韓國數學敎育學會誌<數學敎育>
1983. 12. Vol. XXII, No. 1.

Journal of the Korea Society of Mathematical Education
Dec. 1983. Vol. XXII, Nc. 1.

順序統計量의 分布에 관하여

梨花女子大學校 宋 順 姬

I. 序 論

統計的 推測은 標本調査를 하여 얻은 知識을 토대로 母集團에 관한 정보(母數)를 귀납적으로 推測해 내는 것이 그 중심 과제이다.

順序統計量은 이 統計的 推測의 한 방법으로 중요한 역할을 하고 있다. 예를들면 母數에 관한 信賴區間, 確率標本의 범위(Range of random sample), 確率標本의 中位數, 確率標本의 中間범위(mid-range)의 分布를 생각할 때 確率標本의 分布에 의하지 않고 順序統計量의 分布에 의하여 유도함으로써 편리한 점이 많다. 특히 順序統計量의 確率密度函數에 관한 유도과정을 硏究함으로써 順序統計量의 理解에 도움을 제시해 보려 한다.

II. 順序統計量의 定義와 結合密度函數

〔定義〕 $X_1, X_2, \cdots, X_n$을 독립이고 恒等的인 確率密度函數 $f(x)$와 累積分布函數 $F(x)$를 갖는 確率標本이라 하자. Y_1을 이들 $X_i\,(i=1, 2, \cdots, n)$ 중에서 제일 작은 것, Y_2를 그 다음 작은것, $\cdots$ Y_n을 제일 큰 것이라고 하자. 그러면 $Y_1<Y_2<\cdots<Y_n$이 된다. 이때 $Y_1, Y_2, \cdots, Y_n$을 確率標本 $X_1, X_2, \cdots, X_n$의 n차 順序統計量(order statistics)이라고 한다.

〔定理〕 $X_1, X_2, \cdots, X_n$을 독립이고 恒等的인 確率密度函數 $f(x)$, $a<x<b$와 累積分布函數 $F(x)$를 갖는 確率標本이라 하자. 順序統計量 $Y_1, Y_2, \cdots, Y_n$의 結合密度函數는 다음과 같다.

$$g(y_1, y_2, \cdots, y_n) = \begin{cases} n!f(y_1)f(y_2)\cdots f(y_n), & a<y_1<y_2<\cdots<y_n<b \\ 0 & \text{, 딴곳에서} \end{cases}$$

〔증명〕 표본의 크기 $n=3$일때를 증명하겠으나 이를 확장하면 일반적인 경우도 마찬가지로 증명된다.

X_1, X_2, X_3의 結合密度函數는 $f(x_1)f(x_2)f(x_3)$이다. 이 3차원 確率空間은 X_1, X_2, X_3가 적어도 2개는 같은 모든점(x_1, x_2, x_3)에서 確率이 零이 되도록 정의할 수 있는 집합들과, $f(x_1)f(x_2)f(x_3)>0$로 되는 다음과 같은 6개의 서로 배반되는 집합들의 合集合으로 된다.

$$A_1=\{(x_1, x_2, x_3):\ a<x_1<x_2<x_3<b\}$$
$$A_2=\{(x_1, x_2, x_3):\ a<x_1<x_3<x_2<b\}$$
$$A_3=\{(x_1, x_2, x_3):\ a<x_2<x_1<x_3<b\}$$
$$A_4=\{(x_1, x_2, x_3):\ a<x_2<x_3<x_1<b\}$$
$$A_5=\{(x_1, x_2, x_3):\ a<x_3<x_1<x_2<b\}$$
$$A_6=\{(x_1, x_2, x_3):\ a<x_3<x_2<x_1<b\}$$

$y_1=(x_1, x_2, x_3$의 최소치), $y_2=(x_1, x_2, x_3$의 크기의 中間치), $y_3=(x_1, x_2, x_3$의 최대치)이다.

