

1. 1. 확률변수 X 의 적률생성함수 $M(t)$ 가 존재한다고 가정할 때, 다음을 보이시오.

$$\begin{cases} \Pr(X \geq a) \leq e^{-ta} M(t), & t > 0 \\ \Pr(X \leq a) \leq e^{-ta} M(t) \end{cases}$$

2. X_1, \dots, X_n 이 $\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$ 를 만족하는 분포에서 나온 임의표본일 때, $S_n = \sum_{i=1}^n X_i$ 에 대하여 다음이 성립함을 1을 이용하여 보이시오.

$$\Pr(S_n \geq a) \leq e^{-ta} e^{nt^2/2}, \quad t > 0$$

3. 어떤 도박꾼이 매 배팅에서 같은 확률로 1원을 따거나 잃는 게임을 하려고 한다고 하자. 매 배팅에서의 결과는 서로 독립이라 할 때, 처음 10번의 배팅에서 적어도 8번 이상 이기는 확률의 상한을 2를 이용하여 구하시오. 또, 이 확률의 정확한 값도 계산하시오.

Solution:

1. 이를 *Chernoff's bound*라고 한다. 마코프 부등식(Markov inequality)로부터 간단하게 도출할 수 있다.

$$\Pr(X \geq a) = \begin{cases} \Pr(tX \geq at) = \Pr(e^{tX} \geq e^{at}), & t > 0 \\ \Pr(tX \leq at) = \Pr(e^{tX} \leq e^{at}), & t < 0 \end{cases} \quad (1)$$

$$= \begin{cases} \Pr(X \geq a) \leq e^{-ta} \mathbb{E}(e^{tX}), & t > 0 \\ \Pr(X \leq a) \leq e^{-ta} \mathbb{E}(e^{tX}), & t < 0 \end{cases} \quad (2)$$

(1)에서 (2)로 넘어가는 것은 마코프 부등식이다.

2. 이 문제는 표준정규분포를 가정하는 것보다 정확한 분포를 알고 그 분포의 적률생성함수를 이용하는 것이 더 *tight*한 bound를 준다는 것을 증명하는 것이다. 왜냐하면 표준정규확률변수 n 개를 더한 변수의 적률생성함수가 $e^{nt^2/2}$ 이기 때문이다.

$$X \mid \begin{matrix} 1 & -1 \end{matrix} \quad (3)$$

$$e^{tX} \mid \begin{matrix} e^t & e^{-t} \end{matrix} \quad (4)$$

$$\Pr(X = x) \mid \begin{matrix} \frac{1}{2} & \frac{1}{2} \end{matrix} \quad (5)$$

이로서 다음과 같은 사실을 얻을 수 있다.

$$E(e^{tX}) = \frac{1}{2}(e^t + e^{-t}) \implies E(e^{tS_n}) = \left(\frac{e^t + e^{-t}}{2}\right)^n \quad (6)$$

따라서 $2^{-n}(e^t + e^{-t})^n \leq e^{nt^2/2}$ 임을 보이면 된다.

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} \quad (7)$$

$$e^{-t} = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} \quad (8)$$

$$e^{t^2/2} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{t^2}{2}\right)^k = \sum_{k=0}^{\infty} \frac{t^{2k}}{k! \times 2^k} \quad (9)$$

(7)과 (8)을 더하면 k 가 홀수인 부분은 소거되고 짝수인 부분만 두 배가 된다. 따라서

$$\frac{1}{2}(e^t + e^{-t}) = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \quad (10)$$

분모를 비교하면 $k! \times 2^k < (2k)!$ 이므로 증명끝.

