STA355
UTSG

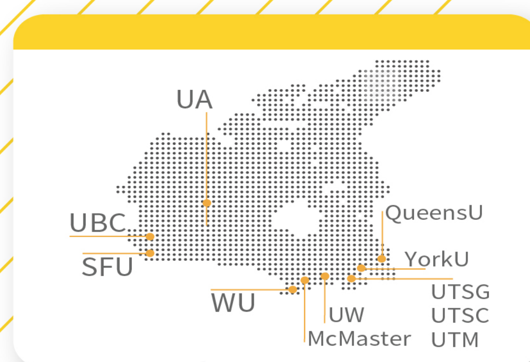
Midterm Review 2

EASY GROUP 易维集团 | Powered by Easy Group EASY EDU 易维集团旗下公司

关于Easy Edu

Easy Education Inc（易途教育）为Easy Group旗下品牌，是加拿大最专业，最具规模及影响力的华人教育培训机构，在多伦多、滑铁卢、温哥华、阿尔伯塔等地均设有培训基地。

截至目前共开设**十一大校区**，提供近三百门大学科目培训辅导。自2014年成立至今，帮助过数万名海外留学生度过学术难关。



累计线下学员 **10万+**

线上学员人数 **16万+**

线上关注人数 **17万+**

想要获得更多加拿大留学生相关资讯
还请大家**关注**我们哦

这里不仅仅有**超多有趣**的小故事
还有 **许多干货**和**最新发布** 的一手消息
涵盖全加拿大**吃喝玩乐学**
让你再也不用费力查找资讯

不仅如此，我们也会不定期放出**超多福利**
甭管你是本科在读、找工作、申研
在**Easy Edu**都能获得最专业的帮助

Easy Edu
为大家的成长、学习保驾护航
在这里，**你永远不是一座孤岛**

4.0

商务合作洽谈 请联系

Disclaimer

This complementary study package is provided by Easy Education Inc. and its affiliated mentors. This study package seeks to support your study process and should be used as a complement, **NOT** substitute to course material, lecture notes, problem sets, past tests and other available resources.

We acknowledge that this package contains some materials provided by professors and staff of the University of Toronto, and the sources of these materials are cited in details wherever they appear.

This package is distributed for free to students participating in Easy Education's review seminars, and are not for sale or other commercial uses whatsoever. We kindly ask you to refrain from copying or selling in part or in whole any information provided in this package.

Thank you for choosing Easy Education. We sincerely wish you the best of luck in all of your exams.

Easy Education Inc.

今日内容:

1. 教授 practice problem
2. Past Tests

Practice Problem

Suppose that X is a positive continuous random variable with cdf $F(x)$ and pdf $f(x)$. The Lorenz curve is defined as $\mathcal{L}_F(t) = \frac{1}{E_F(X)} \int_0^t F^{-1}(s) ds$.

(a) Suppose you have a sample $X_1, \dots, X_n$ from F where F is unknown. Find an estimator of $\theta = \mathcal{L}_F(0.8)$.

Solution:

EASY EDU

(b) Let $E_F(X) = 4$ and $\mathcal{L}_F(t) = t^3$. Find the median of F.

Solution:

Practice Problem

Let X be a continuous random variable with cdf

$$F(x) = \begin{cases} 0 & x < \theta \\ \dfrac{x}{\theta} & 0 < x < \theta \\ 1 & x > \theta \end{cases}$$

Compute the expression of the Lorenz curve $\mathcal{L}_F(p) = \frac{1}{E_F(X)} \int_0^p F^{-1}(t) dt$.

Solution:

Practice Problem

Let $X_1, X_2, \cdots$ be a sequence of independent random variables with cdf

$$F_{X_n}(x) = \begin{cases} \left(1 - \dfrac{1}{1+nx}\right)^n & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the limiting distribution of X_n.

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_n$ are independent random variables with pdf

$$f(x) = \begin{cases} \frac{1}{\theta} & \theta < x < 2\theta \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$.

(a) Find the method of moments estimator of θ.

Solution:

(b) Show that $\frac{X_{(n)}-\theta}{\theta}$ is a pivot for θ where $X_{(n)} = \max\{X_1, \cdots, X_n\}$.

Solution:

Practice Problem

Suppose that $X_1, X_2, \cdots$ is a sequence of independent random variables with mean μ and variance $\sigma^2 < \infty$; define $\bar{X}_n = n^{-1}(X_1 + \cdots + X_n)$. Describe the limiting behaviour (that is, either convergence in probability or convergence in distribution as well as the limit as $n \to \infty$) of the following random variables.

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

Solution:

(b) $\sqrt{n}(\bar{X}_n - \mu)/S_n$.

Solution:

(c) $\sqrt{n}(\exp(\bar{X}_n) - \exp(\mu))/S_n$.

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_n$ are independent positive random variables whose hazard function is

$$h(x) = \begin{cases} x^{-1} & \text{for } x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

and define $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ to be the order statistics.

(a) Show that the τ quantile of each X_i is $(1-\tau)^{-1}$.

Solution:

(b) Suppose that $n = 1000$ and define $D = X_{(501)} - X_{(499)}$. The distribution of D can be approximated by a distribution whose mean is μ. Which one of the following ((i)-(iv)) is the best approximation of μ ? Justify your answer to receive full marks.

(i) $2/1000$
(ii) $4/1000$
(iii) $8/1000$
(iv) $2/\sqrt{1000}$

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_n$ are independent random variables with pdf

$$f(x;\theta) = \left(\frac{1}{2\pi x^3}\right)^{1/2} \exp\left(-\frac{(x-\theta)^2}{2\theta^2 x}\right) \text{ for } x \geq 0$$

where $\theta > 0$ is an unknown parameter.