$A_1, A_2, \cdots, A_6$의 각각은 順序대로 놓은 같은 집합 $\{(y_1, y_2, y_3):\ a<y_1<y_2<y_3<b\}$ 안에 사상하는 1대 1의 변환을 정의한다. 逆函數는 A_1에 있는 점에 대해서 $x_1=y_1$, $x_2=y_2$, $x_3=y_3$, A_2에 있어서는 $x_1=y_1$, $x_2=y_3$, $x_3=y_2$, A_3에 있어서는 $x_1=y_2$, $x_2=y_1$, $x_3=y_3, \cdots$나머지 3개 A_4, A_5, A_6 등에 있는 점에 대해서도 마찬가지이다. 이들의 Jacobian들은 다음과 같다.

$$J_1=\begin{vmatrix}1&0&0\\0&1&0\\0&0&1\end{vmatrix}=1,\quad J_2=\begin{vmatrix}1&0&0\\0&0&1\\0&1&0\end{vmatrix}=-1, \cdots$$

$$J_6=\begin{vmatrix}0&0&1\\0&1&0\\1&0&0\end{vmatrix}=-1.$$

이 6개의 Jacobian들의 절대값은 1이다. 그러

므로 3개의 順序統計量 Y_1, Y_2, Y_3의 結合密度函數는 다음과 같다.

$$g(y_1, y_2, y_3) = \begin{cases} |J_1|f(y_1)f(y_2)f(y_3)+|J_2|f(y_1)f(y_2)f(y_3)+ \\ \cdots+|J_6|f(y_1)f(y_2)f(y_3)=3!f(y_1)f(y_2)f(y_3) \\ \qquad a<y_1<y_2<y_3<b \\ 0 \qquad ,\text{딴 곳에서} \end{cases}$$

標本의 크기를 n으로 확장하면 標本空間은 적어도 2개는 같은 모든 점 $(x_1, x_2, \cdots, x_n)$에서 확률이 零이 되도록 정의할 수 있는 집합들과 $n!$개의 서로 배반되는 집합들의 合집합으로 된다. $n!$개의 서로 배반되는 집합 $A_1, A_2, \cdots, A_{n!}$의 각각은 順序대로 놓은 같은 집합 $\{(y_1, y_2, \cdots, y_n): a<y_1<y_2<\cdots<y_n<b\}$안에 사상하는 1대 1의 변환을 정의하고, 이들 집합들에서 逆函數의 Jacobian의 절대값들은 1이 되므로 順序統計量 $Y_1, Y_2, \cdots, Y_n$의 結合密度函數는 다음과 같다.

$$g(y_1, y_2, \cdots, y_n) = \begin{cases} n!f(y_1)f(y_2)\cdots f(y_n), & a<y_1<y_2<\cdots<y_n<b \\ 0 & ,\text{딴곳에서} \end{cases}$$

III. 順序統計量의 確率密度函數

1. 確率密度函數와 累積分布函數

本 硏究에서는 連續的 確率變數인 경우만 생각하기로 한다.

k번째 順序統計量의 確率密度函數 $g_k(y_k)$는 다음과 같은 重積分을 계산함으로써 구할 수 있으며 그 결과는 잘 알려진 사실이다.

$$g_k(y_k)=\int_{y_k}^{b}\int_{y_{k+1}}^{b}\cdots\int_{a}^{y_3}\int_{a}^{y_2} n!f(y_1)f(y_2)\cdots f(y_n)\,dy_1dy_2\cdots dy_{k-1}dy_ndy_{n-1}\cdots dy$$
$$=\frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k) \cdots\cdots\cdots\cdots ①$$

本 硏究에서는 $g_k(y_k)$를 수학적 귀납법으로 구할 수 있다는 것과 그 결과가 잘 알려진 重積分에 의하여 구한 결과와 같음을 보일려고 한다.

k번째 順序統計量 Y_k를 x로 표시하면 식 ①은 다음과 같이 된다.

$$g_k(x)=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots ②$$

確率標本 $X_1, X_2, \cdots, X_n$은 독립이고 항등적인 確率密度函數(probability density function) $f(x)$와 累積分布函數(cumulative distribution function) $F(x)$를 갖는 連續的인 確率變數의 分布로부터 論하여졌으므로 p.d.f. $f(x)$와 c.d.f. $F(x)$와의 관계는

$$f(x)=\frac{d}{dx}F(x)$$

이고, $X_1, X_2, \cdots, X_n$을 順序統計量으로 변환한 $Y_1<Y_2<\cdots<Y_n$의 k번째 順序統計量 Y_k의 p.d.f. $g_k(y_k)$와 c.d.f. $G_k(y_k)$와의 관계도 다음과 같다.

$$g_k(y_k)=\frac{d}{dy_k}G_k(y_k)$$

2. 順序統計量의 p.d.f.식 ②의 수학적 귀납법에 의한 증명

$k=n$일때

$$G_n(x)=P[Y_n\leq x]=P[X_1\leq x,\ X_2\leq x, \cdots, X_n\leq x]$$
$$=P[X_1\leq x]\cdot P[X_2\leq x]\cdot\cdots\cdot P[X_n\leq x]$$
$$=[F(x)]^n.$$

$$g_n(x)=\frac{d}{dx}[G_n(x)]=\frac{d}{dx}[F(x)]^n$$
$$=n[F(x)]^{n-1}f(x) \cdots\cdots\cdots\cdots\cdots\cdots ③$$

