백분위수

- 이번에는 백분위수에 대하여 다룬다.
- 참고교재는 김우철 수리통계학책.

정리 4.4.3

- ullet X가 연속형 확률변수라고 하자. F(x)를 X의 cdf라고 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 확률 변수는 $F^{-1}(U)$ 는 X와 분포가 같음이 알려져 있다. 즉

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

이다.

__oo&f&X

~~~~~~

• $U_1, \ldots, U_n \stackrel{iid}{\sim} U(0,1)$ 라고 하자. 그리고 $U_{(1)}, \ldots, U_{(n)}$ 을 U_1, \ldots, U_n 의 순서통계량이라고 하자. 정리 4·4·3 에 의해서

$$F^{-1}(U_{(1)}) := X_1^* \sim F$$

 $F^{-1}(U_{(2)}) := X_2^* \sim F$
...

$$F^{-1}(U_{(n)}) := X_n^* \sim F$$

이다.

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

아래가 성립한다. (왜??)

$$egin{pmatrix} X_1^* \\ \dots \\ X_n^* \end{pmatrix} \stackrel{d}{=} egin{pmatrix} X_{(1)} \\ \dots \\ X_{(n)} \end{pmatrix}$$
이는 X 의 cdf F 를 알고 있을 경우 X 의 순서통계량을 어떻게 생성할지

알려준다.

• 어떤분포의 순서통계량: $X_1,\ldots,X_n\stackrel{iid}{\sim}F\implies X_{(1)},\ldots,X_{(r)}.$

정리 4.4.3.

- 균등분포의 순서통계량: $U_1,\ldots,U_n\stackrel{iid}{\sim}U(0,1)\implies U_{(1)},\ldots,U_{(n)}.$
- 지수분포의 순서통계량: $Y_1,\ldots,Y_n \stackrel{iid}{\sim} Exp(1) \implies Y_{(1)},\ldots,Y_{(n)}$.
- 계가 있다. (예제 4.3.4.)

균등분포의 순서통계량과 지수분포의 순서통계량에는 아래와 같은 관

• 그런데 임의의
$$r=1,2,\ldots,n$$
에 대하여 지수분포의 순서통계량 $Y_{(r)}$ 은 아래를 만족한다. (예제 4·3·3·)

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

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

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

• 순서통계량 $X_{(r)}$ 은 아래와 같이 얻을 수 있다.

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

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

따라서 $h(\bigstar)=F^{-1}\big(1-e^{-\bigstar}\big)$ 라고 정의하면 $X_{(r)}\stackrel{d}{=}h\left(\frac{1}{n}Z_1+\cdots+\frac{1}{n-r+1}Z_r\right)$

• $s_n \sim \beta n \iff r_n/n \to \beta$.

가 성립한다.

<u>note:</u> 위의정의는 예제 5.2.7 에 나와있다.

• $r_n \sim \alpha n \iff r_n/n \to \alpha$.

ユ리고 0 < α < β < 1.

• 편의상 아래를 가정하자.

- (1) r_n 을 그냥 r로 쓰자. 여기에서 r은 n개의 sample 중 하위25%에 번
- 째에 해당하는 수 이다. 즉 $r/n \approx 1/4$. 동일한 논리로 s_n 도 그냥 s라고 쓰자.
- (2) $r/n \to \frac{1}{4} \text{ and } s/n \to \frac{3}{4}.$
- (3) α, β 를 그냥 α_r, α_s 로 정의하자. 이는 분위수의 느낌을 좀더 강조하기 위해서이다. α_r 은 r에 해당하는 분위수라는 뜻임.
- $X_{(r)}$ 과 $X_{(s)}$ 의 join pdf를 구하라.

(sol)

• 우선 순서통계량 문제이므로 아래와 같은 확률변수를 가정하자.

$$Y_1,\ldots,Y_n \stackrel{iid}{\sim} Exp(1).$$

• 아래와 같은 벡터를 가정하자.

$$\begin{bmatrix} Y_{(r)} \\ Y_{(s)} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} Z_1 + \dots + \frac{1}{n-r+1} Z_r \\ \frac{1}{n} Z_1 + \dots + \frac{1}{n-s+1} Z_s \end{bmatrix}$$

• 전체적인 그림.

