6장 추정

1. 많이 쓰이는 추정량의 정의와 성질
2. 점추정량
3. 신뢰구간의 추정
4. 분산의 신뢰구간
5. 평균의 신뢰구간 〈모분산을 모르는 경우〉
6. 최소분산 불편추정량과 충분 통계량
7. 추정량의 기타 성질

1. 많이 쓰이는 추정량의 정의와 성질

[정의 6.1-1]

모수를 추정하는 통계량을 **추정량** (Estimator)이라 하고 추정량의 값을 **추정값** 혹은 **추정치** (Estimate)라고 한다.

[정의 6.1-2]

θ의 추정량 $\hat{\theta}$에 대하여 $E(\hat{\theta}) = \theta$인 경우 $\hat{\theta}$을 θ의 **불편 추정량** (Unbiased estimator)이라고 한다. $E(\hat{\theta}) \neq \theta$인 경우에는 $\hat{\theta}$을 θ의 **편의 추정량** (Biased estimator)이라고 한다.

$E(\hat{\theta}) - \theta$를 θ의 **편의** (Bias)라고 하며 $Bias(\hat{\theta})$으로 표기한다. 위의 정의에서 $\hat{\theta}$이 θ의 불편 추정량이면 $Bias(\hat{\theta}) = 0$ 임은 자명하다.

예 6.1-1 동전을 n회 던지는 실험에서 앞면이 나타나는 회수를 X라고 하자. 앞면이 나타나는 비율 p의 추정량으로서 $\hat{p} = \frac{X}{n}$를 사용한다고 하면 X는 이항분포 $B(n, p)$를 따르므로

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n} \cdot np = p$$

이다. 즉, $\hat{p}$는 p의 불편 추정량이 된다. 한편 p의 추정량으로서 $\tilde{p} = \frac{X+1}{n+2}$를 사용한다고 하면

$$E(\tilde{p}) = E\left(\frac{X+1}{n+2}\right) = \frac{1}{n+2}\{E(X)+1\} = \frac{1}{n+2}(np+1) \neq p$$

이므로 $\tilde{p}$는 p의 편의 추정량이 된다. 편의는 다음과 같다.

$$Bias(\tilde{p}) = E(\tilde{p}) - p = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}.$$

예 6.1-2 $X_1, X_2, \cdots, X_n$을 평균이 μ이고 분산이 σ^2인 분포에서의 확률표본이라고 하자. σ^2의 추정량으로서 다음의 두 가지 추정량이 자주 이용된다.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

두 추정량의 불편성을 조사해 본다.

$\sigma^2 = Var(X_i) = E(X_i^2) - \mu^2$이므로 $E(X_i^2) = \mu^2 + \sigma^2$ 이고,

$Var(\overline{X}) = \frac{1}{n}\sigma^2 = E(\overline{X}^2) - \mu^2$이므로 $E(\overline{X}^2) = \mu^2 + \frac{1}{n}\sigma^2$이다.

이를 이용하면

$$\begin{aligned} E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] &= E[\sum X_i^2 - n\overline{X}^2] = \sum E(X_i^2) - nE(\overline{X}^2) \\ &= n(\sigma^2 + \mu^2) - n\left(\mu^2 + \frac{1}{n}\sigma^2\right) \\ &= (n-1)\sigma^2 \end{aligned}$$

이므로 $E[S^2] = \sigma^2$, $E[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2$임을 알 수 있다. 즉, S^2은 σ^2의 불편 추정량이나 $\hat{\sigma}^2$는 σ^2의 불편 추정량이 아니다.

[정의 6.1-3]

θ의 추정량 $\hat{\theta}$과 θ의 거리 $|\hat{\theta} - \theta|$를 θ의 **추정오차** (Error of estimate)라 하고 $E[|\hat{\theta} - \theta|^2] = E(\hat{\theta} - \theta)^2$을 θ의 **평균제곱오차** (Mean square error)라고 하며 $MSE(\hat{\theta})$로 표기한다.

정리 6.1-1 추정량 $\hat{\theta}$에 대한 다음의 등식이 성립한다.

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2.$$

증명 $\mu = E(\hat{\theta})$라 두면

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta}-\theta)^2 = E(\hat{\theta}-\mu+\mu-\theta)^2 \\ &= E(\hat{\theta}-\mu)^2 + 2E(\hat{\theta}-\mu)(\mu-\theta) + E(\mu-\theta)^2 \\ &\quad \langle \mu\text{와 } \theta\text{는 상수이므로} \rangle \\ &= E(\hat{\theta}-\mu)^2 + 2(\mu-\theta)E(\hat{\theta}-\mu) + (\mu-\theta)^2 \\ &\quad \langle E(\hat{\theta}-\mu) = E(\hat{\theta}) - \mu = 0\text{이므로} \rangle \\ &= E(\hat{\theta}-\mu)^2 + (Bias(\hat{\theta}))^2. \end{aligned}$$

예 6.1-3 ([예 6.1-1]의 계속) [예 6.1-1]에서 동전을 독립적으로 n회 던져서 앞면이 나오는 회수를 X라고 할 때, 앞면이 나오는 비율 p의 추정량으로서 $\hat{p} = \frac{X}{n}$, $\tilde{p} = \frac{X+1}{n+1}$을 제시하고 $\hat{p}$는 p의 불편 추정량이며 $\tilde{p}$는 편의 추정량으로서 $\tilde{p}$의 편의를 구하였다. 이제 $\hat{p}$과 $\tilde{p}$의 평균제곱오차를 구해보도록 한다. X는 이항분포 $B(n, p)$를 따르므로 $E(X) = np$, $Var(X) = np(1-p)$이며, $Bias(\hat{p}) = 0$, $Bias(\tilde{p}) = \frac{1-2p}{n+2}$ 이므로