3. 8번 이기면 2번은 진 것이므로 $S_n = 6$ 이고 9번 이기면 1번은 진 것이므로 $S_n = 8$, 10번 모두 이기면 $S_n = 10$ 이다. 2번을 이용하면

$$\Pr(S_n \geq 6) \leq e^{5t^2 - 6t} \quad (11)$$

이므로 최대값은

$$5\left(t^2 - \frac{6}{5}t + \frac{9}{25}\right) - \frac{9}{5} = 5\left(t - \frac{3}{5}\right)^2 - \frac{9}{5} \quad (12)$$

$t = 3/5$ 에서 $e^{-9/5}$ 가 된다. 정확한 확률값은

$$\Pr(S_n = 6) + \Pr(S_n = 8) + \Pr(S_n = 10) = \left(\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right) \left(\frac{1}{2}\right)^{10} \quad (13)$$

$$= \frac{7}{128} \quad (14)$$

2. Let X_1, X_2, \dots, X_n be a random sample from a uniform distribution on $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. Here the unknown parameters are μ and σ , which are the population mean and standard deviation.

1. Obtain the probability density function of X_i .
2. Obtain the likelihood function of μ and σ .

3. Obtain the maximum-likelihood estimators (MLEs) of μ and σ .
4. Obtain the method-of-moments estimators of μ and σ .

Solution:

1. $f_{X_i}(x_i) = (2\sqrt{3}\sigma)^{-1}$ for $\mu - \sqrt{3}\sigma < x_i < \mu + \sqrt{3}\sigma$

2. 가능도함수의 경우에는 가변수를 통해 support를 명시해 주어야 한다.

$$L(\mu, \sigma; x_1, \dots, x_n) = (2\sqrt{3}\sigma)^{-n} \prod_{i=1}^n I_{(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)}(x_i) \quad (15)$$

$$= (2\sqrt{3}\sigma)^{-n} I_{(-\infty, x_{(1)})}(\mu - \sqrt{3}\sigma) I_{(x_{(n)}, \infty)}(\mu + \sqrt{3}\sigma) \quad (16)$$

가능도함수 자체는 σ 에만 의존하고 단조감소함수이므로 σ 가 최대한 작아야 한다. σ 의 범위가

$$\frac{1}{\sqrt{3}}(\mu - x_{(1)}) < \sigma \quad (17)$$

$$\frac{1}{\sqrt{3}}(x_{(n)} - \mu) < \sigma \quad (18)$$

이기 때문에 부등식의 영역을 통해 σ 의 최솟값은 $\mu = 2^{-1}(x_{(1)} + x_{(n)})$ 일 때이고 그 값은

$$\hat{\sigma} = \frac{1}{2\sqrt{3}}(x_{(n)} - x_{(1)}) \quad (19)$$

이다. 요약하자면

$$\hat{\mu} = \frac{1}{2}(X_{(1)} + X_{(n)}) \quad (20)$$

$$\hat{\sigma} = \frac{1}{2\sqrt{3}}(X_{(n)} - X_{(1)}) \quad (21)$$

3. $X \sim \text{Unif}(a, b)$ 이라 할 때

$$E(X) = \frac{a+b}{2} \quad (22)$$

$$\text{Var}(X) = \frac{(b-a)^2}{12} \quad (23)$$

이므로

$$\frac{1}{n} \sum_{i=1}^n X_i = 2\mu \quad (24)$$

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{(2\sqrt{3}\sigma)^2}{12} + 4\mu^2 \quad (25)$$

정리하면

$$\hat{\mu}^{\text{MME}} = \frac{1}{2n} \sum_{i=1}^n X_i \quad (26)$$

$$\hat{\sigma}^{\text{MME}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \quad (27)$$

이고 $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ 이다.

3. Let X_1, X_2, \dots, X_n be a random sample from a Poisson distribution with density

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!},$$

where $\lambda \geq 0$ and $x = 0, 1, 2, \dots$. Let $T = \sum_{i=1}^n X_i$.

1. Show that T is complete and sufficient for λ .
2. Show that T/n is the uniformly minimum variance estimator (UMVUE) of λ .
3. Show that $T(T-1)\cdots(T-k+1)/n^k$ is the UMVUE of λ^k for any positive k .

Solution:

1. 포아송 분포에서 표본의 합이 완전통계량(complete statistic)이 됨은 2가지 방법으로 증명할 수 있다.

