연습문제 5.12.

- $X_1 \dots X_n \sim Poi(\lambda)$.
- $\sqrt{n}\big(g(\bar{X}) - g(\lambda)\big) \xrightarrow{d} Z, \quad Z \sim N(0,1)$ 을 만족하는 변환 g를 구하라.

- CLT를 쓰면
$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$$
- $g(\cdot)$이 미분가능한 함수라고 하자.
- $g(\cdot)$는 미분가능한 어떠한 형태도 가능하다. 예를들어 아래와 같은 형태들이 가능하다고 하자.
$$g(x) = x^2$$
혹은
$$g(x) = 2\sqrt{x}$$
- 예를들어서 $g(x) = 2\sqrt{x}$와 같은 형태라고 하자. 그렇다면 $x = a$에서 $g(x)$를 테일러 전개하면 아래와 같이 표현가능하다.
$$g(x) = g(a) + (x-a)g'(a) + (x-a)^2 O(1)$$
- x대신에 $\bar{X}_n$ a대신에 λ를 대입하면
$$g(\bar{X}_n) = g(\lambda) + (\bar{X}_n - \lambda)g'(\lambda) + (\bar{X}_n - \lambda)^2 O(1).$$
양변에 $\sqrt{n}$을 곱하고 정리하면
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) = \sqrt{n}(\bar{X}_n - \lambda)g'(\lambda) + \sqrt{n}(\bar{X}_n - \lambda)^2 O(1).$$
그런데 $\sqrt{n}(\bar{X}_n - \lambda)^2 O(1) = O_p(1)o_p(1)O(1) = o_p(1)$ 이므로
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) = \sqrt{n}(\bar{X}_n - \lambda)g'(\lambda) + o_p(1)$$
그런데 $g(x) = 2\sqrt{x}$이므로 $g'(x) = 1/\sqrt{x}$이다. 따라서 $g'(\lambda) = 1/\sqrt{\lambda}$.
- 정리하면
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) = \sqrt{n}(\bar{X} - \lambda)1/\sqrt{\lambda} + o_p(1)$$
CLT의 결과를 쓰면
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N(0, \lambda)1/\sqrt{\lambda}$$
따라서
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N(0, 1)$$

(결론) 만약에 $g(x) = 2\sqrt{x}$로 설정한다면
$$\sqrt{n}\big(g(\bar{X}) - g(\lambda)\big)$$
의 극한이 표준정규분포로 수렴하므로 문제에서 원하는 답을 구할 수 있다.

- 이는 바꾸어 말하면 포아송분포에 한정하여 (1) 관측한 확률변수에서 제곱근을 취한뒤 2배를 하고 (2) 그상태에서 CLT를 쓴다면 표준정규분포로 수렴한다는 것이다.
- 이 경우 그냥 포아송분포자체에서 CLT를 쓰는것보다 이득되는 경우가 있다. (일단 이렇게만 알아둘것)

- 그렇다면 어떻게 $g(x) = 2\sqrt{x}$임을 알 수 있을까?
- g가 미분가능하다면
$$g(x) \approx g(a) + (x-a)g'(a)$$
이고 다시 정리하면
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \approx \sqrt{n}(\bar{X}_n - \lambda)g'(\lambda)$$
따라서
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N(0, \lambda)g'(\lambda)$$
이고 정규분포의 특성에 따라서
$$\sqrt{n}\big(g(\bar{X}_n) - g(\lambda)\big) \xrightarrow{d} N\big(0, \lambda(g'(\lambda))^2\big)$$
이다. 따라서
$$\lambda(g'(\lambda))^2 = 1$$
이어야 할 조건은
$$g'(\lambda) = \frac{1}{\sqrt{\lambda}}$$
일 조건이다. 즉 $g(x)$의 형태는 모르겠지만 $g'(x)$는 아래와 같아야 함을 알 수 있다.
$$g'(x) = \frac{1}{\sqrt{x}}$$
이러부터 $g(x)$를 유추하면
$$g(x) = \int \frac{1}{\sqrt{x}} dx = 2\sqrt{x}.$$

- 이제 λ의 점근적인 신뢰구간을 구해보자. 아래의 식을 관찰하자.
$$2\sqrt{n}\big(\sqrt{\bar{X}_n} - \sqrt{\lambda}\big) \xrightarrow{d} N(0,1)$$
이때 $2\sqrt{n}\big(\sqrt{\bar{X}_n} - \sqrt{\lambda}\big)$자체를 하나의 통계량 Y_n으로 본다면
$$Y_n \xrightarrow{d} N(0,1)$$
Y_n이 정규분포를 따른다면 아래가 성립한다.
$$P\big(-Z_{\alpha/2} < Y_n < Z_{\alpha/2}\big) = 1 - \alpha$$
Y_n이 정규분포는 아니지만 점근적으로 정규분포를 따른다면 아래가 성립한다.
$$\lim_{n\to\infty} P\big(-Z_{\alpha/2} < Y_n < Z_{\alpha/2}\big) = 1 - \alpha$$
따라서
$$\lim_{n\to\infty} P\left(-Z_{\alpha/2} < 2\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\lambda}) < Z_{\alpha/2}\right) = 1 - \alpha$$

