Daeyoung Lim

Grad School

Year of 2008, late

1. 다음 물음에 답하시오.

   1. 크기 n인 표본 $\{x_1, \ldots, x_n\}$에서 반복을 허용하여 n번 재추출할 때, 특정 자료값 (예컨대 x_1)이 재표본(크기 n)에 포함될 확률은? 이 확률은 $n \to \infty$에 따라 어느 값에 수렴하는가?
   2. 100년에 99년 정도 규모의 홍수에 대비하여 댐이 건설되었다고 하자. 따라서 이 댐은 100년에 1년 정도 규모의 큰 홍수에는 견디지 못하고 붕괴된다. 이 댐이 100년 이내에 붕괴될 확률은?

Solution:

1. 특정 자료값을 x_1이라 놓았을 때, 1번 재추출 시 그 자료값이 뽑힐 확률은 n^{-1}이다. 따라서 n번 뽑을 경우 x_1이 한번도 뽑히지 않을 확률은

$$\left(1 - \frac{1}{n}\right)^n$$

   이 되고 한 번이라도 뽑히면 재표본에 포함되는 것이므로 그럴 확률은

$$1 - \left(1 - \frac{1}{n}\right)^n$$

   이 된다. $n \to \infty$할 때 이 확률은

$$\lim_{n\to\infty} \left(1 - \left(1 - \frac{1}{n}\right)^n\right) = 1 - e^{-1}$$

   이다.

2. 거지같이 써놔도 찰떡같이 알아먹는지 보려는 문제이다. 그러니까 댐이 붕괴될 확률이 1/100이라는 것이니 100년 이내에 붕괴될 확률은 다음과 같다.

$$\frac{1}{100} + \frac{99}{100}\frac{1}{100} + \cdots + \left(\frac{99}{100}\right)^{99}\frac{1}{100} = \sum_{n=0}^{99}\left(\frac{99}{100}\right)^n \frac{1}{100}$$

   등비수열의 합공식을 쓰면

$$1 - \left(\frac{99}{100}\right)^{100}$$

   이 된다.

2. 다음 물음에 답하시오.

   1. $X_1, \ldots, X_n$을 베르누이(p)로부터 구한 랜덤표본이라고 할 때, 성공확률 p에 대한 크레이머-라오 (Cramer-Rao) 정보부등식에 의한 비편향 추정량들이 가질 수 있는 분산의 하한값을 구하시오.

2. $X_1, \ldots, X_n$을 베르누이(p)로부터 구한 랜덤표본이라고 할 때, 1에서 구한 크레이머-라오 하한값을 이용하여 표본평균 $\overline{X}_n = (X_1 + \cdots + X_n)/n$이 균일 최소분산 추정량이 됨을 설명하시오.
3. $X_1, \ldots, X_n$을 베르누이(p)로부터 구한 랜덤표본이라고 할 때 p와 $\log(p/(1-p))$에 대한 최대가능도(maximum likelihood) 추정량을 각각 구하시오.
4. $X_1, \ldots, X_n$이 일양분포 $\text{Unif}(0, \theta)$로부터 구한 랜덤표본인 경우 크레이머-라오 하한값보다 분산이 작은 비편향 추정량이 존재하는가? 답에 대한 이유를 설명하시오.

Solution:

1. 크레이머-라오 하한값을 구하기 위해서는 피셔 정보량을 구해야 한다.

$$\mathcal{I}(p) = -\text{E}\left(\frac{d^2}{dp^2}\ln f(x|p)\right) \tag{1}$$
$$= \frac{1}{p(1-p)} \tag{2}$$

   따라서 크레이머-라오 정보부등식은 다음과 같다.

$$\text{Var}(T(X_1, \ldots, X_n)) \geq \frac{\{\text{E}'(T(X_1, \ldots, X_n))\}^2}{n\mathcal{I}(p)}$$

   여기서 분자에 들어있는 통계량은 p의 비편향추정량이라 했으므로 미분하게 되면 1이 나온다. 따라서 하한값은 $n^{-1}p(1-p)$이다.

2. 표본평균의 분산은 $n^{-1}p(1-p)$이므로 크레이머-라오 하한값과 동일해진다. 이를 most efficient estimator이라 하며 자동으로 UMVBUE가 된다. (모든 UMVUE가 반드시 크레이머-라오 하한과 같은 것은 아니고 같아질 경우에만 most efficient estimator이란 이름이 붙는다.)

