추정
- 점추정
  - 적률추정법
  - 최대가능도 추정법
- 구간추정
  - 추축변량
  - 베이즈 추정방법.

미지의 모수를 포함하지 않는 관찰값 $X_1,\dots,X_n$의 함수를 통계량이라고 한다.

모수 θ의 함수 $g(\theta)$를 추정하기 위하여 사용되는 통계량 $T(X)=T(X_1,\dots,X_n)$을 $g(\theta)$의 추정량이라고 한다.

$X_1=x_1, X_2=x_2, \dots, X_n=x_n$을 대입하여 구한 추정량의 특정값, $T(x)=T(x_1,\dots,x_n)$을 추정값이라고 한다.

E 4.1

$\hat{\mu}=T_1(X_1,\cdots X_n)=\frac{\sum_{i=1}^n X_i}{n}$, $\sigma^2=T_2(X_1,\cdots,X_n)=\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1}$ 을 추정량으로 사용할 수 있다.

이때, 표본이 주어지면 그들의 값 $T_1(x_1,\cdots,x_n)=\frac{\sum_{i=1}^n x_i}{n}$, $T_2(x_1,\cdots,x_n)=\frac{\sum_{i=1}^n (x_i-\bar{x}_n)^2}{n-1}$ 을 추정값이라고 한다.

4.1. 적률추정법.

모적률과 표본적률의 차이를 줄이는 방식으로 모수를 추정한다.

r차 모적률 (r^{th} population moment) : $\mu_r' = E(X^r)$

r차 표본적률 (r^{th} sample moment) : $m_r' = \frac{\sum_{i=1}^n X_i^r}{n}$ 이고

μ_r'은 모수벡터 $\theta=(\theta_1,\cdots,\theta_k)$의 함수이므로 k개의 연립방정식을 풀면 $\hat{\theta}_1,\cdots,\hat{\theta}_k$를 구할 수 있을 것이다.

즉, 모수의 개수만큼 적률방정식을 세우고 표본적률과 모적률을 일치시키는 방정식을 풀면 모수벡터를 추정한다.

E 4.2

$X_1,\cdots,X_n \sim N(\mu,\sigma^2)$ Random Sample. μ,σ^2을 적률추정법을 사용하여 추정하라.

$$\left.\begin{array}{ll} \mu_1'=E(X) & m_1'=\frac{\sum_{i=1}^n X_i}{n}=\mu \\ \mu_2'=E(X^2) & m_2'=\frac{\sum_{i=1}^n X_i^2}{n}=\sigma^2+\mu^2 \end{array}\right\}$$ 연립방정식 사용.

$\hat{\mu}=\frac{\sum_{i=1}^n X_i}{n}$, $\hat{\sigma}^2 = m_2'-(m_1')^2 = \sum_{i=1}^n \frac{X_i^2}{n} - \left(\sum_{i=1}^n \frac{X_i}{n}\right)^2 = \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - n(\bar{X})^2\right) = \frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n}$

E.4.3

$X_1, \cdots, X_n \sim GAM(k, \theta)$ Random Sample.

$m_1' = \frac{\sum_{i=1}^n X_i}{n} = \mu_1'(k,\theta) = k\theta$

$m_2' = \frac{\sum_{i=1}^n X_i^2}{n} = \mu_2'(k,\theta) = k\theta^2 + k^2\theta^2$) 연립방정식

$\bar{X} = k\theta$, $k\theta^2 = m_2' - (m_1')^2 = \sum_{i=1}^n \frac{X_i^2}{n} - \left(\sum_{i=1}^n \frac{X_i}{n}\right)^2 = \sum_{i=1}^n \frac{X_i^2}{n} - (\bar{X})^2 = \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$

$\theta = \frac{\bar{X}}{k}$ 이므로 $k \times \frac{\bar{X}^2}{k^2} = \frac{\bar{X}^2}{k} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \Rightarrow k = \frac{n(\bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$

$\theta = \frac{\bar{X}}{\frac{n(\bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}} = \frac{\bar{X} \cdot \sum_{i=1}^n (X_i - \bar{X})^2}{n(\bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n\bar{X}}$

4.2 최대가능도 추정법

성공률이 p(0.1 or 0.9) 인 베르누이 분포에서 크기가 10인 랜덤표본을 뽑을 때, 6이라면 0.9로 예측한다.

가능도함수는 $L(\theta)$로 쓰고 $f(x_1, \cdots, x_n; \theta)$이다. 즉 $x_1, \cdots, x_n$을 알 때, 모수 θ를 찾는 것이다.

위의 예에서 모수는 p이고 $x_1, \cdots, x_n$은 6에 해당하는 것이다.

가능도 함수는 모수(θ)의 함수이므로 확률밀도함수가 아니다.

가능도함수는 특정한 시점에서 모수를 추정하기에 편향된 모수값을 추정할 수 있다.

→ 그러므로 표본을 많이 뽑아서 $n \to \infty$ 이라면 추정량의 변동이 줄어들고, 모집단의 참값에 수렴한다.

이를 일치성 (consistency)라고 한다.

$L(\theta; x_1, \cdots, x_n)$을 최대화하는 θ의 값을 $\hat{\theta} = \hat{\theta}(x_1, \cdots, x_n) \in \Omega$라고 할 때,

$\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$을 모수 θ의 최대가능도추정량이라고 한다.

'실제로 관측된' 자료가 얻어질 확률을 가장 크게 만드는 (즉, 주어진 관측값을 가장 잘 설명하는)

θ 값을 모수 θ의 추정량으로 삼는 것이다.

$$L(\theta; x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$$

계산의 편의를 위하여 log를 붙여 사용한다. log는 단조 증가함수이므로 붙여도 상관없다.

E.4.5

$X_1, \cdots, X_n \sim EXP(\theta)$ Random Sample.

$$L(\theta; x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$$

$$\log L(\theta) = -\frac{\sum_{i=1}^{n} x_i}{\theta} - n\log\theta \qquad \frac{\partial l(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} \Rightarrow \frac{n}{\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta^2} \qquad \hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}_n$$

$\theta = \bar{x}_n$일 때, 이차미분값도 0보다 작으므로 이 값에서 로그가능도함수는 최대가 된다.

E.4.6

$X_1, \cdots, X_n \sim poi(\lambda)$ Random Sample.

$$L(\lambda; x_1, \cdots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \qquad l(\lambda) = \sum_{i=1}^{n} x_i \cdot \log\lambda - n\lambda - \log\left(\prod_{i=1}^{n} x_i!\right)$$

$$l'(\lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}_n$$

이차미분해서 $\lambda = \bar{x}_n$ 이라면 음수이기에 최대이다.