Find the MLE of θ based on $X_1, \cdots, X_n$ and give an estimator of its standard error based on the observed Fisher information. (You may assume all the regularity conditions for asymptotic normality are satisfied and you do not need to show your estimator maximizes the likelihood function.)

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_n$ are independent random variables on the interval $[0, \theta]$ with pdf

$$f(x;\theta) = \begin{cases} 3x^2/\theta^3 & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) The cumulative distribution function of $X_{(n)} = \max(X_1, \cdots, X_n)$ is

$$G(x;\theta) = P_\theta\left(X_{(n)} \le x\right) = \left(\frac{x}{\theta}\right)^{3n} \quad \text{for } 0 \le x \le \theta$$

(You do not need to show this.)

Show that $X_{(n)}/\theta$ is a pivot for θ.

Solution:

(b) Use the pivot $X_{(n)}/\theta$ to find an exact $100p\%$ confidence interval of the form $\left[X_{(n)}, aX_{(n)}\right]$ where $a > 1$ will depend on p and n.

Solution:

Practice Problem

Consider the following data

$$1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0$$

which we assume are a random sample (outcomes of independent random variables $X_1, \cdots, X_{10}$) from a Bernoulli distribution whose pmf is

$$f(x;\theta) = \theta^x(1-\theta)^{1-x} \text{ for } x = 0,1$$

where θ is an unknown parameter with $0 < \theta < 1$.

(a) Given the data, compute the MLE of θ and give an estimate of its standard error.

Solution:

(b) Suppose that we know that θ can take only two possible values: $\theta = 1/2$ and $\theta = 3/4$ so that $\Theta = \{1/2, 3/4\}$. What is the MLE of θ in this case?

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_n$ are independent continuous random variables with pdf

$$f(x;\theta) = \frac{1}{\theta} g(x/\theta) \text{ for } x \geq 0$$

where $\theta > 0$ is an unknown parameter but the function g (which is itself a pdf) is known.

(a) Show that $(X_1 + \cdots + X_n)/\theta$ is an exact pivot for θ. (Hint: Find the pdf of X_i/θ.)

Solution:

(b) Even though the random variable in part (a) is a pivot, its distribution is not necessarily easy to compute. Define

$$\mu_g = \int_0^\infty x g(x) dx \text{ and } \sigma_g^2 = \int_0^\infty (x - \mu_g)^2 g(x) dx.$$

Show that

$$\frac{1}{\sigma_g \sqrt{n}} \left(\frac{X_1 + \cdots + X_n}{\theta} - n\mu_g \right)$$

is approximately Normal with mean 0 and variance 1 when n is sufficiently large and use this to construct an approximate $100p\%$ confidence interval for θ, giving all appropriate details.

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_n$ are independent continuous random variables with common pdf $f(x-\theta)$ (where $f(x) = f(-x)$). θ can be estimated using the Winsorized mean:

$$\hat{\theta} = \frac{1}{n}\left\{rX_{(r)} + \sum_{i=r+1}^{n-r} X_{(i)} + rX_{(n-r+1)}\right\}$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics

(a) Suppose we are given an R function winsor where winsor (x, r) computes the estimate of θ for data x for all given value of r. The R code below computes a jackknife variance estimate of $\hat{\theta}$ based on 100 observations using $r = 10$.

```
> r <- 10
> thetahat <- winsor(x,r)
> thetahat
[1] 0.07854802
> thetai <- NULL # leave-one-out estimates
> for (i in 1:100) {
+    thetai <- c(thetai,winsor(x[-i],r))
+    }
> thetadot <- mean(thetai)
> jackvar <- 99*sum((thetai-thetadot)^2)/100
> jackvar
[1] 0.01301084
```

Assuming that the data come from a symmetric distribution, give an approximate 95% confidence interval for θ. (The 0.975 quantile of a standard normal distribution is 1.96.)

Solution:

(b) Suppose that $X_1, \cdots, X_n$ are independent random variables from a continuos distribution F with quantiles $F^{-1}(t)$ for $0 < t < 1$. If we define $\hat{\theta}_n$ to be the Winsorized mean with $r/n = \tau$ then

$$\hat{\theta}_n \xrightarrow{p} \theta(F) = \alpha\{F^{-1}(\tau) + F^{-1}(1-\tau)\} + \beta \int_{\tau}^{1-\tau} F^{-1}(s) ds$$

Find expresssions for α and β in terms of τ.

Solution:

Practice Problem

Suppose that $X_1, \cdots, X_{100}$ are independent Exponential random variables whose pdf is

$$f(x;\lambda) = \lambda \exp(-\lambda x) \quad \text{for } x \geq 0$$

where $\lambda > 0$ is unknown. However, rather than observing all 100 observations, we observe only the smallest 50 observations, in other words, the order statistics $X_{(1)} \leq \cdots \leq X_{(50)}$.

(a) Define $\hat{\lambda} = \ln(2)/X_{(50)}$. From lecture, we know that we can approximate the distribution of $\hat{\lambda}$ by a $\mathcal{N}(\lambda, \sigma^2(\lambda))$ distribution. Find an expression for $\sigma^2(\lambda)$.

Solution:

(b) Find the MLE of λ based on $X_{(1)}, \cdots, X_{(50)}$. (Hint: You may use the fact (without proof) that $100X_{(1)}, 99\left(X_{(2)} - X_{(1)}\right), \cdots, 51\left(X_{(50)} - X_{(49)}\right)$ are independent Exponential random variables with parameter λ.)

Solution: