5.1. 확률변수 $X$가 자유도 32인 카이제곱분포 $\chi^2(32)$를 따를 때 중심극한정리를 이용하여 확률 $P(20<X<40)$의 근사값을 구하고, R을 이용하여 구한 값과 비교하여라.

CLT : $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\bar{X}=\frac{1}{n}\sum_i^n X_i$, $\mu=E[X]$, $\sigma=\sqrt{Var(X)}$

if $X\sim\chi^2(32)$, $\mu=32$, $\sigma^2=64$, $\sigma=8$

$P(20<X<40) = P\left(\sum_i^n 20 < \sum_i^n X_i < \sum_i^n 40\right) = P\left(\frac{1}{n}\sum_i^n 20 < \frac{1}{n}\sum X_i < \frac{1}{n}\sum_i^n 40\right)$

$= P(20<\bar{X}<40) = P\left(\frac{20-32}{8/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{40-32}{8/\sqrt{n}}\right)$

we know that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ $\therefore P(20<X<40) \simeq P\left(\frac{20-32}{8/\sqrt{n}} < Z < \frac{40-32}{8/\sqrt{n}}\right)$, $Z\sim N(0,1)$

Let $n=1$. $P(20<X<40) \simeq P(-1.5<Z<1) = P(Z<1) - P(Z<-1.5)$

$= \Phi(1) + \Phi(1.5) - 1$

$= 0.8413 + 0.9332 - 1$

$= 0.7745$

R Studio : $P(20<X<40) = 0.7947465$

---

5.2. 확률변수 $X$가 포아송분포 Poisson(25)를 따를 때 중심극한정리를 이용하여 다음 확률의 근사값을 구하고, R을 이용하여 구한 값과 비교하여라.

(a) $P(19<X\le 33)$ (b) $P(19.5<X<33.5)$

$\mu=\sigma^2=\lambda$ $\therefore \mu=25$, $\sigma=5$

(a) $P(19<X\le 33) = P\left(\frac{19-25}{5/\sqrt{1}} < \frac{X-25}{5/\sqrt{1}} \le \frac{33-25}{5/\sqrt{1}}\right) \simeq P(-1.2<Z\le 1.6)$

$= 0.9452 - (1-0.8849)$

$= 0.8301$

(b) $P(19.5<X<33.5) = P\left(\frac{19.5-25}{5/\sqrt{1}} < \frac{X-25}{5/\sqrt{1}} < \frac{33.5-25}{5/\sqrt{1}}\right) \simeq P(-1.1<Z<1.7)$

$= \Phi(1.7) - \Phi(-1.1)$

$= 0.9554 - (1-0.8643)$

$= 0.8197$

R Studio : (a) 0.8166448, (b) 0.8166448

* Takeaways: (a) and (b) has the same probabilities. Yet its normal approximation via CLT show different results.

---

5.3. 확률밀도함수가 $f(x;\theta) = e^{-(x-\theta)} I_{(\theta,+\infty)}(x)$ $(\theta\in\mathbb{R})$

인 모집단으로부터의 랜덤표본 $X_1,\cdots,X_n$에 대하여 $Y_n = \min_{1\le i\le n} X_i$ 이라고 할 때,

$$\operatorname*{plim}_{n\to\infty} Y_n = \theta$$

임을 보여라.

$\operatorname*{plim}_{n\to\infty} Y_n=\theta \Leftrightarrow \forall\epsilon>0, \exists K(\epsilon)$ s.t. if $n>K(\epsilon)$, $P(|Y_n-\theta|>\epsilon)=0$

$F(x) = \int_\theta^x e^{-(x-\theta)}dx = \left[-e^{-(x-\theta)}\right]_\theta^x = 1-e^{-(x-\theta)}$

$\therefore$ pdf of $X_{(1)}$ : $\frac{n!}{1!(n-1)!} f_X(x)\left[1-F_X(x)\right]^{n-1}$

$= n\cdot e^{-(x-\theta)}\cdot\left[1-(1-e^{-(x-\theta)})\right]^{n-1} = ne^{-(x-\theta)}e^{-(x-\theta)\cdot(n-1)}$

$= ne^{-n(x-\theta)} I_{[\theta,\infty)}(x)$

$E[Y_n] = \int_\theta^\infty y\cdot ne^{-n(y-\theta)}dy = \int_0^\infty (z+\theta)\cdot n\cdot e^{-nz}dz$ (let $z=y-\theta$)

$= \underbrace{\int_0^\infty zne^{-nz}dz}_{①} + \underbrace{\int_0^\infty \theta\cdot n\cdot e^{-nz}dz}_{②}$

① $\int_0^\infty zne^{-nz}dz = n\int_0^\infty ze^{-nz}dz$

let $u=z$, $v'=e^{-nz}$

$u'=1$, $v=-\frac{1}{n}e^{-nz}$

$\int uv' = [uv] - \int u'v$ $\therefore \int_0^\infty ze^{-nz}dz = \left[-\frac{z}{n}e^{-nz}\right]_0^\infty - \int -\frac{1}{n}e^{-nz}dz$

$= 0 - \left[\frac{1}{n^2}e^{-nz}\right]_0^\infty = \frac{1}{n^2}$

$\therefore ① = \frac{1}{n}$

② : $\int_0^\infty \theta\cdot ne^{-nz}dz = \left[-\theta e^{-nz}\right]_0^\infty = \theta$

$\therefore E[Y_n] = ① + ② = \frac{1}{n} + \theta$

$Var(Y_n) = E[Y_n^2] - E[Y_n]^2$

$E[Y_n^2] = \int_\theta^\infty y^2\cdot ne^{-n(y-\theta)}dy = \int_0^\infty (z+\theta)^2\cdot n\cdot e^{-nz}dz$

$= n\cdot\left[\underbrace{\int_0^\infty z^2e^{-nz}dz}_{①} + \underbrace{\int_0^\infty 2z\cdot\theta\cdot e^{-nz}dz}_{②} + \underbrace{\int_0^\infty \theta^2e^{-nz}dz}_{③}\right]$

② : $2\theta\int_0^\infty ze^{-nz}dz = 2\theta\cdot\frac{1}{n^2}$

③ : $\theta^2/n$

① : $\left[-\frac{z^2}{n}e^{-nz}\right]_0^\infty - \int_0^\infty -\frac{2z}{n}e^{-nz}dz = \frac{2}{n}\cdot\frac{1}{n^2}$

$\therefore n[①+②+③] = \frac{2}{n^2} + \frac{2\theta}{n} + \theta^2$ $\because Var(Y_n) = \frac{2}{n^2} + \frac{2\theta}{n} + \theta^2 - (\theta+\frac{1}{n})^2$

$= 1/n^2$

$\therefore \lim_{n\to\infty} E[Y_n] = \theta$ and $\lim_{n\to\infty} Var(Y_n) = 0$ $\therefore$ plim $Y_n = \theta$ ✓✓

Textbook Solution

$n(Y_n - \theta) \sim Exp(1)$

$E[Y_n] = \theta + 1/n$ , $Var(Y_n) = 1/n^2$

$\therefore \lim_{n\to\infty} E[Y_n] = \theta$ , $\lim_{n\to\infty} Var(Y_n) = 0$

$\therefore p\lim_{n\to\infty} Y_n = \theta$

---

5.4. 베타분포 Beta(1,d)로부터의 랜덤표본 $X_1, \cdots, X_n$에 대하여 $Y_n = \max_{1\le i\le n} X_i$ 이라고 할 때 $n^{1/a}(1-Y_n)$의 극한분포를 구하여라.

$$pdf_X(x) = \frac{\Gamma(a+1)}{\Gamma(1)\Gamma(a)} x^{1-1}(1-x)^{a-1} I_{(0,1)}(x)$$

$$= a(1-x)^{a-1}$$

$$F_X(x) = \int_0^x a(1-x)^{a-1}dx = \left[-(1-x)^a\right]_0^x = 1-(1-x)^a$$

$$f_Y(y) = \frac{n!}{(n-1)!1!} F_X(y)^{n-1} f_X(y) = n\left[1-(1-y)^a\right]^{n-1} \cdot a(1-y)^{a-1}$$

$$F_Y(y) = \int_0^y n\left[1-(1-y)^a\right]^{n-1} \cdot a(1-y)^{a-1}dy = \left[1-(1-y)^a\right]^n$$

Let $Z_n = n^{1/a}(1-Y_n) \Rightarrow z n^{-1/a} = 1-Y_n \Rightarrow Y_n = 1 - z n^{-1/a}$

$$P(Z_n < y) = P(z n^{-1/a} < y n^{-1/a}) = P(-z n^{-1/a} > -y n^{-1/a}) = P(1 - z n^{-1/a} > 1 - y n^{-1/a})$$
$$= P(Y > 1 - y n^{-1/a}) = 1 - P(Y < 1 - y n^{-1/a})$$

$$= 1 - \left[1-(1-1+y n^{-1/a})^a\right]^n$$
$$= 1 - \left[1 - y^a n^{-1}\right]^n = 1 - \left(1 - \frac{y^a}{n}\right)^n$$

$$\lim_{n\to\infty} P(Z_n < y) = \lim_{n\to\infty} 1 - \left(1 - \frac{y^a}{n}\right)^n = 1 - e^{-y^a}$$

$$\frac{\partial}{\partial y} P(Z_n < y) = a y^{a-1} e^{-y^a} I_{(0,+\infty)}(y)$$

[Textbook Solution]

$$P(Y_n \le y) = \left(P(X_i \le y)\right)^n = \left(1-(1-y)^a\right)^n, \quad 0 < y < 1$$

$$\therefore \lim_{n\to\infty} P\left[n^{1/a}(1-Y_n) \le y\right] = 1 - \exp[-y^a], \quad y > 0$$

$$\therefore f(y) = a y^{a-1} \exp\left\{-y^a\right\} I_{(0,+\infty)}(y)$$

---

5.5. 확률밀도함수가 $f(x) = a x^{-a-1} I_{[1,+\infty)}(x)$

인 모집단으로부터의 랜덤표본 $X_1, \cdots, X_n$에 대하여 $Y_n = \max_{1\le i\le n} X_i$ 이라고 할 때 $n^{-1/a} Y_n$의 극한분포를 구하여라.

$$F_X(x) = \int_1^x a x^{-a-1} dx = \left[-x^{-a}\right]_1^x = 1 - x^{-a}$$

$$P(Y_n < y) = \left(P(X_1 < y)\right)^n = (1-y^{-a})^n$$

$$P(n^{-1/a} Y_n < y) = P(Y_n < n^{1/a} y) = \left(1 - \frac{y^{-a}}{n}\right)^n$$

$$\therefore \lim_{n\to\infty} P(n^{-1/a} Y_n < y) = \lim_{n\to\infty} \left(1 - \frac{y^{-a}}{n}\right)^n = \exp\left\{-y^{-a}\right\} I_{(0,\infty)}(y)$$

$$f_Y(y) = a y^{-a-1} \exp\left\{-y^{-a}\right\} I_{(0,\infty)}(y)$$ ✓✓

* key takeaways from Q5.4.~5.6.
순서통계량의 극한분포를 구할 때 순서통계량의 pdf → cdf (적분을 이용)를 거치지 않고 바로 순서통계량의 cdf를 찾는 방법이 빠르다.

