다변량 최대우도추정

정의

- 개요

1. 스칼라 함수에서의 MLE를 다변량 함수로 확정하는 것
2. 스칼라 함수의 성질이 그대로 적용된다

- 다변량 모수의 최대우도추정

1. $X_1, \dots, X_n$을 공통 pdf $f(x;\theta)$를 갖는 iid이고,$\theta \in R^p$인 행렬이라고 할 때

1) 그 우도함수 $L(X_i;\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i;\theta)$ 이고

2) 로그우도함수는 $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} logf(x_i;\theta)$이다.

3) 이 때, 알고 있는 모수 $\theta_1 \dots \theta_k$에 대해

(1) $$\partial \begin{bmatrix} \frac{1}{\partial\theta_1} & \cdots & \frac{1}{\partial\theta_1} \\ \vdots & \ddots & \vdots \\ \frac{1}{\partial\theta_p} & \cdots & \frac{1}{\partial\theta_p} \end{bmatrix} \cdot \partial log \begin{bmatrix} f(x_1;\boldsymbol{\theta}) \\ \dots \\ f(x_n;\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial logf(X_1;\boldsymbol{\theta})}{\partial\theta_1} \\ \dots \\ \sum_{i=1}^{n} \frac{\partial logf(X_1;\boldsymbol{\theta})}{\partial\theta_k} \end{bmatrix} = \begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix}$$

(2) 즉, $\partial \begin{bmatrix} \frac{1}{\partial\theta_1} & \cdots & \frac{1}{\partial\theta_1} \\ \vdots & \ddots & \vdots \\ \frac{1}{\partial\theta_p} & \cdots & \frac{1}{\partial\theta_p} \end{bmatrix} \cdot \partial log \begin{bmatrix} f(x_1;\boldsymbol{\theta}) \\ \dots \\ f(x_n;\boldsymbol{\theta}) \end{bmatrix}$ 의 영공간을 구하면 그것이 MLE 추정량의 해가 된다.

정의

- 다변량 모수의 효율 한계

1. 다변량 모수추정에서의 피셔 정보 행렬 $I(\boldsymbol{\theta})$ 는

1) $\nabla logf(X;\boldsymbol{\theta}) = \left[\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1} \quad \cdots \quad \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}\right]$ 인 확률벡터의 공분산 행렬이다. 즉

2) $I(\boldsymbol{\theta}) = cov\left[\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1} \quad \cdots \quad \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}\right]$ 이고, 이는

(1) $\begin{bmatrix} var(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}) & \cdots & cov(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}, \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}) \\ \vdots & \ddots & \vdots \\ cov(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}, \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}) & \cdots & var(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}) \end{bmatrix}$ 과 같다.

정의

- 다변량 모수의 효율 한계

1. 다변량 모수추정에서의 피셔 정보는

2) I_{jk}성분을 조금 더 단순화하면

(1) 스칼라와 마찬가지로, 0 = $\int_{-\infty}^{\infty}\left[\frac{dlogf(x;\theta)}{d\theta_j}\right]f(x;\theta)\,dx = \mathrm{E}\left[\frac{dlogf(x;\theta)}{d\theta_j}\right]$ 이므로

(2) 위 식을 θ_k에 대해서 한번 더 미분하면

- $\int_{-\infty}^{\infty}\left[\frac{d^2logf(x;\theta)}{d\theta_j\theta_k}\right]f(x;\theta)\,dx + \int_{-\infty}^{\infty}\left[\frac{dlogf(x;\theta)}{d\theta_j}\cdot\frac{dlogf(x;\theta)}{d\theta_k}\right]f(x;\theta)\,dx$

- $-E\left[\frac{d^2logf(x;\theta)}{d\theta_j\theta_k}\right] = \mathrm{E}\left[\frac{dlogf(x;\theta)}{d\theta_j}\cdot\frac{dlogf(x;\theta)}{d\theta_k}\right]$

(3) 한편, 공분산은 cov(x,y) = E(XY) – E(X)E(Y)에서 E(X)E(Y) = 0이고.

$\mathrm{E(XY)} = \mathrm{E}\left[\frac{dlogf(x;\theta)}{d\theta_j}\cdot\frac{dlogf(x;\theta)}{d\theta_k}\right] = -E\left[\frac{d^2logf(x;\theta)}{d\theta_j\theta_k}\right]$ 이므로,

정의

- 다변량 모수의 효율 한계

1. 다변량 모수추정에서의 피셔 정보는

2) I_{jk}성분을 조금 더 단순화하면

- $cov(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}, \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}) = -E(\frac{\partial^2 logf(X;\theta)}{\partial\theta_p\partial\theta_1})$ 란 사실에서 피셔 정보 행렬은