E.4.7

$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$

$$L(\mu, \sigma^2; x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \cdot e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

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

$$\frac{\partial l(\mu, \sigma^2)}{\partial \mu} = \frac{\cancel{2}\sum_{i=1}^{n}(x_i-\mu)}{\cancel{2}\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} = 0 \rightarrow \sum_{i=1}^{n} x_i = n\mu \qquad \mu = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}_n$$

$$\frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{\cancel{2}} \cdot \frac{\cancel{2\pi}}{\cancel{2\pi}\sigma^2} + \frac{2\sum_{i=1}^{n}(x_i-\mu)^2}{4\sigma^4} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} = 0$$

$$\rightarrow n = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} \rightarrow \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}{n}$$

$$\therefore \hat{\mu} = \bar{X}_n \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}{n}$$

E.4.8

$X_1, \cdots, X_n \sim$ Logistic $\quad f(x;\theta) = \dfrac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$, $-\infty < x < \infty$, $-\infty < \theta < \infty$ 이다.

$$l(\theta) = \sum_{i=1}^{n} \log\left(\frac{e^{-(x_i-\theta)}}{(1+e^{-(x_i-\theta)})^2}\right) = \sum_{i=1}^{n}\left(-x_i+\theta - 2\log(1+e^{-x_i+\theta})\right) = -\sum_{i=1}^{n} x_i + n\theta - 2\sum_{i=1}^{n}\log(1+e^{-x_i+\theta})$$

$l'(\theta) = n - 2\sum_{i=1}^{n} \dfrac{e^{-x_i+\theta}}{1+e^{-x_i+\theta}} = 0$ 간단한 해는 존재하지 않으나 단조감소함수이고 $\theta \to -\infty$일 때, n, $\theta \to \infty$일 때 $-n$으로 접근하므로 유일해를 가진다는 것을 알 수 있다.

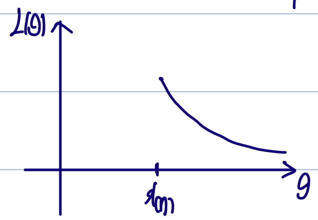
. 가능도 함수는 언제나 미분가능한 것은 아니다. 따라서 최대가능도 추정값이 항상 미분방정식을 풀어서 구해지는 것은 아니다.

E.4.9

$X_1, \cdots, X_n \sim U(0, \theta)$

$$L(\theta; x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \frac{1}{\theta^n} & 0 \le x_i \le \theta \\ 0 & \text{otherwise} \end{cases}$$

) 미분이 불가능한 케이스이다.



$x_1, \cdots, x_n$ 중에 하나라도 $0 \le x_i \le \theta$를 만족하지 않는다면 0이므로 $x_1, \cdots, x_n$ 중에 가장 큰 $x_{(n)}$이 θ보다 작아야 한다.

$\frac{1}{\theta^n}$은 감소함수이기에 $x_{(n)}$에 따라 $L(\theta)$값이 정해진다.

E.4.10

$X_1, \cdots, X_n \sim poi(\lambda)$

$$l(\lambda) = \sum_{i=1}^{n} \log\left(\frac{\lambda^x e^{-\lambda}}{x!}\right) = \sum_{i=1}^{n}\left(x_i\log\lambda - \lambda - \log(x!)\right) = \log\lambda\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n}\log(x!)$$

$l'(\lambda) = \dfrac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0 \Rightarrow \lambda = \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{x}_n$

$\bar{x}_n = 0$이라면 $l'(\lambda)$가 항상 음의 값을 가진다.
그리고 $\lambda = 0$이어야하는데 λ는 0보다 크므로 $\bar{x}_n = 0$ 일 때는
최대가능도 추정값은 존재하지 않는다

E.4.11

$X_1, \cdots, X_n \sim U(\theta-1, \theta+1)$ Random Sample.

$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{2} I(\theta-1 \le x_i \le \theta+1) = (\frac{1}{2})^n I(\theta-1 \le x_{(1)}, x_{(n)} \le \theta+1) = 2^{-n} I(x_{(n)}-1 \le \theta \le x_{(1)}+1)$

구간 $x_{(n)}-1$, $x_{(1)}+1$ 의 모든 점에서 최댓값 2^{-n}을 가진다.

즉 구간 내 모든 점이 최대가능도 추정량이 된다.

E.4.12

$(N_1, N_2, N_3) \sim MULT(n, p_1, p_2, p_3)$

$p_1 = \theta^2 \quad p_2 = 2\theta(1-\theta) \quad p_3 = (1-\theta)^2 \quad 0<\theta<1$ 일 때, θ의 최대가능도추정량은?

$N_1 = n_1$, $N_2 = n_2$, $N_3 = n_3$

$$L(\theta) = \frac{n!}{n_1! n_2! n_3!} \theta^{2n_1} (2\theta(1-\theta))^{n_2} \cdot (1-\theta)^{2n_3}$$

$$l(\theta) = \log\left(\frac{n!}{n_1! n_2! n_3!}\right) + 2n_1 \log\theta + n_2 \log(2\theta(1-\theta)) + 2n_3 \log(1-\theta)$$

$$= \log\left(\frac{n!}{n_1! n_2! n_3!}\right) + 2n_1 \log\theta + n_2 \log 2\theta + n_2 \log(1-\theta) + 2n_3 \log(1-\theta)$$

$l'(\theta) = \frac{2n_1}{\theta} + \frac{2n_2}{2\theta} - \frac{n_2}{1-\theta} - \frac{2n_3}{1-\theta} = 0$ 이므로 양변에 $\theta(1-\theta)$를 곱하면

$$2n_1(1-\theta) + n_2(1-\theta) - \theta n_2 - 2\theta n_3 = 2n_1 - 2n_1\theta + n_2 - \theta n_2 - \theta n_2 - 2\theta n_3 = 2n_1 + n_2 - 2n_1\theta - 2n_2\theta - 2n_3\theta$$

$= 2n_1 + n_2 - 2\theta(n_1+n_2+n_3) = 0 \quad \theta = \frac{2n_1+n_2}{2(n_1+n_2+n_3)} = \frac{2n_1+n_2}{2n} \quad (n_1+n_2+n_3 = n)$

2차미분은 음수이므로 최대가능도 추정량은 $\underline{n_1 \ne n}$, $\underline{n_3 \ne n}$ 인 경우에 $\hat{\theta} = \frac{2N_1+N_2}{2N}$

$n_1 = n$이면 $\hat{\theta} = 0$ $\quad$ $n_3 = n$이면 $\hat{\theta} = 1$

· 최대가능도 추정량의 불변성

θ을 추정하는 어떤 추정량이 $T(X)$일 때, θ^2의 추정량은 $T(X)^2$가 되고 $\theta+c$는 $T(X)+c$이다.

즉 θ의 함수 $g(\theta)$의 추정량이 $g(T(X))$가 되는 성질을 불변성 원리 (invariance property) 라고 한다.

미분이 쉬운 θ를 찾고 다른 형태의 모수를 쉽게 구할 수 있다.

<증명>

$\eta = g(\theta)$ $\quad L^*(g(\theta)) = L^*(\eta) = \prod_{i=1}^{n} f(x_i; g^{-1}(\eta)) = L(g^{-1}(\eta)) = L(\theta)$ 이고

$\max_{\eta \in g(\Omega)} L^*(\eta) = \max_{\eta \in g(\Omega)} L(g^{-1}(\eta)) = \max_{\theta \in g(\Omega)} L(\theta)$ 이므로

$L^*(\hat{\eta}) = \max_{\eta \in g(\Omega)} L^*(\eta) = \max_{\theta \in g(\Omega)} L(\theta) = L(\hat{\theta}) = L^*(g(\hat{\theta}_n))$ 이다.

즉 $\hat{\eta} = g(\hat{\theta}_n)$ 이다.

E. 4.13

$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$ 인데 $g(\sigma^2) = \sigma$ 의 최대가능도 추정량은 불변성에 의해 $g(\hat{\sigma}^2) = \hat{\sigma}$ 가 된다.

E. 4.14

$X_1, \cdots, X_n \sim Exp(\lambda)$ Random Sample.