---

5.6. 표준지수분포 Expl(1)로부터의 랜덤표본 $X_1, \cdots, X_n$에 대하여 $Y_n = \max_{1\le i\le n} X_i$ 이라고 할 때 $Y_n - \log n$의 극한분포를 구하여라.

$$f_X(x) = e^{-x} I_{(0,+\infty)}(x) \quad , \quad F(x) = (1-e^{-x}) I_{(0,+\infty)}(x)$$

$$P(Y_n < y) = \prod_{i=1}^n P(X_i < y) = \left(P(X_i < y)\right)^n = (1-e^{-y})^n$$

$$P(Y_n - \log n < y) = P(Y_n < y + \log n) = \left(1 - e^{-(y+\log n)}\right)^n$$
$$= \left[1 - e^{-y} \cdot n^{-1}\right]^n = \left[1 - e^{-y}/n\right]^n$$

$$\therefore \lim_{n\to\infty} P(Y_n - \log n < y) = \lim_{n\to\infty} (1 - e^{-y}/n)^n = \exp\left\{-e^{-y}\right\}$$

$$\therefore f_Y(y) = \frac{\partial}{\partial y} F_Y(y) = e^{-y} \cdot \exp\left\{-e^{-y}\right\} I_{(-\infty,+\infty)}(y)$$

* Textbook has the same answer but the domain of y in the textbook is incorrect.

* The limiting distribution of $Y_n - \log n$ is the standard Gumbell distribution. The Gumbel distribution is used to model the probability of extreme maximum events, such as floods, rainfall, wind speeds, and structural loads - ChatGPT

5.7. 베타분포 Beta(a,1) (a>0) 으로부터 랜덤표본 $X_1, \cdots, X_n$ 에 대하여

$Y_n = \min_{1\le i\le n} X_i$ 이라고 할 때 $n^r Y_n$ $(r>0)$의 극한분포가 존재하기 위한 r의 값을 구하고, 그 극한분포를 구하여라.

$$f_X(x) = \frac{\Gamma(a+1)}{\Gamma(a)\Gamma(1)} x^{a-1}(1-x)^{1-1} I_{(0,1)}(x)$$

$$= a x^{a-1} I_{(0,1)}(x)$$

$$F_X(x) = \int_0^x a x^{a-1} dx = \left[ x^a \right]_0^x = x^a I_{(0,1)}(x)$$

$$P(Y_n < y) = 1 - P(Y_n > y) = 1 - \prod_{i=1}^n P(X_i > y) = 1 - \left[ 1 - x^a \right]^n$$

$$P(n^r Y_n < y) = P(Y_n < n^{-r} y) = 1 - \left[ 1 - (n^{-r} y)^a \right]^n$$

Let $r = \frac{1}{a}$, then $\lim_{n\to\infty} P(n^r Y_n < y) = \lim_{n\to\infty} 1 - \left[ 1 - \frac{y^a}{n} \right]^n = 1 - \exp\{-y^a\}$

$\therefore$ pdf of $n^{\frac{1}{a}} Y_n$ : $a y^{a-1} \exp\{-y^a\} I_{(0,\infty)}(y)$ ✓✓

if $r$: $0 < r < 1/a$ : $\lim_{n\to 0} P(n^r Y_n < y) \xrightarrow{d} 0$ ✓✓

+ $$\lim_{n\to\infty} P(n^r Y_n \le y) = \begin{cases} 1, & 0 < ra < 1 \\ 1 - \exp[-y^a], & ra = 1 \\ 0, & ra > 1 \end{cases}$$

---

5.8. 균등분포 U(0,1)로부터의 랜덤표본 $U_1, \cdots, U_n$ 에 대하여

$$X_n = \min_{1\le i\le n} U_i \ , \quad Y_n = \max_{1\le i\le n} U_i \ , \quad R_n = Y_n - X_n$$

이라고 할 때 다음에 답하여라.

(a) $\lim_{n\to\infty} (X_n, Y_n)^t = (0,1)^t$ 임을 보여라.

(b) $n(1-R_n)$의 극한분포를 구하여라.

$$\text{pdf } f_U(u) = 1 \cdot I_{(0,1)}(u_i) \ , \quad F_U(u) = \int_0^u 1 \cdot du = u\, I_{(0,1)}(u_i)$$

$$P(X_n < x) = 1 - P(X_n > x) = 1 - \prod_{i=1}^n P(U_i > x) = 1 - (1-x)^n$$

$$\therefore \lim_{n\to\infty} P(X_n < x) = \lim_{n\to\infty} 1 - (1-x)^n = 1$$

$$\frac{\partial}{\partial x} \lim_{n\to\infty} P(X_n < x) = \frac{\partial}{\partial x} 1 = 0 \quad \therefore X_n \xrightarrow[n\to\infty]{p} 0$$

$$P(Y_n < y) = \prod_{i=1}^n P(U_i < y) = y^n I_{(0,1)}(y)$$

$$\lim_{n\to\infty} P(Y_n < y) = \lim_{n\to\infty} y^n = 0 \quad \forall y \in (0,1)$$

$$\therefore X_n \xrightarrow[n\to\infty]{p} 0 \ , \quad Y_n \xrightarrow[n\to\infty]{p} 1 .$$

As $n\to\infty$, the cdf of $X_n$ and $Y_n$ becomes degenerate at 0 and 1 respectively. A degenerate distribution has no pdf.

* we can find the pdf by deriving the cdf if the distribution is:
① non-degenerate
② absolutely continuous. (Assures that there are no jumps in the cdf)

(b) $R_n = Y_n - X_n$

$$n(1-R_n) = n(1 - Y_n - (X_n - 0))$$

$$P(n(1-R_n) < y) = P(1 - R_n < y/n) = P(1 - y/n \le R_n)$$

$$= 1 - P(R_n < 1 - y/n) = 1 - P(Y_n - X_n < 1 - y/n)$$

$$\text{pdf of } (X_n, Y_n)^t : \frac{n!}{1!(n-2)!1!} f_U(x)\left[F_U(y) - F_U(x)\right]^{n-2} f_U(y)$$

$$= n(n-1)\left[y - x\right]^{n-2} I_{(0<x<y<1)}$$

Let $u = \begin{cases} R_n = Y_n - X_n \\ W_n = X_n \end{cases}$ $\quad u^{-1} = \begin{cases} X_n = W_n \\ Y_n = R_n + W_n \end{cases}$

1. Transformation is 1-1
2. Domain of transformation is an open set in $\mathbb{R}^2$
3. $|J_{u^{-1}}| = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = 1 \neq 0$

$$\therefore f_{R_n, W_n}(r, w) = n(n-1)[r]^{n-2} I_{(0<w)} I_{(0<r)} I_{(r+w<1)}$$

$$f_{R_n}(r) = \int_0^{1-r} n(n-1) r^{n-2} dw = n(n-1) r^{n-2}(1-r)$$

$$= \left[ n(n-1) r^{n-2} - n(n-1) r^{n-1} \right] I_{(0,1)}(r)$$

$$F_{R_n}(r) = \int_0^r n(n-1) r^{n-2} - n(n-1) r^{n-1} dr = \left[ n r^{n-1} - (n-1) r^n \right]_0^r$$

$$= n r^{n-1} - (n-1) r^n = r^{n-1}(n - nr + r)$$

$$\therefore 1 - F_{R_n}(1 - y/n) = 1 - (1 - y/n)^{n-1}\left(n - n(1 - y/n) + (1 - y/n)\right)$$

$$= 1 - (1 - y/n)^n \cdot (1 - y/n)^{-1}\left(n - n + y + 1 - y/n\right)$$

$$\therefore \lim_{n\to\infty} 1 - F_{R_n}(1 - y/n) = 1 - e^{-y} \cdot 1 \cdot (y + 1)$$

$$\therefore \frac{\partial}{\partial y} \lim_{n\to\infty} 1 - F_{R_n}(1 - y/n) = e^{-y} y - e^{-y} + e^{-y}$$

$$= y e^{-y} I_{(0,1)}(y) \ \sim \ \text{Gamma}(2,1)$$

[Textbook Solution]

(a) Beta$(\alpha, \beta)$의 평균 : $\frac{\alpha}{\alpha+\beta}$ , 분산 : $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$X_n \sim \text{Beta}(1, n)$ , $Y_n \sim \text{Beta}(n, 1)$

(i) $E[X_n] = \frac{1}{n+1}$ , $\text{Var}(X_n) = \frac{n}{(n+1)^2(n+2)}$

$\lim_{n\to\infty} E[X_n] = 0$ , $\lim_{n\to\infty} \text{Var}(Y_n) = 0$

(ii) $E[Y_n] = \frac{n}{n+1}$ , $\text{Var}(X_n) = \frac{n}{(n+1)^2(n+2)}$

$\lim_{n\to\infty} E[Y_n] = 1$, $\lim_{n\to\infty} \text{Var}(Y_n) = 0$

$\therefore \operatorname{plim}_{n\to\infty} X_n = 0$ , $\operatorname{plim}_{n\to\infty} Y_n = 1$ 즉, $\operatorname{plim}_{n\to\infty} (X_n, Y_n)^t = (0,1)^t$

(b) (4.16)에서 $R_n$의 확률밀도함수는

$$pdf_{R_n}(y) = n(n-1)y^{n-2}(1-y)I_{(0,1)}(y)$$

$$\therefore P(R_n \le y) = ny^{n-1}(1-y) + y^n, \quad 0 < y < 1$$

$$\therefore \lim_{n\to\infty} P(n(1-R_n) \le y) = 1 - ye^{-y} - e^{-y}, \quad y > 0$$

극한분포의 확률밀도함수는 $ye^{-y}I_{(0,\infty)}(y)$

---

☆☆☆

5.9. 확률변수 $X_n$과 $Z$에 대하여

$$X_n \xrightarrow[n\to\infty]{d} Z$$

이고, $Z$의 누적분포함수가 모든 실수에서 연속인 함수일 때 다음이 성립함을 밝혀라.[84]

$$\lim_{n\to\infty} \sup_x |P(X_n \le x) - P(Z \le x)| = 0$$

[84] 이러한 균등수렴의 결과로부터 $P(X_n \le x_n) \simeq P(Z \le x_n)$과 같이 $n$에 의존하는 값 $x_n$ 까지의 누적확률의 근사계산이 가능한 것이다.

- Gemini 3.0.

This is a classic result in probability theory known as Polya's theorem. It states that if a sequence of cumulative distribution functions (CDFs) converges pointwise to a continuous limiting CDF, the convergence is actually uniform over the entire line $\mathbb{R}$.

① Setup and Notation

Let $F_n(x) = P(X_n \le x)$ and $F(x) = P(Z \le x)$

we are given two facts:

1. pointwise convergence: For every $x \in \mathbb{R}$, $\lim_{n\to\infty} F_n(x) = F(x)$.
2. Continuity: $F(x)$ is continuous on $\mathbb{R}$.

we want to prove Uniform Convergence:

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0$$

② The partition Strategy

The logic relies on the fact that since F is continuous and bounded between 0 and 1, we can chop the y-axis into small pieces and map those back to the x axis.

Let $\epsilon > 0$ be an arbitrary positive number. Choose an integer K large enough such that $\frac{1}{K} < \epsilon$.

Since $F(x)$ is a valid cdf, we know:

- $\lim_{x\to-\infty} F(x) = 0$
- $\lim_{x\to+\infty} F(x) = 1$
- $F(x)$ is non-decreasing and continuous.

We can select K-1 points on the real line $x_1 < x_2 < \cdots < x_{K-1}$ such that

$$F(x_j) = \frac{j}{K} \quad \text{for } j = 1, \cdots, K-1$$

For completeness, let's define the endpoints as $x_0 = -\infty$ and $x_K = +\infty$. This partitions the real line intervals $(x_{j-1}, x_j]$.

Key Property of the Partition: For any interval $(x_{j-1}, x_j]$, the change in the target function F is small:

$$F(x_j) - F(x_{j-1}) = \frac{j}{K} - \frac{j-1}{K} = \frac{1}{K} < \epsilon$$

(Note: With $F(x_0) = 0$ and $F(x_K) = 1$)

③ Applying Pointwise Convergence.

We now apply the given pointwise convergence at our specific, finite set of points $\{x_1, \cdots, x_{K-1}\}$

Since there are only finitely many points, we can find an integer N such that for all $n \ge N$:

$$|F_n(x_j) - F(x_j)| < \epsilon \quad \forall j = 1, \cdots, K-1$$

④ Bounding the Error for any x

Now consider any arbitrary $x \in \mathbb{R}$. This x must fall into one of the intervals say $x \in (x_{j-1}, x_j]$ for some $j \in \{1, \cdots, K\}$

We use the monotonicity (non-decreasing nature) of the CDFs to "squeeze" the function.

Since $x_{j-1} < x \le x_j$, we know:

$$F_n(x_{j-1}) \le F_n(x) \le F_n(x_j)$$
$$F(x_{j-1}) \le F(x) \le F(x_j)$$

We want to bound the difference $F_n(x) - F(x)$.

- Upper Bound: $F_n(x) - F(x) \le F_n(x_j) - F(x_{j-1})$
  $= [F_n(x_j) - F(x_j)] + [F(x_j) - F(x_{j-1})]$

using our chosen bounds:

- $|F_n(x_j) - F(x_j)| < \epsilon$ (from Step 3)
- $F(x_j) - F(x_{j-1}) < \epsilon$ (from Step 2)

Thus, $F_n(x) - F(x) < \epsilon + \epsilon = 2\epsilon$

- Lower Bound:

$$F_n(x) - F(x) = F_n(x_{j-1}) - F(x_j)$$
$$= [F_n(x_{j-1}) - F(x_{j-1})] + [F(x_{j-1}) - F(x_j)]$$

using the bounds:

- $|F_n(x_{j-1}) - F(x_{j-1})| > -\epsilon$
- $F(x_{j-1}) - F(x_j) > -\epsilon$

Thus: $F_n(x) - F(x) > -\epsilon - \epsilon = -2\epsilon$

Note: If $j=1$ or $j=K$, we rely on the fact that $F_n(-\infty)=F(-\infty)=0$ and $F_n(\infty)=F(\infty)=1$, so the bounds still hold.

(5) Conclusion

Combing the upper and lower bounds, for any $x \in \mathbb{R}$ and any $n \geq N$:

$$|F_n(x)-F(x)| < 2\epsilon$$

Since this holds for any $x$, it holds for the supremem:

$$\sup_x |F_n(x)-F(x)| < 2\epsilon$$

As $\epsilon$ was arbitrary, the limit must be zero:

$$\lim_{n\to\infty} \sup_x |P(X_n \leq x) - P(Z \leq x)| = 0$$

〈Steps of proof〉

1. Select $x_1 < \cdots < x_n$ on the real line such that $F(x_j) = \frac{j}{K}$

2. Show that $F(x_j) - F(x_{j-1}) = \frac{1}{K}$

3. Show that $\exists N \in \mathbb{N}$ s.t. if $n \geq N$, $|F_n(x_j) + F(x_j)| < \epsilon \ \forall j = 1, \cdots, K-1$

4. $F_n(x_{j-1}) \leq F_n(x) \leq F_n(x_j)$
   $F(x_{j-1}) \leq F(x) \leq F(x_j)$
   } upper & lower bound of $F_n(x)$ & $F(x)$

5. $F_n(x_{j-1}) + F(x_j) \leq F_n(x) - F(x) \leq F_n(x_j) + F(x_{j-1})$

6. $F_n(x_{j-1}) - F(x_{j-1}) + F(x_{j-1}) - F(x_j) \leq$

   $F_n(x) - F(x) \leq$

   $F_n(x_j) - F(x_j) + F(x_j) - F(x_{j-1})$

   $\sup |F_n(x) - F(x)| \leq 2\epsilon + \frac{1}{K}$

∴ $\lim_{n\to\infty} \sup |F_n(x) - F(x)| \leq 0$

5.10. 확률변수의 열 $X_n\ (n=1,2,\cdots)$에 대하여

$$\lim_{K\to\infty} \sup_n P(|X_n|>K) = 0$$

이 성립하면 $X_n\ (n=1,2,\cdots)$이 확률적으로 유계 (bounded in probability) 라고 하고 기호로는

$$X_n = O_p(1),\quad n\to\infty$$

로 나타낸다. 한편 $p\lim_{n\to\infty} Z_n = 0$ 인 $Z_n\ (n=1,2,\cdots)$에 대하여는

$$Z_n = o_p(1),\quad n\to\infty$$

와 같이 나타낸다. 이제

$$X_n = O_p(1),\quad Y_n = O_p(1),\quad Z_n = o_p(1),\quad n\to\infty$$

일 때 다음이 성립함을 보여라.

(a) $X_n + Y_n = O_p(1)$ $n\to\infty$ ← Big O

(b) $X_nY_n = O_p$, $n\to\infty$ ← big O

(c) $Z_nX_n = o_p(1)$, $n\to\infty$ ← little O

(d) $Z_n = O_p(1)$, $n\to\infty$ ← big O

(a) $\lim_{K\to\infty} \sup_n P(|X_n+Y_n|>K) = 0$

$$(|X_n+Y_n|>K) \subseteq (|X_n|+|Y_n|>K) \subseteq (|X_n|>K/2)\cup(|Y_n|>K/2) \quad *$$

$$\therefore\ P(|X_n+Y_n|>K) \le P(|X_n|>K/2) + P(|Y_n|>K/2)$$

$$\therefore\ 0 \le \lim_{K\to\infty}\sup_n P(|X_n+Y_n|>K) \le \lim_{K\to\infty}\sup_n P(|X_n|>K/2) + \lim_{K\to\infty}\sup_n P(|Y_n|>K/2)$$

$$= 0+0 = 0$$

If $|X_n+Y_n|>K$, at least one of the events must occur:
$|X_n|>K/2$ or $|Y_n|>K/2$

(b) $X_nY_n = O_p(1)$, $n\to\infty$

$$(|X_nY_n|>K) \subseteq (|X_n|>\sqrt{K}) \cup (|Y_n|>\sqrt{K})$$

$$\therefore\ P(|X_nY_n|>K) \le P(|X_n|>\sqrt{K}) + P(|Y_n|>\sqrt{K})$$

$$\therefore\ 0 \le \lim_{K\to\infty}\sup_n P(|X_nY_n|>K) \le \lim_{K\to\infty}\sup_n P(|X_n|>\sqrt{K}) + \lim_{K\to\infty}\sup_n P(|Y_n|>\sqrt{K}) = 0$$

$$\therefore\ \lim_{K\to\infty}\sup_n P(|X_nY_n|>K) = 0$$

(c) $Z_nX_n = o_p(1)$, $n\to\infty$

$p\lim Z_n = 0 \iff \forall\epsilon>0, \exists\delta>0$ s.t. if $n>\delta$, $P(|Z_n|>\epsilon) = 0$

WTS: if $n>\delta$, $P(|Z_nX_n|>\epsilon) = 0$

$$P(|Z_nX_n|>\epsilon) = P(|Z_nX_n|>\epsilon, |X_n|>K) + P(|Z_nX_n|>\epsilon, |X_n|<K)$$
$$\le \sup_n P(|X_n|>K) + P(|Z_n|>\epsilon/K)$$

$$\therefore\ \lim_{n\to\infty} P(|Z_nX_n|>\epsilon) \le \lim_{n\to\infty}\sup_n P(|X_n|>K) + \lim_{n\to\infty} P(|Z_n|>\epsilon/K) \xrightarrow{K\uparrow\infty} 0+0$$

$$\therefore\ \lim_{n\to\infty} P(|Z_nX_n|>\epsilon) = 0$$

(d) $Z_n = O_p(1)$, $n\to\infty$

WTS: $\lim_{K\to\infty}\sup_n P(|Z_n|>K) = 0$

$p\lim_{n\to\infty} Z_n = 0 \iff \forall\epsilon>0\ \exists\ \delta$ s.t. if $n>\delta$, $P(|Z_n|>\epsilon) = 0$

$$\sup_n P(|Z_n|>K) \ge P(|Z_n|>K)$$

$\lim_{K\to\infty}\sup_n P(|Z_n|>K) \ge \lim_{K\to\infty} P(|Z_n|>K) \xrightarrow[n\to\infty]{p} 0$ ··· lower bound

$$\sup_n P(|Z_n|<K) \ge P(|Z_n|<K)$$

$\lim_{K\to\infty}\sup_n P(|Z_n|<K) \ge \lim_{K\to\infty} P(|Z_n|<K) \xrightarrow[n\to\infty]{p} 1$

$\lim_{K\to\infty}\sup_n P(|Z_n|>K) = 1 - \lim_{K\to\infty}\sup_n P(|Z_n|<K) \le 1 - \lim_{K\to\infty} P(|Z_n|<K) \xrightarrow[n\to\infty]{p} 1-1 = 0$

↳ ··· upper bound.

$$\therefore\ 0 \le \lim_{K\to\infty}\sup_n P(|Z_n|>K) \le 0$$

[Textbook]

$\lim_{x\to\infty} P(|X_1|>x) = 0, \cdots, \lim_{x\to\infty} P(|X_N|>x) = 0$

$\Rightarrow \forall\epsilon>0, \exists K_1 : P(|X_1|>K_1)<\epsilon, \cdots, \exists K_N : P(|X_N|>K_N)<\epsilon$

$K = \max\{K_0, K_1, \cdots, K_n\}$ 이라고 하면,

$\max_{1\le n\le N} P(|X_n|>K) \le \max_{1\le n\le N} P(|X_n|>K_n) < \epsilon$

$\sup_{n\ge N} P(|X_n|>K) \le \sup_{n\ge N} (|X_n|>K_0) < \epsilon$

$\therefore\ K\ge K \to \sup_{n\to\infty} P(|X_n|>K) \le \sup_n P(|X_n|>K) < \epsilon$

$\max_{1\le n\le N} P(|X_n|>K) \le \max_{1\le n\le N} P(|X_n|>K_n) < \epsilon$

$\sup_{n\ge N} P(|X_n|>K) \le \sup_{n\ge N} (|X_n|>K_0) < \epsilon$

$\therefore K \ge K \to \sup_{n\to\infty} P(|X_n|>K) \le \sup_n P(|X_n|>K) < \epsilon$

[Gemini 3.0. proof]

1. Definitions:

$Z_n \xrightarrow[n\to\infty]{P} 0$ : $\forall \varepsilon>0 \ \exists K(\varepsilon)$ s.t. if $n>k(\varepsilon)$, $P(|Z_n|>\varepsilon)=0$

$Z_n = O_p(1)$ : $\forall \varepsilon>0 \ \exists M$ s.t. $\sup_n P(|Z_n|>M) < \varepsilon$

↳ In simpler terms, for any probability tolerance $\delta$, we can find a bound $M$ s.t. the probability of $Z_n$ falling outside $[-M, M]$ is uniformly small for all $n$.

2. The proof:

We need to show that for any given $\varepsilon>0$, we can find a constant $M$ such that $P(|Z_n|>M) < \varepsilon$ for all $n$

Step 1: Fix the tolerance.
Let $\varepsilon>0$ be arbitrary.

Step 2: Handle the "tail" of the sequence (large $n$)
Size $Z_n \xrightarrow{P} 0$, we know that $P(|Z_n|>\eta) \to 0$ for any $\eta>0$.
Let's arbitrarily choose $\eta=1$. Because the limit is 0, there exists an integer $N$ such that for all $n>N$:

$$P(|Z_n|>1) < \varepsilon$$

This handles all the terms in the sequence after $N$. For these terms, a bound of $M=1$ is sufficient to keep the probability below $\varepsilon$.

Step 3: Handle the "beginning" of the sequence (finite $n$) Now we must consider the finite set of random variables $\{Z_1, Z_2, \cdots, Z_N\}$. For any single random variable $X$, $\lim_{x\to\infty} P(|X|>x)=0$. Therefore, for each $Z_i$ $(i=1,\cdots,N)$, we can find a constant $M_i$ such that:

$$P(|Z_i|>M_i) < \varepsilon$$

Step 4: Construct the global bound $M$
To ensure the condition holds for all $n$, we take the maximum of the bound required for tail (1) and the bounds required for the first $N$ terms:

$$M = \max\{1, M_1, M_2, \cdots, M_N\}$$

Step 5: Verify the bound:
Now, let's check the condition $P(|Z_n|>M)$ for any $n$:

- Case 1: $n \le N$. Since $M \ge M_n$, the event $\{|Z_n|>M\}$ is a subset of $\{|Z_n|>M_n\}$.
- Case 2: $n>N$. Since $M \ge 1$, the event $\{|Z_n|>M\}$ is a subset of $\{|Z_n|>1\}$

$$P(|Z_n|>M) \le P(|Z_n|>1) < \delta$$

Conclusion:
For any $\delta>0$, we have constructed an $M$ such that $\sup_n P(|Z_n|>M) < \delta$.
Therefore $Z_n$ is bounded in probability.

Intuitive note: If a sequence converges to 0, the probability mass is concentrating tighter and tighter around 0. Since the mass is collapsing inward, it cannot be "escaping" to infinity. Therefore, it must be bounded.

---

☆☆

5.11 확률변수 $X_n$과 $Z$의 누적분포함수가 모든 실수에서 연속인 경우에

$$X_n \xrightarrow[n\to\infty]{d} Z$$

이면 $X_n = O_p(1)$, $n\to\infty$ 임을 밝혀라.

$X_n \xrightarrow[n\to\infty]{d} Z$ : $\lim_{n\to\infty} P(X_n \le x) = P(Z \le x)$

$X_n = O_p(1), n\to\infty$ : $\forall \varepsilon, \exists M$ s.t. if $m>M$, $\sup_n P(|X_n|>m) < \varepsilon$

$P(|X_n|>M) = 1 - P(|X_n|<M) = 1 - P(-M<X_n<M)$

The cdf is continuous on all points.
$\therefore \forall \varepsilon>0, \ \exists M$ s.t. $P(|Z|>M) < \varepsilon$ … ①
$\therefore \forall \varepsilon>0, \ \exists M$ s.t. $P(|X_n|>M) < \varepsilon$
$\Rightarrow \exists m_1 : P(|X_1|>m_1) < \varepsilon, \cdots, \exists m_N : P(|X_N|>m_N) < \varepsilon$
Let $M = \max\{m_1, \cdots, m_N\}$

Then $\forall n : 1 \le n \le N$, $P(|X_n|>M) < \varepsilon$
Meanwhile, if $n>N$, $\exists N$ s.t. if $n>N$, $|P(X_n<x) - P(Z<x)| < \varepsilon$

$\Rightarrow |P(X_n>x) - P(Z>x)| < \varepsilon \Rightarrow |P(X_n>M)| - |P(Z>M)| < \varepsilon$
$\Rightarrow |P(X_n>M)| < \varepsilon + |P(Z>M)| < 2\varepsilon$
$\Rightarrow |P(X_n>M)| < 2\varepsilon$

* Key takeaway:
① CDF는 $x\to+\infty$ 일 때 1로 수렴. ⇒ for large $k$, $P(Z<K) < \varepsilon$
② $X_n \xrightarrow[n\to\infty]{d} Z$ : 모든 $\varepsilon>0$에 대하여 $n \ge N(\varepsilon, k)$ 일 경우, $|P(X_n<k) - P(Z<k)| < \varepsilon$ 이다.
③ $N$을 기준으로 $1\le n\le N$, $n>N$ 2가지 경우로 나누어 생각.
④ $\sup_n$ 때문에 $1\le n\le N$은 각각의 경우로 나누어서 생각.
$n>N$은 충분히 경계를 잡을

5.12. 포아송분포 Poisson($\lambda$)로부터의 랜덤표본을 이용한 표본 평균 $\bar{x}_n$에 대하여

$$\sqrt{n}\{g(\bar{x}_n)-g(\lambda)\} \xrightarrow{d} Z,\ Z\sim N(0,1)$$

가 성립하는 분산안정변환 $g$를 구하고, 이 결과를 이용하여 $\lambda$에 관한 점근신뢰구간을 구하는 방법을 설명하여라.