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n},$$

$$Var(\tilde{p}) = Var\left(\frac{X+1}{n+2}\right) = \frac{1}{(n+2)^2}np(1-p) = \frac{np(1-p)}{(n+2)^2}$$

이다. 따라서 구하는 평균제곱오차는 다음과 같다.

$$MSE(\hat{p}) = Var(\hat{p}) = \frac{p(1-p)}{n},$$

$$\begin{aligned} MSE(\tilde{p}) &= Var(\tilde{p}) + (Bias(\tilde{p}))^2 = \frac{np(1-p)}{(n+2)^2} + \frac{(1-2p)^2}{(n+2)^2} \\ &= \frac{1+(n-4)p-(n-4)p^2}{(n+2)^2}. \end{aligned}$$

예 6.1-4 ([예 6.1-2]의 계속) [예 6.1-2]에서 $X_1, X_2, \cdots, X_n$이 정규분포 $N(\mu, \sigma^2)$에서의 확률표본이라고 할 때, σ^2의 추정량으로서 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$과 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$의 불편성을 다루었다. 이제는 평균제곱오차를 구해보도록 한다.

〈주의 : $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2$는 자유도가 $n-1$인 카이제곱분포 $\chi^2(n-1)$을 따른다는 성질을 이용할 것임. 이는 다음 장에서 증명하도록 함.〉

카이제곱분포의 기대값은 자유도와 같고 분산은 자유도의 2배임을 이미 앞에서 공부하였다. 이를 이용하면

$$E[\sum(X_i-\overline{X})^2] = (n-1)\sigma^2,$$
$$Var[\sum(X_i-\overline{X})^2] = 2(n-1)\sigma^4$$

이므로 $Var(S^2) = \frac{2}{n-1}\sigma^4$ 이다. 한편, $S^2 = \frac{1}{n-1}\sum(X_i-\overline{X})^2$은 σ^2의 불편 추정량이며 $\hat{\sigma}^2 = \frac{1}{n}\sum(X_i-\overline{X})^2 = \frac{n-1}{n}S^2$이므로

$$Var(\hat{\sigma}^2) = Var\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2\left(\frac{2}{n-1}\right)\sigma^4 = \frac{2(n-1)}{n^2}\sigma^4$$

이다. 따라서 평균제곱오차는 아래와 같이 구해진다.

$$MSE(S^2) = Var(S^2) = \frac{2}{n-1}\sigma^4,$$

$$MSE(\hat{\sigma}^2) = Var(\hat{\sigma}) + (Bias(\hat{\sigma}^2))^2 = \frac{2(n-1)}{n^2}\sigma^4 + \frac{1}{n^2}\sigma^4 = \frac{2n-1}{n^2}\sigma^4.$$

[참고] 위에서 구한 평균제곱오차는 $n>1$이면 $MSE(\hat{\sigma}^2) < MSE(S^2)$인 관계가 성립하므로 평균제곱오차의 관점에서는 $\hat{\sigma}^2$이 S^2보다 더 좋은 추정량이라고 할 수 있다.

[정의 6.1-4]

T_1과 T_2를 θ의 추정량이라고 하자. $MSE(T_1) \le MSE(T_2)$이면 T_1이 T_2보다 **효율적** (Efficient)인 추정량이라고 하며

$$e(T_1, T_2) = \frac{MSE(T_2)}{MSE(T_1)}$$

을 T_2에 대한 T_1의 **상대효율** (Relative efficiency)이라고 한다.

예 6.1-5 [예 6.1-4]에서 $n \ge 1$인 경우에는 $MSE(\hat{\sigma}^2) \le MSE(S^2)$이므로 $\hat{\sigma}^2$가 S^2보다 효율적이다. S^2에 대한 $\hat{\sigma}^2$의 상대효율은 아래와 같다.

$$e(\hat{\sigma}^2, S^2) = \frac{MSE(S^2)}{MSE(\hat{\sigma}^2)} = \frac{2n^2}{(n-1)(2n-1)} > 1.$$

예 6.1-6 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$, $x>0$ (지수분포의 pdf임은 알고 있죠?)인 분포에서 추출한 확률표본이라고 하자. $E(X_i) = \mu$, $Var(X_i) = \mu^2$임은 이미 알고 있는 사실이다. 이제 $\sum_{i=1}^{n} a_i = 1$을 만족하는 a_i에 대하여 통계량 $U = a_1X_1 + \cdots + a_nX_n$의 성질에 관하여 살펴보자. $E(U) = a_1E(X_1) + \cdots + a_nE(X_n) = (a_1 + \cdots + a_n)E(X) = \mu$이므로 U는 μ의 불편 추정량임을 알 수 있다. a_i의 값에 따른 다음의 두 가지 통계량에 대한 분산을 비교해 보자.

(i) $a_1 = 1$, $a_2 = \cdots = a_n = 0$인 경우 :

$U_1 = X_1$으로 둘 수 있으므로 $Var(U_1) = Var(X_1) = Var(X) = \mu^2$이다.