방정식 ②에 $k=n$을 대입하면

$$g_n(x)=\frac{n!}{(n-1)!(n-n)!}[F(x)]^{n-1}[1-F(x)]^{n-n}f(x)$$
$$=n[F(x)]^{n-1}f(x) \cdots\cdots\cdots\cdots\cdots\cdots ④$$

식 ③은 重積分에 의하여 구한 결과인 식 ④와 같으므로 $k=n$일때 식 ②는 成立된다.

다음 $k=m+1(m\leq n-1)$일때 식 ②가 成立된다고 가정하고 $k=m$에 대하여 成立됨을 밝히려고 한다.

$$G_m(x)=P[Y_m\leq x]$$
$$=P[Y_{m+1}\leq x]+P[Y_m\leq x<Y_{m+1}]$$

위 식의 2번째항 $P[Y_m\leq x<Y_{m+1}]$은 $X_1, X_2, \cdots, X_n$ 중 x보다 작든지 같은 것이 m개 존재하는 확률임을 의미한다. 만일 $X_k\leq x$이면 n회 독립시행에서 k번째 시행이 성공한 二元確率(binomial probability)이라고 생각할 수 있다.

k번째 시행에서 성공의 확률은 $F(x)$이다. $P[Y_m \leqq x < Y_{m-1}]$은 n회 시행에서 m회 성공한 확률이 되므로

$$P[Y_m \leqq x < Y_{m+1}] = {}_nC_m[F(x)]^m[1-F(x)]^{n-m}$$
$$= \frac{n!}{m!(n-m)!}[F(x)]^m[1-F(x)]^{n-m}$$

그리하여 $G_m(x)$는 다음과 같이 된다.

$$G_m(x) = G_{m+1}(x) + \frac{n!}{m!(n-m)!}[F(x)]^m[1-F(x)]^{n-m}$$

양변을 미분하면

$$g_m(x) = g_{m+1}(x) + \frac{n!}{m!(n-m)!}m[F(x)]^{m-1}f(x)[1-F(x)]^{n-m} - \frac{n!}{m!(n-m)!}[F(x)]^m(n-m)[1-F(x)]^{n-m-1}f(x)$$
$$= \frac{n!}{m!(n-m-1)!}[F(x)]^m[1-F(x)]^{n-m-1}f(x) + \frac{n!}{(m-1)!(n-m)!}[F(x)]^{m-1}[1-F(x)]^{n-m}f(x) - \frac{n!}{m!(n-m-1)!}[F(x)]^m[1-F(x)]^{n-m-1}f(x)$$
$$= \frac{n!}{(m-1)!(n-m)!}[F(x)]^{m-1}[1-F(x)]^{n-m}f(x) \quad \cdots\cdots\cdots\cdots ⑤$$

이제 식 ②에 $k=m$을 대입하면 다음과 같이 된다.

$$g_m(x) = \frac{n!}{(m-1)!(n-m)!}[F(x)]^{m-1}[1-F(x)]^{n-m}f(x) \quad \cdots\cdots\cdots\cdots ⑥$$

식 ⑤는 重積分에 의하여 구한식 ⑥과 같으므로 $k=m$일때 成立한다. 그러므로 $k=1, 2, \cdots, n$ 모두에 대하여 成立한다.

IV. 順序統計量의 分布에 관한 例

〔보기 1〕 Y_1, Y_2, Y_3, Y_4, Y_5를 p.d.f.

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{딴곳에서} \end{cases}$$

를 갖는 分布로부터의 크기 5인 確率標本의 順序統計量이라하자 이 확률표본의 中位數 즉 Y_3의 p.d.f.를 구하여 보자.

3번째 順序統計量의 p.d.f. $g_3(y_3)$ 즉 $g_3(x)$를 구하는 것과 같다.

