2012 S. T. Yau College Math Contests Oral Exam on Statistics

Saturday, August 4, morning

Problem 1. Let $X_1, \cdots, X_n$ be n independent and identically distributed observations from the exponential distribution with density function $f(x) = \frac{1}{\beta} e^{-x/\beta}$, $x \geq 0$.

a) Let T be an unbiased estimator of the scale parameter β. Prove that

$$\mathrm{Var}(T) \geq \frac{\beta^2}{n}.$$

b) Can you find an unbiased estimator T that attains the lower bound in part a)? If yes, please construct one. If no, please show why such an estimator does not exist.

Solutions to 2012 S. T. Yau College Math Contests Oral Exam on Statistics

Saturday, August 4, morning

Problem 1. Let $X_1, \cdots, X_n$ be n independent and identically distributed observations from the exponential distribution with density function $f(x) = \frac{1}{\beta} e^{-x/\beta}$, $x \geq 0$.

a) Let T be an unbiased estimator of the scale parameter β. Prove that

$$\text{Var}(T) \geq \frac{\beta^2}{n}.$$

Solution: The above lower bound on the variance of an unbiased estimator T of the scale parameter β is given by the Cramér-Rao bound $1/I(\beta)$. The log-likelihood function is

$$\ell(\beta) = \sum_{i=1}^{n} \left\{ -\log \beta - \frac{X_i}{\beta} \right\},$$

which leads to

$$\ell'(\beta) = \sum_{i=1}^{n} \left\{ -\frac{1}{\beta} + \frac{X_i}{\beta^2} \right\} \quad \text{and} \quad \ell''(\beta) = \sum_{i=1}^{n} \left\{ \frac{1}{\beta^2} - \frac{2X_i}{\beta^3} \right\}.$$

Thus the Fisher information is

$$I(\beta) = -E\ell''(\beta) = \frac{n}{\beta^2}.$$

b) Can you find an unbiased estimator T that attains the lower bound in part a)? If yes, please construct one. If no, please show why such an estimator does not exist.

Solution: The answer is yes. The maximum likelihood estimator $\widehat{\beta}$, which solves the score equation $\ell'(\beta) = 0$, is identical to the sample mean $\frac{1}{n} \sum_{i=1}^{n} X_i$. It is easy to show that such an estimator is unbiased and attains the lowest variance.

2012 S. T. Yau College Math Contests Oral Exam on Statistics

Saturday, August 4, afternoon

Problem 1. Let $X_1, \cdots, X_n$ be n independent and identically distributed observations from the Cauchy distribution with density function $f(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$, $x \in \mathbb{R}$.

a) Let T be an unbiased estimator of the location parameter θ. Prove that

$$\mathrm{Var}(T) \geq \frac{2}{n}.$$

b) Can you find an unbiased estimator T that attains the lower bound in part a)? If yes, please construct one. If no, please show why such an estimator does not exist.

c) Can you provide an estimator T that can attain the lower bound on $\mathrm{Var}(T)$ in part a) asymptotically, by removing the constraint of unbiasedness?

Solutions to 2012 S. T. Yau College Math Contests Oral Exam on Statistics

Saturday, August 4, afternoon

Problem 1. Let $X_1, \cdots, X_n$ be n independent and identically distributed observations from the Cauchy distribution with density function $f(x) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$, $x \in \mathbb{R}$.

a) Let T be an unbiased estimator of the location parameter θ. Prove that

$$\mathrm{Var}(T) \geq \frac{2}{n}.$$

Solution: The above lower bound on the variance of an unbiased estimator T of the location parameter θ is given by the Cramér-Rao bound $1/I(\theta)$. The log-likelihood function is

$$\ell(\theta) = \sum_{i=1}^{n} \left\{ -\log \pi - \log \left[1 + (X_i - \theta)^2 \right] \right\},$$

which leads to

$$\ell'(\theta) = \sum_{i=1}^{n} \frac{2(X_i - \theta)}{1 + (X_i - \theta)^2} \quad \text{and} \quad \ell''(\theta) = \sum_{i=1}^{n} \frac{-2 + 2(X_i - \theta)^2}{[1 + (X_i - \theta)^2]^2}.$$

Thus the Fisher information is

$$I(\theta) = -E\ell''(\theta) = \frac{n}{2}.$$

b) Can you find an unbiased estimator T that attains the lower bound in part a)? If yes, please construct one. If no, please show why such an estimator does not exist.