(ii) $a_1 = \cdots = a_n = \frac{1}{n}$인 경우 :

$U_n = \overline{X} = \frac{1}{n}(X_1 + \cdots X_n)$로 둘 수 있으므로

$Var(U_n) = Var(\overline{X}) = \frac{1}{n} Var(X) = \frac{1}{n}\mu^2$이다.

따라서, (i)과 (ii)를 비교해 보면 U_n이 U_1보다 더 효율적인 통계량이라고 할 수 있으며 상대효율 $e(U_n, U_1) = \frac{Var(U_1)}{Var(U_n)} = n$이다.

예 6.1-7 $X_1, X_2, \cdots, X_n$은 균일분포 $U(0, \theta)$로부터 추출한 확률표본이라고 하자. θ의 두 통계량으로서 $\hat{\theta} = X_{(n)}$과 $\tilde{\theta} = 2\overline{X}$를 생각해 본다. 밀도함수가 $f(x) = \frac{1}{\theta}$, $0 \le x \le \theta$이므로 먼저 확률표본의 평균과 분산을 구하면 아래와 같다.

$$E(X_i) = E(X) = \int_0^\theta x \cdot \frac{1}{\theta} dx = \frac{\theta}{2},$$

$$E(X^2) = \int_0^\theta x^2 \cdot \frac{1}{\theta} dx = \frac{1}{3}\theta^2,$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{3}\theta^2 - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}.$$

$\tilde{\theta}$의 평균, 분산, 평균제곱오차는 아래와 같이 계산된다.

$$E(\tilde{\theta}) = E(2\overline{X}) = 2E(\overline{X}) = 2E(X) = \theta,$$

$$Var(\tilde{\theta}) = Var(2\overline{X}) = 4 \cdot \frac{1}{n} Var(X) = \frac{\theta^2}{3n},$$

$$MSE(\tilde{\theta}) = Var(\tilde{\theta}) + (Bias(\tilde{\theta}))^2 = \frac{\theta^2}{3n} + 0^2 = \frac{\theta^2}{3n}.$$

이제 $\hat{\theta}$의 평균, 분산, 평균제곱오차를 구해보자. 먼저, 계산을 간편하게 하기 위하여 $X_{(n)} = Y$라 두면 Y의 pdf는 $f(y) = \frac{ny^{n-1}}{\theta^n}$, $0 \le y \le \theta$ 이므로 아래의 계산 결과를 얻을 수 있다.

$$E(\hat{\theta}) = E(Y) = \frac{n}{\theta^n}\int_0^\theta y \cdot y^{n-1} dy = \frac{n}{n+1}\theta,$$

$$E(Y^2) = \frac{n}{\theta^2}\int_0^\theta y^2 \cdot y^{n-1} dy = \frac{n}{n+2}\theta^2,$$

$$Var(\hat{\theta}) = Var(Y) = E(Y^2) - (E(Y))^2 = \left\{\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right\}\theta^2,$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2$$

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

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

따라서 상대효율을 구하면

$$e(\hat{\theta}, \tilde{\theta}) = \frac{MSE(\tilde{\theta})}{MSE(\hat{\theta})} = \frac{(n+1)(n+2)}{6n}$$

이다. n의 값이 2보다 크거나 같으면 $e(\hat{\theta}, \tilde{\theta}) \ge 1$이므로 $\hat{\theta}$가 $\tilde{\theta}$보다 효율적임을 알 수 있다.

이 절의 마지막으로 추정량의 일치성에 대하여 공부하도록 한다.

[정의 6.1-5]

T가 θ의 추정량이라고 할 때, 임의의 $\varepsilon > 0$에 대하여 $\lim_{n\to\infty} P[|T-\theta| < \varepsilon] = 1$을 만족하면 T를 θ의 **일치 추정량** (Consistent estimator)이라고 한다. 특히, $\lim_{n\to\infty} MSE(T) = \lim_{n\to\infty} E(T-\theta)^2 = 0$을 만족하는 경우에는 T를 **평균제곱오차 일치추정량** (MSE consistent estimator)이라고 한다.

[참고] 체비세프의 부등식을 이용하면 0보다 큰 모든 r에 대하여

$$P[|T-\theta| > r] \le \frac{1}{r^2} E[T-\theta]^2 = \frac{1}{r^2} MSE(T)$$

이므로 MSE 일치 추정량은 일치 추정량이 된다. 그러나 역은 성립하지 않는다.

예 6.1-8 동전을 n회 독립적으로 던지는 실험에서 앞면이 나오는 회수를 X라고 할 때, 앞면이 나오는 비율 p에 대한 추정량으로서 $\hat{p}=\frac{X}{n}$, $\tilde{p}=\frac{X+1}{n+2}$을 앞에서 다루었다. 이제 또 다른 추정량 $p^*=\frac{1}{2}$에 대하여 생각해 보자.

$$E(p^*)=\frac{1}{2}, \quad Var(p^*)=Var\left(\frac{1}{2}\right)=0,$$

$$MSE(p^*)=Var(p^*)+(Bias(p^*))^2=\left(\frac{1}{2}-p\right)^2$$

이므로 $\lim_{n\to\infty} MSE(p^*)=\left(\frac{1}{2}-p\right)^2$이다. 즉, p^*는 p의 일치 추정량이 아님을 알 수 있다. 한편 다른 두 추정량에 대해서는

$$\lim_{n\to\infty} MSE(\hat{p})=\lim_{n\to\infty}\frac{p(1-p)}{n}=0,$$

$$\lim_{n\to\infty} MSE(\tilde{p})=\lim_{n\to\infty}\frac{1+(n-4)p-(n-4)p^2}{(n+2)^2}=0$$

이므로 $\hat{p}$와 $\tilde{p}$는 p의 일치 추정량임을 알 수 있다.

