter value (i.e., θ) is unknown.
  - Recall that: $P(X = x|\theta) = \dfrac{\theta^x e^{-\theta}}{x!}$
- They assign you the task of figuring out the probability that 6 claims will be made tomorrow given that today 4 were received.
- Q: What can we learn from the respective likelihood function?

The Likelihood Function

- We can start by first plugin in the value $x = 4$ into $P(X=4|\theta) = (1/24)\ \theta^4\ e^{-\theta}$ → $L(\theta) = (1/24)\ \theta^4\ e^{-\theta}$

- There are many values of θ that we can choose, so instead, we can plot $L(\theta)$:

According to the plot, the value $\theta = 4$ is most likely to produce x =4



The Likelihood Function

- Example (2 observations): Same problem as before, but this time we observe X = 4 and X =2, i.e., $\vec{x} = (4, 2)$.

- We need to compute P(X=4|θ) and P(X=2|θ):

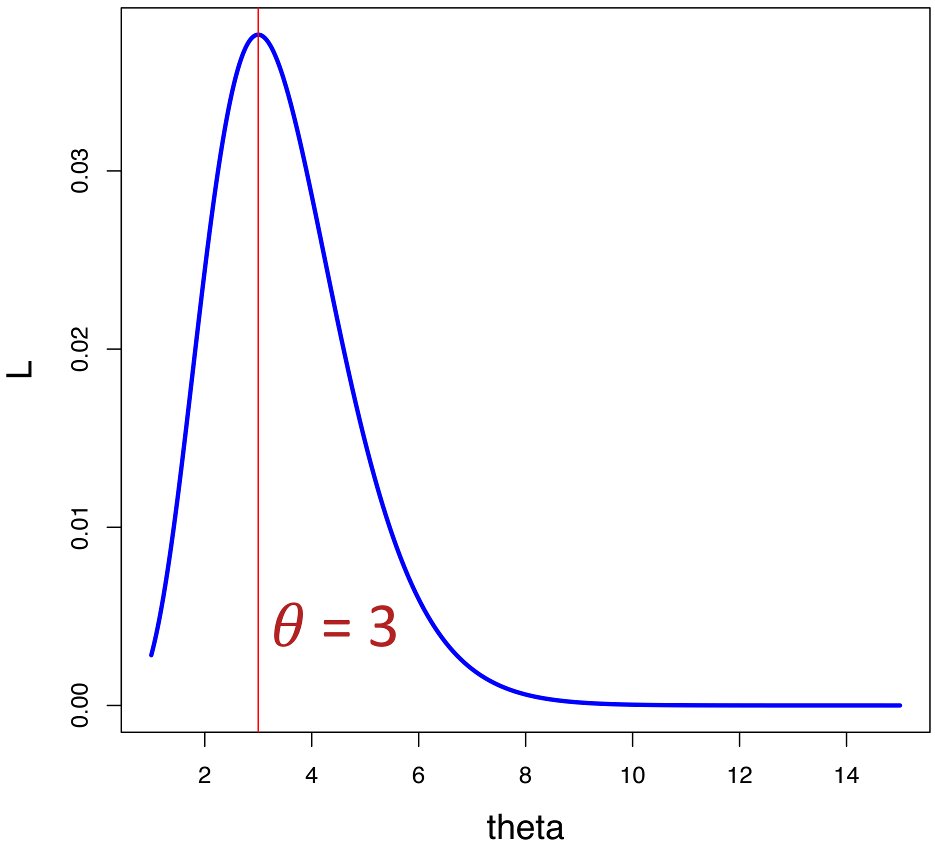
→ $P(X = 4|\theta) = \dfrac{\theta^4 e^{-\theta}}{4!}$ and $P(X = 2|\theta) = \dfrac{\theta^2 e^{-\theta}}{2!}$

→
$$
\begin{aligned}
P(\mathbf{X} = (4, 2)|\theta) &= P(X = 4|\theta) P(X = 2|\theta) \\
&= \left(\frac{\theta^4 e^{-\theta}}{4!}\right)\left(\frac{\theta^2 e^{-\theta}}{2!}\right) \\
&= \frac{\theta^6 e^{-2\theta}}{48} \qquad \rightarrow \boxed{L(\theta|(4,2)) = \frac{\theta^6 e^{-2\theta}}{48}}
\end{aligned}
$$

Can you guess the value of θ?

The Likelihood Function

According to our result, the value θ = 3 maximizes L(θ).



Q1: How is θ=3 related to our data (4,2)?

Q2: How would you generalize the previous result if you had $\boldsymbol{X}$ =(x_1,…, x_n) observations?

$$\rightarrow L(\theta | (x_1, \ldots, x_n)) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!}$$

Maximum Likelihood Estimation (MLE)

Introduction

- MLE consists of finding the parameter value(s) $\hat{\theta}$ that maximize the likelihood function L(θ), given the data.

- Note: Since MLE requires taking derivatives of products, for convenience we often take the log of L(θ) and differentiate a sum instead.

- Def: Log-Likelihood = $l(\theta) = \ln(L(\theta))$

Maximum Likelihood Estimation (MLE)

Steps for MLE

1. Find L(θ)
2. Compute $l(\theta) = \ln(\theta)$
3. Compute $S(\theta|x) \; = \; d/d\theta \; (l(\theta))$
   – Note: S(θ|x) is known as the Score Function
4. Solve the score equation $S(\theta|x) \;\; = \;\; 0$ for θ
   – The solution gives us the maximum likelihood estimator of θ, we call this $\hat{\theta}$
   – As a sanity check, we should verify that $d^2/d\theta^2 \, (l(\theta)) < 0$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ Poisson(θ), find the MLE of θ.

Step 1: Find $L(\theta)$ → Start with the PDF $P(X = x|\theta) = \dfrac{\theta^x e^{-\theta}}{x!}$

$$\rightarrow L(\theta) = \left(\frac{\theta^{x_1} e^{-\theta}}{x_1!}\right) \cdots \left(\frac{\theta^{x_n} e^{-\theta}}{x_n!}\right)$$

$$= \frac{\theta^{x_1 + \cdots + x_n} e^{-n\theta}}{x_1! \cdots x_n!}$$

$$= \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{x_1! \cdots x_n!}$$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ Poisson(θ), find the MLE of θ.

Step 2: Compute $l(\theta) = \ln(L(\theta)) \rightarrow$

$$\ln(L(\theta)) = \ln\left(\frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{x_1! \cdots x_n!}\right)$$

$$= \ln\left(\theta^{\sum_{i=1}^{n} x_i}\right) + \ln\left(e^{-n\theta}\right) - \ln\left(x_1! \cdots x_n!\right)$$

$$= \ln(\theta) \sum_{i=1}^{n} x_i - n\theta - \ln\left(x_1! \cdots x_n!\right)$$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ Poisson(θ), find the MLE of θ.