- $\begin{bmatrix} var(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}) & \cdots & cov(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}, \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}) \\ \vdots & \ddots & \vdots \\ cov(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}, \frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_1}) & \cdots & var(\frac{\partial logf(X;\boldsymbol{\theta})}{\partial\theta_p}) \end{bmatrix} = \begin{bmatrix} -E(\frac{\partial^2 logf(X;\theta)}{\partial^2\theta_1}) & \cdots & -E(\frac{\partial^2 logf(X;\theta)}{\partial\theta_1\partial\theta_p}) \\ \vdots & \ddots & \vdots \\ -E(\frac{\partial^2 logf(X;\theta)}{\partial\theta_p\partial\theta_1}) & \cdots & -E(\frac{\partial^2 logf(X;\theta)}{\partial^2\theta_p}) \end{bmatrix}$

정의

- 다변량 모수의 효율 한계

1. 다변량 모수추정에서의 피셔 정보는

3) 이 때, $X_1, \dots, X_n$을 X에서 추출한 확률변수라고 하고, 각각의 표본에 대한 피셔 정보를 위 식에 반영하면

(1) $\nabla logL(X_i;\boldsymbol{\theta}) = \left[\sum_{i=1}^{n} \frac{\partial logf(x_i;\theta)}{\partial\theta_1} \quad \cdots \quad \sum_{i=1}^{n} \frac{\partial logf(x_i;\theta)}{\partial\theta_p}\right]$

(2) 스칼라에서와 마찬가지로, 각각의 피셔정보는 하나의 확률변수의 n배가 되므로, 이때 피셔정보는

$$- \begin{bmatrix} -nE(\frac{\partial^2 logf(X;\theta)}{\partial^2\theta_1}) & \cdots & -nE(\frac{\partial^2 logf(X;\theta)}{\partial\theta_1\partial\theta_p}) \\ \vdots & \ddots & \vdots \\ -nE(\frac{\partial^2 logf(X;\theta)}{\partial\theta_p\partial\theta_1}) & \cdots & -nE(\frac{\partial^2 logf(X;\theta)}{\partial^2\theta_p}) \end{bmatrix} = \mathrm{n} \begin{bmatrix} -E(\frac{\partial^2 logf(X;\theta)}{\partial^2\theta_1}) & \cdots & -E(\frac{\partial^2 logf(X;\theta)}{\partial\theta_1\partial\theta_p}) \\ \vdots & \ddots & \vdots \\ -E(\frac{\partial^2 logf(X;\theta)}{\partial\theta_p\partial\theta_1}) & \cdots & -E(\frac{\partial^2 logf(X;\theta)}{\partial^2\theta_p}) \end{bmatrix} = \mathrm{n}I(\boldsymbol{\theta})$$ 이다.

정의

- 다변량 모수의 효율 한계

2. 이 때, 효율 한계는

1) $\frac{1}{\mathrm{n}I(\boldsymbol{\theta})} \leq Var(Y)$ 에서 이 값이 해당 통계량의 최저 하한일 때

2) 마찬가지로 효율 추정량이라고 한다.

정의

- MLE 추정량의 근사 행태

1. $X_1, \ldots, X_n$을 공통 pdf $f(x;\theta)$를 갖는 iid라고 할 때,

1) MLE 추정량 $\frac{\partial \mathrm{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ = 0은 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} \xrightarrow{p} \boldsymbol{\theta}$ 인 $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$를 가진다.

2) 위를 만족하는 모든 열은$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}) \xrightarrow{D} N_p[0, I^{-1}(\boldsymbol{\theta})]$ 이다.

- G가 $1 \leq k \leq p$에서 변환

1. g(θ) = $\begin{bmatrix} g_1(\theta) \\ \cdots \\ g_k(\theta) \end{bmatrix}$ 이고, 이 때

2. B = $\begin{bmatrix} \frac{\partial g_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial g_1(\theta)}{\partial \theta_p} \\ \cdots & \ddots & \cdots \\ \frac{\partial g_k(\theta)}{\partial \theta_1} & \cdots & \frac{\partial g_k(\theta)}{\partial \theta_p} \end{bmatrix}$ 인 편미분행렬이 존재한다면

1) $\widehat{\boldsymbol{\gamma}}$= g$(\widehat{\boldsymbol{\theta}})$ 이면 $\widehat{\boldsymbol{\gamma}}$는 $\boldsymbol{\gamma}$= $\mathbf{g}(\boldsymbol{\theta})$ 의 MLE 추정량이다.

2) 이 때, $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \gamma) \xrightarrow{D} N_p[0, BI^{-1}(\boldsymbol{\theta})B^T]$ 이고, 따라서 정보행렬 $\mathbf{I}(\boldsymbol{\gamma})$ =$(BI^{-1}(\boldsymbol{\theta})B^T)^{-1}$을 가진다.