2. 점추정량

먼저 가장 오래된 점추정량으로서 K. Pearson이 1894년에 발표한 **적률방법** (Method of moments)을 소개하도록 한다. 이 방법은 모집단의 r차 적률을 표본에서의 r차 적률로 추정하는 것이다. 간단하게 수식으로 표현하면 $E(X^r) \to \frac{1}{n}\sum X_i^r$로 추정하는 것이다. 예를 통하여 설명하도록 한다.

예 6.2-1 $X_1, X_2, \cdots, X_n$이 베르누이 시행 $B(1, p)$에서의 확률표본이라고 하자. 베르누이의 시행에서 $E(X)=p$를 표본에서의 1차 적률 $\frac{1}{n}\sum X_i = \overline{X}$로 추정하므로 p의 적률방법에 의한 추정량은 $\hat{p}=\overline{X}$이다.

예 6.2-2 $X_1, X_2, \cdots, X_n$을 정규분포 $N(\mu, \sigma^2)$에서 추출한 확률표본이라고 하자. μ와 σ^2의 적률방법에 의한 추정량을 구해보자. X의 1차 및 2차 적률은 $E(X)=\mu$, $E(X^2)=Var(X)+(E(X))^2=\sigma^2+\mu^2$이고 $E(X)$에 대응하는 표본의 적률은 $\frac{1}{n}\sum X_i=\overline{X}$이며 $E(X^2)$에 대응하는 표본에서의 적률은 $\frac{1}{n}\sum X_i^2$이다. 따라서, μ와 σ^2의 적률방법에 의한 추정량은 아래와 같다.

$$\hat{\mu}=\overline{X}=\frac{1}{n}\sum X_i,$$

$$\widehat{Var(X)}=\hat{\sigma}^2=\frac{1}{n}\sum X_i^2-\overline{X}^2=\frac{1}{n}\sum(X_i-\overline{X})^2.$$

예 6.2-3 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x)=\lambda e^{-\lambda x}$, $\lambda>0$, $x>0$인 지수분포에서 추출한 확률표본이라고 하자. λ의 적률방법에 의한 추정량을 구해보자. $E(X)=\frac{1}{\lambda}$이고 $E(X)$에 대응하는 표본에서의 적률은 $\frac{1}{n}\sum X_i=\overline{X}$이므로 λ의 적률방법에 의한 추정량은 $\hat{\lambda}=\frac{1}{\overline{X}}$이다.

예 6.2-4 $X_1, X_2, \cdots, X_n$은 밀도함수가 $f(x)=\theta x^{\theta-1}$, $\theta>0$, $0<x<1$인 분포로부터 추출한 확률표본이라고 하자. θ의 적률방법에 의한 추정량을 구해보자. $E(X)=\int_0^1 x\cdot\theta x^{\theta-1}dx=\frac{\theta}{\theta+1}$이고 $E(X)$에 대응하는 표본에서의 적률은 $\frac{1}{n}\sum X_i=\overline{X}$이므로 θ의 적률방법에 의한 추정량은 $\hat{\theta}=\frac{\overline{X}}{1-\overline{X}}$임을 알 수 있다.

적률방법에 의한 추정량은 매우 직관적이며 구하기 쉬운 장점이 있는 반면에 때로는 정확하지 않은 추정량을 제공할 때도 있다. 모수가 여러 개 있을 때 모수의 수 만큼 적률이 존재하지 않는 경우도 있기 때문에 적률방법에 의한 추정량을 구하기가 쉽지 않다. 일반적인 경우에 널리 적용이 가능한 R. A. Fisher의 **최우 추정법**을 소개하고자 한다.

((주의)) 모수가 θ인 밀도함수 $f(x)$를 여기서는 $f(x:\theta)$로 표기한다.

[정의 6.2-1]

모수가 θ인 결합밀도함수 $L(\theta)=\prod_{i=1}^{n} f(x_i:\theta)$를 θ의 **우도함수** (Likelihood function)라 하고 우도함수 $L(\theta)$를 최대로 하는 θ의 값 $\hat{\theta}$를 θ의 **최우 추정량** (Maximum likelihood estimator : MLE)이라고 한다.

$L(\theta)$의 값은 당연히 0보다 크며, $\ln L(\theta)$는 $L(\theta)$의 단조증가함수이므로 $L(\theta)$를 최대로 하는 θ는 $\ln L(\theta)$를 최대로 한다. 그래서, 많은 경우에 $\ln L(\theta)$를 최대로 하는 θ의 값을 최우 추정량으로 택한다.

예 6.2-5 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:p)=p^x(1-p)^{1-x}$, $x=0, 1$, $0\le p\le 1$인 베르누이 시행에서의 확률표본이라고 하자. 모수 p의 최우 추정량을 구해보자. 우도함수를 구해보면

$$L(p)=\prod f(x:p)=p^{\sum x_i}(1-p)^{n-\sum x_i}$$

이므로 $L(p)$의 자연대수를 취하면

$$\ln L(p)=\sum x_i \ln p+(n-\sum x_i)\ln(1-p)$$

이다. 이를 p에 대하여 미분하여 0으로 두면 $\ln L(\theta)$를 최대로 하는 p의 최우 추정량을 구하게 된다.

$$\frac{\partial \ln L(p)}{\partial p}=\frac{1}{p}\sum x_i-\frac{1}{1-p}(n-\sum x_i)=0,$$
$$\hat{p}=\frac{1}{n}\sum X_i=\overline{X}.$$

p의 최우 추정량은 표본평균이 된다.