$l(\lambda) = \sum_{i=1}^{n} \log\left(\frac{1}{\lambda} e^{-\frac{x_i}{\lambda}}\right) = \sum_{i=1}^{n}\left(-\log\lambda - \frac{x_i}{\lambda}\right) = -n\log\lambda - \sum_{i=1}^{n}\frac{x_i}{\lambda}$

$l'(\lambda) = -\frac{n}{\lambda} + \frac{\sum_{i=1}^{n} x_i}{\lambda^2} = 0 \quad \frac{n}{\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda^2} \quad \lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}_n$ 이므로 $\hat{\lambda} = \bar{X}_n$

만약 확률변수가 1보다 큰 확률에 관심이 있다면 $g(\lambda) = P(X \geq 1) = \int_1^\infty \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = \left[-e^{-\frac{x}{\lambda}}\right]_1^\infty = e^{-\frac{1}{\lambda}}$

이므로 $g(\hat{\lambda}) = e^{-\frac{1}{\bar{X}_n}}$ 이 된다.

4.3 추정의 기준.

확률분포를 살리는 것은 추정량의 성격을 파악하게 도움을 준다.

E. 4.15

$X_1, \cdots, X_n \sim U(0, \theta)$ Random Sample.

θ의 최대가능도 추정량은 $X_{(n)}$이다. $X_{(n)}$의 pdf는 $\theta = 1, n = 5$인 경우

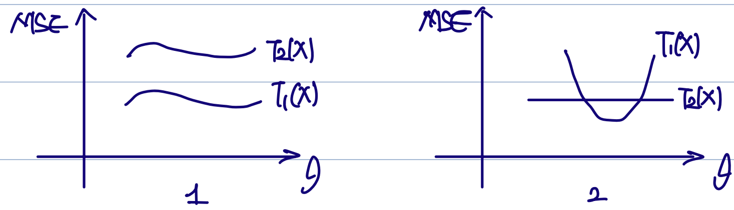
1. $f(X_{(n)}) = n \cdot [F(x_{(n)})]^{n-1} \cdot f(x_{(n)}) = 5 \cdot \left[\frac{x}{\theta}\right]^4 \cdot \frac{1}{\theta} = \frac{5}{\theta^5} x^4 = 5x^4 \ (0 < x < 1)$

2. $F_{X_{(5)}}(x) = P(X_{(5)} \leq x) = \underline{P(X_1 \leq x, \cdots, X_5 \leq x) = \{P(X \leq x)\}^5} = (F_X(x))^5 = x^5 \ (0 < x < 1)$

$f_{X_{(5)}}(x) = 5x^4 \ (0<x<1)$ ↳ 표본최댓값이 x보다 작으면 나머지도 모두 작고, 모든 관측값이 x보다 작으면 표본 최댓값도 x보다 작다.

Loss function : $(T(X)-g(\theta))^2$ 이고 Loss function 또한 확률변수이므로 이것을 최소화하는 추정량 $T(X)$를 찾는 것은 불가능하다. 따라서 손실함수의 기댓값을 기준으로 삼는다.

즉 $MSE = E(T(X)-g(\theta))^2$를 사용한다.



1의 경우는 $T_1(X)$는 항상 $T_2(X)$보다 좋지만 2는 아니기에 어떤 추정량($T(X)$)를 사용할 지 명확하지 않다. 이렇게 MSE를 모든 $\theta\in\Omega$에 대해 균일하게 최소화하는 θ는 존재 X.

편향적인 추정량을 제외하기 위한 비편향 추정량을 고려하자

$T(X)$를 $g(\theta)$의 추정량이라고 할 때, $E(T(X))-g(\theta)$를 편향이라고 하고

$E(T(X))-g(\theta)=0$ 이면 $T(X)$는 $g(\theta)$의 비편향추정량이다.

E.4.17

$X_1,\cdots,X_n \sim N(\mu,\sigma_0^2)$ Random Sample.

$T_1(X)=\bar{X}_{10}$ $T_2(X)=\frac{X_1+X_2}{2}$는 $E(T_1(X))=E(T_2(X))=\mu$ 이므로 모두 모평균 μ의 비편향 추정량이다.

E.4.18

$U(0,\theta)$에서 고려한 최대가능도 추정량 $X_{(n)}$은 θ보다 클 수 없다. 즉 기댓값은 θ보다 작다.

확인해보면 $E(X_{(n)}) = \int_0^\theta x\cdot\frac{n}{\theta^n}\cdot x^{n-1}dx = \int_0^\theta n\cdot\frac{x^n}{\theta^n}dx = \frac{n}{n+1}\cdot\frac{1}{\theta^n}[x^{n+1}]_0^\theta = \frac{n}{(n+1)\theta^n}\cdot\theta^{n+1} = \frac{n}{n+1}\theta$

$\frac{n}{n+1}\theta$는 θ보다 작다. 즉, 비편향 추정량이 아니다.

$T_1(X) = \frac{n+1}{n} \cdot X_{(n)}$은 $E(T_1(X)) = \frac{n+1}{n} \cdot \frac{n}{n+1} \cdot \theta = \theta$ 이기에 $T_1(X)$는 비편향추정량이다.

$g(\theta)$의 추정량 $T(X)$의 MSE는 $MSE = Var(T(X)) + (bias)^2$ 이다.

<증명>

$MSE = (T(X) - g(\theta))^2 = E(T(X) - E(T(X)) + E(T(X)) - g(\theta))^2$

$= E([T(X) - E(T(X))]^2) + E[E(T(X)) - g(\theta)]^2 + 2E([T(X) - E(T(X))][E(T(X)) - g(\theta)])$

$= Var(T(X)) + (E(T(X)) - g(\theta))^2 + 2\underbrace{\cancel{E\{T(X) - E(E(T(X)))\}}}_{0}\{E(T(X)) - g(\theta)\}$

$= Var(T(X)) + (Bias)^2$

비편향 추정량 중에서 더 좋은 것을 따지고 싶을 때는 분산비(Variance Ratio)를 이용한다.

$T_1(X), T_2(X)$가 비편향추정량일 때, $\frac{Var(T_2(X))}{Var(T_1(X))}$를 $T_1(X)$의 $T_2(X)$에 대한 상대효율이라고 한다.

E. 4.19

E. 4.17에서 $T_1(X) = \bar{X}_{10}$과 $T_2(X) = \frac{X_1 + X_2}{2}$의 분산은 각각 $Var(T_1(X)) = \frac{\sigma_0^2}{10}$ $Var(T_2(X)) = \frac{\sigma_0^2}{2}$

$\frac{Var(T_2(X))}{Var(T_1(X))} = \frac{\frac{\sigma_0^2}{2}}{\frac{\sigma_0^2}{10}} = \frac{10 \times \sigma_0^2}{2 \times \sigma_0^2} = 5$ 이므로 $Var(T_2(X))$가 $Var(T_1(X))$보다 효율이 낮은 추정량이다

E. 4.20

$U(0,\theta)$의 θ에 대한 비편향추정량은 $\hat{\theta} = T_1(X) = \frac{n+1}{n} X_{(n)}$이다. $X_{(n)} = \frac{n}{n+1} T_1(X)$

$f_{T_1}(t) = f_{X_{(n)}}\left(\frac{n}{n+1}t\right)\left|\frac{d\frac{n}{n+1}t}{dt}\right| = n \cdot \left(\frac{\frac{n}{n+1}t}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} \cdot \frac{n}{n+1} = \frac{n}{\theta^n} \cdot \left(\frac{n}{n+1}\right)^n \cdot t^{n-1} \; I\left(0 \le t \le \frac{n+1}{n}\theta\right)$

이제 $X_{(1)}$의 비편향추정량을 구해서 비교.