Step 3: Compute S(θ|x) →

$$S(\theta|x) = \frac{d}{d\theta}\ln(\theta)$$

$$= \frac{d}{d\theta}\left(\ln(\theta)\sum_{i=1}^{n} x_i - n\theta - \ln\left(x_1!\cdots x_n!\right)\right)$$

$$= \frac{1}{\theta}\sum_{i=1}^{n} x_i - n$$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ Poisson(θ), find the MLE of θ.

Step 4: Solve S(θ|x) = 0 →

$$\rightarrow \frac{1}{\theta}\sum_{i-1}^{n} x_i - n = 0$$

$$\rightarrow \theta = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\rightarrow \widehat{\theta} = \overline{x}$$

Note: Sanity check → $\frac{d^2}{d\theta^2} l(\theta) = -\frac{n}{\overline{x}} < 0$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ Exp(λ), find the MLE of λ.

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

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

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

$$\rightarrow \quad \widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{X}}$$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ N(μ, σ^2), find the MLE of μ and σ.

$$L\left(\mu,\sigma^2\right)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x_i-\mu)^2/\left(2\sigma^2\right)}$$

$$=\frac{1}{\left(2\pi\sigma^2\right)^{n/2}}e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ln L\left(\mu,\sigma^2\right)=\frac{-n}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

$$\frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)=0$$

$$\frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial\left(\sigma^2\right)}=\frac{-n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)^2=0$$

$$\hat{\mu}=\bar{X} \quad \text{and} \quad \hat{\sigma}^2=\frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n}$$

Maximum Likelihood Estimation (MLE)

- Example: Assume X ~ Weibull (δ,β), find the MLE of δ and β.

$$L(\delta,\beta)=\prod_{i=1}^{n} f(t)=\prod_{i=1}^{n}\frac{\beta}{\delta}\left(\frac{x_i}{\delta}\right)^{\beta-1} e^{-\left(\frac{x_i}{\delta}\right)^{\beta}}$$

$$\ln L\left(\delta,\beta\right)=n\ln\beta-\beta n\ln\delta+\sum_{i=1}^{n}(\beta-1)\ln x_i-\sum_{i=1}^{n}\left(\frac{x_i}{\delta}\right)^{\beta}$$

$$\frac{\partial \ln L\left(\delta,\beta\right)}{\partial\delta}=-\beta n+\frac{\beta}{\delta^{\beta}}\sum_{i=1}^{n}x_i^{\,\beta}=0$$

$$\frac{\partial \ln L\left(\delta,\beta\right)}{\partial\beta}=\frac{n}{\beta}-n\ln\delta+\sum_{i=1}^{n}\ln x_i-\sum_{i=1}^{n}\left(\frac{x_i}{\delta}\right)^{\beta}\ln\left(\frac{x_i}{\delta}\right)=0$$

$$\beta=\left[\frac{\sum_{i=1}^{n}x_i^{\beta}\ln x_i}{\sum_{i=1}^{n}x_i^{\beta}}-\frac{\sum_{i=1}^{n}\ln x_i}{n}\right]^{-1}$$

$$\delta=\left[\frac{\sum_{i=1}^{n}x_i^{\beta}}{n}\right]^{1/\beta}$$

Maximum Likelihood Estimation (MLE)

- Example (when it doesn't work): Assume X ~ Gamma (r,λ), find the MLE of r and λ.

$$\ln L(r,\lambda) = \ln\left(\prod_{i=1}^{n} \frac{\lambda^r x_i^{r-1} e^{-\lambda x_i}}{\Gamma(r)}\right)$$

$$= nr\ln(\lambda) + (r-1)\sum_{i=1}^{n}\ln(x_i) - n\ln\left[\Gamma(r)\right] - \lambda\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L(r,\lambda)}{\partial r} = n\ln(\lambda) + \sum_{i=1}^{n}\ln(x_i) - n\frac{\Gamma'(r)}{\Gamma(r)}$$

$$\frac{\partial \ln L(r,\lambda)}{\partial \lambda} = \frac{nr}{\lambda} - \sum_{i=1}^{n} x_i$$

$\widehat{\lambda} = \frac{\widehat{r}}{\overline{X}}$ (1) and $n\ln(\widehat{\lambda}) + \sum_{i=1}^{n}\ln(x_i) = n\frac{\Gamma'(\widehat{r})}{\Gamma(\widehat{r})}$ (2)

→ There is no closed form solution!

Maximum Likelihood Estimation (MLE)

- Textbook Example:

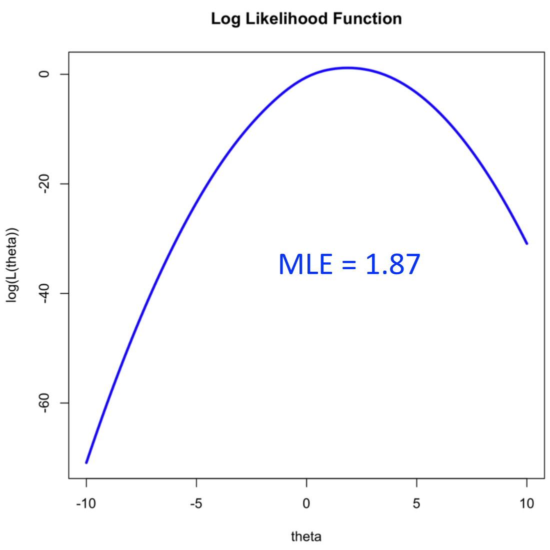
6.2.17 A likelihood function is given by $\exp(-(\theta-1)^2/2) + 3\exp(-(\theta-2)^2/2)$ for $\theta \in R^1$. Numerically approximate the MLE by evaluating this function at 1000 equispaced points in $(-10, 10]$. Also plot the likelihood function.

