분위수

정의

- X를 연속형 CDF를 갖는 확률변수라고 할 때 0 < P < 1에 대하여

1. $p_{분위수} = F^{-1}(p)$로 정의하자.

1) 확률표본 $X_1, \ldots, X_n$의 순서통계량 $Y_1, \ldots, Y_k, \ldots, Y_n$중 k가 누적 확률 P에 해당하는 $p_{분위수}$ 인 경우,

(1) 그 기댓값을 구하면 E[F(Y_k)] = $\int_a^b F(Y_k)g(Y_k)\,dy_k$

(2) 한편, $g(Y_k) = \int_{y_k}^{b} \cdots \int_{y_{n-1}}^{b} \int_a^{y_k} \cdots \int_a^{y_2} \mathrm{f}(y_1) \cdots \mathrm{f}(y_n)\,dy_1 \cdots dy_{k-1}\,dy_n \cdots dy_{k+1}$

$= \frac{n!}{(k-1)!(n-k)!}[\mathrm{F}(y_k)]^{k-1}\,[1-\mathrm{F}(y_k)]^{n-k}\mathrm{f}(y_k)$

(3) $F(Y_k)$ = u로 치환하면, $\frac{du}{dY_k}$ = f(y_k) , du = $g(y_k)dY_k$ 이므로

(4) $\int_a^b F(Y_k)g(Y_k)\,dy_k = \int_a^b u\frac{n!}{(k-1)!(n-k)!}u^{k-1}\,[1-u]^{n-k}\,du$

$= \frac{n!}{(k-1)!(n-k)!}\int_a^b u^k\,[1-u]^{n-k}\,du = \frac{n!(n-k)!}{(k-1)!(n-k)!(n+1)!} = \frac{k}{n+1}$

정의

- q-q 플롯

1. X의 CDF F(x)는 알려졌으나 모수 a,b는 모른다고 가정하자.

1) 이 때, z = $\frac{x-a}{b}$ 꼴의 관계를 갖는 확률변수를 정의하면, cdf(z) = f($\frac{x-a}{b}$) 이다.

2) $\varepsilon_{x,P}$가 X의 $p_{분위수}$ 이고, $\varepsilon_{z,P}$가 z의 $p_{분위수}$ 라고 한다면

(1) $p= p[\mathrm{x} \leq \varepsilon_{x,P}] = \mathrm{p}\left[\mathrm{z} \leq \frac{\varepsilon_{x,P}-a}{b}\right]$ 이다.

3) 한편, 순서통계량 $Y_1, \ldots, Y_k, \ldots, Y_n$이 존재할 때, Y_k는 ε_{x,p_k}에 대한 추정량이고, 같은 분위수 k에 대해 ε_{z,p_k} = $F^{-1}(p_k)$로 정의할 때

4) Y_k에 대한 ε_{z,p_k}의 그래프를 q-q 플롯 이라고 한다.

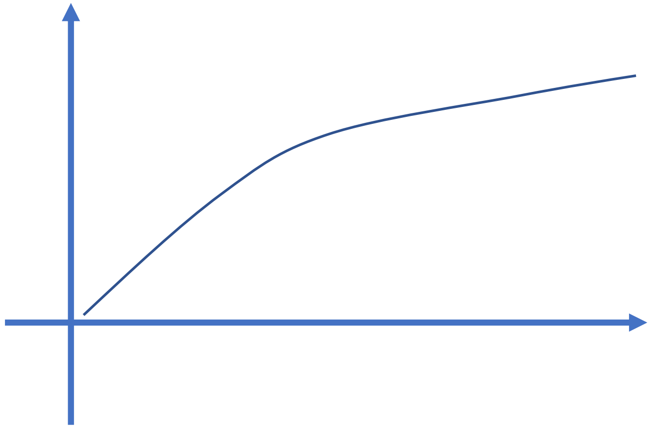
2. 이 때 , $\mathrm{p}\left[\mathrm{z} \leq \frac{\varepsilon_{x,P}-a}{b}\right]$ 에 대해서 ε_{x,p_k} = $\mathrm{b}\varepsilon_{z,p_k} + a$ 를 유추할 수 있다. 즉, Y_k와 ε_{z,p_k}는 보통 선형관계이다.

정의

- 분위수와 신뢰구간

1. 성공 실패를 판별하는 이항분포의 pdf를 활용하여 신뢰구간을 정의한다.
2. X를 f(x)를 갖는 연속확률변수라고 할 때, ε_p 를 $F(\varepsilon_p) = P$를 만족하는 p분위수라 하자.

1) 이 때, $Y_1, \dots, Y_k, \dots, Y_n$ 에서 Y_k를 ε_p 에 대한 점추정량이라고 할 때



2) $Y_i < \varepsilon_p < Y_j$ 일때, 적어도 i개가 ε_p 보다 작아야 하고 ε_p 를 포함한다 하더라도 J개를 넘지 않는다.

3) (1-a) = $p(Y_i < \varepsilon_p < Y_j)$ 일 때,이는 n회 시행에서 i와 j 사이의 성공을 얻는 이항 확률 실험과 같다. 즉

$$a = \sum_{w=i}^{j-1} \frac{n!}{(n-w)!} \cdot p^w (1-p)^{(n-w)}$$