예제

- 정규 모형하에서의 최대우도 추정

1. $X_1, \ldots, X_n$이 iid N(μ, σ^2) 이라고 할 때, $\boldsymbol{\theta}$=$\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$이고, Ω= $\begin{bmatrix} (-\infty, \infty) \\ (0, \infty) \end{bmatrix}$라고 한다면

1) 로그우도 함수는 $\mathrm{l}(\mu, \sigma^2) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{26^2}\Sigma(x_i - \mu)^2$

(1) 이 때, $\begin{bmatrix} \mu_{mle} \\ \sigma_{mle}^2 \end{bmatrix}$ = log$[f(X_1), \ldots, f(X_n)] \cdot \partial \begin{bmatrix} \mu & \sigma^2 \\ \ldots & \ldots \\ \mu & \sigma^2 \end{bmatrix}$ 이므로

(2)$\mu_{mle} \Rightarrow \frac{\partial \Sigma log f(x_i;\theta)}{\partial\mu} = \frac{1}{\sigma^2}\Sigma(x_i - \mu) = 0,\ {\sigma^2}_{mle} \Rightarrow = \frac{\partial \Sigma log f(x_i;\theta)}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x_i - \mu)^2$=0

(3) 위 mle 추정식을 정리하면 $\mu_{mle} = \frac{\Sigma(x_i)}{n} = \overline{x},\ \sigma_{mle} = \frac{\sqrt{\Sigma(x_i-\mu)^2}}{n}$

예제

- 정규 pdf의 피셔정보행렬

1. X~N(μ, σ^2) 이라고 할 때, $\boldsymbol{\theta}=\begin{bmatrix}\mu \\ \sigma^2\end{bmatrix}$이고, $\Omega=\begin{bmatrix}(-\infty,\infty) \\ (0,\infty)\end{bmatrix}$라고 한다면

1) 로그우도 함수는 $\mathrm{l}(\mu,\sigma^2) = -\frac{1}{2}\log 2\pi - \log\sigma - \frac{1}{26^2}(x_i-\mu)^2$

(1) 피셔정보를 구하기 위해 이계 미분하면

- $\frac{\partial^2 \mathrm{l}(\mu,\sigma^2)}{\partial\mu^2} = -\frac{1}{26^2}$, $\frac{\partial^2 \mathrm{l}(\mu,\sigma^2)}{\partial(\sigma^2)^2} = \frac{1}{6^2} - \frac{3}{6^4}(\mathrm{X}-\mu)^2$, $\frac{\partial^2 \mathrm{l}(\mu,\sigma^2)}{\partial\mu\partial\sigma^2} = \frac{\partial\left[\frac{1}{\sigma^2}(\mathrm{X}-\mu)\right]}{\partial\sigma^2} = \frac{2}{6^3}(\mathrm{X}-\mu)$

(2) 기댓값을 구하면

- $-\mathrm{E}(\frac{1}{6^2}) = \frac{1}{6^2}$,
- $-\mathrm{E}[\frac{1}{6^2} - \frac{3}{6^4}(\mathrm{X}-\mu)^2] = -(\frac{1}{6^2} - \frac{3}{6^4}6^2) = \frac{2}{6^2}$
- $-\mathrm{E}[\frac{2}{6^3}(\mathrm{X}-\mu)] = \frac{2}{6^3}\cdot 0 = 0$

(3) 따라서 피셔정보행렬은

- $\begin{bmatrix}\frac{1}{6^2} & 0 \\ 0 & \frac{2}{6^2}\end{bmatrix}$ 이다.

예제

- $\widehat{\boldsymbol{\gamma}}$= g($\widehat{\boldsymbol{\theta}}$) 꼴의 피셔정보(정규분포 분산에 대한 정보)

1. 위 예제에서 정규 분포의 피셔 정보 행렬은 $\begin{bmatrix} \frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu^2} & \frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu\partial\sigma^2} \\ \frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu\partial\sigma^2} & \frac{\partial^2 l(\mu,\sigma^2)}{\partial(\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{6^2} & 0 \\ 0 & \frac{2}{6^2} \end{bmatrix}$

1) 이 때, 변환 g($[\mu, \sigma]$) = σ^2을 고려하면

(1) 편미분행렬 B = $[\frac{\partial g([\mu,\sigma])}{\partial\mu} \quad \frac{\partial g([\mu,\sigma])}{\partial\sigma}]$ = [0,2σ] 이고.

(2) σ^2의 정보는 $\mathbf{I}(\boldsymbol{\gamma})$ =$(BI^{-1}(\boldsymbol{\theta})B^T)^{-1}$ 에서

- $\{[0,2\sigma] \cdot \begin{bmatrix} \frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu^2} & \frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu\partial\sigma^2} \\ \frac{\partial^2 l(\mu,\sigma^2)}{\partial\mu\partial\sigma^2} & \frac{\partial^2 l(\mu,\sigma^2)}{\partial(\sigma^2)^2} \end{bmatrix}^{-1} \cdot [0,2\sigma]^T\}^{-1} = \frac{1}{2\sigma^4}$

(3) 따라서 , σ^2의 라오-크래머 하한은 $\frac{2\sigma^4}{n}$ 이다.

(4) 한편, σ^2의 불편추정량인 표본분산 $\frac{\Sigma(x_i-\mu)^2}{(n-1)}$의 분산은 $\frac{2\sigma^4}{n-1}$ 이므로, 불편추정량은 효율추정량이 아니다.