예 6.2-6 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:\theta)=\frac{1}{\theta}e^{-\frac{1}{\theta}x}$, $x>0$, $0<\theta<\infty$인 지수분포로부터 추출한 확률표본이라고 할 때, θ의 최우 추정량을 구해보자. 먼저 우도함수는

$$L(\theta)=\frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum x_i}$$

이며 우도함수의 로그값은

$$\ln L(\theta)=-n\ln\theta-\frac{1}{\theta}\sum x_i$$

이므로 θ에 대한 미분 값을 0으로 두고 최우 추정량을 구하면 아래와 같다.

$$\frac{\partial \ln L(\theta)}{\partial \theta}=-\frac{n}{\theta}+\frac{1}{\theta^2}\sum x_i=0,$$
$$\therefore \hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}.$$

예 6.2-7 $X_1, X_2, \cdots, X_n$을 정규분포 $N(\theta, \sigma^2)$으로부터 추출한 확률표본이라고 할 때, θ와 σ^2의 최우 추정량을 구해보자. 먼저 우도함수는

$$L(\theta:\sigma^2)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\sigma^2}\right]$$

이며 우도함수의 로그값은

$$\ln L(\theta:\sigma^2)=-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\sigma^2}-\frac{n}{2}\ln\sigma^2-\frac{n}{2}\ln 2\pi$$

이므로 θ와 σ^2에 대한 편미분을 0으로 놓고 최우 추정량을 구하면 아래와 같은 결과를 얻는다.

$$\frac{\partial \ln L(\theta:\sigma^2)}{\partial \theta}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta)=0$$
$$\Rightarrow \hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X},$$

$$\frac{\partial \ln L(\theta:\sigma^2)}{\partial \sigma^2}=\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\sigma^4}-\frac{n}{2\sigma^2}=0$$
$$\Rightarrow \hat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

예 6.2-8 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:\theta)=I_{\left[\theta-\frac{1}{2},\,\theta+\frac{1}{2}\right]}(x)$, $-\infty<\theta<\infty$인 균일분포 $U\left(\theta-\frac{1}{2},\,\theta+\frac{1}{2}\right)$에서 추출한 확률표본이라고 할 때, θ의 최우 추정량을 구해보자. 먼저 우도함수는

$$L(\theta)=\prod_{i=1}^{n} f(x_i:\theta)=\prod_{i=1}^{n} I_{\left[\theta-\frac{1}{2},\,\theta+\frac{1}{2}\right]}(x_i)=I_{\left[x_{(n)}-\frac{1}{2},\,x_{(1)}+\frac{1}{2}\right]}(\theta)$$

이므로 $X_{(n)}-\frac{1}{2}\le\hat{\theta}\le X_{(1)}+\frac{1}{2}$을 만족하는 $\hat{\theta}$는 최우 추정량이 된다. 이 예를 통하여 우리는 최우 추정량은 유일하게 결정되는 값이 아님을 알 수 있고 보통 최우 추정량으로서

$$\hat{\theta}=\frac{1}{2}\left(X_{(n)}-\frac{1}{2}+X_{(1)}+\frac{1}{2}\right)=\frac{1}{2}(X_{(n)}+X_{(1)})$$

즉, 표본의 **중간 범위수** (Sample midrange)를 흔히 사용한다.

예 6.2-9 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}$, $x=0, 1, 2, \cdots$인 포아송 분포로부터 추출한 확률표본이라고 할 때, λ의 최우 추정량을 구해보자. 우도함수는

$$L(\lambda)=\frac{\lambda^{\sum x_i}e^{-n\lambda}}{x_1!x_2!\cdots x_n!}$$

이므로 우도함수의 로그값은

$$\ln L(\lambda)=(\sum x_i)\ln\lambda-n\lambda-\ln(x_1!x_2!\cdots x_n!)$$

이다. $\ln L(\lambda)$를 λ에 대하여 미분한 값을 0으로 놓고 풀면 최우 추정량은 다음과 같다.

$$\frac{\partial \ln L(\lambda)}{\partial\lambda}=\frac{1}{\lambda}\sum x_i-n=0$$

$$\Rightarrow \hat{\lambda}=\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i.$$

예 6.2-10 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:\theta)=\theta x^{\theta-1}$, $0<x<1$, $0<\theta<\infty$인 분포로부터 추출한 확률표본이라고 할 때, θ의 최우 추정량을 구해보자. 먼저 우도함수는

$$L(\theta)=\theta^n x_1^{\theta-1}x_2^{\theta-1}\cdots x_n^{\theta-1}=\theta^n(x_1x_2\cdots x_n)^{\theta-1}$$

이므로 우도함수의 로그값은

$$\ln L(\theta)=n\ln\theta+(\theta-1)\ln\prod_{i=1}^{n} x_i$$

이다. $\ln L(\theta)$의 θ에 대한 미분 값을 0으로 놓고 풀면 θ의 최우 추정량은 다음과 같다.

$$\frac{\partial \ln L(\theta)}{\partial\theta}=\frac{n}{\theta}+\sum\ln x_i=0$$

$$\Rightarrow \hat{\theta}=-\frac{n}{\sum\ln x_i}=-\frac{1}{\frac{1}{n}\sum\ln x_i}.$$