이는
$$\lim_{n\to\infty} P\left(-Z_{\alpha/2} < 2\sqrt{n}(\sqrt{\lambda}-\sqrt{\bar{X}_n}) < Z_{\alpha/2}\right) = 1-\alpha$$
정리하면
$$\lim_{n\to\infty} P\left(\frac{-Z_{\alpha/2}}{2\sqrt{n}}+\sqrt{\bar{X}_n} < \sqrt{\lambda} < \frac{Z_{\alpha/2}}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right) = 1-\alpha$$
따라서
$$\lim_{n\to\infty} P\left(\left(\frac{-Z_{\alpha/2}}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right)^2 < \lambda < \left(\frac{Z_{\alpha/2}}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right)^2\right) = 1-\alpha$$
이 성립한다.

- 위의 식이 의미하는것은 λ가 구간
$$\left(\left(\frac{-Z_{\alpha/2}}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right)^2, \left(\frac{Z_{\alpha/2}}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right)^2\right)$$
에 포함될 확률이 점근적으로 $1-\alpha$에 수렴한다는 뜻이다. 이는 바꾸어 말하면 매우 큰 n에 대하여 $1-\alpha$의 확률로 모수가 위의 구간에 있음을 확신할 수 있다는 의미이다. 이를 $1-\alpha$의 점근적 신뢰구간이라고 부른다.

- 예를들어 λ의 점근적인 95퍼센트 신뢰구간은 아래와 같이 구할 수 있다.
$$\left(\left(\frac{-1.96}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right)^2, \left(\frac{1.96}{2\sqrt{n}}+\sqrt{\bar{X}_n}\right)^2\right)$$

연습문제 5.14.

- $\begin{pmatrix}X_1\\Y_1\end{pmatrix},\dots,\begin{pmatrix}X_n\\Y_n\end{pmatrix} \overset{idd}{\sim} F$ with

(1) $EX_1 = EY_1 = 0$,

(2) $EX_1^2 = EY_1^2 = 1$,

(3) $EX_1Y_1 = \rho$,

(4) $EX_1^4 < \infty$, $EY_1^4 < \infty$.

- $\hat{\rho}_n = \frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2}\sqrt{\sum_{i=1}^n (Y_i-\bar{Y})^2}}$.

- 아래와 같은 기호를 약속하자.

(1) $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

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

(3) $\overline{XY} = \frac{1}{n}\sum_{i=1}^n X_iY_i$.

(a) $\sqrt{n}\bar{X}\bar{Y} \overset{p}{\to} 0$ 임을 보여라.

(sol)

- $\sqrt{n}\bar{X} = O_p(1)$, $\bar{Y} = o_p(1)$ 이므로 자명함. 여기에서 $\sqrt{n}\bar{X} = O_p(1)$ 는 CLT에 의해서 성립하고 $\bar{Y} = o_p(1)$ 는 큰수의 법칙 즉 WLLN에 의해서 성립한다.

note: 추가적으로 $\bar{XY} - \bar{X}\bar{Y} = \bar{XY} + o_p(\frac{1}{\sqrt{n}})$ 임도 알 수 있다.

(b) 아래가 성립함을 보여라.
$$\sqrt{m_2 - m_1^2} - \sqrt{m_2} = \frac{-m_1^2}{\sqrt{m_2-m_1^2}+\sqrt{m_2}}$$
이때 $m_1 = \bar{X}$, $m_2 = \bar{X^2}$ 라고 하자.

(sol) (자명하다)

note: 추가적으로 아래를 확인가능하다.
$$\sqrt{m_2 - m_1^2} = \sqrt{m_2} + \frac{-m_1^2}{\sqrt{m_2-m_1^2}+\sqrt{m_2}} = \sqrt{m_2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

(pf)

- $\sqrt{n}m_1^2 = \sqrt{n}m_1m_1 = O_p(1)o_p(1) = o_p(1) \quad \therefore m_1^2 = o_p(\frac{1}{\sqrt{n}})$.

- $m_2 = 1 + o_p(1)$ 이다. 즉 $m_2 \overset{p}{\to} 1$이다. 또한 $m_1^2 = o_p(\frac{1}{\sqrt{n}})$ 이다. 따라서
$$\frac{-m_1^2}{\sqrt{m_2-m_1^2}+\sqrt{m_2}} = o_p\left(\frac{1}{\sqrt{n}}\right)$$
이다. 왜냐하면 분모는 1로 확률수렴하고 (continuous mappting thm) 분자는 $o_p(\frac{1}{\sqrt{n}})$이기 때문.