$f_{X_{(1)}}(x_{(1)}) = n \cdot \left(1 - \frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n}{\theta}\left(1 - \frac{x}{\theta}\right)^{n-1}, \; 0 < x < \theta$

$E(X_{(1)}) = \int_0^\theta x \cdot \frac{n}{\theta} \cdot \left(1 - \frac{x}{\theta}\right)^{n-1} dx = \frac{n}{\theta}\int_0^\theta x\left(1 - \frac{x}{\theta}\right)^{n-1} dx = \frac{\cancel{n}}{\theta} \cdot \frac{\theta^{\ast}}{\cancel{n}(n+1)} = \frac{\theta}{n+1}$

즉 비편향추정량으로 $T_2(X)=(n+1)X_{(1)}$을 생각할 수 있다.

$Var(T_1(X))$, $Var(T_2(X))$를 구하기 위해 $E(T_1(X)^2)$, $E(T_2(X)^2)$을 구하자.

$$E(T^2)=\int_0^{\frac{n+1}{n}\theta} t^2\cdot\frac{n}{\theta^n}\cdot\left(\frac{n}{n+1}\right)^n\cdot t^{n-1}\,dt=\frac{n}{\theta^n}\cdot\left(\frac{n}{n+1}\right)^n\int_0^{\frac{n+1}{n}\theta} t^{n+1}\,dt=\frac{\cancel{n}}{\cancel{\theta^n}}\cdot\left(\frac{\cancel{n}}{\cancel{n+1}}\right)^{\cancel{n}}\cdot\frac{1}{n+2}\cdot\left(\frac{n+1}{n}\right)^{\cancel{n+2}}\cdot\cancel{\theta^{n+2}}\,\theta^2$$

$$=\frac{(n+1)^2}{n}\cdot\frac{\theta^2}{n+2}=\frac{(n+1)^2}{n(n+2)}\theta^2$$

$Var(T_1(X))=\frac{(n+1)^2}{n(n+2)}\theta^2-\theta^2=\frac{\theta^2}{n(n+2)}$ 이다.

$$f_{T_2}(t)=f_{X_{(1)}}\left(\frac{t}{n+1}\right)\left|\frac{d\frac{t}{n+1}}{dt}\right|=n\cdot\left(1-\frac{\frac{t}{n+1}}{\theta}\right)^{n-1}\cdot\frac{1}{\theta}\cdot\frac{1}{n+1}=n\left(\frac{\theta-\frac{t}{n+1}}{\theta}\right)^{n-1}\cdot\frac{1}{\theta}\cdot\frac{1}{n+1}=\frac{n}{n+1}\cdot\frac{1}{\theta^n}\cdot\left(\theta-\frac{t}{n+1}\right)^{n-1} I(0\le t\le (n+1)\theta)$$

$$E(T_2^2)=\int_0^{(n+1)\theta} t^2\cdot\frac{1}{\theta^n}\cdot\frac{n}{n+1}\cdot\left(\theta-\frac{t}{n+1}\right)^{n-1}dt=\frac{n}{\theta^n(n+1)}\int_0^{(n+1)\theta} t^2\left(\theta-\frac{t}{n+1}\right)^{n-1}dt$$

$$=\left(\frac{1}{\theta(n+1)}\right)^n\cdot n\int_0^{(n+1)\theta} t^2(n\theta+\theta-t)^{n-1}dt=\frac{1}{\cancel{\theta^n(n+1)^n}}\cdot n\cdot\frac{2}{n(n+1)(n+2)}\cancel{\theta^{n+2}}\theta^2\cancel{(n+1)^{n+2}}=\frac{(n+1)^2\cdot 2\cancel{n}\cdot\theta^2}{\cancel{n}\cancel{(n+1)}(n+2)}=\frac{2(n+1)\theta^2}{n+2}$$

$E(T_2)^2=\theta^2$

$$Var(T_2)=\frac{2(n+1)\theta^2}{n+2}-\theta^2=\left(\frac{2n+2}{n+2}-\frac{n+2}{n+2}\right)\theta^2=\frac{n}{n+2}\theta^2$$

$\frac{Var(T_1(X))}{Var(T_2(X))}=\frac{\frac{1}{n(n+2)}\theta^2}{\frac{n}{n+2}\theta^2}=\frac{n+2}{n^2(n+2)}=\frac{1}{n^2}$ $n>1$ 일 때, T_2는 T_1에 비해 효율이 안좋다.

여기서 T_2는 T_1과 같은 분산을 갖기위해 표본의 크기를 n^2배 더 가져야한다.

그러면 비편향추정량은 2개보다 많이 있어 어떻게 해야 효율이 가장 좋은 추정량을 찾을 건가?

우리는 효율이 100%인 비편향추정량 즉, 최소분산 비편향추정량을 찾으려고 한다.

4.4 최소분산 비편향 추정량 (Minimum Variance Unbiased Estimator, MVUE)

비편향 추정량 중에서 최소분산을 가지는 추정량에 대해 다룬다.

1. $T^*(X)$는 $g(\theta)$의 비편향추정량이다.

2. $g(\theta)$의 임의의 비편향추정량 $T(X)$에 대하여 $Var(T^*(X))\le Var(T(X))$를 만족한다.

어떻게 MVUE를 구하는가?

1. 크래머-라오 하한값 : 분산이 적절한 조건하에서 비편향 추정량이 가질 수 있는 분산의 하한값.

2. 라오 블랙웰의 정리, 레만 쉐페의 정리 by 완비충분통계량

1. 크래머-라오

크래머-라오의 정보부등식은 비편향추정량이 가질 수 있는 분산의 하한을 제공한다.

이 하한을 크래머-라오 하한값이라고 한다.

어떤 비편향 추정량도 이보다 작은 분산도 가질 수 없다.

즉 어떤 비편향추정량이 이 값을 가진다면 그것은 MVUE가 된다.

정보부등식에 사용되는 피셔의 정보 $I(\theta)$는 $I(\theta) = E[(\frac{\partial}{\partial\theta}\log f(X;\theta))^2]$으로 정의된다.

E. 4.21

$X \sim N(\mu, \sigma_0^2)$일 때, $f(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\mu)^2}{2\sigma_0^2}}$

$I(\mu) = E[(\frac{\partial}{\partial\mu}\log f(x;\mu))^2] = E[(\frac{X-\mu}{\sigma_0^2})^2] = E[\frac{1}{\sigma_0^4}\cdot(X-\mu)^2] = \frac{1}{\sigma_0^4}E[(X-\mu)^2] = \frac{\sigma_0^2}{\sigma_0^4} = \frac{1}{\sigma_0^2}$

σ_0가 커질수록 관찰값 X가 모평균 μ에 대하여 가지는 정보의 양은 작아진다.

그리고 $I(\mu)$는 μ의 값에 의존하지 않는다.

$X \sim N(\mu, \sigma_1^2)$ $Y \sim N(\mu, \sigma_2^2)$에서 $\sigma_1^2 < \sigma_2^2$라면 μ를 추정할 때, X를 쓰는게 좋다.

E. 4.22

성공확률이 p인 베르누이 분포.