$X\sim \text{Poisson}(\lambda)$, $E[X]=\lambda$, $\text{Var}(\lambda)$

By CLT, $\sqrt{n}(\bar{x}-\lambda)\sim N(0,\lambda)$

$\sqrt{n}\{g(\bar{x})-g(\lambda)\} \longrightarrow \dot{g}(\lambda)Z\sim N(0,\dot{g}(\lambda)^2\lambda)$

$\therefore \dot{g}(\lambda)=\frac{1}{\sqrt{\lambda}}=\lambda^{-1/2} \Rightarrow \dot{g}(x)=x^{-1/2} \Rightarrow g(x)=\int x^{-1/2}dx=2x^{\frac{1}{2}}$

$\therefore \sqrt{n}(2\sqrt{\bar{x}}-2\sqrt{\lambda})\sim N(0,1)$

$100(1-\alpha)\%$ CI: $\left(-z_{\alpha/2} < \sqrt{n}(2\sqrt{\bar{x}}-2\sqrt{\lambda}) < \frac{1}{\sqrt{n}}z_{\alpha/2}\right)$

$\Rightarrow \left(-\frac{1}{\sqrt{n}}z_{\alpha/2} < 2(\sqrt{\lambda}-\sqrt{\bar{x}}) < \frac{1}{2\sqrt{n}}z_{\alpha/2}\right)$

$\Rightarrow \left(-\frac{1}{2\sqrt{n}}z_{\alpha/2} < \sqrt{\lambda}-\sqrt{\bar{x}} < \frac{1}{2\sqrt{n}}z_{\alpha/2}\right)$

$\Rightarrow \left(\sqrt{\bar{x}}-\frac{1}{2\sqrt{n}}z_{\alpha/2} < \sqrt{\lambda} < \sqrt{\bar{x}}+\frac{1}{2\sqrt{n}}z_{\alpha/2}\right)$

$\Rightarrow \left(\left(\sqrt{\bar{x}}-\frac{1}{2\sqrt{n}}z_{\alpha/2}\right)^2 < \lambda < \left(\sqrt{\bar{x}}+\frac{1}{2\sqrt{n}}z_{\alpha/2}\right)^2\right)$ ✓✓

---

5.13. 베르누이분포 Bernoulli($p$)로부터의 랜덤표본 $X_1,\cdots,X_n$을 이용한 표본비율 $\hat{p}=\sum_{i=1}^{n}X_i/n$에 대하여

$$\sqrt{n}\{g(\hat{p}_n)-g(p)\} \xrightarrow[n\to\infty]{d} Z,\ Z\sim N(0,1)$$

가 성립하는 분산안정변환 $g$를 구하고, 이 결과를 이용하여 $p$에 관한 점근신뢰구간을 구하는 방법을 설명하여라.

$X\sim \text{Bernoulli}(p)$, $E[X]=p$, $\text{Var}(X)=p(1-p)$

$f_X(x)=p^x(1-p)^{1-x}\quad x=0,1$

$E[e^{tx}]=\sum_{x=0,1}e^{tx}p^x(1-p)^{1-x}=(1-p)+e^tp$

$cgf_X(t)=\log(1-p+e^tp)$, $c'(t)=\frac{e^tp}{1-p+e^tp}$, $c''(t)=\frac{e^tp}{1-p+e^tp}-\frac{(e^tp)^2}{1-p+e^tp}$

$c''(t=0)=p-p^2=p(1-p)$

$\sqrt{n}(\hat{p}_n-p)\longrightarrow Z,\ Z\sim N(0,p(1-p))$

$\sqrt{n}\{g(\hat{p}_n)-g(p)\}\longrightarrow \dot{g}(p)Z,\ \dot{g}(p)Z\sim N(0,\dot{g}(p)^2p(1-p))$

$\dot{g}(p)=(p(1-p))^{-1/2}=\left(\frac{1}{p(1-p)}\right)^{1/2}=\left(\frac{1}{p}+\frac{1}{1-p}\right)^{1/2}$

Let $p=\sin^2\theta$, $1-p=1-\sin^2\theta=\cos^2\theta$

$dp=2\sin\theta\cos\theta\,d\theta \quad\therefore \frac{d\theta}{dp}=\frac{1}{2\sin\theta\cos\theta}$

$\int \dot{g}(p)dp=\int (\sin^2\theta\cos^2\theta)^{-1/2}\cdot 2\sin\theta\cos\theta\,d\theta=\int 2\,d\theta=2\theta$

$=2\sin^{-1}(p)+C$

$\therefore 2\sqrt{n}\{\sin^{-1}(\hat{p}_n)-\sin^{-1}(p)\}\to Z(0,1)$

$100(1-\alpha)\%$ CI: $\left(-z_{\alpha/2} < 2\sqrt{n}\{\sin^{-1}(\hat{p}_n)-\sin^{-1}(p)\} < z_{\alpha/2}\right)$

$: \left(\sin^{-1}(\hat{p}_n)-\frac{1}{2\sqrt{n}}z_{\alpha/2} < \sin^{-1}(p) < \sin^{-1}(\hat{p}_n)+\frac{1}{2\sqrt{n}}z_{\alpha/2}\right)$

$: \left(\sin\left[\sin^{-1}(\hat{p}_n)-\frac{1}{2\sqrt{n}}z_{\alpha/2}\right] < p < \sin\left[\sin^{-1}(\hat{p}_n)+\frac{1}{2\sqrt{n}}z_{\alpha/2}\right]\right)$

$\therefore$ $100(1-\alpha)\%$ CI (asymptotic):

$\left(\sin\left[\sin^{-1}(\hat{p}_n)-\frac{1}{2\sqrt{n}}z_{\alpha/2}\right],\ \sin\left[\sin^{-1}(\hat{p}_n)+\frac{1}{2\sqrt{n}}z_{\alpha/2}\right]\right)$

---

5.14. 서로 독립이고 성공률이 $p(0<p<1)$인 베르누이 시행 $X_1,\cdots,X_n,\cdots$에서 $r$번째 성공까지의 시행회수를 $W_r$이라고 할 때

$$\hat{p}_r=r/W_r$$

에 대한 다음 질문에 답하여라.

(a) $r$이 한없이 커질 때 $\sqrt{r}(\hat{p}_r-p)$의 극한분포를 구하여라.

(b) $r$이 한없이 커질 때 $\sqrt{r}\{g(\hat{p}_r)-g(p)\}\xrightarrow[n\to\infty]{d}N(0,1)$

(a) $W_r=W_1\oplus W_2-W_1\oplus W_3-W_2\oplus\cdots\oplus W_r-W_{r-1}$, $W_i-W_{i-1}\sim Geo(p)$

$\frac{W_r}{r}=\frac{1}{r}\sum_{i=1}^{r}X_i$, $X_i\sim Geo(p)$, $E[X]=\frac{1-p}{p}$, $Var(X)=\frac{1-p}{p^2}$

$\therefore \sqrt{r}(\bar{x}-E[X])\sim N(0,(1-p)\cdot p^{-2})$

$\Leftrightarrow \sqrt{r}\left(\frac{W_r}{r}-E\left[\frac{W_r}{r}\right]\right)\sim N(0,(1-p)p^{-2})$

$g_1(x)=x^{-1}$, $\dot{g}_1(x)=-x^{-2}$ $\quad \dot{g}_1\left(\frac{W_r}{r}\right)=-\left(\frac{W_r}{r}\right)^{-2}=-\left(\frac{r}{W_r}\right)^2=-\hat{p}^2$

$\therefore \dot{g}_1(E[\frac{W_r}{r}])=-p^2$

$\Leftrightarrow \sqrt{r}(\hat{p}-p)\sim N(0,\dot{g}_1(E[\frac{W_r}{r}])^2\cdot(1-p)p^{-2})\equiv N(0,(1-p)p^2)$ ✓✓

(b) $\sqrt{r}\{g_2(\hat{p})-g_2(p)\}\xrightarrow[n\to\infty]{d}\dot{g}_2(p)Z$, $\dot{g}_2(p)Z\sim N(0,\dot{g}_2(p)^2(1-p)p^2)$

$\therefore \dot{g}_2(p)^2=\frac{1}{(1-p)p^2}\quad\Leftrightarrow\quad g_2(p)=\frac{1}{\sqrt{(1-p)p^2}}=\frac{1}{p\sqrt{1-p}}$

let $p=\cos^2(\theta)$, $dp=-2\sin(\theta)\cos(\theta)$ $\qquad 1-p=\sin^2\theta\ \therefore\sqrt{1-p}=\sin\theta$

$\int\frac{1}{\sqrt{1-p}\,p}dp=\int\frac{1}{\sin\theta\cdot\cos^2\theta}\cdot-2\sin\theta\cos\theta\,d\theta=-2\int\frac{1}{\cos\theta}d\theta$

$1-p=t^2\quad -dp=2t\,dt$, $dp=-2t\,dt$

$p=1-t^2$

$\int\frac{1}{\sqrt{1-p}\cdot p}dp=\int\frac{1}{\sqrt{t^2}(1-t^2)}\cdot-2t\,dt=\int\frac{-2}{1-t^2}dt=\int\frac{-2}{(1-t)(1+t)}dt$

$=\int-\left(\frac{1}{1-t}+\frac{1}{1+t}\right)dt=\log|1-t|-\log|1+t|=\log\left|\frac{1-t}{1+t}\right|$

$=\log\left(\frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}\right)$ ✓✓

$100(1-\alpha)\%$ asymptotic CI : $\left(-z_{\alpha/2} < \log\left(\frac{1-\sqrt{1-\hat{p}}}{1+\sqrt{1-\hat{p}}}\right) - \log\left(\frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}\right) < z_{\alpha/2}\right)$

Textbook does not detail the method of obtaining a confidence interval.

* Key Takeaways

· 저번 연습필요! 5.12 - 5.14

let $y = \log\left(\frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}\right) \Leftrightarrow e^y = \frac{1-\sqrt{1-p}}{1+\sqrt{1-p}}$

$e^y + e^y\sqrt{1-p} = 1 - \sqrt{1-p}$

$e^y\sqrt{1-p} + \sqrt{1-p} = 1 - e^y$

$\sqrt{1-p}\,(1+e^y) = 1 - e^y$

$\sqrt{1-p} = (1-e^y)/(1+e^y)$

$\therefore p = 1 - \left[\frac{1-e^y}{1+e^y}\right]^2$

$\therefore g^{-1}(x) = 1 - \left[\frac{1-e^x}{1+e^x}\right]^2$

$\therefore$ CI : $\left[\, g^{-1}(L),\ g^{-1}(U)\,\right]$, $L = g(\hat{p}) - \frac{z_{\alpha/2}}{\sqrt{n}}$, $U = g(\hat{p}) + \frac{z_{\alpha/2}}{\sqrt{n}}$

---

5.15. 〈예 5.3.8〉에서 표본분위수의 극한분포를 구하는 과정에서

$$Y_n = \frac{1}{n}z_1 + \cdots + \frac{1}{n-r_n+1}z_{r_n},\quad z_i \overset{iid}{\sim} \mathrm{Exp}(1)\ (i=1,2,\cdots,n)$$

$$W_n = \sqrt{n}\,\frac{Y_n - (-\log(1-a))}{\sqrt{a/(1-a)}}$$

으로 정의된 확률변수들에 대하여 다음이 성립함을 설명하여라.

