5. 순서통계량

SOOKHEE KWON

[정의 5.5-1]

순서 통계량(Order statistics)은 가장 작은 값부터 가장 큰 값까지 크기 순으로 배열한 확률표본이다.

순서 통계량은 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$의 형태로 표시하며 $X_{(1)}$은 $X_1, X_2, \cdots, X_n$ 중에서 가장 작은 값을 나타내며, $X_{(n)}$은 가장 큰 값을 나타낸다. 앞에서 언급했던 표본의 범위를 $R = X_{(n)} - X_{(1)}$로 간단히 나타낼 수 있다.

예 5.5-1 $X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)} < X_{(5)}$를 밀도함수가 $f(x)=2x,\ 0<x<1$ 인 분포에서 표본의 크기가 $n=5$인 확률표본 X_1, X_2, X_3, X_4, X_5의 순서 통계량이라고 할 때, $P\left(X_{(4)} \le \frac{1}{2}\right)$를 구하라.

풀이 $X_{(4)} \le 1/2$의 뜻을 정확하게 알아둘 필요가 있다. 이는 5개의 확률표본 X_1, X_2, X_3, X_4, X_5 가운데 적어도 4개는 1/2보다 작거나 같은 값을 갖는다는 뜻이다. 사건 $X_i \le 1/2,\ i=1, 2, 3, 4, 5$를 성공이라고 하면 성공할 확률은 $P\left(X_i \le \frac{1}{2}\right) = \int_0^{\frac{1}{2}} 2x\,dx = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$이 된다. 따라서, 5번의 독립시행에서 4번 이상의 성공을 하게 되는 확률을 구하면 된다.

$$P\left(X_{(4)} \le \frac{1}{2}\right) = \binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)^5 = 0.0156.$$

5. 순서통계량

이제 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$을 누적분포함수가 $F(x)$, 확률밀도함수가 $F'(x) = f(x)$인 분포로부터 얻은 순서 통계량이라고 하자. r번째 순서 통계량 $X_{(r)}$, $r = 1, 2, \cdots, n$의 분포함수와 밀도함수를 구해보자. $X_{(r)} \le x$는 x보다 작거나 같은 표본이 적어도 r개 이상임을 의미한다. 따라서, 분포함수는

$$F_{(r)}(x) = P[X_{(r)} \le x] = \sum_{k=r}^{n} \binom{n}{k} [F(x)]^k [1-F(x)]^{n-k}$$
$$= \sum_{k=r}^{n-1} \binom{n}{k} [F(x)]^k [1-F(x)]^{n-k} + [F(x)]^n$$

이며 밀도함수는 위의 분포함수의 도함수를 구하면 된다.

$$f_{(r)}(x) = F_{(r)}{}'(x) = \sum_{k=r}^{n-1} \binom{n}{k} k[F(x)]^{k-1} f(x) [1-F(x)]^{n-k}$$
$$+ \sum_{k=r}^{n-1} \binom{n}{k} [F(x)]^k (n-k) [1-F(x)]^{n-k-1} (-f(x)) + n[F(x)]^{n-1} f(x) .$$

위 식의 우변에서 $\binom{n}{k} k = \frac{n!}{(k-1)!(n-k)!}$, $\binom{n}{k}(n-k) = \frac{n!}{k!(n-k-1)!}$ 임을 이용하면 서로 상쇄되는 항들이 나오므로 밀도함수는 다음과 같이 간단하게 표현된다.

$$f_{(r)}(x) = \frac{n!}{(r-1)!(n-r)!} [F(x)]^{r-1} [1-F(x)]^{n-r} f(x) .$$

[순서 통계량의 pdf에 관한 성질]

(1) $X_{(1)}$의 확률밀도함수 :