 $Y_{(r)}$ 과 $Y_{(s)}$ 번째 순서통계량의 분포 $\Longrightarrow X_1,\ldots,X_n \overset{iid}{\sim} F$ 인 임의의 분포에서 $X_{(r)}$ 과 $X_{(s)}$ 의 분포

• 아래가 성립한다. (예제 5.2.7)

$$E(Y_{(r)}) = E\left(\frac{1}{n}Z_1 + \dots + \frac{1}{n-r+1}Z_r\right) = \frac{1}{n} + \dots + \frac{1}{n-r+1}$$
$$= \frac{1}{n}\sum_{k=1}^{r-1} \frac{1}{1-k/n} \approx \int_0^{\alpha_r} \frac{1}{1-x} dx = -\log(1-\alpha_r)$$

• 아래도 성립한다. (예제 5.2.7)

$$V(Y_{(r)}) = V\left(\frac{1}{n}Z_1 + \dots + \frac{1}{n-r+1}Z_r\right) = \frac{1}{n^2} + \dots + \frac{1}{(n-r+1)^2}$$
$$= \frac{1}{n^2} \sum_{k=1}^{r-1} \frac{1}{(1-k/n)^2} \approx \int_0^{\alpha_r} \frac{1}{(1-x)^2} dx = \frac{1}{n} \frac{\alpha_r}{1-\alpha_r}$$

• 따라서

$$\sqrt{n} \left(\begin{bmatrix} Y_{(r)} \\ Y_{(s)} \end{bmatrix} - \begin{bmatrix} -\log(1-\alpha_r) \\ -\log(1-\alpha_s) \end{bmatrix} \right) \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\alpha_r}{1-\alpha_r} & \frac{\alpha_r}{1-\alpha_r} \\ \frac{\alpha_r}{1-\alpha_r} & \frac{\alpha_s}{1-\alpha_s} \end{bmatrix} \right).$$

그런데

$$\begin{bmatrix} X_{(r)} \\ X_{(s)} \end{bmatrix} = \begin{bmatrix} h(Y_{(r)}) \\ h(Y_{(s)}) \end{bmatrix} = \boldsymbol{h} \left(\begin{bmatrix} Y_{(r)} \\ Y_{(s)} \end{bmatrix} \right)$$

• 따라서

$$\sqrt{n} \times \left(\begin{bmatrix} X_{(r)} \\ X_{(s)} \end{bmatrix} - \begin{bmatrix} ? \\ ? \end{bmatrix} \right) \\
= \sqrt{n} \times \left(\boldsymbol{h} \left(\begin{bmatrix} Y_{(r)} \\ Y_{(s)} \end{bmatrix} \right) - \boldsymbol{h} \left(\begin{bmatrix} -\log(1 - \alpha_r) \\ -\log(1 - \alpha_s) \end{bmatrix} \right) \right) \\
= \sqrt{n} \times \left(\begin{bmatrix} h(Y_{(r)}) \\ h(Y_{(s)}) \end{bmatrix} - \begin{bmatrix} h(-\log(1 - \alpha_r)) \\ h(-\log(1 - \alpha_s)) \end{bmatrix} \right) \\
\stackrel{d}{\longrightarrow} \begin{bmatrix} ?? & ?? \\ ?? & ?? \end{bmatrix}^T N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\alpha_r}{1 - \alpha_r} & \frac{\alpha_r}{1 - \alpha_r} \\ \frac{\alpha_r}{1 - \alpha_r} & \frac{\alpha_s}{1 - \alpha_s} \end{bmatrix} \right)$$

 $\underline{note:}$ 여기에서 $h: \mathbb{R}^2 \to \mathbb{R}^2$ 이므로 $\begin{bmatrix} ?? & ?? \\ ?? & ?? \end{bmatrix}$ 와 같이 2×2 매트릭스가 나왔다.

면 $\begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial y_1} & \frac{\partial y_2}{\partial x_1} & \frac{\partial y_3}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_3}{\partial x_2} \end{bmatrix}$

note: 만약에 $h: \mathbb{R}^2 \to \mathbb{R}^3$ 이었다면 즉 $h(x_1, x_2) = (y_1, y_2, y_3)$ 꼴이었다

일것이다.