(a) $Y_n$의 누율생성함수 $cgf_{Y_n}(s)$에 대하여

$$cgf_{Y_n}(s) \underset{n\to\infty}{\sim} -s\log(1-a) + \frac{s^2}{2n}\frac{a}{1-a}$$

(b) $W_n$의 누율생성함수 $cgf_{W_n}(t)$에 대하여

$$cgf_{W_n}(t) = cgf_{Y_n}\left(\sqrt{n}\,t/\sqrt{a/(1-a)}\right) + \sqrt{n}\,t\log(1-a)/\sqrt{a/(1-a)}$$

(c) $cgf_{W_n}(t) \underset{n\to\infty}{\sim} \frac{1}{2}t^2$

(a) $M_{Y_n}(s) = E[\exp\{s\cdot Y_n\}] = E\left[\exp\left\{s\left(\frac{1}{n}z_1 + \cdots + \frac{1}{n-r_n+1}z_{r_n}\right)\right\}\right]$

$= E\left[\exp\left\{s\cdot\frac{1}{n}z_1\right\}\right]\cdots E\left[\exp\left\{s\cdot\frac{1}{n-r_n+1}z_{r_n}\right\}\right]$

$= M_{z_1}(s/n)\cdots M_{z_{r_n}}(s/(n-r_n+1))$

$\mathrm{mgf}\ \mathrm{Exp}(1) = (1-t)^{-1}$

$\therefore$ mgf of $Y_n$ : $(1-\frac{s}{n})^{-1}(1-\frac{s}{n-1})^{-1}\cdots(1-\frac{s}{n-r_n+1})^{-1}$

cgf of $Y_n$ : $\sum_{k=0}^{r_n-1} -\log\left(1-\frac{s}{n-k}\right)$, $\because n\to\infty$, $a = r_n/n$

$= C(0) + \frac{C'(0)}{1!}s + \frac{C''(0)}{2!}s^2 + R_n(s)$, $R_n(s)\to 0$

$C(0) = 0$

$C'(0) = E[Y_n] = -\log(1-a)$

$C''(0) = \mathrm{Var}(Y_n) = \frac{1}{n}\frac{a}{1+a}$

$\therefore C_{Y_n}(s) = -s\log(1-a) + \frac{s^2}{2n}\frac{a}{1-a}$ $(n\to\infty)$

---

(b) $cgf_{W_n}(t) = \log E[\exp\{W_n\cdot t\}] = \log E\left[\exp\left\{\sqrt{n}\,\frac{Y_n - (-\log(1-a))}{\sqrt{a/(1-a)}}t\right\}\right]$

$= \log E\left[\exp\left\{Y_n\cdot\frac{\sqrt{n}}{\sqrt{a/(1-a)}}t + \frac{\sqrt{n}\log(1-a)}{\sqrt{a/(1-a)}}t\right\}\right]$

$= \log E\left[\exp\left\{Y_n\left(\frac{\sqrt{n}}{\sqrt{a/(1-a)}}t\right)\right\}\right] + \frac{\sqrt{n}\log(1-a)}{\sqrt{a/(1-a)}}t$

$= cgf_{Y_n}\left(\sqrt{n}\,t/\sqrt{a/(1-a)}\right) + \sqrt{n}\,t\log(1-a)/\sqrt{a/(1-a)}$

(c) $cgf_{W_n}(t) = \underbrace{\left(\frac{\sqrt{n}\,t}{\sqrt{a/(1-a)}}\right)\log(1-a) + \left(\frac{nt^2}{a/(1-a)}\right)\cdot\frac{1}{2n}\cdot\frac{a}{1-a}}_{\text{cancel out}} + \sqrt{n}\,t\log(1-a)/\sqrt{a/(1-a)}$

$= \frac{1}{2}t^2$

$\therefore cgf_{W_n}(t) \underset{n\to\infty}{\sim} \frac{1}{2}t^2$

---

★★★★

5.16. 〈예 5.3.8〉에서와 같은 가정하에서 $r_n \underset{n\to\infty}{\sim} an$, $s_n \underset{n\to\infty}{\sim} \beta n$ $(0<\alpha<\beta<1)$인 경우에 표본분위수 $X_{(r_n)}$과 $X_{(s_n)}$의 결합분포에 대한 극한분포를 유도하고, 이로부터 표본 사분위수범위인

$$X_{(s_n)} - X_{(r_n)}\ \left(r_n \underset{n\to\infty}{\sim} n/4,\ s_n \underset{n\to\infty}{\sim} 3n/4\right)$$

의 극한분포를 유도하여라.

Let $f_X(x)$ : pdf of $X$, $F_X(x)$ : cdf of $X$. Let $X_{(r_n)} = X_r$, $X_{(s_n)} = X_s$

$$\sqrt{n}\left(X_{(r_n)} - F^{-1}(\alpha)\right) \xrightarrow[n\to\infty]{d} N\left(0,\ \alpha(1-\alpha)/[f(F^{-1}(\alpha))]^2\right)$$

$$\sqrt{n}\left(X_{(s_n)} - F^{-1}(\beta)\right) \xrightarrow[n\to\infty]{d} N\left(0,\ \beta(1-\beta)/[f(F^{-1}(\beta))]^2\right)$$

$$Y_n = \begin{pmatrix} \frac{1}{n}z_1 + \cdots + \frac{1}{n-r_n+1}z_{r_n} \\ \frac{1}{n}z_1 + \cdots + \frac{1}{n-s_n+1}z_{s_n} \end{pmatrix}$$

$E[Y_{n1}] \sim -\log(1-a)$, $\mathrm{Var}(Y_{n1}) \sim \frac{1}{n}\frac{a}{1-a}$

$E[Y_{n2}] \sim \log(1-\beta)$, $\mathrm{Var}(Y_{n2}) \sim \frac{1}{n}\frac{\beta}{1-\beta}$

$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Var}(Y_1) \sim \frac{1}{n}\frac{a}{1-a}$

$X_{(r_n)} = h\left(\frac{1}{n}z_1 + \cdots + \frac{1}{n-r_n+1}z_{r_n}\right)$, $h(y) = F^{-1}(1-e^{-y})$

From Q.5.15, we know that $(Y_{n1} - E[Y_{n1}])/\mathrm{Var}(Y_{n1}) \sim N(0,1)$

$$\therefore \frac{Y_{n1} - (-\log(1-a))}{\sqrt{a/n(1-a)}} \sim N(0,1) \Leftrightarrow \sqrt{n}\,(Y_{n1} - (-\log(1-a))) \xrightarrow[n\to\infty]{d} N(0,\ a/(1-a))$$

$$\sqrt{n}\left(h(Y_{n1}) - h(-\log(1-a))\right) \xrightarrow[n\to\infty]{d} h'(-\log(1-a))\,N(0,\ a/(1-a))$$

let $z = F^{-1}(1-e^{-y}) \Rightarrow F(z) = 1-e^{-y} \Rightarrow f(z)dz = e^{-y}dy$ $\therefore \frac{dz}{dy} = e^{-y}/f(z)$

$\therefore h'(y) = e^{-y}/f(F^{-1}(1-e^{-y}))$

$h'(-\log(1-a)) = (1-a)/f(F^{-1}(a))$. let $h(-\log(1-a)) = \xi_{r_n}$

$\therefore \sqrt{n}\,(X_{(r_n)} - \xi_{r_n}) \xrightarrow[n\to\infty]{d} N\left(0,\ a(1-a)/[f(F^{-1}(a))]^2\right)$ □

we can follow similar steps for $Y_{n2}$. (생략)

$$\sqrt{n}\left(\begin{pmatrix}Y_{n1}\\Y_{n2}\end{pmatrix}-\begin{pmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{pmatrix}\right)\xrightarrow[n\to\infty]{d} N(0,\Sigma),\quad \Sigma=\begin{pmatrix}\mathrm{Var}(Y_{n1}) & \mathrm{Cov}(Y_{n1},Y_{n2})\\ \mathrm{Cov}(Y_{n1},Y_{n2}) & \mathrm{Var}(Y_{n2})\end{pmatrix}$$

$$\sqrt{n}\left(h\begin{pmatrix}Y_{n1}\\Y_{n2}\end{pmatrix}-h\begin{pmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{pmatrix}\right)\xrightarrow[n\to\infty]{d} \dot{h}\begin{pmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{pmatrix}\cdot N(0,\Sigma)$$

$$\overset{d}{=} N\left(0,\ \dot{h}\begin{pmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{pmatrix}^{t}\Sigma\,\dot{h}\begin{pmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{pmatrix}\right)$$

Let $\xi_\alpha = F^{-1}(\alpha)$ and $F^{-1}(\beta)=\xi_\beta$

$$\dot{h}[-\log(1-\alpha)] = (1-\alpha)/[f(F^{-1}(\alpha))] = (1-\alpha)/[f(\xi_\alpha)]$$
$$\dot{h}[-\log(1-\beta)] = (1-\beta)/[f(F^{-1}(\beta))] = (1-\beta)/[f(\xi_\beta)]$$

$$\therefore\ \dot{h}\left(\begin{bmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{bmatrix}\right)^{t}\Sigma\,\dot{h}\left(\begin{bmatrix}-\log(1-\alpha)\\-\log(1-\beta)\end{bmatrix}\right)$$

$$= \begin{bmatrix}\frac{1-\alpha}{f(\xi_\alpha)} & 0\\ 0 & \frac{1-\beta}{f(\xi_\beta)}\end{bmatrix}^{t}\begin{bmatrix}\mathrm{Var}(Y_{n1}) & \mathrm{Var}(Y_{n1})\\ \mathrm{Var}(Y_{n1}) & \mathrm{Var}(Y_{n2})\end{bmatrix}\begin{bmatrix}\frac{1-\alpha}{f(\xi_\alpha)} & 0\\ 0 & \frac{1-\beta}{f(\xi_\beta)}\end{bmatrix}$$