3. 최대가능도추정량(MLE)는 invariance property가 있다. 즉, $\hat{\theta}$가 θ에 대한 MLE라면, 임의의 함수 T에 대하여 $T(\theta)$의 MLE는 $T\left(\hat{\theta}\right)$이다. 따라서 p의 MLE만 알면 $\ln(p/(1-p))$는 자동으로 알게 된다.

$$L(p|X_1, \ldots, X_n) = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} \tag{3}$$
$$\ln L(p|X_1, \ldots, X_n) = \ln p \cdot \sum_i^n X_i + \left(n - \sum_{i=1}^n X_i\right)\ln(1-p) \tag{4}$$
$$\frac{d\ln L(p|X_1, \ldots, X_n)}{dp} = \frac{\sum_{i=1}^n X_i}{p} - \frac{n - \sum_{i=1}^n X_i}{1-p} = 0 \tag{5}$$
$$\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i \tag{6}$$

그러므로 $\ln(p/(1-p))$의 MLE는 $\ln(\hat{p}/(1-\hat{p}))$이다.

4. 크레이머-라오 정보부등식은 다음의 세 가지 가정이 있다.

   - 모수 공간이 열린 집합(open set)이어야 한다. (Θ is an open set.) 혹은 $\theta \in \Theta^{\circ}$, (여기서 Θ°은 Θ의 interior임.)
   - $\mathcal{Y}_\theta = \{y \in \mathcal{Y} | f_Y(y|\theta) > 0\}$이 모든 $\theta \in \Theta$에 대하여 같은 support을 가져야 한다.
   - Dominated Convergence Theorem을 만족하는 함수 g가 존재해야 한다.(적분과 미분의 순서를 바꾸기 위함)

   균일 분포는 일단 두 번째 가정이 깨지므로 크레이머-라오 부등식이 성립하지 않는다. 더 심각한 것은 균일분포는 score statistic (로그가능도함수를 모수에 대해 미분한 것)의 기댓값이 0이 되지 않는다는 점이다.

$$\mathrm{E}\left(\frac{\partial}{\partial\theta}\ln f_Y(y|\theta)\right) = \int_0^\theta \frac{\partial}{\partial\theta}\frac{1}{\theta}\,dy \neq 0$$

   아무튼 안 된다.

3. Let $X_1, \ldots, X_n$ be a random sample from $f(x;\lambda) = \lambda\exp(-\lambda x)$, where $0 < x < \infty$. The parameter λ is λ_0 or λ_1, which is a known fixed number and $\lambda_1 > \lambda_0$. We want to test $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda = \lambda_1$. Obtain the most powerful test, including the distribution of the test statistic.

Solution:

4. Consider the linear model

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}$$

with $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, y_1, y_2 being scalar random variables, $\boldsymbol{X}$ is $\boldsymbol{X} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, with $\mathrm{E}(e) = \mathrm{E}\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\mathrm{E}(\boldsymbol{e}\boldsymbol{e}') = \boldsymbol{\Sigma}_{ee}$ is known and invertible.

   1. With $\boldsymbol{\Sigma}_{ee} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, determine the circumstances under which the BLUE(Best Linear Unbiased Estimator) of β is equal to y_1. Hint: The BLUE of β, $\hat{\beta}_{\mathrm{BLUE}} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$, is obtained by solving the normal equation $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\hat{\beta}_{\mathrm{BLUE}} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$. Find the relationship among σ_{ij}'s such that the normal equation is free from y_2.

2. With $\boldsymbol{\Sigma}_{ee} = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$ and $\beta = \mu$, find the BLUE of μ and its variance. Also derive the mean and variance of the arithmetic mean $\dfrac{y_1 + y_2}{2}$. Compare the variance of two estimators and state your conclusion.