(c) 아래가 성립함을 보여라.
$$\sqrt{m_2} - 1 = \frac{m_2-1}{2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

(sol)

- 아래식을 관찰하라.
$$\sqrt{m_2} - 1 = \frac{m_2-1}{\sqrt{m_2}+1} = \frac{m_2-1}{2} + \left(\frac{1}{\sqrt{m_2}+1} - \frac{1}{2}\right)(m_2-1)$$

- 이제 아래를 보이면 된다.

$$\left(\frac{1}{\sqrt{m_2}+1}-\frac{1}{2}\right)(m_2-1) := r_{3n} = o_p\left(\frac{1}{\sqrt{n}}\right)$$

따라서 아래를 보이면 된다.

$$\sqrt{n}\left(\frac{1}{\sqrt{m_2}+1}-\frac{1}{2}\right)(m_2-1) = o_p(1)$$

- 그런데 CLT에 의해서 $\sqrt{n}(m_2-1) = O_p(1)$ 가 성립한다.
- continuous mapping thm에 의해서 아래가 성립한다.

$$m_2 \xrightarrow{p} 1 \quad \Longrightarrow \quad \left(\frac{1}{\sqrt{m_2}+1}-\frac{1}{2}\right) \xrightarrow{p} 0$$

- 따라서

$$\left(\frac{1}{\sqrt{m_2}+1}-\frac{1}{2}\right) = o_p(1)$$

- 따라서 증명이 끝난다.

(d) 아래를 보여라.

$$c_1 - m_1 s_1 - \rho\sqrt{m_2-m_1^2}\sqrt{s_2-s_1^2} = c_1 - \rho\frac{m_2+s_2}{2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

이때 $c_1 = \bar{XY}$, $s_1 = \bar{Y}$, $s_2 = \bar{Y}^2$이다.

(sol)

- (b)의 결과로부터 아래가 성립한다.

$$\sqrt{m_2-m_1^2} = \sqrt{m_2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

- (c)의 결과로부터 아래가 성립한다.

$$\sqrt{m_2}-1 = \frac{m_2-1}{2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

- 둘을 종합하면

$$\sqrt{m_2-m_1^2} = 1 + \frac{m_2-1}{2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

- 따라서

$$\sqrt{m_2-m_1^2}\sqrt{s_2-s_1^2}$$

는 아래식들의 합이다.

(1) $1+\frac{s_2-1}{2}+o_p(\frac{1}{\sqrt{n}})$

(2) $\frac{m_2-1}{2}\left(1+\frac{s_2-1}{2}+o_p(\frac{1}{\sqrt{n}})\right) = \frac{m_2-1}{2}+O_p(\frac{1}{\sqrt{n}})O_p(\frac{1}{\sqrt{n}})+o_p(\frac{1}{\sqrt{n}})$

note: $O_p(\frac{1}{\sqrt{n}})O_p(\frac{1}{\sqrt{n}}) = \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}O_p(1) = \frac{1}{\sqrt{n}}o_p(1) = o_p(\frac{1}{\sqrt{n}})$

(2) 따라서 결국 $\frac{m_2-1}{2}\left(1+\frac{s_2-1}{2}+o_p(\frac{1}{\sqrt{n}})\right) = \frac{m_2-1}{2}+o_p(\frac{1}{\sqrt{n}})$.

(3) $o_p(1/\sqrt{n})\left(1+\frac{s_2-1}{2}+o_p(1/\sqrt{n})\right) = o_p(\frac{1}{\sqrt{n}})$

- 따라서

$$\sqrt{m_2-m_1^2}\sqrt{s_2-s_1^2} = 1+\frac{s_2-1}{2}+\frac{m_2-1}{2}+o_p\left(\frac{1}{\sqrt{n}}\right)$$

정리하면

$$\sqrt{m_2-m_1^2}\sqrt{s_2-s_1^2} = \frac{s_2+m_2}{2}+o_p\left(\frac{1}{\sqrt{n}}\right)$$

- 한편 (a) 로부터

$$m_1 s_1 = \bar{X}\bar{Y} = o_p\left(\frac{1}{\sqrt{n}}\right)$$

- 따라서 증명이 끝난다.