$f_X(x;p) = p^x(1-p)^{1-x}$ $(x = 0, 1,\ 0<p<1)$ $\log f_X(X;p) = X\log p + (1-X)\log(1-p)$

$\frac{\partial \log f_X(X;p)}{\partial p} = \frac{X}{p} - \frac{1-X}{1-p} = \frac{X-p}{p(1-p)}$

$E[(\frac{\partial \log f_X(X;p)}{\partial p})^2] = E[(\frac{X-p}{p(1-p)})^2] = \frac{E((X-p)^2)}{p^2(1-p)^2} = \frac{E(X^2) - 2pE(X) + E(p^2)}{p^2(1-p)^2} = \frac{p(1-p)}{p^2(1-p)^2} = \frac{1}{p(1-p)}$

$p = \frac{1}{2}$일 때, 피셔의 정보가 가장 작다.

· 정보부등식 (Information Inequality)

(1) 모수가 다르면 pdf가 다르다.

(2) $A=\{x : f(x;\theta)>0\}$는 모수 θ에 의존하지 않으며, 모든 $x\in A$, $\theta\in\Omega$에 대하여

$\log f(x;\theta)$는 θ에 대하여 두 번 미분가능하고, 도함수가 연속이다.

(3) $T(X)$가 모든 $\theta\in\Omega$에 대해 $E(T(X))<\infty$ 라면

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n$$ 이다.

이제 모든 $\theta\in\Omega$에 대하여 $Var(T(X))<\infty$, $E(T(X))=g(\theta)$, $0<I(\theta)<\infty$ 라고 하면

$g(\theta)$는 미분가능하며 $Var(T(X))\ge\dfrac{(g'(\theta))^2}{nI(\theta)}$ 이 성립한다.

<증명>

$T(X)$는 $g(\theta)$의 비편향추정량이므로 $E(T(X))=g(\theta)$이다.

$$(E(T(X)))'=g'(\theta)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n-g(\theta)\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n$$

($\because$ $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n=1$ 이므로 $\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n=0$)

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[T(x_1,\cdots,x_n)-g(\theta)]\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[T(x_1,\cdots,x_n)-g(\theta)]\left[\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}f(x_i;\theta)\right]\times\prod_{i=1}^{n}f(x_i;\theta)\,dx_1\cdots dx_n$$

$$=E\left[(T(x_1,\cdots,x_n)-g(\theta))\cdot\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}f(x_i;\theta)\right)\right]$$ 이 성립한다.

코시-슈바르츠 부등식
$E(X^2)E(Y^2)\ge E(XY)^2$

코시 슈바르츠 정리에 의하여

$$[g'(\theta)]^2\le E[(T(X)-g(\theta))^2]\cdot E\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}f(X_i;\theta)\right)^2\right]=Var(T(X))\times E\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}f(X_i;\theta)\right)^2\right]$$

또는 $$Var(T(X))=\frac{[g'(\theta)]^2}{E\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n}f(X_i;\theta)\right)^2\right]}$$ 이 성립한다.

그런데 $E\left[\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right]=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\log f(X_i;\theta)\,f(x;\theta)\,dx=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x;\theta)\,dx=\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(x;\theta)\,dx$

$=\frac{\partial}{\partial\theta}\cdot 1=0$ 이므로

$$E\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(X_i;\theta)\right)^2\right] = E\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right]E\left[\frac{\partial}{\partial\theta}\log f(X_j;\theta)\right]$$

$$= nE\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right]$$ 이다.

(X는 i와 j가 다를 때는 서로 독립임을 이용했다)

즉, $T(X)$가 θ의 비편향추정량이라고 한다면, $Var(T(X)) \geq \frac{1}{nI(\theta)}$이 성립한다.

E. 4.23

$X_1, \ldots, X_n \sim Poi(\lambda)$ Random Sample.

$I(\lambda) = E\left(\left(\frac{\partial}{\partial\lambda}\log f(X;\lambda)\right)^2\right) = E\left(\left(\frac{X}{\lambda}-1\right)^2\right) = E\left(\left(\frac{X-\lambda}{\lambda}\right)^2\right) = \frac{1}{\lambda}$

그런데 $Var(\bar{X}_n) = \frac{\lambda}{n}$이고 $\frac{1}{nI(\lambda)} = \frac{\lambda}{n}$으로 같기에 $\bar{X}_n$은 최소분산 비편향 추정량이다.

E. 4.24

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ Random sample.

$I(\mu) = E\left[\left(\frac{\partial}{\partial\mu}\log f(X;\mu)\right)^2\right] = \frac{1}{\sigma^2}$이고 $\frac{1}{nI(\theta)} = \frac{\sigma^2}{n}$이다. $Var(\bar{X}_n) = \frac{\sigma^2}{n}$이므로 $\bar{X}_n$은 MVUE이다.

· 꼭 하한을 만족해야 MVUE는 아니다. 하한보다 크더라도 MVUE가 될 수 있다.

· 완비충분통계량. (MVUE를 구하는 2번째 방법)

충분통계량

$S(X) = (S_1(X), \ldots, S_l(X))$를 l개 통계량의 벡터라고 하자.

이때, 조건부 확률변수 $(X_1, \ldots, X_n)|S(X)$의 분포가 모수 $\theta = (\theta_1, \ldots, \theta_k)$에 의존하지 않으면

$S(X)$를 결합충분통계량(jointly sufficient statistic)이라고 한다.

일변량 $(l=1)$의 경우는 $S(X)$가 θ의 충분통계량이라고 한다.

· 충분통계량에 대한 설명.

충분통계량의 값이 알려지면 표본의 결합분포는 더이상 모수에 의존하지 않는다.

즉 표본은 모수에 관한 어떤 정보도 가지고 있지 않다는 의미이다.

표본이 모수추정에 사용되는 이유는 확률변수 $X_1,\dots,X_n$ 각각은 모수의 함수가 아니지만

그 분포가 모수에 의존하기 때문이다.

하지만 충분통계량의 값이 조건부로 알려지면 더이상 분포가 모수에 의존하지 않으므로

모수추정에 사용할 근거를 잃어버린다.

따라서 반대로 $X_1,\dots,X_n$이 관찰되었을 때, 모수에 대한 정보는 모두 충분통계량 $S(x)$에

포함되어 있다고 할 수 있다.

그러므로 자료의 축소가 가능하다. 모수 추정이 목적이라면 충분통계량만 가지고 있어도

같은 효율을 가지는 추정값을 얻을 수 있다.

E. 4.25

$X_1,\dots,X_n \sim Ber(p)$ $\quad S=\sum_{i=1}^{n} X_i$

$f_S(s;p)=\binom{n}{s}p^s(1-p)^{n-s}\ (s=0,1,\dots,n)$ $\quad S=\sum_{i=1}^{n}X_i$가 p에 대한 충분통계량.

$$f(x_1,\dots,x_n\mid s)=p(X_1=x_1,\dots,X_n=x_n\mid S=s)=\frac{p(X_1=x_1,\dots,X_n=x_n,S=s)}{p(S=s)}=\begin{cases}\dfrac{p(X_1=x_1,\dots,X_n=x_n,S=s)}{p(S=s)}, & s=\sum_{i=1}^{n}x_i\\ 0 & \end{cases}$$

$$f(x_1,\dots,x_n\mid s)=\frac{f(x_1,\dots,x_n\mid s)}{f(s)}=\frac{p^s(1-p)^{n-s}}{\binom{n}{s}p^s(1-p)^{n-s}}=\frac{1}{\binom{n}{s}},\quad s=\sum_{i=1}^{n}x_i$$

$x_1,\dots,x_n$을 모두 알기에 순서에 따른 경우의 수는 사라진다.