Solution:

1. Let $\boldsymbol{G} = \boldsymbol{\Sigma}^{-1/2}$. Then the model becomes

$$\boldsymbol{G}\boldsymbol{y} = \boldsymbol{G}\boldsymbol{X}\beta + \boldsymbol{G}\boldsymbol{e}$$

and $\text{Var}(\boldsymbol{G}\boldsymbol{e}) = \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^T = \boldsymbol{I}_2$. By LSE, BLUE is

$$\text{argmin}_\beta \left(\boldsymbol{G}\boldsymbol{e}\right)^T \left(\boldsymbol{G}\boldsymbol{e}\right)$$

which, by plugging in the relationship, becomes

$$\text{argmin}_\beta \left(\boldsymbol{y} - \boldsymbol{X}\beta\right)^T \boldsymbol{G}^T\boldsymbol{G} \left(\boldsymbol{y} - \boldsymbol{X}\beta\right).$$

Since $\boldsymbol{G}^T\boldsymbol{G} = \boldsymbol{G}^2 = \boldsymbol{\Sigma}^{-1}$, the parts that include β are

$$\beta^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\beta - 2\beta^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

Differentiating w.r.t. β yields

$$2\beta^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} - 2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \boldsymbol{0}$$

and arranging the result,

$$\hat{\beta}_{\text{BLUE}} = \left(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}.$$

이것을 문제에 주어진 것을 대입하여 전개하면 $\sigma_{11} = \sigma_{12}$이고 $\sigma_2 \neq \sigma_{12}$일 때 BLUE가 y_1이 된다.

2. BLUE of μ is

$$\hat{\mu}_{\text{BLUE}} = \frac{\sigma_{22}}{\sigma_{11}+\sigma_{22}}\left(\frac{1}{\sigma_{11}}y_1 + \frac{1}{\sigma_{22}}y_2\right) \tag{7}$$

$$= \frac{\sigma_{22}}{\sigma_{11}+\sigma_{22}}y_1 + \frac{\sigma_{11}}{\sigma_{11}+\sigma_{22}}y_2. \tag{8}$$

BLUE turns out to be the weighted mean of y_1 and y_2. Therefore, the expected value of BLUE is the parameter itself; hence, unbiased. The variance of BLUE is

$$\text{Var}\left(\hat{\mu}_{\text{BLUE}}\right) = \left(\frac{\sigma_{22}}{\sigma_{11}+\sigma_{22}}\right)^2\sigma_{11} + \left(\frac{\sigma_{11}}{\sigma_{11}+\sigma_{22}}\right)^2\sigma_{22} \tag{9}$$

$$= \frac{\sigma_{11}\sigma_{22}}{\sigma_{11}+\sigma_{22}}. \tag{10}$$

And

$$\text{E}\left(\frac{y_1+y_2}{2}\right) = \frac{1}{2}\left(\text{E}\left(y_1\right)+\text{E}\left(y_2\right)\right) = \mu \text{Var}\left(\frac{y_1+y_2}{2}\right) \quad = \frac{1}{4}\left(\sigma_{11}+\sigma_{22}\right) \tag{11}$$

Both are unbiased but the variance of BLUE is smaller.

5. X_n이 평균이 $n\lambda_0$인 포아송 분포를 따른다고 하고,

$$Z_n = \frac{X_n - a_n}{b_n}$$

이라고 하자. n이 무한대로 접근함에 따라 Z_n의 분포가 표준정규분포에 수렴하기 위한 수열 a_n과 b_n을 구하고, 이에 필요한 정리를 정확히 설명하시오.

Solution: 중심극한정리는 분산이 유한한 확률변수가 여러개의 독립인 확률변수들의 합 꼴로 표현될 때 그 극한분포가 정규분포가 됨을 증명한 정리이다. $X_n \sim \text{Poi}\left(n\lambda_0\right)$라면 $Y_i \overset{\text{iid}}{\sim} \text{Poi}\left(\lambda_0\right),\ 1 \leq i \leq n$을 이용하여 $X_n \equiv Y_1 + \cdots + Y_n$로 재표현할 수 있다. 이때

$$\sqrt{n}\frac{\left(\overline{X}_n - \text{E}\left(Y_1\right)\right)}{\sqrt{\text{Var}\left(Y_1\right)}} \sim \mathcal{N}\left(0,1\right)$$

이므로

$$\sqrt{n}\frac{\left(\overline{X}_n - \text{E}\left(Y_1\right)\right)}{\sqrt{\text{Var}\left(Y_1\right)}} = \frac{X_n - n\lambda_0}{\sqrt{n\lambda_0}}$$

로 $a_n = n\lambda_0$, $b_n = \sqrt{n\lambda_0}$임을 알 수 있다.