順序統計量의 結合密度函數의 定理에 의하여 結合密度函數는 다음과 같다.

$$g(y_1, y_2, y_3, y_4, y_5) = \begin{cases} \frac{5!}{\theta^5}, & 0 < y_1 < y_2 < y_3 < y_4 < y_5 < \theta \\ 0, & \text{딴곳에서} \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{딴곳에서} \end{cases}$$

이므로

$$F(x) = \frac{x}{\theta}$$

이다. 식 ②에서 $k=3$인 경우이므로

$$g_3(x) = \frac{5!}{(3-1)!(5-3)!}\left(\frac{x}{\theta}\right)^{3-1}\left(1-\frac{x}{\theta}\right)^{5-3}\frac{1}{\theta}$$
$$= \frac{5!}{2!\,2!}\,\frac{x^2}{\theta^3}(1-\frac{x}{\theta})^2 \quad \cdots\cdots\cdots\cdots ⑦$$

한편 重積分으로 계산하면 다음과 같다.

$$g_3(y_3) = \int_{y_3}^{\theta}\int_{y_4}^{\theta}\int_0^{y_3}\int_0^{y_2}\frac{5!}{\theta^5}dy_1dy_2dy_5dy_4$$
$$= \int_{y_3}^{\theta}\int_{y_4}^{\theta}\int_0^{y_3}\frac{5!}{\theta^5}y_2dy_2dy_5dy_4$$
$$= \int_{y_3}^{\theta}\int_{y_4}^{\theta}\frac{5!}{\theta^5}\,\frac{y_3^2}{2}\,dy_5dy_4$$
$$= \int_{y_3}^{\theta}[\frac{5!}{2}\,\frac{y_3^2}{\theta^5}(\theta-y_4)]dy_4$$
$$= \left[\frac{5!}{2}\,\frac{y_3^2}{\theta^5}(\theta y_4-\frac{y_4^2}{2})\right]_{y_3}^{\theta}$$
$$= \frac{5!}{2!\,2!}\,\frac{y_3^2}{\theta^3}(1-\frac{y_3}{\theta})^2.$$

따라서 $g_3(x) = \frac{5!}{2!\,2!}\,\frac{x^2}{\theta^3}\left(1-\frac{x}{\theta}\right)^2 \cdots\cdots ⑧$

식 ②에 $k=3$을 대입하여 얻은식 ⑦과 重積分 계산을 하여 얻은식 ⑧은 같다.

〔보기 2〕 보기 1에서 미지의 모수 θ에 대한 95%의 신뢰구간을 順序統計量 Y_5를 사용하여 구하여 보자.

5개의 順序統計量 Y_1, Y_2, Y_3, Y_4, Y_5의 結合密度函數는 다음과 같다.

$$g(y_1, y_2, y_3, y_4, y_5) = \begin{cases} \frac{5!}{\theta}, & 0<y_1<y_2<y_3<y_4<y_5<\theta \\ 0, & \text{딴곳에서} \end{cases}$$

따라서 Y_5의 주변밀도는 다음과 같다.

$$g_5(y_5) = \begin{cases} \int_0^{y_5}\int_0^{y_4}\int_0^{y_3}\int_0^{y_2} \frac{5!}{\theta^5} dy_1 dy_2 dy_3 dy_4, & 0<y_5<\theta \\ 0, & \text{딴곳에서} \end{cases}$$

$$= \begin{cases} \frac{5y_5^4}{\theta^5}, & 0<y_5<\theta \\ 0, & \text{딴곳에서} \end{cases}$$

$P(c_1\theta<Y_5<c_2\theta)=0.95$가 되도록 $0<c_1<c_2\leqq 1$를 택하는 방법은 얼마든지 많이 있다. 이 많은 방법들은 θ에 대한 95% 신뢰구간 $(\frac{y_5}{c_2}, \frac{y_5}{c_1})$가 되게 할 것이다.

$$P(c_1\theta<Y_5<c_2\theta)=\int_{c_1\theta}^{c_2\theta} \frac{5y_5^4}{\theta^5} dy_5$$

$$= \left[\frac{y_5^5}{\theta^5} \right]_{c_1\theta}^{c_2\theta} = c_2^5 - c_1^5$$

$$=0.95$$

만일 $c_2=1$라면 $c_1=\sqrt[5]{0.05}$가 되어 하나의 Y_5에 의한 母數의 95% 신뢰구간 $(Y_5, \frac{Y_5}{\sqrt[5]{0.05}})$을 얻는다.

〔보기 3〕 $X_1, X_2, \cdots, X_n$을 독립이고 恒等的인 p.d.f. $f(x)$, $-\infty<x<\infty$를 갖는 확률표본이라고 하자. 順序統計量 $Y_1, Y_2, \cdots, Y_n$의 범위(Range) $Y_n-Y_1=R$의 確率密度函數를 구하여 보자.