$\frac{1}{\binom{n}{s}}$는 모수 p에 의존하지 않으므로 $S=\sum_{i=1}^{n}X_i$는 p의 충분통계량이다.

· $(S_1,\dots,S_q)$이 결합충분통계량이면 $(S_1,\dots,S_q)$의 1:1 함수 역시도 결합충분통계량이다.

But! 통계량이 충분한지를 따질 때, 복잡한 조건부 pdf를 구하는 과정이 필요하므로

대신 인수분해 정리(factorization theorem)를 이용하면 보다 간단히 확인할 수 있다.

또한 어떤 통계량이 충분통계량인지 결합 확률밀도함수로부터 찾을 수 있다.

인수분해 정리.

$X_1, \cdots, X_n$ 의 jpdf $f(x_1, \cdots, x_n ; \theta)$ 일 때, $S(X) = (S_1(X), \cdots, S_k(X))$를 k개의 통계량

이때 S가 결합 충분통계량일 필요충분조건은 jpdf가 s와 θ만의 함수인 g와, $(x_1, \cdots, x_n)$만의 함수 h와

곱으로 나타내어지는 것이다.

$$f(x_1, \cdots, x_n ; \theta) = g(s(x) ; \theta) \times h(x_1, \cdots, x_n)$$

<증명>

S가 결합충분통계량이고, S의 pdf를 $f_S(s;\theta)$로 표기하면

$$f(x_1, \cdots, x_n ; \theta) = f_S(s;\theta)\underline{f_{X|S}(x_1, \cdots, x_n | S)} = g(s;\theta)h(x_1, \cdots, x_n)$$

↳ S를 알면 더이상 θ에 의존하지 않는다.

E. 4.26

$X_1, \cdots, X_n \sim Bernoulli(\theta)$ Random Sample.

$$f(x_1, \cdots, x_n ; p) = \prod_{i=1}^{n} p^{x}(1-p)^{1-x} \; I(x=0,1) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} I_{\{0,1\}}(x_i)$$

$S = \sum_{i=1}^{n} X_i$ 일 때, $h(x_1, \cdots, x_n) = \prod_{i=1}^{n} I_{\{0,1\}}(x_i) = \begin{cases} 1 & x_i = 0,1 \quad i = 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$

$$g(s;p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} = p^{s}(1-p)^{n-s}$$

$f(x_1, \cdots, x_n ; p) = g(s;p)h(x_1, \cdots, x_n)$ 의 꼴로 표현가능하다. 따라서 $\sum_{i=1}^{n} X_i$는 충분통계량이다.

E. 4.27

$X_1, \dots, X_n \sim U(\theta_1, \theta_2)$ Random Sample.

$$f(x_1, \dots, x_n ; \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} I_{(\theta_1, \theta_2)}(x_i) = \left(\frac{1}{\theta_2 - \theta_1}\right)^n \prod_{i=1}^{n} I_{(\theta_1, \theta_2)}(x_i)$$

$$= \left(\frac{1}{\theta_2 - \theta_1}\right)^n I(\theta_1 \le x_{(1)} \text{ and } x_{(n)} \le \theta_2)$$

E. 4.28

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ Random Sample.

$f(x_1, \dots, x_n ; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2)}$ 이다.

$f(x_1, \dots, x_n ; \mu, \sigma^2) = g(S(x) ; \mu, \sigma^2) \cdot h(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2)} \times 1$ 이다.

모수가 2개 이므로 $S = \sum_{i=1}^{n} x_i$, $S^2 = \sum_{i=1}^{n} x_i^2$ 가 결합충분통계량이다. 그리고 $\bar{X}_n$, $\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$도 함수꼴이므로 충분통계량이다.

E. 4.29

$X_1, \dots, X_n \sim f(x ; \theta)$ Random Sample.

순서통계량은 결합충분통계량이다. 주어진 $x_{(1)}, \dots, x_{(n)}$에 대하여 $(X_1, \dots, X_n)$이 가질 수 있는 값은 $n!$개이다.

그리고 확률도 $\frac{1}{n!}$이므로 모수 θ에 의존하지 않는다.

충분통계량은 유일하지 않다. 그렇다면 어떤 충분통계량을 사용?

↳ 모수의 정보를 잃지 않으면서 자료의 축소를 많이 하는 것!

즉 모수와 관련없는 정보를 가장 적게 가지고 있는 것을 선호!

↳ 그런 것을 최소 충분통계량이라고 한다.

S, U 모두 충분통계량일 때,

S가 U의 함수라면 S가 최소 충분통계량이 되고

U를 알면 S를 알기 때문에 U는 S보다 더 큰 정보를 가진다.

라오-블랙웰 정리 : 충분통계량을 활용하는 추정방법의 원리

모수의 함수 $g(\theta)$에 대해서 S가 충분통계량이고, $T(X)$가 그의 한 비편향추정량이라고 하자.

이때, $\delta(S) = E(T(X)|S)$ 라고 하면 $\delta(S)$도 역시 $g(\theta)$의 비편향 추정량이며,

모든 θ에 대하여 $Var(\delta(S)) = E[(\delta(S) - g(\theta))^2] \le E[(T(X) - g(\theta))^2] = Var(T(X))$ 가 성립한다.

<증명>

$S(s)$는 모수에 의존X, $E(S(s)) = E(T(X)) = g(\theta)$ 이므로 비편향추정량이다.

(정리. 2.13) $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$ 이므로 $Y=T(X)$, $X=S$라고 하면

$$Var(T(X)) = E(\underbrace{Var(T(X)|S)}_{E[(T(X)-E(T(X)|S))^2|S] = E[(T(X)-S(S))^2|S]}) + Var(E(T(X)|S))$$

$$= E[E(T(X)-S(S))^2|S] + Var(S(S)) \geq Var(S(S))$$

위의 내용은 U.E $T(X)$가 존재하면 충분통계량 S에 대한 조건부 기댓값 $E(T(X)|S)$ 역시 UE이며, $T(X)$보다 작거나 같은 분산을 가진다.

따라서 MVUE를 찾을 때, 그 후보를 충분통계량의 함수 꼴로 국한시킬 수 있다.

E. 4.30

$X_1, \cdots, X_n \sim Ber(p)$ Random Sample.

X_1는 모수 p의 U.E이다. $(E(X_1)=p)$ 그리고 충분통계량 $S=\sum_{i=1}^n X_i$ 가 주어졌을 때,

X_1의 조건부 확률분포를 살피면서 이보다 작은 분산을 갖는 U.E를 찾아보자.