• 잠시 시간을 투자하여 $\begin{bmatrix} ?? & ?? \\ ?? & ?? \end{bmatrix}$ 를 계산하자. $\begin{bmatrix} ?? & ?? \\ ?? & ?? \end{bmatrix} = \begin{bmatrix} \frac{\partial h(x_1)}{\partial x_1} & \frac{\partial h(x_2)}{\partial x_1} \\ \frac{\partial h(x_1)}{\partial x_1} & \frac{\partial h(x_2)}{\partial x_2} \end{bmatrix}$

$$\left[\frac{\partial h(x_1)}{\partial x_2} \frac{\partial h(x_2)}{\partial x_2} \right]$$

$$h(x_1) = F^{-1}(1 - e^{-x_1})$$

이므로

이다. 그런데

$$F(h(x_1)) = 1 - e^{-x_1}$$

그러므로

$$f(h(x_1))\frac{\partial h(x_1)}{\partial x_1} = e^{-x_1} \Longleftrightarrow \frac{\partial h(x_1)}{\partial x_1} = \frac{e^{-x_1}}{f(h(x_1))}$$

 $\frac{\partial}{\partial x_1} F(h(x_1)) = e^{-x_1}$

.. -1] /]/

따라서

$$x_1$$
대신에 $-\log(1-\alpha_r)$ 를 대입하자.
$$e^{-x_1}=e^{\log(1-\alpha_r)}=1-\alpha_r$$

 $f(h(x_1)) = f(h(-\log(1-\alpha_r))) = f(F^{-1}(1-e^{\log(1-\alpha_r)})) = f \circ F^{-1}(\alpha_r)$ 따라서

$$\begin{bmatrix} ?? & ?? \\ ?? & ?? \end{bmatrix} = \begin{bmatrix} \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & 0 \\ 0 & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix}$$

 $\underline{note:}$ X가 연속확률변수이므로 F^{-1} 도 연속함수이다. 따라서 $\left[\frac{1-\alpha_r}{f\circ F^{-1}(\alpha_r)} \quad 0\right]$

$$\begin{bmatrix} \frac{1-\alpha_r}{f\circ F^{-1}(\alpha_r)} & 0 \\ 0 & \frac{1-\alpha_s}{f\circ F^{-1}(\alpha_s)} \end{bmatrix}$$
의 각 원소가 모두 연속이다. 따라서 모든 편미분이 존재하고 그것이 연

• 따라서 수렴하는 분포는

$$\begin{bmatrix} \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & 0\\ 0 & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix}^T N \left(\begin{bmatrix} 0\\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\alpha_r}{1-\alpha_r} & \frac{\alpha_r}{1-\alpha_r}\\ \frac{\alpha_r}{1-\alpha_r} & \frac{\alpha_s}{1-\alpha_s} \end{bmatrix} \right)$$

와 같다. 따라서 수렴하는 분포의 분산은

$$\begin{bmatrix} \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & 0 \\ 0 & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix}^T \begin{bmatrix} \frac{\alpha_r}{1-\alpha_r} & \frac{\alpha_r}{1-\alpha_r} \\ \frac{\alpha_r}{1-\alpha_r} & \frac{\alpha_s}{1-\alpha_s} \end{bmatrix} \begin{bmatrix} \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & 0 \\ 0 & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & 0 \\ 0 & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix} \begin{bmatrix} \frac{\alpha_r}{1-\alpha_r} & \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & \frac{\alpha_r}{1-\alpha_r} & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \\ \frac{\alpha_r}{1-\alpha_r} & \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & \frac{\alpha_s}{1-\alpha_s} & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \\ \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} & \frac{1-\alpha_r}{f \circ F^{-1}(\alpha_r)} & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} & \frac{1-\alpha_s}{f \circ F^{-1}(\alpha_s)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{(1-\alpha_r)\alpha_r}{f \circ F^{-1}(\alpha_r) \vee f \circ F^{-1}(\alpha_s)} & \frac{\alpha_r(1-\alpha_s)}{f \circ F^{-1}(\alpha_r) \vee f \circ F^{-1}(\alpha_s)} \\ \frac{\alpha_r(1-\alpha_s)}{f \circ F^{-1}(\alpha_r) \vee f \circ F^{-1}(\alpha_s)} & \frac{(1-\alpha_s)\alpha_s}{(f \circ F^{-1}(\alpha_s))^2} \end{bmatrix}$$

• 결론적으로 아래와 같이 주장할 수 있다.

$$\sqrt{n} \times \left(\begin{bmatrix} X_{(r)} \\ X_{(s)} \end{bmatrix} - \begin{bmatrix} ? \\ ? \end{bmatrix} \right)
\xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{(1-\alpha_r)\alpha_r}{(f \circ F^{-1}(\alpha_r))^2} & \frac{\alpha_r(1-\alpha_s)}{f \circ F^{-1}(\alpha_r) \times f \circ F^{-1}(\alpha_s)} \\ \frac{\alpha_r(1-\alpha_s)}{f \circ F^{-1}(\alpha_r) \times f \circ F^{-1}(\alpha_s)} & \frac{(1-\alpha_s)\alpha_s}{(f \circ F^{-1}(\alpha_s))^2} \end{bmatrix} \right)$$

여기에서

$$\begin{bmatrix}
? \\
?