順序統計量 $Y_1, Y_2, \cdots, Y_n$의 結合密度函數는

$$g(y_1, y_2, \cdots, y_n)=n!f(y_1)f(y_2)\cdots y(y_n),$$
$$-\infty<y_1<y_2<\cdots y_n<\infty$$

이므로 y_1과 y_n의 結合密度函數는 다음과 같다.

$$h(y_1, y_n)=\int_{y_1}^{y_n}\cdots\int_{y_1}^{y_4}\int_{y_1}^{y_3} g(y_1, y_2, \cdots, y_n)$$
$$dy_2 dy_3\cdots dy_{n-1}, \quad -\infty<y_1<y_n<\infty$$

또는

$$h(y_1, y_n)=n!f(y_1)f(y_n)\int_{y_1}^{y_n}\cdots\int_{y_1}^{y_4}\int_{y_1}^{y_3}$$
$$f(y_2)f(y_3)\cdots f(y_{n-1})dy_2 dy_3\cdots dy_{n-1} \cdots\cdots ⑨$$

$$\int_{y_1}^{y_3} f(y_2)dy_2 = F(y_3)-F(y_1)$$

$$\int_{y_1}^{y_4} [F(y_3)-F(y_1)]f(y_3)dy_3$$

$$=\frac{[F(y_4)-F(y_1)]^2}{2}$$

$$\int_{y_1}^{y_5} \frac{1}{2}[F(y_4)-F(y_1)]^2 f(y_4)dy_4$$

$$=\frac{1}{3!}[F(y_5)-F(y_1)]^3$$

식 ⑨를 계속하여 계산하면 다음과 같다.

$$h(y_1, y_n)=n(n-1)[F(y_n)-F(y_1)]^{n-2}$$
$$f(y_1)f(y_n),$$
$$-\infty<y_1<y_n<\infty$$

$R=Y_n-Y_1$, $S=Y_1$으로 치환하면

$$q(R, S)=n(n-1)[F(R+S)-F(S)]^{n-2}$$
$$f(S)f(R+S) \cdots\cdots\cdots\cdots\cdots\cdots ⑩$$

로 된다. 따라서 Range $R=Y_n-Y_1$의 密度函數는 $q(R, S)$를 S에 관하여 적분하므로써 얻을 수 있다. 실예를 들면 $X_1, X_2, \cdots, X_6$을 독립이고 항등적인 p.d.f. $f(x)=1$, $0<x<1$을 갖는 확률분포라 하자. Range $R=Y_6-Y_1$의 p.d.f.를 구하여 보자. 累積分布函數 $F(x)$는

$$F(x)=\int_0^x dt = x$$

이므로

$$F(R+S)=R+S, \quad F(S)=S$$

이다. $R=Y_6-Y_1$, $S=Y_1$으로 치환하면 R과 S의 結合密度函數는 식 ⑩에 의하여 다음과 같다.

$$q(R, S)=6\cdot 5(R+S-S)^4=30R^4.,$$
$$0<S<1-R, \quad 0<R<1$$

따라서 범위(Range) R의 密度函數는 다음과 같다.

$$M(R)=30\int_0^{1-R} R^4 dS = 30R^4(1-R),$$
$$0<R<1.$$

V. 結 論

$X_1, X_2, \cdots, X_n$은 독립이고 항등적인 確率密度函數 $f(x)$, $a<x<b$와 累積分布函數 $F(x)$를 갖는 確率標本일때 k번째 順序統計量의 確率密度函數 $g_k(x)$는 다음과 같다.

$$g_k(x)=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}$$
$$[1-F(x)]^{n-k} f(x)$$

本 硏究에서는 順序統計量의 k번째 確率密度函數를 수학적 귀납법에 의하여 유도하였고 그 결과가 잘 알려진 重積分에 의하여 구한 것과 같음을 보였다. 그리고 順序統計量의 分布에 의하여 確率標本의 中位數, 母數에 관한 信賴區間 確率標本의 범위등을 구할 수 있는 예를 보임으로써 順序統計量이 統計推測의 한 方法으로 이용될 수 있음을 보였다.

참 고 문 헌

1. A. M. Mood, *Introduction to the theory of statistics*, McGraw-Hill Book Co. Inc., New York, pp. 240-242, (1963).
2. H.A. David, *Order Statistics*, Wiley, New York, p. 140, (1970).
3. R.V. Hogg and A.T. Craig, *Introduction to Mathematical Statistics*, Macmillan, New York, p. 186 and p. 148, (1971).
4. 崔至薰, 金榮敎, 數理統計, 東明社, 서울, pp. 101-107, (1963).
5. 鄭英鎭, 數理統計學, 塔出版社, 서울, pp. 31-45, (1982).