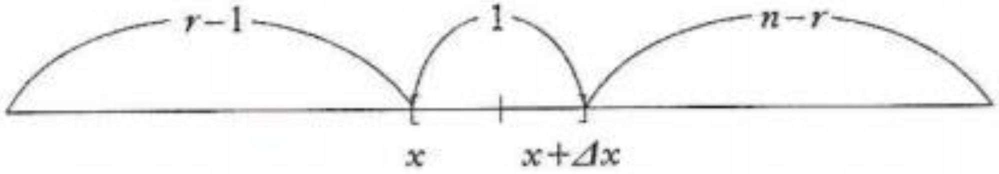
$$f_{(1)}(x) = \frac{n!}{(n-1)!}[1-F(x)]^{n-1}f(x)$$
$$= n[1-F(x)]^{n-1}f(x).$$

(2) $X_{(n)}$의 확률밀도함수 :

$$f_{(n)}(x) = n[F(x)]^{n-1}f(x).$$

5. 순서통계량

(3) 기하학적 해석



위의 그림은 r번째의 순서 통계량이 x와 $x+\Delta x$ 사이에 있고 x보다 작은 확률표본이 $r-1$개 있고 $x+\Delta x$보다 큰 확률표본이 $n-r$개 위치하고 있음을 보여준다. $P[X > x+\Delta x] = 1-F(x+\Delta x) \approx 1-F(x)$, $P[x < X \le x+\Delta x] \approx f(x)\Delta x$, $P(X \le x) = F(x)$로 표현할 수 있으므로 r번째 순서 통계량의 밀도함수 $f_{(r)}(x)$는 다항분포의 형태로 얻어진다고 볼 수 있다. 따라서,

$$
\begin{aligned}
f_{(r)}(x)\Delta x &= P[x < X_{(r)} \le x+\Delta x] \\
&= \frac{n!}{(r-1)!(n-r)!}[F(x)]^{r-1}[1-F(x)]^{n-r} f(x)\Delta x
\end{aligned}
$$

를 얻는다.

5. 순서통계량

예 5.5-2 ([예 5.5-1]의 계속) $X_{(4)}$의 분포함수는 각 확률표본이 성공할 확률이 $P(X_i \le x) = \int_0^x 2t\,dt = x^2$인 5번의 독립적인 시행에서 적어도 4번 이상 성공할 확률을 나타내므로 다음과 같다.

$$F_{(4)}(x) = P(X_{(4)} \le x) = \binom{5}{4}(x^2)^4 \cdot (1-x^2) + (x^2)^5 .$$

$X_{(4)}$의 pdf는 분포함수 $F_{(4)}(x)$를 미분함으로써 쉽게 얻어진다.

$$\begin{aligned} f_{(4)}(x) &= \binom{5}{4} 4 \cdot (x^2)^3 \cdot (2x) \cdot (1-x^2) \\ &\quad + \binom{5}{4}(x^2)^4 \cdot (-2x) + 5 \cdot (x^2)^4 \cdot (2x) \\ &= \frac{5!}{3!1!}(x^2)^3(1-x^2)(2x) \\ &= \frac{5!}{3!1!}[F(x)]^3[1-F(x)]f(x) . \end{aligned}$$

5. 순서통계량

예 5.5-3 $X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)}$를 표준균일분포 (pdf가 $f(x)=1$, $0<x<1$임) 에서 표본의 크기가 $n=4$인 확률표본의 순서 통계량이라고 하자. $F(x)=x$이므로 $X_{(3)}$의 pdf는 다음과 같다.

$$f_{(3)}(x) = \frac{4!}{2!1!}x^2(1-x), \quad 0<x<1.$$

따라서 $P\left(\frac{1}{3} < X_{(3)} < \frac{2}{3}\right) = \int_{1/3}^{2/3} 12x^2(1-x)dx = \frac{13}{27}$임을 알 수 있다.

5. 순서통계량

[참고] 확률변수 X의 분포함수 $F(x)$가 연속이면 확률변수 $F(x)$의 분포는 표준균일분포임을 앞에서 이미 설명하였다. 이제 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$을 확률표본 X_1, $X_2, \cdots, X_n$의 순서 통계량이라고 하면 분포함수 F는 증가함수이므로 $F(X_{(1)}) < F(X_{(2)}) < \cdots < F(X_{(n)})$이 성립한다. 간단히 표현하기 위하여 $F(X_{(r)}) = Z_{(r)}$이라 두면 $Z_{(1)} < Z_{(2)} < \cdots < Z_{(n)}$가 되며 $Z_{(r)}$은 표준균일분포 $U(0, 1)$을 따르는 확률표본의 r번째 순서 통계량이 된다. 이제 활용도가 높은 순서 통계량의 통계적 성질을 정리해 본다.