R Code:

```
L<-function(theta){
return((exp(-(theta-1)^2)/2)+3*exp(-((theta-2)^2)/2))
}
#Generate 1000 equispaced points
theta<-seq(-10, 10, length=1000)

#MLE
theta[which(L(theta)==max(L(theta)))]

#Plot:
plot(theta, log(L(theta)), type='l', main="Log Likelihood Function")
```



Inferences Based on the MLE

Standard Errors, Bias, and Consistency

- Q: Now that we have our MLE of θ, what do we do with it?
- A: Typically we use it to find functions of θ, such as the mean, median, variance, quartiles, etc. We denote them in general as $\psi(\theta)$.
- These functions represents characteristics of the underlying population that we wish to estimate, therefore, we denote such estimates as: $\widehat{\psi}(x) = \psi(\widehat{\theta}(x))$

Inferences Based on the MLE

Standard Errors, Bias, and Consistency

- Q: How reliable is $\widehat{\psi}(x) = \psi(\widehat{\theta}(x))$?
- A: We can look at the sampling distribution of $\widehat{\psi}(x)$ for every value of θ.
- Theorem: Let $\widehat{\theta}$ be an MLE of θ, and let g(θ) be a function of θ. Then an MLE of g(θ) is $g(\widehat{\theta})$.
- Def: The Mean Squared Error (MSE) of the estimator T(x) of $\psi(\theta)$ is given by $\mathrm{MSE}_\theta(T) = E_\theta[(T - \psi(\theta))^2]$.

- Note: In practice we instead evaluate $\mathrm{MSE}_{\widehat{\theta}(x)}(T)$

Inferences Based on the MLE

Standard Errors, Bias, and Consistency

- Def: The **bias** in the estimator T of $\psi(\theta)$ is given by $E_\theta(T) - \psi(\theta)$ whenever $E_\theta(T)$ exists. When the bias in an estimator T is 0 for every θ, we call T an ***unbiased estimator*** of ψ, i.e., T is unbiased whenever $E_\theta(T) = \psi(\theta)$.
- Note: $\mathrm{MSE}_{\widehat{\theta}(x)}(T) = \mathrm{Var}_{\widehat{\theta}(x)}(T)$ (for unbiased estimators) and the respective ***standard error*** is given by $\mathrm{SD}_{\widehat{\theta}(x)} = \sqrt{\mathrm{Var}_{\widehat{\theta}(x)}(T)}$
- Def: A sequence of estimates $T_1, T_2, \ldots$ is said to be ***consistent*** (in probability) for $\psi(\theta)$ if $T_n \xrightarrow{P_\theta} \psi(\theta)$ as $n \to \infty$ for every $\theta \in \Omega$.
  - As we increase the amount of data we collect, the sequence of estimates should converge to the true value of $\psi(\theta)$

Inferences Based on the MLE

- Let $\widehat{\Theta}$ be a point estimator of a parameter θ:

- Consistency: the more data we get, the sequence of estimates should converge to the true value of $\psi(\theta)$

- Bias of the estimator: $E[\widehat{\Theta}] - \theta$

- Efficiency: An estimator is efficient if it has the lowest possible variance among all unbiased estimators (MSE is the lowest)

- Variance of the estimator: $V(\widehat{\Theta}) = E\left[\widehat{\Theta} - E[\widehat{\Theta}]\right]^2$

- MSE of the estimator: $MSE(\widehat{\Theta}) = E\left[\widehat{\Theta} - \theta\right]^2$

$$= V(\widehat{\Theta}) + \left[E[\widehat{\Theta}] - \theta\right]^2$$

Inferences Based on the MLE

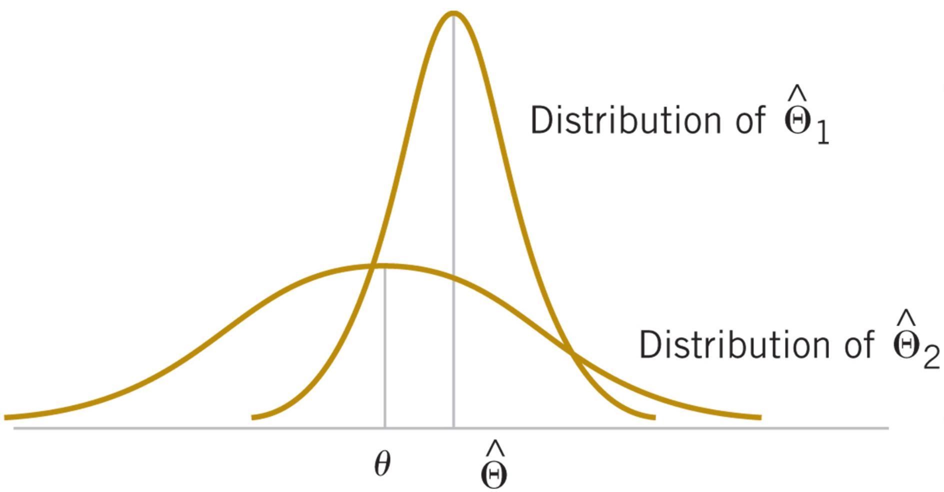
- The MSE is an important criterion for comparing two estimators.

Def: Relative Efficiency $= \dfrac{MSE(\widehat{\Theta}_1)}{MSE(\widehat{\Theta}_2)}$

- If the relative efficiency is less than 1, we conclude that the 1st estimator is superior than the 2nd estimator.

Inferences Based on the MLE

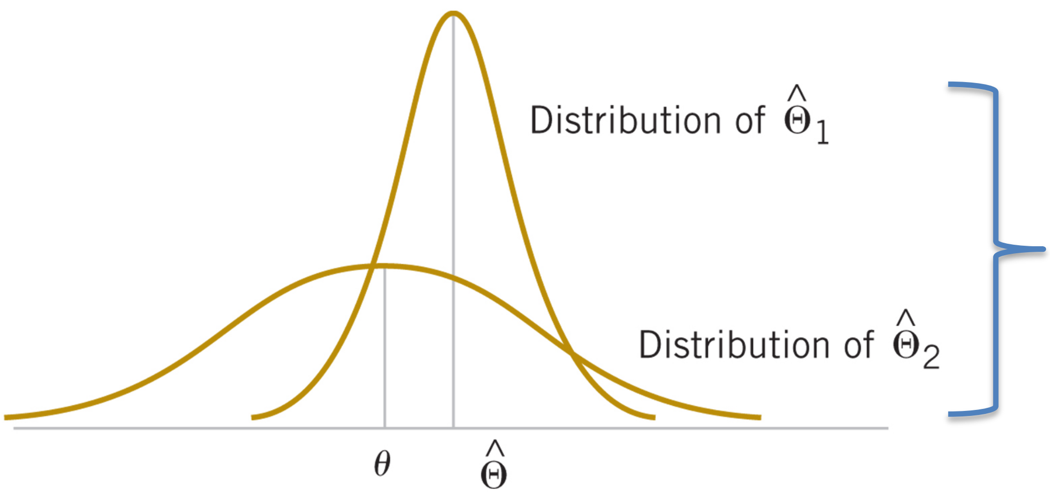
- A biased estimator can be preferred than an unbiased estimator if it has a smaller MSE.
- *Biased estimators are occasionally used in linear regression.*
- An estimator whose MSE is smaller than that of any other estimator is called an **optimal estimator.**



A biased estimator $\widehat{\Theta}_1$ that has a smaller variance than the unbiased estimator $\widehat{\Theta}_2$.