예 6.2-11 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:\theta)=\frac{1}{\theta^2}xe^{-\frac{x}{\theta}}$, $0<x<\infty$, $0<\theta<\infty$인 분포로부터 추출한 확률표본이라고 할 때, θ의 최우 추정량을 구해보자. 먼저 우도함수는

$$L(\theta)=\prod_{i=1}^{n}f(x_i:\theta)=\left(\frac{1}{\theta^2}\right)^n(x_1x_2\cdots x_n)e^{-\frac{1}{\theta}\sum x_i}$$

이므로 우도함수의 로그값은

$$\ln L(\theta)=-2n\ln\theta+\sum\ln x_i-\frac{1}{\theta}\sum x_i$$

이다. $\ln L(\theta)$의 θ에 대한 미분값을 0으로 놓고 풀면 아래와 같은 최우 추정량이 구해진다.

$$\frac{\partial\ln L(\theta)}{\partial\theta}=-\frac{2n}{\theta}+\frac{1}{\theta^2}\sum x_i=0$$
$$\Rightarrow\hat{\theta}=\frac{1}{2n}\sum X_i=\frac{\overline{X}}{2}.$$

예 6.2-12 $X_1, X_2, \cdots, X_n$을 밀도함수가 $f(x:\theta)=\frac{1}{\theta}$, $0\le x_i\le\theta$인 균일분포 $U(0,\theta)$로부터 추출한 확률표본이라고 할 때, θ의 최우 추정량을 구해보자. 먼저 우도함수는

$$L(\theta)=\prod_{i=1}^{n}f(x_i:\theta)=\left(\frac{1}{\theta}\right)^n,\quad 0\le x_1\le\theta,\quad\cdots,\quad 0\le x_n\le\theta$$
$$=\left(\frac{1}{\theta}\right)^n,\quad\theta\ge x_{(n)},\quad\theta>0$$

으로 주어진다. 이 우도함수 $L(\theta)$는 θ의 값이 작을수록 커진다. 그러나 θ는 항상 $x_{(n)}$보다 크거나 같기 때문에 $L(\theta)$의 값을 최대로 하는 θ의 값은 $x_{(n)}$이다. 그러므로 θ의 최우 추정량 $\hat{\theta}$은 $X_{(n)}$이다.

3. 신뢰구간의 추정

θ의 추정량을 $\hat{\theta}$라고 할 때 $\hat{\theta}$와 θ의 거리 $|\hat{\theta}-\theta|$가 지정된 값 ε보다 작게 될 확률이 $1-\alpha$라고 하면

$$P[|\hat{\theta}-\theta| \le \varepsilon] = 1-\alpha$$

로 표현할 수 있고 다음과 같이 나타낼 수도 있다.

$$P[\hat{\theta}-\varepsilon \le \theta \le \hat{\theta}+\varepsilon] = 1-\alpha.$$

이때, 구간 $[\hat{\theta}-\varepsilon,\ \hat{\theta}+\varepsilon]$을 θ의 $100(1-\alpha)\%$ **신뢰구간** (Confidence interval)이라고 한다. θ를 추정할 때, 하나의 값으로 추정하는 **점추정** (Point estimation)과는 상대적으로 θ가 어떤 구간에 포함될 확률로 추정하는 것을 **구간추정** (Interval estimation)이라고 한다. 구간추정을 할 때, 확률 $1-\alpha$를 **신뢰계수** (Confidence coefficient)라 부른다.

예 6.3-1 (단일표본의 신뢰구간)

$X_1, X_2, \cdots, X_n$을 정규분포 $N(\mu, \sigma^2)$으로부터 추출한 확률표본이라고 하자. 모분산 σ^2은 알려져 있다고 가정한다. 중심극한정리에 의하여 $\overline{X}$는 정규분포 $N\left(\mu, \frac{\sigma^2}{n}\right)$을 따르게 된다. 이를 표준화하면 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$는 표준정규분포를 따르게 되므로

$$P\left[-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right] = 1-\alpha$$

이다. 이를 μ에 대하여 표현하면 다음과 같다.

$$P\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

예 6.3-2 (두 표본의 경우)

$X_1, X_2, \cdots, X_m$은 정규분포 $N(\mu_X, \sigma_X^2)$에서 추출한 확률표본이고, $Y_1, Y_2, \cdots, Y_n$은 정규분포 $N(\mu_Y, \sigma_Y^2)$에서 추출한 확률표본이며 X_i와 Y_j는 서로 독립이라고 하자. $\overline{X}$의 분포는 $N\left(\mu_X, \frac{\sigma_X^2}{m}\right)$이고 $\overline{Y}$의 분포는 $N\left(\mu_Y, \frac{\sigma^2}{n}\right)$이므로 $\overline{X}-\overline{Y}$의 적률생성함수는 다음과 같이 계산된다.

$$\begin{aligned} E[e^{t(\overline{X}-\overline{Y})}] &= E[e^{t\overline{X}}]\cdot E[e^{-t\overline{Y}}] \\ &= \exp\left[t\mu_X + \frac{t^2\left(\frac{\sigma_X^2}{m}\right)}{2}\right]\cdot \exp\left[-t\mu_Y + \frac{t^2\left(\frac{\sigma_Y^2}{n}\right)}{2}\right] \\ &= \exp\left[t(\mu_X-\mu_Y) + \left(\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right)\frac{t^2}{2}\right] \end{aligned}$$