① $Z_{(r)}$의 확률밀도함수 : $f_{(r)}(z) = \dfrac{n!}{(r-1)!(n-r)!} z^{r-1}(1-z)^{n-r}, \quad 0 < z < 1.$

② $Z_{(r)}$의 기대값 :

$$\begin{aligned} E(Z_{(r)}) &= \int_0^1 z \cdot \frac{n!}{(r-1)!(n-r)!} z^{r-1}(1-z)^{n-r} dz \\ &= \frac{n!}{(r-1)!(n-r)!} \int_0^1 z^r(1-z)^{n-r} dz \\ &= \frac{n!}{(r-1)!(n-r)!} \frac{r!(n-r)!}{(n+1)!} \\ &= \frac{r}{n+1}. \end{aligned}$$

위의 세 번째 등호에서는 앞에서 이미 배운 베타함수(Beta function)의 성질 $Be(a, b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 z^{a-1}(1-z)^{b-1} dz$를 이용하였다.

③ $E(Z_{(r)}) = \dfrac{r}{n+1}$ 로부터 다음의 값은 쉽게 얻어진다.

$$E(Z_{(1)}) = \frac{1}{n+1}, \qquad E(Z_{(n)}) = \frac{n}{n+1},$$

$$E(Z_{(r)} - Z_{(r-1)}) = E(Z_{(r)}) - E(Z_{(r-1)}) = \frac{r}{n+1} - \frac{r-1}{n+1} = \frac{1}{n+1},$$

$$E(1 - Z_{(n)}) = 1 - E(Z_{(n)}) = 1 - \frac{n}{n+1} = \frac{1}{n+1}.$$

④ $Z_{(r)}$의 이차 적률은

$$\begin{aligned} E(Z_{(r)}^2) &= \int_0^1 z^2 \cdot \frac{n!}{(r-1)!\,(n-r)!} z^{r-1}(1-z)^{n-r} dz \\ &= \frac{n!}{(r-1)!\,(n-r)!} \int_0^1 z^{r+1}(1-z)^{n-r} dz \\ &= \frac{n!}{(r-1)!(n-r)!} \frac{\Gamma(r+2)\Gamma(n-r+1)}{\Gamma(n+3)} \\ &= \frac{r(r+1)}{(n+1)(n+2)} \end{aligned}$$

이므로 $Z_{(r)}$의 분산은 아래와 같다.

$$Var(Z_{(r)}) = E(Z_{(r)}^2) - [E(Z_{(r)})]^2 = \frac{r(r+1)}{(n+1)(n+2)} - \left(\frac{r}{n+1}\right)^2 = \frac{r(n-r+1)}{(n+1)^2(n+2)}.$$

5. 순서통계량

예 5.5-4 모수(λ)가 1인 지수분포에서 확률표본 $X_1, X_2, \cdots, X_n$을 얻었다. 다음을 계산해 본다. 모수가 1인 표준지수분포의 밀도함수는 $f(x)=e^{-x}$, $x>0$이고 분포함수는 $F(x)=1-e^{-x}$, $x>0$ 임은 이미 앞에서 다루었다. 최대값과 최소값의 분포함수는

$$F_{(n)}(x) = (F(x))^n = (1-e^{-x})^n, \quad x>0$$
$$F_{(1)}(x) = 1-(1-F(x))^n = 1-e^{-nx}, \quad x>0$$

이고, 밀도함수는

$$f_{(n)}(x) = ne^{-x}(1-e^{-x})^{n-1}, \quad x>0$$

$$f_{(1)}(x) = ne^{-nx}, \quad x>0$$

이다. $n=3$인 경우에 최대값과 최소값의 밀도함수는

$$f_{(3)}(x) = 3e^{-x}(1-e^{-x})^2 = 3(e^{-x}-2e^{-2x}+e^{-3x}),$$
$$f_{(1)}(x) = 3e^{-3x}, \quad x>0$$