Solution: The answer is no. From the proof of the Cramér-Rao theorem, we see that the above lower bound on variance can be attained only if the following Cauchy-Schwarz inequality becomes an equation

$$\left(E \left\{ \ell'(\theta)(T - \theta) \right\} \right)^2 \leq E \left\{ \ell'(\theta) \right\}^2 E(T - \theta)^2.$$

It is well known that the equation holds only when

$$T - \theta = (\text{some constant}) \cdot \ell'(\theta) = (\text{some constant}) \cdot \sum_{i=1}^{n} \frac{2(X_i - \theta)}{1 + (X_i - \theta)^2},$$

which entails that

$$T = \theta + (\text{some constant}) \cdot \sum_{i=1}^{n} \frac{2(X_i - \theta)}{1 + (X_i - \theta)^2}.$$

The above representation shows that such an "optimal" estimator T should always depend on the location parameter θ, which cannot be an estimator in the first place.

c) Can you provide an estimator T that can attain the lower bound on $\mathrm{Var}(T)$ in part a) asymptotically, by removing the constraint of unbiasedness?

Solution: The answer is yes by the classical asymptotic theory of the maximum likelihood estimator (MLE). The MLE $\widehat{\theta}$, which solves the score equation $\ell'(\theta) = 0$, is known to be asymptotically normal with mean θ and variance $1/I(\theta) = \frac{2}{n}$.

2012 S. T. Yau College Math Contests Oral Exam on Statistics

Sunday, August 5, morning

Problem 1. Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_1, \cdots, y_n)^T$ is an n-dimensional vector of response, $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_p)$ is an $n \times p$ design matrix, $\boldsymbol{\beta}_0 = (\beta_{0,1}, \cdots, \beta_{0,p})^T$ is a p-dimensional vector of regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T$ is an n-dimensional vector of independent and identically distributed noise with mean 0 and variance σ^2. It is well known that the ordinary least-squares estimator becomes unstable or even inapplicable when p is large compared to n. One idea for remedying this issue is the ridge regression which gives the ridge estimator

$$\widehat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^T\mathbf{y},$$

where $\lambda > 0$ is called the ridge parameter.

a) Calculate the mean of $\widehat{\boldsymbol{\beta}}_{\text{ridge}}$.

b) Calculate the covariance matrix of $\widehat{\boldsymbol{\beta}}_{\text{ridge}}$.

Solutions to 2012 S. T. Yau College Math Contests Oral Exam on Statistics

Sunday, August 5, morning

Problem 1. Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_1, \cdots, y_n)^T$ is an n-dimensional vector of response, $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_p)$ is an $n \times p$ design matrix, $\boldsymbol{\beta}_0 = (\beta_{0,1}, \cdots, \beta_{0,p})^T$ is a p-dimensional vector of regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T$ is an n-dimensional vector of independent and identically distributed noise with mean 0 and variance σ^2. It is well known that the ordinary least-squares estimator becomes unstable or even inapplicable when p is large compared to n. One idea for remedying this issue is the ridge regression which gives the ridge estimator

$$\widehat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^T\mathbf{y},$$

where $\lambda > 0$ is called the ridge parameter.

a) Calculate the mean of $\widehat{\boldsymbol{\beta}}_{\text{ridge}}$.

Solution:

$$E\widehat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0 - \lambda(\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\boldsymbol{\beta}_0.$$

b) Calculate the covariance matrix of $\widehat{\boldsymbol{\beta}}_{\text{ridge}}$.

Solution:

$$\begin{aligned}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{\text{ridge}}) &= (\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^T\mathrm{Cov}(\mathbf{y})\mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\\ &= \sigma^2(\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda I_p)^{-1}.\end{aligned}$$

2012 S. T. Yau College Math Contests Oral Exam on Statistics

Sunday, August 5, afternoon

Problem 1. Let $X_i \sim N(\theta_i, \frac{1}{n})$, $i = 1, \cdots, n$, be independent. Find an estimator $\widehat{T}$ of $T = \sum_{i=1}^{n} \theta_i^2$ and calculate $E(\widehat{T} - T)^2$.