\end{bmatrix} = \begin{bmatrix}
h(EY_{(r)}) \\
h(EY_{(s)})
\end{bmatrix} = \begin{bmatrix}
h(-\log(1 - \alpha_r)) \\
h(-\log(1 - \alpha_s))
\end{bmatrix} \\
= \begin{bmatrix}
F^{-1} \circ (1 - e^{\log(1 - \alpha_r)}) \\
F^{-1} \circ (1 - e^{\log(1 - \alpha_r)})
\end{bmatrix} = \begin{bmatrix}
F^{-1}(\alpha_r) \\
F^{-1}(\alpha_s)
\end{bmatrix}$$

따라서

$$\sqrt{n} \times \left(\begin{bmatrix} X_{(r)} \\ X_{(s)} \end{bmatrix} - \begin{bmatrix} F^{-1}(\alpha_r) \\ F^{-1}(\alpha_s) \end{bmatrix} \right)
\xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{(1-\alpha_r)\alpha_r}{(f \circ F^{-1}(\alpha_r))^2} & \frac{\alpha_r(1-\alpha_s)}{f \circ F^{-1}(\alpha_r) \times f \circ F^{-1}(\alpha_s)} \\ \frac{\alpha_r(1-\alpha_s)}{f \circ F^{-1}(\alpha_r) \times f \circ F^{-1}(\alpha_s)} & \frac{(1-\alpha_s)\alpha_s}{(f \circ F^{-1}(\alpha_s))^2} \end{bmatrix} \right)$$

• 마지날을 구해보자.

$$\sqrt{n}(X_{(r)} - F^{-1}(\alpha_r)) \stackrel{d}{\to} N\left(0, \frac{(1 - \alpha_r)\alpha_r}{(f \circ F^{-1}(\alpha_r))^2}\right)$$

$$\sqrt{n}(X_{(s)} - F^{-1}(\alpha_s)) \stackrel{d}{\to} N\left(0, \frac{(1 - \alpha_s)\alpha_s}{(f \circ F^{-1}(\alpha_s))^2}\right)$$
note: 참고로 $F^{-1}(\alpha_r)$ 는 α_r -percentile의 정의가 된다. 그리고 $X_{(r)}$ 는

sample α_r -percentile 이라 볼 수 있다. **note:** 따라서 위의 식을 관찰하면 표본백분위수는 백분위수로 확률수렴

함을 알 수 있다.

note: 추가적으로 표본 백분위수의 분포도 알 수 있다.

note: 또한 아래와 같이 범위(백분위수간의 차이)의 분포도 알 수 있다.

ullet $R=X_{(s)}-X_{(r)}$ 의 분포를 구해보자. 정규분포의 차는 다시 정규분포를

따르므로 $\sqrt{n}(R-\mu) \stackrel{d}{\to} N(0,\sigma^2)$

$$\mu = F^{-1}(\alpha_s) - F^{-1}(\alpha_r)$$

.

여기에서

$$\sigma^2 = \frac{(1 - \alpha_s)\alpha_s}{(f \circ F^{-1}(\alpha_s))^2} + \frac{(1 - \alpha_r)\alpha_r}{(f \circ F^{-1}(\alpha_r))^2} + 2\frac{\alpha_r(1 - \alpha_s)}{f \circ F^{-1}(\alpha_r) \times f \circ F^{-1}(\alpha_s)}$$
• 위의식에서 $\alpha_r = \frac{1}{4}$, $\alpha_s = \frac{3}{4}$ 를 대입하면

 $\mu = F^{-1}(3/4) - F^{-1}(1/4)$

$$\sigma^{2} = \frac{(1 - 3/4)3/4}{(f \circ F^{-1}(3/4))^{2}} + \frac{(1 - 1/4)1/4}{(f \circ F^{-1}(1/4))^{2}} + 2\frac{1/4(1 - 3/4)}{f \circ F^{-1}(1/4) \times f \circ F^{-1}(3/4)}$$

$$= \frac{1}{16} \left(\frac{3}{(f \circ F^{-1}(3/4))^{2}} + \frac{3}{(f \circ F^{-1}(1/4))^{2}} + \frac{2}{f \circ F^{-1}(1/4) \times f \circ F^{-1}(3/4)} \right)$$