이므로 $\overline{X}-\overline{Y}$의 분포는 $N\left(\mu_X-\mu_Y, \frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right)$임을 알 수 있다. 따라서 이를 표준화하면 $\frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}}$은 표준정규분포 $N(0,1)$을 따르게 되므로

$$P\left[-z_{\alpha/2} \le \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}} \le z_{\alpha/2}\right] = 1-\alpha$$

로 둘 수 있으므로 $\mu_X-\mu_Y$에 관하여 풀면 아래와 같다.

$$P\left[\overline{X}-\overline{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}} \le \mu_X-\mu_Y \le \overline{X}-\overline{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}\right] = 1-\alpha.$$

예 6.3-3 (백분위수의 신뢰구간)

$X_{(1)}<X_{(2)}<X_{(3)}<X_{(4)}<X_{(5)}$를 표본의 크기가 $n=5$인 확률표본의 순서 통계량이라고 하자. 표본의 중앙값은 $X_{(3)}$이며 모집단의 중앙값은 $\pi_{0.5}$로 표기하자. 이제 구간 $(X_{(1)}, X_{(5)})$가 $\pi_{0.5}$를 포함하는 확률 $P[X_{(1)}<\pi_{0.5}<X_{(5)}]$를 계산해 보자. 각 표본 X가 $\pi_{0.5}$보다 작으면 성공이라고 할 때, 독립시행에서 성공할 확률은 $P[X<\pi_{0.5}]=0.5$이다. $X_{(1)}$이 $\pi_{0.5}$보다 작고 $X_{(5)}$가 $\pi_{0.5}$보다 크기 위해서는 적어도 성공이 한번은 있어야 하지만 다섯 번 모두 성공은 아니다. 이를 식으로 표시하면 다음과 같다.

$$P[X_{(1)}<\pi_{0.5}<X_{(5)}] = \sum_{k=1}^{4}\binom{5}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{5-k}$$
$$= 1-\left(\frac{1}{2}\right)^5-\left(\frac{1}{2}\right)^5 = \frac{15}{16}.$$

즉, 구간 $(X_{(1)}, X_{(5)})$가 $\pi_{0.5}$를 포함하는 확률은 $\frac{15}{16}\approx 0.94$이다. 이와 같은 방법으로 표본의 크기가 n인 일반적인 경우를 생각해 보면

$$P[X_{(1)}<\pi_{0.5}<X_{(n)}] = \sum_{k=1}^{n-1}\binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k}$$
$$= 1-\left(\frac{1}{2}\right)^n-\left(\frac{1}{2}\right)^n = 1-\left(\frac{1}{2}\right)^{n-1}$$

이며 $\pi_{0.5}$가 구간 $(X_{(i)}, X_{(j)})$에 포함될 확률은 다음과 같이 표현된다.

$$P[X_{(i)}<\pi_{0.5}<X_{(j)}] = \sum_{k=i}^{j-1}\binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k}.$$

만일 성공할 확률을 $P[X<\pi_p]=p$로 두고 아래의 확률을 계산하면 신뢰계수에 적합한 신뢰구간을 얻게 된다.

$$P[x_{(i)}<\pi_p<x_{(j)}] = \sum_{k=i}^{j-1}\binom{n}{k}p^k(1-p)^{n-k}.$$

[참고] 모분산 σ^2을 모르더라도 표본의 크기가 매우 큰 경우 ($n\geq 30$)를 살펴보자. 앞에서 우리는 표본의 분산으로 주로

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right)$$

을 사용하였다. $E(X_i^2)=\mu^2+\sigma^2$, $E(\overline{X}^2)=\mu^2+\frac{\sigma^2}{n}$이므로

$$E(S^2) = \frac{1}{n}\left[\sum_{i=1}^{n}E(X_i^2) - nE(\overline{X}^2)\right]$$
$$= \frac{1}{n}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2) - n\left(\mu^2+\frac{\sigma^2}{n}\right)\right]$$
$$= \frac{1}{n}[n\mu^2+n\sigma^2-n\mu^2-\sigma^2]$$
$$= \frac{n-1}{n}\sigma^2.$$

즉, $\frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$는 σ^2의 불편 추정량이 된다. 미지의 모분산 대신에 이 불편 추정량을 사용하면

$$\frac{\overline{X}-\mu}{\sqrt{\frac{n}{n-1}S^2/n}} = \frac{\overline{X}-\mu}{S/\sqrt{n-1}}$$

는 표준정규분포 $N(0, 1)$을 따르게 되므로 μ의 신뢰구간은 다음과 같이 구해진다.

$$P\left[\overline{X}-z_{\alpha/2}\frac{S}{\sqrt{n-1}} \leq \mu \leq \overline{X}+z_{\alpha/2}\frac{S}{\sqrt{n-1}}\right] = 1-\alpha.$$

((주의)) 모분산 σ^2을 모를 때, 표본의 크기가 작은 경우 ($n<30$)에는 정규분포를 사용하는 것이 아니라 t-분포를 사용해야 한다. 나중에 자세히 언급하도록 한다.