Inferences Based on the MLE

- A biased estimator can be preferred than an unbiased estimator if it has a smaller MSE.
- *Biased estimators are occasionally used in linear regression.*
- An estimator whose MSE is smaller than that of any other estimator is called an **optimal estimator.**



A biased estimator $\widehat{\Theta}_1$ that has a smaller variance than the unbiased estimator $\widehat{\Theta}_2$.

Inferences Based on the MLE

- Example: Sample Mean is Unbiased estimator of μ.

  X is a random variable with mean μ and variance σ^2.
  Let $X_1, X_2, \ldots, Xn$ be a random sample of size n.

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$$

$$= \frac{1}{n}[\mu + \mu + \ldots + \mu] = \frac{n\mu}{n} = \mu$$

Inferences Based on the MLE

- Example: Sample Variance (S^2) is Unbiased unbiased estimator of σ²

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

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

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\sigma^2/n\right)\right]$$

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

Distribution Free Methods

Advantages of MLE

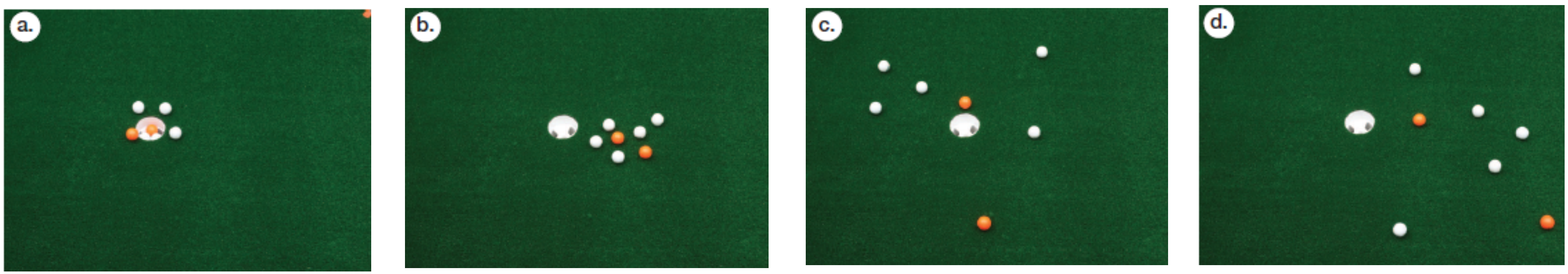
- Often yields good estimates, especially for large sample size.
- Invariance property of MLEs
- Asymptotic distribution of MLE is Normal.
- Most widely used estimation technique.
- Usually they are consistent estimators.

Disadvantages of MLE

- Requires that the PDF or PMF is known except the value of parameters.
- MLE may not exist or may not be unique.
- MLE may not be obtained explicitly (numerical or search methods may be required.). It is sensitive to the choice of starting values when using numerical estimation.
- MLEs can be heavily biased for small samples.

Inferences Based on the MLE

Accuracy and Precision



a. Both accurate and precise.
b. Precise but not accurate.
c. Accurate but not precise.
d. Neither accurate nor precise.

Inferences Based on the MLE

Accuracy and Precision, Bias and Standard Error

- **Bias** is a measure of the **accuracy**.
  - If only basketball players are measured to estimate the proportion of Americans who are taller than 6 feet, then there is a bias for a larger proportion.

- **Standard Error** is a measure of **precision**.
  - If the sample size is only three, the estimate of the proportion of tall people using the sample is likely to be far from the proportion of tall people in the US. The standard error will be large.

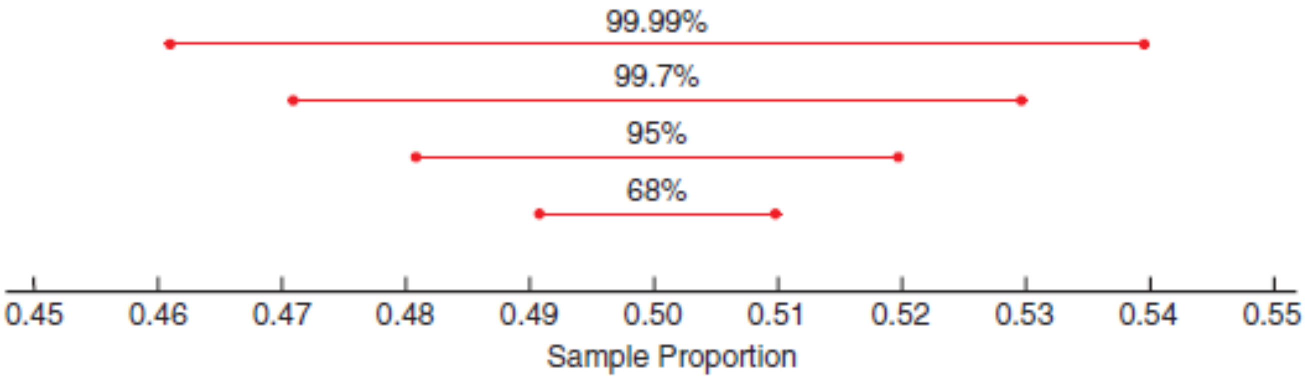
Inferences Based on the MLE

Confidence Intervals

- Q: Now that we have our $\psi(\theta)$, can we find a lower and an upper bound for it with some degree of confidence?
  - Note: Tradeoff between the length of the interval and our degree of confidence that the interval will contain the true value $\psi(\theta)$.
  - The length of the interval is a measure of how accurately the data allows us to know the true value of $\psi(\theta)$.

Inferences Based on the MLE

Confidence Intervals



- Increasing the level of confidence increases the margin of error.

- Decreasing the level of confidence decreases the margin of error.