X_1은 1 아니면 0을 가지기에 이산형 확률변수.

$$p(X_1=0 | \textstyle\sum_{i=1}^n X_i = s) = \frac{p(X_1=0, \sum_{i=1}^n X_i = s)}{p(\sum_{i=1}^n X_i = s)} = \frac{p(X_1=0, \sum_{i=2}^n X_i = s)}{p(\sum_{i=1}^n X_i = s)} = \frac{p(X_1=0)p(\sum_{i=2}^n X_i = s)}{p(\sum_{i=1}^n X_i = s)}$$

$$= \frac{\cancel{(1-p)}\binom{n-1}{s}\cancel{p^s}\cancel{(1-p)^{n-1-s}}}{\binom{n}{s}\cancel{p^s}\cancel{(1-p)^{n-s}}} = \frac{\frac{(n-1)!}{(n-1-s)!s!}}{\frac{n!}{(n-s)!\cdot s!}} = \frac{\overset{n-s}{\cancel{(n-s)!}}\cancel{s!}\cancel{(n-1)!}}{\underset{n}{\cancel{n!}}\cancel{(n-1-s)!}\cancel{s!}} = \frac{n-s}{n}$$

$$p(X_1=1 | \textstyle\sum_{i=1}^n X_i = s) = \frac{p(X_1=1, \sum_{i=1}^n X_i = s)}{p(\sum_{i=1}^n X_i = s)} = \frac{p(X_1=1, \sum_{i=2}^n X_i = s-1)}{p(\sum_{i=1}^n X_i = s)} = \frac{p(X_1=1)p(\sum_{i=2}^n X_i = s-1)}{p(\sum_{i=1}^n X_i = s)}$$

$$= \frac{\cancel{p}\cdot\binom{n-1}{s-1}\cancel{p^{s-1}}\cdot\cancel{(1-p)^{n-s}}}{\binom{n}{s}\cancel{p^s}\cancel{(1-p)^{n-s}}} = \frac{\binom{n-1}{s-1}}{\binom{n}{s}} = \frac{\frac{(n-1)!}{(n-s)!(s-1)!}}{\frac{n!}{(n-s)!s!}} = \frac{\cancel{(n-s)!}\overset{s}{\cancel{s!}}\cancel{(n-1)!}}{\underset{n}{\cancel{n!}}\cancel{(n-s)!}\cancel{(s-1)!}} = \frac{s}{n}$$

$$E(X_1 | \textstyle\sum_{i=1}^n X_i = s) = 0 \times \frac{n-s}{n} + 1 \times \frac{s}{n} = \frac{s}{n}$$

즉 $\frac{S}{n}$는 p의 새로운 U.E가 된다. 라오-블랙웰 정리에 따르면 $\frac{S}{n} \le p(1-p)$ 이다.

실제로 $\frac{\sum_{i=1}^{n} X_i}{n}$ 의 분산도 $\frac{p(1-p)}{n}$ 이다.

충분통계량의 함수로서 비편향 추정량이 유일한 경우 그것이 유일한 MVUE가 된다는 것을 알 수 있다.

예를 들어 충분통계량 S가 있어서 이 충분통계량의 함수로서 UE $S^*(S)$로 유일하다고 해보자.

또 $S^*(S)$ 보다 더 좋은 UE가 없는지 확인하기 위해 어떤 UE $T(X)$를 가정하자.

만약 $T(X)$가 S의 함수가 아니라면 $E(T(X)|S)$가 더 좋은 U.E가 될 것이고 앞에서 가정한 유일성 때문에

$E(T(X)|S) = S^*(S)$가 될 것이다.

그리고 $T(X)$가 S의 함수라면 그 자체가 $S^*(S)$와 같을 수 밖에 없다.

이 유일성(uniqueness)을 다루는데 있어 중요한 도구가 되는 완비성(Completeness)에 대해 알아보자.

Random Sample $X_1, \cdots, X_n$으로부터 계산된 통계량 $S=S(X_1, \cdots, X_n)$에 대하여

$E(g(S))=0$ 을 모든 $\theta \in \Omega$에 대해 만족하는 θ에 무관한 함수 g가 $g(\cdot)\equiv 0$ 뿐이라면

S를 완비통계량이라고 한다. 만약 S가 θ에 대한 충분통계량이라면, 이를 완비충분통계량이라고 한다.

완비통계량은 <u>통계량에 대한 함수가 모든 모수에 대해 0이 될 때</u>, 그 함수가 항상 0인 경우밖에 없을 때,

↳ 모수에 상관없이 항상 0일 때,

완비성에 대한 설명.

완비성은 추정량을 찾을 때, 불필요한 정보를 제거하는 개념이다.

어떤 통계량 $S(X)$가 완비통계량이라는 것은

"어떤 함수 $g(S)$가 기대값이 0이 되려면, 그 함수가 항상 0일 수밖에 없다"는 뜻이다.

즉, 불필요한 정보는 포함하지 않는 통계량이라는 것이다.

어떤 충분통계량 $S(X)$가 완비통계량이려면, 아무리 복잡한 함수 $g(S)$를 만들어도,

$E[g(T)]=0$ 일 때, 0이 될 수 있는 경우가 $g(t)$가 0이 되는 경우밖에 없다면

완비성을 만족한다.

E.4.31

$X_1, \cdots, X_n \sim Ber(p)$ Random Sample.

예를 들어 $X_2 - X_1$은 완비통계량이 아니다. $E(X_1 - X_2) = 0$ 이지만 $X_1 - X_2$는 항상 0이 아니기 때문이다.

$T = \sum_{i=1}^n X_i$가 완비충분통계량인지 살펴보자.

$T \sim B(n,p)$ 이므로 $E[g(T)] = \sum_{t=0}^n g(t)\binom{n}{t} p^t (1-p)^{n-t} = (1-p)^n \sum_{t=0}^n g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t = 0$ 이 되려면

$\binom{n}{t}\left(\frac{p}{1-p}\right)^t$, $(1-p)^n$은 $0<p<1$ 이므로 절대 0이 될 수 없으므로 결론적으로 $g(t)$가 0이 되어야한다.

즉 $g(t)=0$이다.

E.4.32

$X_1, \cdots X_n \sim U(0,\theta)$ Random Sample.

$f_{X_{(n)}}(y) = n \cdot F(y)^{n-1} f(y) = n \cdot \left(\frac{y}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n y^{n-1}}{\theta^n}$, $0<y<\theta$ 이다.

$E[g(X_{(n)})] = \int_0^\theta g(y) \cdot \frac{n y^{n-1}}{\theta^n} dy = \frac{n}{\theta^n} \int_0^\theta g(y) y^{n-1} dy$ 에서 $E[g(X_{(n)})]=0$ 이라는 것은

$\int_0^\theta g(y) y^{n-1} dy$가 0이어야 한다. 이 식을 θ에 미분하면 $g(\theta)\theta^{n-1}=0$이 모든 $\theta>0$에 대하여 성립함을 의미하고

θ^{n-1}은 0이 될 수 없으므로 $g(\theta)$가 0이 되어야한다.

즉 $X_{(n)}$은 완비성을 만족하고 충분통계량이므로 완비충분통계량이다.

레만 - 쉐페 정리.

모든 θ에 대하여 S가 완비충분통계량이고, $T(X)$가 $g(\theta)$의 비편향추정량이라고 하자.

$\delta(S) = E(T(X)|S)$는 $g(\theta)$의 유일한 최소분산 비편향 추정량 (MVUE) 이다.

라오 블랙웰 정리에 따라서 기존의 불편추정량보다 더 작은 분산추정량을 찾을 수 있지만 그것이 항상 최소분산을 가지는 것은 아니다. 하지만 충분통계량이 추가적으로 완비성이 갖춰지면 비편향추정량은 유일해지며 라오 블랙웰 정리에 따라 유일한 MVUE가 된다.

E. 4.33

$X_1, \cdots, X_n \sim U(0, \theta)$ Random Sample.

$X_{(n)}$이 완비충분통계량임은 보였다. 그리고 $E\left(\frac{n+1}{n} X_{(n)}\right) = \theta$ 이므로 $\frac{(n+1)X_{(n)}}{n}$은 유일한 MVUE이다.