↳ $= \mathrm{Cov}(Y_{n1},Y_{n2})$

$$= \begin{bmatrix}\frac{1-\alpha}{f(\xi_\alpha)}\cdot\mathrm{Var}(Y_{n1}) & \frac{1-\alpha}{f(\xi_\alpha)}\mathrm{Var}(Y_{n1})\\ \frac{1-\beta}{f(\xi_\beta)}\cdot\mathrm{Var}(Y_{n2}) & \frac{1-\beta}{f(\xi_\beta)}\cdot\mathrm{Var}(Y_{n2})\end{bmatrix}\begin{bmatrix}\frac{1-\alpha}{f(\xi_\alpha)} & 0\\ 0 & \frac{1-\beta}{f(\xi_\beta)}\end{bmatrix}$$

$$= \begin{bmatrix}\frac{\alpha(1-\alpha)}{f(\xi_\alpha)^2} & \frac{\alpha(1-\beta)}{f(\xi_\alpha)f(\xi_\beta)}\\ \frac{\alpha(1-\beta)}{f(\xi_\alpha)f(\xi_\beta)} & \frac{\beta(1-\beta)}{f(\xi_\beta)^2}\end{bmatrix} = (\dot{h})^{t}\Sigma\,\dot{h}$$

$$\therefore\ \sqrt{n}\left(\begin{pmatrix}X_{(rn)}\\X_{(sn)}\end{pmatrix}-\begin{pmatrix}\xi_\alpha\\\xi_\beta\end{pmatrix}\right)\xrightarrow[n\to\infty]{d} N(0,(\dot{h})^{t}\Sigma\,\dot{h})$$

$$g\begin{bmatrix}X_1\\X_2\end{bmatrix} = X_2 - X_1$$
$$\dot{g}\begin{bmatrix}X_1\\X_2\end{bmatrix} = \begin{bmatrix}\frac{\partial}{\partial X_1}g & \frac{\partial}{\partial X_2}g\end{bmatrix} = [-1\quad 1]$$

$$\therefore\ \sqrt{n}\left(g\begin{bmatrix}X_{(rn)}\\X_{(sn)}\end{bmatrix}-g\begin{pmatrix}\xi_\alpha\\\xi_\beta\end{pmatrix}\right)\xrightarrow[n\to\infty]{d} N(0,(\dot{g})^{t}(\dot{h})^{t}\Sigma\,\dot{h}\,\dot{g})$$

$$(\dot{g})^{t}(\dot{h})^{t}\Sigma\,\dot{h}\,\dot{g}$$

$$= [-1\quad 1]\begin{bmatrix}\frac{\alpha(1-\alpha)}{f(\xi_\alpha)^2} & \frac{\alpha(1-\beta)}{f(\xi_\alpha)f(\xi_\beta)}\\ \frac{\alpha(1-\beta)}{f(\xi_\alpha)f(\xi_\beta)} & \frac{\beta(1-\beta)}{f(\xi_\beta)^2}\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}$$

$$= \frac{\alpha(1-\alpha)}{f(\xi_\alpha)^2} - 2\frac{\alpha(1-\beta)}{f(\xi_\alpha)f(\xi_\beta)} + \frac{\beta(1-\beta)}{f(\xi_\beta)^2}$$

(insert $\alpha=1/4$, $\beta=3/4$)

$$= \frac{1}{16}\left(\frac{3}{f(\xi_{0.25})^2} - \frac{2}{f(\xi_{0.25})f(\xi_{0.75})} + \frac{3}{f(\xi_{0.75})^2}\right) = \sigma^2_{IQR}$$

$$\sqrt{n}(IQR_n - IQR)\xrightarrow{d} N(0,\sigma^2_{IQR})$$

$$\begin{bmatrix}-1 & 1\end{bmatrix}\begin{bmatrix}a & b\\ b & c\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix} = \begin{bmatrix}b-a & -b+c\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix} = -b+a-b+c = a-2b+c$$

5.17.

확률밀도함수 $f(x;\mu,\sigma) = \frac{1}{\sigma}\frac{e^{-(x-\mu)/\sigma}}{(1+e^{-(x-\mu)/\sigma})^2}\,I_{(-\infty,+\infty)}(x)$

인 로지스틱분포 $L(\mu,\sigma)$, $-\infty<\mu<+\infty$, $\sigma>0$ 인 모형에서의 랜덤표본 $X_1,X_2,\cdots,X_n$ $(n\geq 2)$에 기초한 표본사분위수 $X_{([n/4])}$, $X_{([n/2])}$, $X_{([3n/4])}$ 을 통해 정의되는 다음과 같은 통계량들을 이용하여 $\mu,\sigma$에 대한 추론을 하려고 한다.

$$\hat{\mu}_n \equiv X_{([n/2])},\quad \hat{\sigma}_n \overset{d}{\equiv} (X_{[3n/4]} - X_{([n/4])})/(2\log 3)$$

표본 크기 $n$이 한없이 커질 때 $\sqrt{n}\frac{\hat{\mu}_n-\mu}{\hat{\sigma}_n}$ 의 극한분포를 구하고, 이 결과를 이용하여 $\mu$에 관한 95% 근사신뢰구간을 구하는 방법을 설명하여라.

① $X_{([n/2])} \overset{d}{\equiv} \frac{1}{n}Z_1+\cdots+\frac{1}{n-[\frac{n}{2}]+1}Z_{n/2}$

$\alpha=1/2$, $\beta=3/4$, $\gamma=1/4$

$$Y_\alpha = \frac{1}{n}Z_1+\cdots+\frac{1}{n-\frac{n}{2}+1}Z_{n/2},\quad \begin{cases}E[Y_\alpha]\sim -\log(1-\alpha) = -\log(1/2) = \log 2\\ \mathrm{Var}(Y_\alpha)\sim \alpha/(1-\alpha) = 1\end{cases}$$

From Q5.15. $\sqrt{n}(Y_\alpha - \log 2)\xrightarrow[n\to\infty]{d} N(0,1)$

$$\sqrt{n}(X_{(n/2)} - \xi_\alpha)\xrightarrow[n\to\infty]{d} N(0,\alpha(1-\alpha)/[f(F^{-1}(\alpha))]^2)$$
$$\overset{d}{=} N(0,1/4/[f(\xi_\alpha)]^2),\quad F^{-1}(\alpha)=\xi_\alpha$$

$h(y) = F^{-1}(1-e^{-y})$

$F(h(y)) = 1-e^{-y}$

$f(h)\,dh = e^{-y}dy$

$\therefore dh/dy = e^{-y}/f(h) = e^{-y}/f(F^{-1}(1-e^{-y}))$

$$\therefore\ \sqrt{n}(\hat{\mu}-\mu) = \sqrt{n}(X_{(n/2)} - \xi_\alpha)\xrightarrow[n\to\infty]{d} N(0,1/[4\{f(\xi_\alpha)\}^2])$$

② $\hat{\sigma}^2_n \overset{d}{\equiv} (X_{(3n/4)} - X_{(n/4)})/(2\log 3)$

$$\sqrt{n}\left(X_{(3n/4)} - X_{(n/4)} - (\xi_\beta - \xi_\gamma)\right)\xrightarrow[n\to\infty]{d} N(0,\sigma^2_{IQR})$$

$$\sigma^2_{IQR} = \frac{\beta(1-\beta)}{f(\xi_\beta)^2} - 2\frac{\gamma(1-\beta)}{f(\xi_\beta)f(\xi_\gamma)} + \frac{\gamma(1-\gamma)}{f(\xi_\gamma)^2} = \frac{1}{16}\left(\frac{3}{f(\xi_{0.75})^2} - 2\frac{1}{f(\xi_{0.75})f(\xi_{0.25})} + \frac{3}{f(\xi_{0.25})^2}\right)$$

$$f(x;\mu,\sigma) = \frac{1}{\sigma}\frac{e^{-(x-\mu)/\sigma}}{(1+e^{-(x-\mu)/\sigma})^2}\,I_{(-\infty,+\infty)}(x)$$

$F(x;\mu,\sigma) = (1+e^{-(x-\mu)/\sigma})^{-1}$

$y = (1+e^{-(x-\mu)/\sigma})^{-1}$

$1/y = 1+e^{-(x-\mu)/\sigma}$

$1/y - 1 = e^{-(x-\mu)/\sigma}$

$\log(y^{-1}-1) = -\frac{1}{\sigma}(x-\mu)$

$-\sigma\log(y^{-1}-1) = x-\mu$

$\therefore x = \mu - \sigma\log(y^{-1}-1)$

$\therefore F^{-1}(x) = \mu - \sigma\log(x^{-1}-1)$

$\xi_\beta = F^{-1}(3/4) = \mu - \sigma\log(1/3)$
$\quad = \mu + \sigma\log 3$

$\xi_\gamma = F^{-1}(1/4) = \mu - \sigma\log(3)$

$\therefore\ \xi_\beta - \xi_\gamma = 2\sigma\log(3)$

$\xi_\alpha = F^{-1}(1/2) = \mu - \sigma\log(2-1) = \mu$

$f(\xi_\alpha) = 1/6\cdot 1/4 = 1/4\sigma)^{-1}$

$exp\{-(\xi_\beta-\mu)/\sigma\} = 1/3$

$exp\{-(\xi_\gamma-\mu)/\sigma\} = 3$

$$\therefore f(\xi_\beta) = \frac{1}{\sigma}\frac{1/3}{(1+1/3)^2} = \frac{1}{\sigma}\frac{1}{3}\frac{9}{16} = \frac{3}{16}\sigma^{-1}$$

$$f(\xi_\gamma) = \frac{1}{\sigma}\frac{3}{(1+3)^2} = \frac{1}{\sigma}\frac{3}{16}$$

$$\therefore f(\xi_\beta)^2 = f(\xi_\beta)f(\xi_\gamma) = f(\xi_\gamma)^2 = \left(\frac{3}{16}\right)^2\sigma^{-2}$$

$$\therefore\ \sigma^2_{IQR} = \frac{1}{16}\left(4\cdot\left(\frac{16}{3}\right)^2\sigma^2\right) = \frac{64}{9}\sigma^2$$

↳ unnecessary

$$\sqrt{n}\,(\hat{\mu}_n-\mu)\xrightarrow[n\to\infty]{d} N\left(0,\ 1/4\cdot(4\sigma)^{-2}\right)\equiv N(0,4\sigma^2)$$

$$\therefore\ \sqrt{n}\,(\hat{\mu}_n-\mu)/2\sigma\xrightarrow[n\to\infty]{d} N(0,1)$$

$$\sqrt{n}\,(X_{(3n/4)}-X_{(n/4)}-(\xi_\beta-\xi_\gamma))\xrightarrow[n\to\infty]{d} N(0,\sigma^2_{IQR})\overset{d}{=}N\left(0,\frac{64}{9}\sigma^2\right)$$

$$\therefore\ \sqrt{n}\,(X_{(3n/4)}-X_{(n/4)}-2\sigma\log 3)\xrightarrow[n\to\infty]{d} N\left(0,\frac{64}{9}\sigma^2\right)\cdots *$$

$$\Rightarrow\ \sqrt{n}\,(X_{(3n/4)}-X_{(n/4)}-2\sigma\log 3)/(8\sigma/3)\xrightarrow[n\to\infty]{d} N(0,1)$$