Inferences Based on the MLE

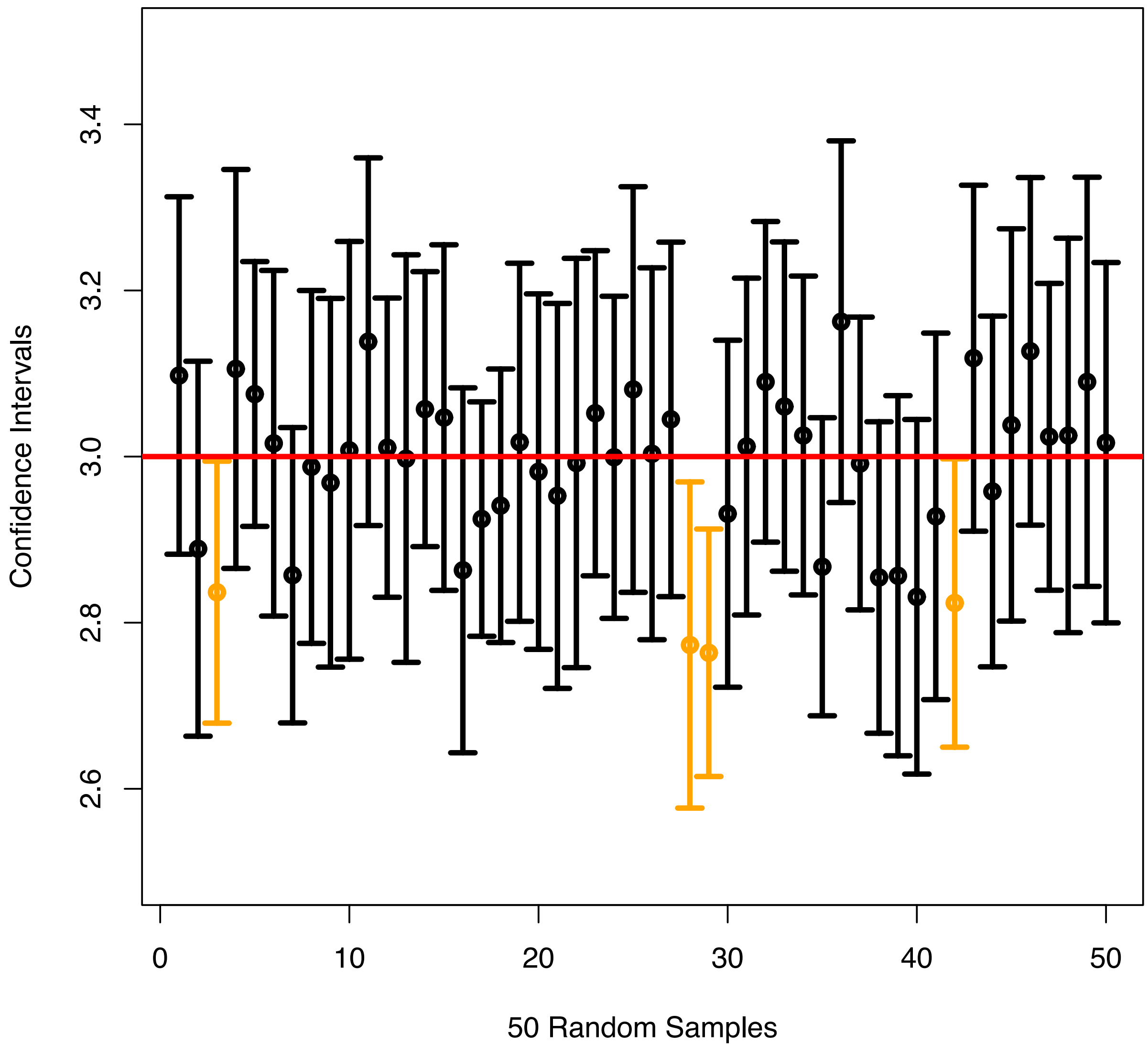
Confidence Interval Interpretation

(e.g., for the population proportion and 95% level of confidence)

- For every random sample that can be taken from a population there corresponds a 95% confidence interval. 95% of these confidence intervals will successfully contain the population proportion and 5% will not.

- **Q:** *Given a 95% CI for a population proportion from a particular sample, what is the probability that it contains the true population proportion?*

Inferences Based on the MLE



Inferences Based on the MLE

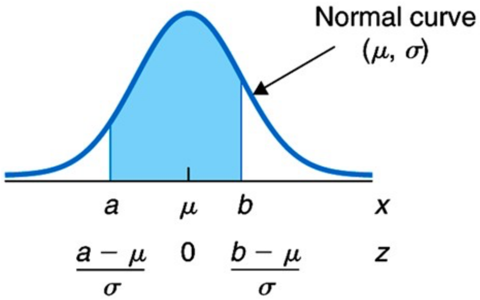
Confidence Intervals

- Def: An interval $C(x) = (l(x), u(x))$ is a γ-confidence interval for $\psi(\theta)$ if $P_\theta(\psi(\theta) \in C(x)) = P_\theta(l(x) \leq \psi(\theta) \leq u(x)) \geq \gamma$ for every $\theta \in \Omega$, where γ is referred to as the **level of confidence**.
- Popular values for γ are 0.99, 0.95, and 0.90
- Let α = 1- $\gamma \rightarrow$ $$CI = \widehat{\theta} \pm \underbrace{(t_{\nu,\alpha/2}, z_{\alpha/2})se(\widehat{\theta})}_{\text{Margin of Error = E}}$$

Inferences Based on the MLE

Confidence Intervals

- Example: $X \sim N(\mu,\sigma^2)$, and Φ = CDF



$$P(X \le a) = P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X > a) = P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = P\left(Z > \frac{a-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Inferences Based on the MLE

Confidence Intervals		
Parameter	**Assumptions**	**Endpoints**
μ	$\mathcal{N}(\mu,\sigma^2)$ or n large, σ^2 known	$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
μ	$\mathcal{N}(\mu,\sigma^2)$, σ^2 unknown	$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$
σ^2	$\mathcal{N}(\mu,\sigma^2)$	$(n-1)s^2/\chi^2_{\alpha/2,n-1}, (n-1)s^2/\chi^2_{1-\alpha/2,n-1}$
σ^2_X/σ^2_Y	$\mathcal{N}(\mu_X,\sigma^2_X)$, $\mathcal{N}(\mu_Y,\sigma^2_Y)$	$\frac{s^2_X/s^2_Y}{F_{\alpha/2,n-1,m-1}}, (s^2_X/s^2_Y)F_{\alpha/2,m-1,n-1}$
p	$b(n,p)$, n is large	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$p_1 - p_2$	$b(n_1,p_1)$, $b(n_2,p_2)$	$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
$\mu_X - \mu_Y$	$\mathcal{N}(\mu_X,\sigma^2_X)$, $\mathcal{N}(\mu_Y,\sigma^2_Y)$ $\sigma^2_X = \sigma^2_y$, known	$\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\sigma^2_X/n + \sigma^2_Y/m}$
$\mu_X - \mu_Y$	$\mathcal{N}(\mu_X,\sigma^2_X)$, $\mathcal{N}(\mu_Y,\sigma^2_Y)$ $\sigma^2_X = \sigma^2_y$, unknown	$\overline{x} - \overline{y} \pm t_{\alpha/2,n+m-2}\, S_p\sqrt{1/n+1/m}$ where $S_p = \sqrt{\frac{(n-1)s^2_x+(m-1)s^2_y}{n+m-2}}$

Comment: X has a sample size n and Y has a sample size m.