- 다음은 완전통계량의 정의를 이용한 방법이다. 통계량 $T(X_1, \dots, X_n)$ 에 대해 $f(t; \theta)$ 가 T 의 밀도함수 혹은 질량함수라 할 때 다음이 성립하면 그 분포족을 ‘완전(complete)’하다 하고, 이와 동치인 명제는 T 가 완전통계량이라는 것이다. 모든 모수값 θ 에 대해

$$E(r(T)) = 0 \implies \Pr(r(T) = 0) = 1.$$

주의해야 하는 것은 ‘완전성(completeness)’는 하나의 분포족(a family of distributions)의 특성이라는 것이다. 문제의 경우 우리의 통계량은

$$T \sim \text{Poi}(n\lambda)$$

을 따르기 때문에 다음을 보여야 한다.

$$E(r(T)) = 0 \implies \Pr(r(T) = 0) = 1 \quad (28)$$

포아송 분포의 특징 상 지수함수는 음수일 수 없으므로

$$E(r(T)) = \sum_{t=0}^{\infty} r(t) \frac{e^{-n\lambda} (n\lambda)^t}{t!} \quad (29)$$

$$= \sum_{t=0}^{\infty} r(t) \frac{(n\lambda)^t}{t!} \quad (30)$$

$$= 0 \quad (31)$$

모든 λ 에 대해 $n\lambda > 0$ 이므로 모든 $r(t) = 0$ 이다.

Measure-theoretic하게는 적분의 vanishing property에 의해

$$\int r(T) dP = 0 \implies r(T) = 0 \text{ } P\text{-a.e} \quad (32)$$

이다.

- 두 번째 증명은 지수족에서의 완전통계량 정리를 이용하는 것이다.

즉, 임의표본 X_1, \dots, X_n 가 지수족에 속하는 분포에서 얻은 것일 (모수도 일반적인 상황을 가정해서 k 개로 한다.) 때 밀도함수는

$$f(x; \theta) = h(x) c(\theta) \exp \left(\sum_{j=1}^k \eta_j(\theta) t_j(x) \right)$$

이고 결합밀도함수는

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \left(\prod_{i=1}^n h(x_i) \right) c(\theta)^n \exp \left(\sum_{j=1}^k \left(\eta_j(\theta) \sum_{i=1}^n t_j(x_i) \right) \right)$$

이며 이때

$$T(X) = \left(\sum_{i=1}^n t_1(X_i), \dots, \sum_{i=1}^n t_k(X_i) \right)$$

는 완전통계량이다. 포아송분포는 결합밀도함수가

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \lambda) = \left(\prod_{i=1}^n \frac{1}{x_i!} \right) e^{-n\lambda} \exp \left(\ln \lambda \cdots \sum_{i=1}^n x_i \right) \quad (33)$$

이기 때문에 $\sum_{i=1}^n X_i$ 이 완전통계량이 된다.

4. 세 변수에 대한 n 개의 관찰값 (x_{i1}, x_{i2}, Y_i) , $(i = 1, \dots, n)$ 에 대하여

$$E(Y_i | x_{i1}, x_{i2}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \text{ 그리고 } \text{Var}(Y_i | x_{i1}, x_{i2}) = \sigma^2$$

이라 하자.

1. 모수 $\beta_0, \beta_1, \beta_2$ 의 최소제곱추정량 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ 을 유도하시오.
2. $s^2 = \frac{1}{n-3} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2$ 이 σ^2 에 대한 불편추정량이 됨을 보이시오.

Solution:

5. X_n 이 자유도가 n 인 카이제곱분포를 따른다고 한다. 이 때

$$Z_n = \frac{X_n - n}{\sqrt{2n}}$$

이라고 하자. n 이 무한대로 접근함에 따라 Z_n 의 분포가 어느 분포에 수렴하는지를 자세히 보아라. (힌트: 자유도가 n 인 카이제곱확률변수는 자유도가 1인 서로 독립인 카이제곱확률변수의 합으로 표현될 수 있다.)