---

$$*\ \ X_{(3n/4)}-X_{(n/4)}\xrightarrow[n\to\infty]{p}\xi_\beta-\xi_\gamma=2\sigma\log 3$$

$$\therefore\ (X_{(3n/4)}-X_{(n/4)})/2\log 3\xrightarrow[n\to\infty]{p}\sigma$$

$$\therefore\ \hat{\sigma}_n\xrightarrow[n\to\infty]{p}\sigma$$

$$\frac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{\sigma}\cdot\frac{\sigma}{\hat{\sigma}_n}\xrightarrow{n\to\infty} N(0,4)\cdot 1\quad(\because \text{Slutsky's Theorem})$$

$$\therefore\ \frac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{\hat{\sigma}_n}\xrightarrow[n\to\infty]{d} N(0,4)\ \Rightarrow\ \frac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{2\hat{\sigma}_n}\xrightarrow[n\to\infty]{d} N(0,1)$$

$100(1-\alpha)\%$ CI : $-z_{\alpha/2}<\dfrac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{2\hat{\sigma}_n}<z_{\alpha/2}$

$$\therefore\ \left(\hat{\mu}_n-2\hat{\sigma}_n\frac{z_{\alpha/2}}{\sqrt{n}},\ \hat{\mu}_n+2\hat{\sigma}_n\frac{z_{\alpha/2}}{\sqrt{n}}\right)$$

95% asymptotic CI : $\hat{\mu}\pm 1.96\cdot\dfrac{2\hat{\sigma}_n}{\sqrt{n}}$ ✓✓

---

5.18. 확률 밀도함수가

$$f(x;\mu,\sigma)=\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}\,I_{(-\infty,+\infty)}(x),\quad -\infty<\mu<+\infty,\ \sigma>0$$

인 이중지수분포 DE$(\mu,\sigma)$ 모형에서의 랜덤표본 $X_1,\cdots,X_n$ $(n\geq 2)$ 에 기초한 표본중앙값과 표본평균절대편차 (mean absolute deviation)를 각각

$$\hat{\mu}_n\equiv X_{([n/2])},\quad \hat{\sigma}_n\equiv\frac{1}{n}\sum_{i=1}^{n}|X_i-X_{([n/2])}|$$

이라고 하자

(a) $\hat{\sigma}_n\xrightarrow[n\to\infty]{p}\sigma$ 임을 밝혀라

(b) 표본크기 $n$이 한없이 커질 때 $\sqrt{n}\,\dfrac{\hat{\mu}_n-\mu}{\hat{\sigma}_n}$ 의 극한분포를 구하고, 이 결과를 이용하여 $\mu$에 관한 95% 근사신뢰구간을 구하는 방법을 설명하여라.

(a) $f(x;\mu,\sigma)=\dfrac{1}{2\sigma}e^{-|x-\mu|/\sigma}$

$$F(x;\mu,\sigma)=\int_{-\infty}^{x}\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}dx=\int_{-\infty}^{x\wedge\mu}\frac{1}{2\sigma}e^{-(\mu-x)/\sigma}dx+\int_{\mu}^{x}\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}dx\,I_{(\mu\leq x)}$$

$$=\left[\frac{1}{2}e^{-(\mu-x)/\sigma}\right]_{-\infty}^{x\wedge\mu}+\left[-\frac{1}{2}e^{-|x-\mu|/\sigma}\right]_{\mu}^{x}I_{(\mu\leq x)}$$

$$=\frac{1}{2}\exp\left\{-(\mu-x\wedge\mu)/\sigma\right\}+\frac{1}{2}\left[1-e^{-(x-\mu)/\sigma}\right]I_{(\mu\leq x)}$$

$$\therefore\ F(x)=\begin{cases}\frac{1}{2}\exp\left\{-(\mu-x)/\sigma\right\} & \text{if } x<\mu\\ 1-\frac{1}{2}\exp\left\{-(x-\mu)/\sigma\right\} & \text{if } x\geq\mu\end{cases}$$

$$y=\frac{1}{2}\exp\left\{-(\mu-x)/\sigma\right\}\ \Rightarrow\ \log(2y)=-(\mu-x)/\sigma$$
$$\Rightarrow\ \mu+\sigma\log(2y)$$

$$\therefore F^{-1}(x)=\mu+\sigma\log(2x)\quad(x<\mu)$$

$$y=1-\frac{1}{2}\exp\left\{-(x-\mu)/\sigma\right\}\ \Rightarrow\ \log(2-2y)=-(x-\mu)/\sigma$$
$$\Rightarrow\ x=\mu-\sigma\log(2(1-y))$$

$$\therefore F^{-1}(x)=\mu-\sigma\log(1-x)\quad(x\geq\mu)$$

$$F^{-1}(x)=\begin{cases}\mu+\sigma\log(2x) & \text{if } x<\mu\\ \mu-\sigma\log(2(1-x)) & \text{if } x\geq\mu\end{cases}$$

$F^{-1}(1/2)=\mu\ \ \therefore\ \xi_{1/2}=\mu$

* 확률수렴의 정의 이용.

$$\hat{\sigma}_n\equiv\frac{1}{n}\sum_{i}^{n}|X_i-X_{([n/2])}|\xrightarrow[n\to\infty]{p}E[|X-\mu|]\ \text{ by the Law of Large Numbers.}$$

$$E[|X-\mu|]=\int_{-\infty}^{\infty}|x-\mu|\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}dx$$

$$=\int_{\mu}^{\infty}|x-\mu|\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}dx+\int_{-\infty}^{\mu}|x-\mu|\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}dx$$

$$=\int_{\mu}^{\infty}(x-\mu)\frac{1}{2\sigma}e^{-(x-\mu)/\sigma}dx+\int_{-\infty}^{\mu}(\mu-x)\frac{1}{2\sigma}e^{-(\mu-x)}dx$$

$$=\int_{\mu}^{\infty}(x-\mu)\frac{1}{2\sigma}e^{-(x-\mu)/\sigma}dx+\int_{\mu}^{\infty}(x-\mu)\frac{1}{2\sigma}e^{-(x-\mu)}dx$$

$$=\int_{\mu}^{\infty}(x-\mu)\frac{1}{\sigma}e^{-(x-\mu)/\sigma}dx=\int_{0}^{\infty}y\frac{1}{\sigma}e^{-y/\sigma}dy=\sigma$$

(b) $\sqrt{n}\,(\hat{\mu}_n-\mu)=\sqrt{n}\,(X_{([\frac{n}{2}])}-\xi_{1/2})\xrightarrow[n\to\infty]{d} N(0,1/2(1-1/2)/[f(F^{-1}(1/2))]^2)$

$$=N(0,(4\cdot f(\mu))^{-2})$$

$f(\mu)=\dfrac{1}{2\sigma}\ \ \therefore\ 4\cdot f(\mu)^2=\sigma^{-2}$

$$\xrightarrow[n\to\infty]{d} N(0,\sigma^2)$$

$$\therefore\ \frac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{\hat{\sigma}}=\frac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{\sigma}\cdot\frac{\sigma}{\hat{\sigma}}\xrightarrow[n\to\infty]{d} N(0,1)\cdot 1=N(0,1)\quad(\because \text{Slutsky's Theorem})$$

$100(1-\alpha)\%$ CI : $-z_{\alpha/2}<\dfrac{\sqrt{n}\,(\hat{\mu}_n-\mu)}{\hat{\sigma}}<z_{\alpha/2}$

95% CI : $\left(\hat{\mu}_n\pm 1.96\dfrac{\hat{\sigma}}{\sqrt{n}}\right)$

5.18. TEXTBOOK SOLUTION.

(a) $|\,|a|-|b|\,| \le |a-b|$ 임을 이용하면

$$|\hat{\sigma}_n - \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|| \le \frac{1}{n}\sum_{i=1}^{n}\left||X_i - X_{([n/2])}|-|X_i-\mu|\right|$$

$$\le \frac{1}{n}\sum_{i=1}^{n}|X_{([n/2])}-\mu| = |X_{([n/2])}-\mu|$$

한편 이경우에 누적분포함수의 역함수는 연속이므로 예 5.2.7에서와 같이 분위수의 수렴성이 성립한다. 즉 $X_{([n/2])} \xrightarrow[n\to\infty]{P} \mu$ 이 성립하고 이로부터

$$\left(\hat{\sigma} - \frac{1}{n}\sum|X_i-\mu|\right) \xrightarrow[n\to\infty]{P} 0$$

그런데 대수의 법칙으로부터 $\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu| \xrightarrow[n\to\infty]{P} E(|X_i-\mu|)=\sigma$

따라서 $\hat{\sigma}_n \xrightarrow[n\to\infty]{P} \sigma$

(b) 이 경우에 확률밀도함수와 누적분포함수는 각각

$$f(x)=\frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right),\ g(z)=\frac{1}{2}e^{-|z|}$$

$$F(x)=G\left(\frac{x-\mu}{\sigma}\right),\ G(z)=\frac{1}{2}e^{z}I_{(z\le 0)}+\left(1-\frac{1}{2}e^{-z}\right)I_{(z>0)}$$

로서 예 5.3.8의 조건을 만족하므로

$$\sqrt{n}(\hat{\mu}_n-\mu) = \sqrt{n}(X_{([n/2])}-\mu) \xrightarrow[n\to\infty]{d} N(0,\pi^2)$$

이제 (a)와 슬럿츠키의 정리로부터 $\sqrt{n}\frac{\hat{\mu}_n-\mu}{\hat{\sigma}_n} \xrightarrow[n\to\infty]{d} N(0,1)$

따라서 $\mu$에 관한 95% 근사신뢰구간은 $\hat{\mu}_n \pm 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}$