Inferences Based on the MLE

Tests of Hypotheses		
Parameter	**Assumptions**	**Critical Region**
$H_o : \mu = \mu_o$, $H_1 : \mu > \mu_o$	$\mathcal{N}(\mu, \sigma^2)$ or n large, σ^2 known	$z = \frac{\bar{x}-\mu_o}{\sigma/\sqrt{n}} \geq z_\alpha$
$H_o : \mu = \mu_o$, $H_1 : \mu > \mu_o$	$\mathcal{N}(\mu, \sigma^2)$, σ^2 unknown	$t = \frac{\bar{x}-\mu_o}{s/\sqrt{n}} \geq t_{\alpha, \nu-1}$
$H_o : p = p_o$, $H_1 : p > p_o$	$b(n, p)$, n large	$z = \frac{\hat{p}-p_o}{\sqrt{p_o(1-p_o)/n}} \geq z_\alpha$
$H_o : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o$	$\mathcal{N}(\mu, \sigma^2)$	$\chi^2 = \frac{(n-1)s^2}{\sigma_o^2} \geq \chi^2_{\alpha,n-1}$
$H_o : p_1 - p_2 = 0$, $H_1 : p_1 - p_2 > 0$	$b(n_1, p_1), b(n_2, p_2)$,	$z = \frac{\hat{p}_1-\hat{p}_2-0}{SE_p} \geq z_\alpha$ where $SE_p = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$, $\hat{p}$ = 'pooled proportion' and $\hat{p} = \frac{m_1+m_2}{n_1+n_2}$, $\hat{q} = 1 - \hat{p}$
$H_o : \mu_X - \mu_Y = 0$, $H_1 : \mu_X - \mu_Y > 0$	$\mathcal{N}(\mu_X, \sigma_X^2)$, σ_X^2 known $\mathcal{N}(\mu_Y, \sigma_Y^2)$, σ_Y^2 known	$z = \frac{\bar{x}-\bar{y}-0}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}} \geq z_\alpha$
$H_o : \mu_X - \mu_Y = 0$, $H_1 : \mu_X - \mu_Y > 0$	$\mathcal{N}(\mu_X, \sigma_X^2)$, $\mathcal{N}(\mu_Y, \sigma_Y^2)$ $\sigma_X^2 = \sigma_Y^2$, unknown	$t = \frac{\bar{x}-\bar{y}-0}{s_p\sqrt{1/n+1/m}} \geq t_{\alpha,n+m-2}$ $s_p = \sqrt{\frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}}$

Comment: X has a sample size n and Y has a sample size m.

Inferences Based on the MLE

Tests of Hypotheses

- A statistical hypothesis is a statement about the parameters of one or more populations.
- Hypothesis-testing procedures rely on using the information in a random sample from the population of interest.
- If this information is consistent with the hypothesis, then we will conclude that the hypothesis is **true**; if this information is inconsistent with the hypothesis, we will conclude that the hypothesis is **false**.

Inferences Based on the MLE

Tests of Hypotheses

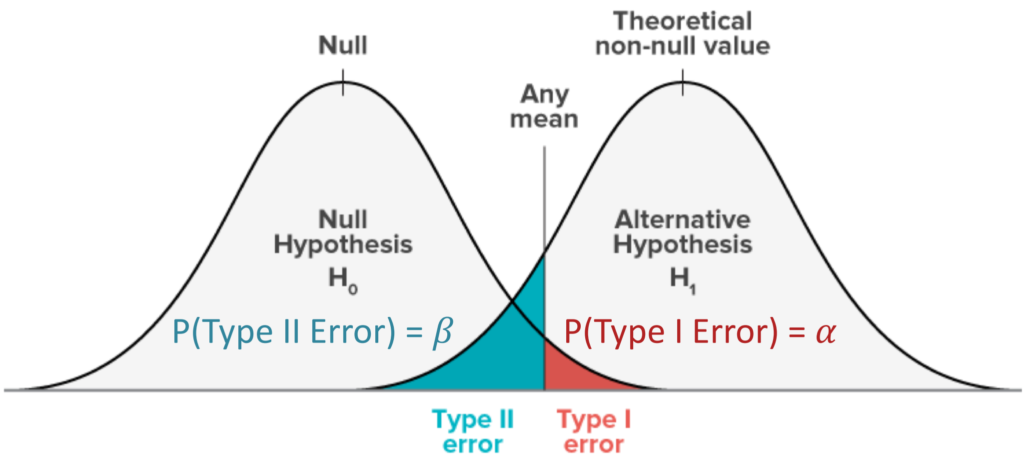
- The hypothesis that the characteristic of interest $\psi(\theta)= \psi_0$, is often referred to as the '**Null Hypothesis**', denoted as $\mathbf{H_0}$. This hypothesis is associated with a treatment having no effect.

- When the observations do not support H_0, we then favor the 'Alternative Hypothesis', denoted as $\mathbf{H_1}$.

Inferences Based on the MLE

Decisions in Hypothesis Testing

	Fail to reject the Null	Reject the Null
H_0 True	Correct	Type I Error Observe a difference when none exists
H_0 False	Type II Error Fail to observe a difference when one exists	Correct



Note: The type I error probability is called the significance level, or the α-error, or the size of the test.

Inferences Based on the MLE

Decisions in Hypothesis Testing

- The **power of a statistical** test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.
- The power is computed as **1 - β**, and power can be interpreted as the probability of correctly rejecting a false null hypothesis.
- The **P-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.
- P-value is the **observed significance level.**

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

- Two-sided hypothesis test on the mean:

  Test H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$

  The test-statistic (**known variance**) is: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

  Decision Rule: Reject H_0 if the observed value of the test statistic z_0 is either: $z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$

  →same as P-value < α

  Fail to reject H_0 if the observed value of the test statistic z_0 is $-z_{\alpha/2} < z_0 < z_{\alpha/2}$

  →same as P-value ≥ α

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

- Finding the **Probability of Type II Error β**
  - Suppose the null hypothesis is false and the true value of the mean is: $\mu = \mu_0 + \delta$, where $\delta > 0$.