예 6.3-4 (단일표본에서 백분율의 신뢰구간)

동전을 던지는 실험에서와 같이 확률변수 X가 항분포 $B(n, p)$를 따른다고 하면 중심극한정리를 사용하여

$$\frac{X-np}{\sqrt{np(1-p)}} = \frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

은 표준정규분포 $N(0, 1)$을 따르게 된다. 따라서, p의 신뢰구간은 아래와 같은 식으로 구해진다.

$$P\left[-z_{\alpha/2} \le \frac{\frac{X}{n}-p}{\sqrt{p(1-p)/n}} \le z_{\alpha/2}\right] = 1-\alpha.$$

위의 p에 관한 부등식은

$$\left|\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \le z_{\alpha/2}$$

로 표현할 수 있고 이 부등식을 풀어서 신뢰구간을 구하면

$$\left(\frac{X}{n}-p\right)^2 - z_{\alpha/2}^2\frac{p(1-p)}{n} \le 0$$

$$\Rightarrow \frac{2X+z_{\alpha/2}^2-\sqrt{4z_{\alpha/2}^2X+z_{\alpha/2}^4-\frac{4X^2z_{\alpha/2}^2}{n}}}{2(n+z_{\alpha/2}^2)} \le p$$

$$\le \frac{2X+z_{\alpha/2}^2+\sqrt{4z_{\alpha/2}^2X+z_{\alpha/2}^4-\frac{4X^2z_{\alpha/2}^2}{n}}}{2(n+z_{\alpha/2}^2)}$$

이므로 계산과정이 매우 복잡하게 된다. 따라서 분모항 $\frac{p(1-p)}{n}$에서 p대신 추정량 $\frac{X}{n}$를 대입하여 신뢰구간을 구하는 것이 훨씬 쉽다.

$$P\left[-z_{\alpha/2} \le \frac{\frac{Y}{n}-p}{\sqrt{\frac{Y}{n}\left(1-\frac{Y}{n}\right)/n}} \le z_{\alpha/2}\right] = 1-\alpha.$$

따라서 p의 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$\left[\frac{X}{n} - z_{\alpha/2}\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}},\ \frac{X}{n} + z_{\alpha/2}\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}}\right].$$

예 6.3-5 (두 표본에서 백분율의 차에 대한 신뢰구간)

성공할 확률이 p_1인 실험을 n_1회 수행하여 성공한 회수가 X_1이라고 하고 이와는 독립적으로 성공할 확률이 p_2인 실험을 n_2회 수행하여 성공한 회수가 X_2라고 하자. 앞에서와 마찬가지로 $\frac{X_1}{n_1}$은 정규분포 $N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$을 따르고, $\frac{X_2}{n_2}$는 정규분포 $N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$를 따르게 된다. 따라서 $\frac{X_1}{n_1}-\frac{X_2}{n_2}$는 정규분포 $N\left(p_1-p_2, \frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}\right)$를 하게 되므로 이를 표준화하면

$$\frac{\frac{X_1}{n_1}-\frac{X_2}{n_2}-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$$

은 표준정규분포 $N(0, 1)$을 갖게 된다. 분모에서 p_1 대신 $\frac{X_1}{n_1}$, p_2 대신 $\frac{X_2}{n_2}$를 대입하면

$$P\left[-z_{\alpha/2} \le \frac{\frac{X_1}{n_1}-\frac{X_2}{n_2}-(p_1-p_2)}{\sqrt{\frac{\frac{X_1}{n_1}\left(1-\frac{X_1}{n_1}\right)}{n_1}+\frac{\frac{X_2}{n_2}\left(1-\frac{X_2}{n_2}\right)}{n_2}}} \le z_{\alpha/2}\right] = 1-\alpha$$

이므로 p_1-p_2의 $100(1-\alpha)\%$ 신뢰구간은 아래와 같이 표현된다.

$$\frac{X_1}{n_1}-\frac{X_2}{n_2} \pm z_{\alpha/2}\sqrt{\frac{\frac{X_1}{n_1}\left(1-\frac{X_1}{n_1}\right)}{n_1}+\frac{\frac{X_2}{n_2}\left(1-\frac{X_2}{n_2}\right)}{n_2}}.$$

[참고] (표본의 크기)

표본의 크기를 얼마로 해야 하는가 하는 문제는 매우 현실적인 문제이다. 임상실험을 해야하는 제약회사의 경우 환자 1명당 실험 비용이 수천 만원이 투입된다고 하면 표본의 크기를 설정하는 문제는 매우 심각하다. 여기서는 간단한 경우만 다루어 보도록 한다.

앞에서 μ에 대한 $100(1-\alpha)\%$의 신뢰구간은 $\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$이었다. 여기서 오차의 한계를 ε으로 정하면 $\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$이므로 구하고자 하는 표본의 크기는 $n = z_{\alpha/2}^2\frac{\sigma^2}{\varepsilon^2}$이 된다.

한편 비율의 구간 추정에서는 오차의 한계를 $\varepsilon = \frac{z_{\alpha/2}\sqrt{p^*(1-p^*)}}{\sqrt{n}}$으로 둘 수 있으므로 표본의 크기는 $n = \frac{z_{\alpha/2}^2\, p^*(1-p^*)}{\varepsilon^2} \le \frac{z_{\alpha/2}^2}{4\varepsilon^2}$ ($p^*(1-p^*)$는 $p^* = \frac{1}{2}$일 때 최대값이 $\frac{1}{4}$이므로)이 된다.

예를 들어 오차의 한계를 10%로 하고 95% 신뢰수준을 갖는다면 표본의 크기는 $n = \left(\frac{1.96}{(2)(0.1)}\right)^2 \approx 96$이 된다. 오차의 한계를 5%로 한다면 표본의 크기는 약 385가 된다. p의 값이 0.2로 알려져 있다면 $\varepsilon = 0.1$인 경우 표본의 크기는 $n = \left(\frac{1.96}{0.1}\right)^2(0.2)(0.8) \approx 62$이다.