이와 같이 완비충분통계량은 점추정에서 매우 유용한 역할을 한다. 그러나 복잡한 계산을 필요로 한다.
그러나 다음에 정의되는 지수족은 완비충분성이 보장되고, 자주 사용되는 많은 확률분포가 지수족에 속하므로 이를 쉽게 활용할 수 있다.

확률밀도함수가 함수 a, b, c_i, t_i $(i=1, \cdots, k)$에 대하여

$$f(x;\theta) = a(\theta)\, b(x)\, e^{\sum_{i=1}^{k} c_i(\theta) t_i(k)}, \quad -\infty < x < \infty$$

$\theta = \theta_1, \cdots, \theta_k$로 표현되면 이를 k개의 모수 $\theta_1, \cdots, \theta_k$를 가진 지수족에 속한다고 한다.

여기서 $x : f(x;\theta) > 0$는 모수 θ에 의존하지 않는다.

E.4.34

$f(x;p) = p^x(1-p)^{1-x} = (1-p)e^{x \cdot \log\left(\frac{p}{1-p}\right)}$ 로 표현되므로 $a(p) = 1-p$, $b(x) = 1$, $c(p) = \log\left(\frac{p}{1-p}\right)$, $t(x) = x$ 이다.

$\{x : f(x;p) > 0\}$는 $\{0, 1\}$로서 모수에 의존하지 않는다.

따라서 베르누이 분포는 모수가 하나인 지수족에 속한다.

$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2, \cdots$ $= e^{-\lambda} \cdot \frac{1}{x!} e^{x \log \lambda}$로 표현 가능.

$a(\lambda) = e^{-\lambda}$, $b(x) = \frac{1}{x!}$, $c(\lambda) = \log \lambda$, $t(x) = x$로 지수족에 속한다.

$\{x : f(x;\lambda)>0\} = \{0,1,2,\cdots\}$로 모수에 의존 X.

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x}$$

$a(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}}$, $b(x)=1$, $c_1(\mu,\sigma^2) = -\frac{1}{2\sigma^2}$, $c_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2}$, $t_1(x)=x^2$, $t_2(x)=x$ 로 표현가능하다.

여기서 $\{x : f(x;\mu,\sigma^2)>0\}$ 실수전체이므로 모수에 의존하지 않는다.

이외에 지수족에 속하는 확률분포들은 지수분포, 베타분포, 감마분포, 음이항분포들이 있다.

그러나 $U(0,\theta)$ 분포같이 $f(x;\theta)>0$ 인 x의 값이 모수 θ에 의존하는 경우는 지수족에 포함되지 않는다.

정리 4.7

지수족에 속하는 pdf $f(x;\theta) = a(\theta)b(x)e^{\sum_{i=1}^{k} c_i(\theta)t_i(x)}$ 로부터 랜덤표본 $X_1,\cdots,X_n$을 얻었다면,

통계량 $S_1 = \sum_{i=1}^{n} t_1(X_i), \cdots, S_k = \sum_{i=1}^{n} t_k(X_i)$는 모수 $\theta_1,\cdots,\theta_k$에 대한 결합완비충분통계량이다.

$S_1,\cdots,S_k$의 함수로 만들어지는 $\theta_1,\cdots,\theta_k$의 비편향추정량은 $\theta_1,\cdots,\theta_k$의 MVUE가 된다.

(완비충분통계량의 1:1 함수도 완비충분통계량이 된다.)

E.4.35

$X_1,\cdots,X_n \sim Ber(p)$ Random Sample.

$f(x_1,\cdots,x_n;p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$ $(x_i = 0,1,\ 0<p<1)$

이때 $S = \sum_{i=1}^{n} X_i$는 모수 p에 대한 완비충분통계량이라는 것을 보였다.

$Var(X) = p(1-p)$의 MVUE를 구하고자 한다. 이때 완비충분통계량의 함수인 $\bar{X}_n(1-\bar{X}_n)$을 사용하면,

$E(\bar{X}_n(1-\bar{X}_n)) = E(\bar{X}_n) - E(\bar{X}_n^2) = p - (p^2 + Var(\bar{X}_n)) = p - p^2 - \frac{p(1-p)}{n} = p(1-p)(1-\frac{1}{n})$이 되므로

$\bar{X}_n(1-\bar{X}_n) \times \frac{1}{1-\frac{1}{n}} = \bar{X}_n(1-\bar{X}_n)\cdot\frac{n}{n-1} = \frac{n\bar{X}_n(1-\bar{X}_n)}{n-1}$ 은 완비충분통계량의 함수이며, 비편향추정량이므로

$Var(X)$의 MVUE이다.

E.4.36

$X_1, \cdots, X_n \sim poI(\lambda)$ Random Sample.

$f(x_1, \cdots, x_n ; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$, $(x_i = 0, 1, 2, \cdots)$ 가 된다.

$\sum_{i=1}^n X_i$는 완비충분통계량이며 그의 1:1 함수 $\bar{X}_n$도 완비충분통계량이다.

λ의 추정량으로 $\bar{X}_n$을 사용한다면 $E(\bar{X}_n) = \lambda$이므로 $\bar{X}_n$은 MVUE가 된다.

E.4.37

$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ Random Sample.

$f(x_1, \cdots, x_n ; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} + \mu \cdot \frac{\sum_{i=1}^n x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}$ 으로 변형되어지므로

$\sum_{i=1}^n x_i^2$, $\sum_{i=1}^n x_i$는 각각 예 4.34, 정리 4.7의 의해 결합 완비충분통계량이다.

이들의 함수인 $\bar{X}_n$과 $\sum_{i=1}^n (X_i - \bar{X}_n)^2 / n-1$은 각각 μ와 σ^2의 비편향추정량이므로 이들은 MVUE이다.

만약 μ가 알려져 있다면 $\sum_{i=1}^n (X_i - \mu)^2$은 완비충분통계량이며, $\sum_{i=1}^n (X_i - \mu)^2 / n$은 σ^2의 MVUE가 된다.

σ^2의 최대가능도 추정량도 $\sum_{i=1}^n (X_i - \mu)^2 / n$이므로 μ를 아는 경우는 같지 알려진 μ를 모르면 같지 다르게 된다.

바수정리 (완비충분통계량의 성질을 나타내는 Basu의 정리를 살펴보자)

$S = (S_1, \cdots, S_k)$가 모두 $(\theta_1, \cdots, \theta_k)$의 결합 완비충분통계량이라고 하자. 이때 어떤 통계량 T의 분포가

모든 $\theta = (\theta_1, \cdots, \theta_k)$에 의존하지 않는다면 S와 T는 독립이다.

확률밀도함수가 θ에 의존하지 않는 통계량 T를 θ에 대한 보조통계량이라고 한다.

바수의 정리를 이용하면, 두 통계량의 독립을 보이는데 있어서 때로는 두 통계량의 복잡한 결합분포를 구하지 않아도 된다.

E.4.38

$X_1, \cdots, X_n$을 $N(\mu, \sigma^2)$ Random Sample.

$\hat{\sigma}^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 / n$ $\hat{\mu} = \bar{X}_n$을 생각해보자.

$\bar{X}_n$은 μ의 완비충분통계량이고, $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-1)}$이므로 이는 μ에 의존하지 않는다.

즉, $\frac{n\hat{\sigma}^2}{\sigma^2}$은 μ에 대한 보조통계량이 되고, 바수의 정리에 따라 $\hat{\mu}=\bar{X}_n$과 $\hat{\sigma}^2=\sum_{i=1}^{n}(X_i-\bar{X}_n)^2/n$은 서로 독립이다.