(e) $\sqrt{n}(\hat{\rho}_n-\rho) = \sqrt{n}\bar{W}+o_p(1)$ 임을 보여라. 단

$$W = XY - \frac{\rho(X^2+Y^2)}{2}.$$

(sol)

- $\bar{W} = \bar{XY} - \frac{\rho(\bar{X}^2+\bar{Y}^2)}{2} = c_1 - \frac{\rho(m_2+s_2)}{2}$.
- $\sqrt{n}(\hat{\rho}_n-\rho) = \sqrt{n}\times\frac{c_1-m_1s_1-\rho\sqrt{m_2-m_1^2}\sqrt{s_2-s_1^2}}{\sqrt{m_2-m_1^2}\sqrt{s_2-s_1^2}}$
- (d)에 의해서 "분자" $= c_1-\frac{\rho(m_2+s_2)}{2}+o_p(\frac{1}{\sqrt{n}})$
- 분모는 1로 확률수렴한다.
- 따라서

$$\sqrt{n}(\hat{\rho}_n-\rho) = \sqrt{n}\times\frac{c_1-\frac{\rho(m_2+s_2)}{2}+o_p(\frac{1}{\sqrt{n}})}{1+o_p(1)} = \frac{\sqrt{n}\bar{W}+o_p(1)}{1+o_p(1)} = \sqrt{n}\bar{W}+o_p(1)$$

정리 4.4.3

- X가 연속형 확률변수라고 하자. $F(x)$를 X의 *cdf*라고 하자.
- $F(x)$도 분명히 함수이므로 $F(X)$를 정의할 수 있다. $F(X)$는 확률변수가 된다.
- 확률변수이므로 분포를 따른텐데 $F(X)$는 아래의 분포를 따름이 알려져 있다.

$$F(X) \sim U(0,1)$$

- 또한 $F^{-1}(U)$ 역시 확률변수가 된다. (단 $U \sim U(0,1)$.) 그런데 이 확률변수는 $F^{-1}(U)$ 는 X와 분포가 같음이 알려져 있다. 즉

$$F^{-1}(U) \overset{d}{=} X$$

이다.

- $U_1,\dots,U_n \overset{iid}{\sim} U(0,1)$ 라고 하자. 그리고 $U_{(1)},\dots,U_{(n)}$을 $U_1,\dots,U_n$의 순서통계량이라고 하자. 정리 4.4.3 에 의해서

$$\begin{aligned} F^{-1}(U_{(1)}) &:= X_1^* \sim F \\ F^{-1}(U_{(2)}) &:= X_2^* \sim F \\ &\dots \\ F^{-1}(U_{(n)}) &:= X_n^* \sim F \end{aligned}$$

이다.

- F가 순증가 함수이므로 $X_1^* < \cdots < X_n^*$이다.

- 아래가 성립한다. (왜??)

$$\begin{pmatrix} X_1^* \\ \dots \\ X_n^* \end{pmatrix} \overset{d}{=} \begin{pmatrix} X_{(1)} \\ \dots \\ X_{(n)} \end{pmatrix}$$

이말은 X의 cdf F를 알고 있을 경우 X의 순서통계량을 어떻게 생성할지 알려준다.

정리 4.4.3.

- 어떤분포의 순서통계량: $X_1,\dots,X_n \overset{iid}{\sim} F \implies X_{(1)},\dots,X_{(r)}$.

- 균등분포의 순서통계량: $U_1,\dots,U_n \overset{iid}{\sim} U(0,1) \implies U_{(1)},\dots,U_{(n)}$.

- 지수분포의 순서통계량: $Y_1,\dots,Y_n \overset{iid}{\sim} Exp(1) \implies Y_{(1)},\dots,Y_{(n)}$.

- 균등분포의 순서통계량과 지수분포의 순서통계량에는 아래와 같은 관계가 있다. (예제 4.3.4.)

$$U_{(r)} \overset{d}{=} 1 - e^{-Y_{(r)}}, \quad r = 1,2,\dots,n$$

- 그런데 임의의 $r = 1,2,\dots,n$에 대하여 지수분포의 순서통계량 $Y_{(r)}$은 아래를 만족한다. (예제 4.3.3.)

$$\forall r,\ \exists Z_1,\dots,Z_r \overset{iid}{\sim} Exp(1)\ s.t. \quad Y_{(r)} \overset{d}{=} \frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r$$

- 순서통계량 $X_{(r)}$은 아래와 같이 얻을 수 있다.

$$X_{(r)} \overset{d}{=} F^{-1}(U_{(r)})$$

- 그런데 (1) $U_{(r)} \overset{d}{=} 1 - e^{-Y_{(r)}}$ 와 (2) $Y_{(r)} \overset{d}{=} \frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r$을 이용하면

$$X_{(r)} \overset{d}{=} F^{-1}\left(1 - e^{-\left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)}\right)$$

따라서 $h(\bigstar) = F^{-1}\big(1 - e^{-\bigstar}\big)$라고 정의하면

$$X_{(r)} \overset{d}{=} h\left(\frac{1}{n}Z_1 + \cdots + \frac{1}{n-r+1}Z_r\right)$$

가 성립한다.