- The test statistic Z_0 is: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{\overline{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \dfrac{\delta\sqrt{n}}{\sigma}$

- Hence, the distribution of Z_0 when H_1 is true is:

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right)$$

and finally, $\beta = \Phi\left(z_{\alpha/2} - \dfrac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \dfrac{\delta\sqrt{n}}{\sigma}\right)$

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

- Example: **(a)** Test the null that the mean hourly wage for recent MBA graduates is \$50. Assume α=0.05, σ=2. From a sample of 25 graduates we obtain $\overline{X} = 51.3$

Six Step Solution

1. Parameter of interest: μ, mean hourly wage for recent MBA graduates
2. Null hypothesis: $H_0 : \mu = 50$
3. Alternative Hypothesis: $H_1 : \mu \neq 50$
4. Test statistic: $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{51.3 - 50}{2/\sqrt{25}} = 3.25$

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

Seven Step Solution

5. Find the critical value $z_{\alpha/2}$: $z_{\alpha/2} = 1.96$

6. Conclusion:

(Method 1) Compare z_0 against $z_{\alpha/2}$:

$z_0 = 3.25 > z_{\alpha/2} = 1.96$ → **Reject H_0**

(Method 2) Compare P-value against α:

P-value = $2[1 - \Phi(3.25)] = 0.0012 < \alpha = 0.05$ → **Reject H_0**

Interpretation: The mean hourly wage for recent MBA graduates differs from $50, based on a sample of 25 individuals.

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

- Example: **(b)** Find the probability of a Type II error (β) for the two-sided test with $\alpha = 0.05$, assuming that μ = \$49. *Note:* Here $\delta = 1$ and $z_{\alpha/2} = 1.96$.

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

$$\rightarrow \beta = \Phi\left(1.96 - \frac{\sqrt{25}}{2}\right) - \Phi\left(-1.96 - \frac{\sqrt{25}}{2}\right)$$

$$= \Phi(-0.54) - \Phi(-4.46) = 0.295$$

$$\approx 0.30$$

  – The probability is about 0.3 that the test will fail to reject the null hypothesis when the true mean hourly wage for recent MBA graduates is \$49.

- **Interpretation:** A sample size of $n = 25$ results in reasonable, but not great power $= 1 - \beta = 1 - 0.3 = 0.70$.

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

- Finding the **Sample Size for a Test**

  - Sample Size for a Two-Sided Test

$$n \approx \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2}$$

  - Sample Size for a One-Sided Test

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\delta^2}$$

where $\delta = \mu - \mu_0$

Inferences Based on the MLE

Application: Two-sided hypothesis test on the mean with known variance

- Example: **(c)** Suppose that an analyst wishes to design the test so that if the true mean hourly wage for recent MBA graduates differs from \$50 by as much as \$1, the test will detect this (i.e., Reject H_0: $\mu = 50$) with a high probability, say, 0.90. How large should the sample size be?
  Note: $\sigma = 2$, $\delta = 51 - 50 = 1$, $\alpha = 0.05$, and $\beta = 0.10$.

- Since $z_{\alpha/2} = z_{0.025} = 1.96$ and $z_\beta = z_{0.10} = 1.28$, the sample size required to detect this departure from H_0: $\mu = 50$ is found by

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2} = \frac{(1.96 + 1.28)^2 2^2}{1^2} \approx 42$$

- The approximation is good here, since

$\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma) = \Phi(-5.20) \approx 0$, which is small relative to β.

- **Interpretation:** To achieve a much higher power of 0.90 we need a considerably large sample size, $n = 42$ instead of $n = 25$.

Distribution Free Methods

Method of Moments

- Let X_1, X_2,…,X_n be a random sample from the probability distribution $f(x)$ (PMF or PDF) with m unknown parameters θ_1, θ_2, …, θ_m.
  - The k^{th} **population moment** is $\mu_k = E[X^k]$, k = 1, 2, …
  - The k^{th} **sample moment** is $M_k = \left(\frac{1}{n}\right) \sum_{i=1}^{n} X_i^k$
- The moment estimators $\widehat{\Theta}_1, \widehat{\Theta}_2, \ldots, \widehat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments, and then solving the resulting simultaneous equations for the unknown parameters.

Distribution Free Methods

Method of Moments

- Example: Suppose that X_1, X_2, …, X_n is a random sample from a Gamma distribution with parameter r and λ where $E(X)$ = r/ λ and $E(X^2)$ = r(r+1)/ λ^2.

$$\frac{r}{\lambda} = E\left(X\right) = \bar{X} \text{ is the mean}$$

$$\frac{r}{\lambda^2} = E\left(X^2\right) - E\left(X\right)^2 \text{ is the variance or}$$

$$\frac{r\left(r+1\right)}{\lambda^2} = E\left(X^2\right) \text{ and now solving for } r \text{ and } \lambda:$$

$$\hat{r} = \frac{\bar{X}^2}{\left(1/n\right)\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

$$\hat{\lambda} = \frac{\bar{X}}{\left(1/n\right)\sum_{i=1}^{n} X_i^2 - \bar{X}^2}$$

Distribution Free Methods

Method of Moments

- Example: For the exponential distribution:

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0$$

Since E[X] = 1/λ,

$$\rightarrow \bar{X} = \frac{1}{\hat{\lambda}},\quad \hat{\lambda} = \frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Distribution Free Methods

Method of Moments

- Example: For the Pareto distribution:

$$f(x) = \frac{\beta}{x^{\beta+1}},\ x > 1 \quad \text{where} \quad \mu = \frac{\beta}{\beta - 1}$$

Since $M_1 = \overline{X}$, then $\mu_1 = M_1 \rightarrow \widehat{\beta} = \frac{\overline{X}}{\overline{X} - 1}$

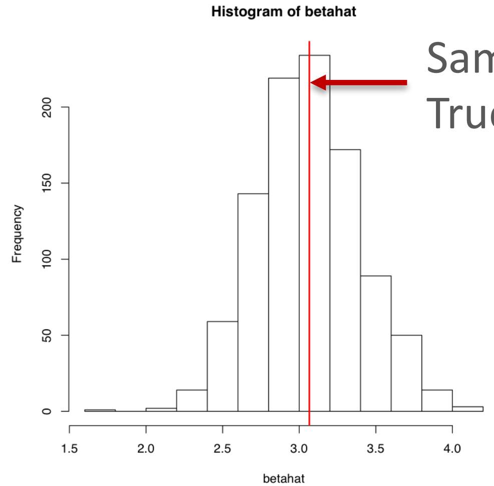
Distribution Free Methods

Method of Moments

- Example: For the Pareto distribution:

$$f(x) = \frac{\beta}{x^{\beta+1}}, \; x > 1 \quad \text{where} \quad \mu = \frac{\beta}{\beta - 1}$$

Since $M_1 = \overline{X}$, then $\mu_1 = M_1$ → $\widehat{\beta} = \dfrac{\overline{X}}{\overline{X} - 1}$



Sample mean beta = 3.05
True value: β = 3.0

Distribution Free Methods

Method of Moments

- Example (when it doesn't work): Consider the uniform distribution, where $X \sim U(a, b)$.

  $\mu_1 = E[X] = (a + b)/2$, and

  $\mu_2 = E[X^2] = (a^2 + ab + b^2)/3$,

  We can solve for (a, b) → $(a, b) = (\mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}, \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)})$

- Assume we obtain the sample {0,0,0,1} from a $U(0, 1)$. Our estimates for e.g., b, based on the sample moments would be: $\widehat{b}$ = 0.94. This suggests it would be impossible to have obtained this sample from $U(0, 1)$.

Distribution Free Methods

Advantages of the Method of Moments

- They are often simple to derive
- They are consistent estimators when $\theta_1, \ldots, \theta_k$ are continuous functions of $\mu_1, \ldots, \mu_k$.
- It is consistent
- They provide starting values in search for maximum likelihood estimates

Disadvantages of the Method of Moments

- They may not be unique in a given set of data. (Multiple solutions to set of equations) Also, they need not exist
- When they exist they are not necessarily constrained to fall in the parameter space. (Variance component estimation is an ex- ample of this situation.)
- They may be inefficient. Sometimes this is because they violate the *Sufficiency Principle*.

Distribution Free Methods

Bootstrapping

- There are many instances when we do not know the properties of the estimator of interest.
- Bootstrap is founded on the idea of performing a simulation on our data to get an estimate of the sampling distribution of the estimator.
- We repeatedly resample the data with replacement and calculate the estimate each time. The distribution of these bootstrap estimates approximates the sampling distribution of the estimate.

Distribution Free Methods

Bootstrapping

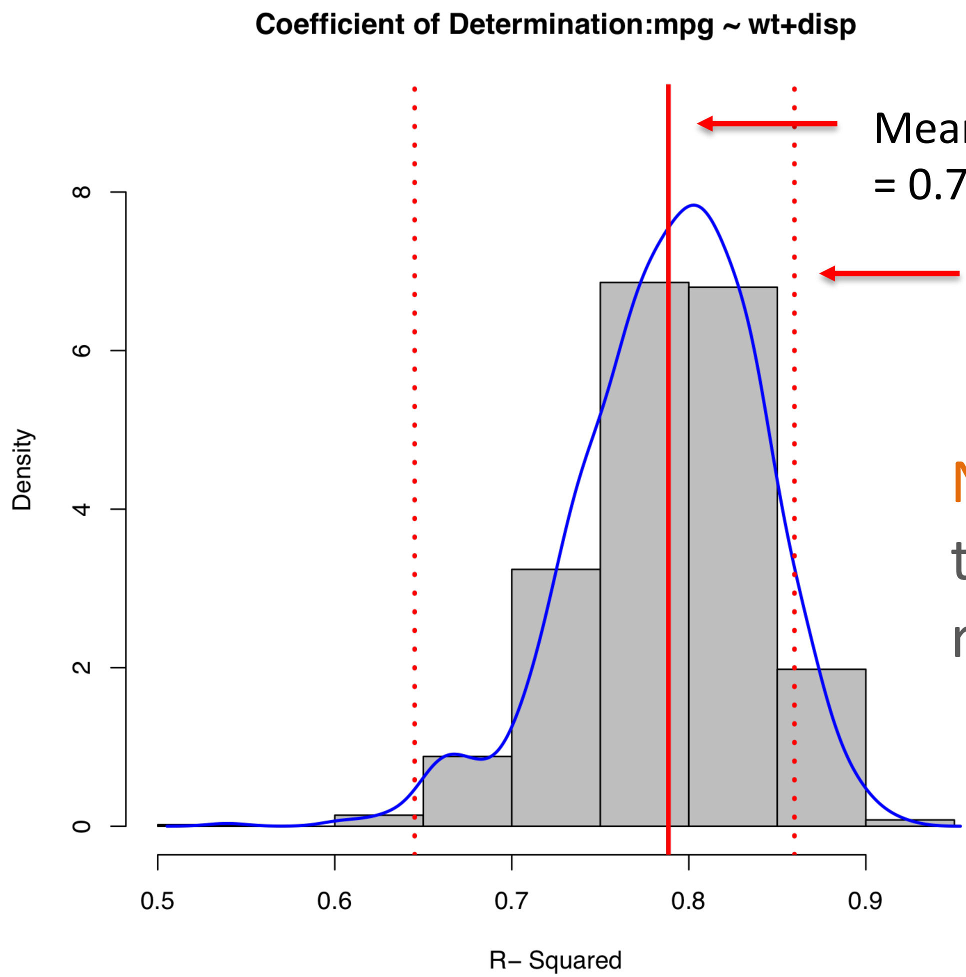
- The bootstrap mean, standard error, and confidence interval: Assume X_1, X_2,…, X_n is a sample from an unknow distribution with CDF Fθ, such that $E_\theta(\widehat{F}(x)) = F_\theta(x)$ and we are interested in estimating $\psi(\theta) = T(F_\theta)$.
- Mean: $\overline{\widehat{\psi}} = \frac{1}{m}\sum_{i=1}^{m}\widehat{\psi}_i$
- Standard Error: $se(\widehat{\psi}) = \sqrt{\widehat{\mathrm{Var}}_{\widehat{F}}(\widehat{\psi})} = \sqrt{\frac{1}{m-1}\left(\sum_{i=1}^{m}\widehat{\psi}_i^2 - \left(\frac{1}{m}\sum_{I=1}^{m}\widehat{\psi}_i\right)^2\right)}$
- Confidence Interval: $\widehat{\psi} \pm t_{(1+\gamma)/2,n-1}\sqrt{\widehat{\mathrm{Var}}_{\widehat{F}}(\widehat{\psi})}$

Distribution Free Methods

Bootstrapping

- Example: Find the distribution of R^2 from the regression: mpg $=\beta_1$ disp + β_2 weight + e
  - Sample 1000 (with replacement)from the original dataset
  - Run the regression, and store R^2 for each one
  - Plot the histogram of R^2's, compute the mean and respective confidence interval across all 1000 estimates of R^2

Distribution Free Methods



Mean from 1000 bootstrapped R^2's = 0.788

95% Confidence Intervals = [0.645, 0.859]

Note: We can easily perform the same estimates for the regression coefficients.

Note: Use the R library 'boot'