이며, $\int_0^\infty ax \cdot e^{-ax}dx = \frac{1}{a}$ 임을 이용하면 최대값과 최소값의 기대값을 아래와 같이 쉽게 얻게 된다.

$$E[X_{(1)}] = \int_0^\infty x \cdot f_{(1)}(x)\,dx = \int_0^\infty 3x \cdot e^{-3x}dx = \frac{1}{3},$$

$$E[X_{(3)}] = \int_0^\infty x \cdot f_{(3)}(x)\,dx = \int_0^\infty 3x \cdot (e^{-x} - 2e^{-2x} + e^{-3x})dx$$
$$= 3\left(1 - \frac{1}{2} + \frac{1}{3} \times \frac{1}{3}\right) = \frac{11}{6}.$$

5. 순서통계량

예 5.5-5 $X_1, X_2, \cdots, X_5$이 표준균일분포 $U(0, 1)$에서 추출된 확률표본일 때, $P\left[X_{(5)} > \frac{1}{2}\right]$와 $P\left[Me \geq \frac{1}{2}\right]$을 구하라. ($Me$는 중앙값, 여기서는 $X_{(3)}$를 뜻함).

풀이 확률변수 X의 분포함수는 $F(x) = x,\ 0 < x < 1$이므로 $X_{(5)}$의 분포함수는 $F_{(5)}(x) = x^5,\ 0 < x < 1$이며 $X_{(3)}$의 분포함수는 정의에 의하여

$$F_{(3)}(x) = \sum_{j=3}^{5} \binom{5}{j} x^j (1-x)^{5-j} = 1 - \sum_{j=0}^{2} \binom{5}{j} x^j (1-x)^{5-j}$$

이다. 따라서 구하는 확률은 아래와 같다.

$$P\left[X_{(5)} > \frac{1}{2}\right] = 1 - F_{(5)}\left(\frac{1}{2}\right) = 1 - \left(\frac{1}{2}\right)^5 = 0.96875,$$

$$P\left[X_{(3)} \geq \frac{1}{2}\right] = 1 - F_{(3)}\left(\frac{1}{2}\right) = 1 - \left(1 - \sum_{j=0}^{2} \binom{5}{j} \left(\frac{1}{2}\right)^j \left(1 - \frac{1}{2}\right)^{5-j}\right) = 0.5.$$

5. 순서통계량

정리 5.5-1 밀도함수가 $f(x)$인 분포에서 추출한 확률표본의 최대값을 Y, 최소값을 X라고 하면 X와 Y의 결합확률밀도함수는 아래와 같다.

$$f_{X,Y}(x, y) = n(n-1)(F(y)-F(x))^{n-2}f(x)f(y), \quad x \le y.$$

증명 본 정리의 개략적인 증명을 살펴보도록 하자. 먼저, X와 Y의 결합밀도함수는

$$f_{X,Y}(x, y)\Delta x\Delta y = P[x < X \le x+\Delta x,\ y < Y \le y+\Delta y]$$

으로 표현할 수 있는데 이는 최소값은 구간 $(x, x+\Delta x]$에, 최대값은 구간 $(y, y+\Delta y]$에 있고 나머지 $n-2$개의 표본은 구간 $(x+\Delta x, y]$에 있게 되는 확률이다. 그리고, 하나의 표본이 구간 $(x, x+\Delta x]$에서 나타날 확률은 $f(x)\Delta x$, 구간 $(y, y+\Delta y]$에서 나타날 확률은 $f(y)\Delta y$, 구간 $(x+\Delta x, y]$에서 나타날 확률은 $F(y)-F(x+\Delta x)$로 둘 수 있으므로 다항분포를 이용하면 $P[x < X \le x+\Delta x,\ y < Y \le y+\Delta y]$는 아래와 같다.

$$\begin{aligned} &P[x < X \le x+\Delta x,\ y < Y \le y+\Delta y] \\ &\qquad \approx \frac{n!}{1!1!(n-2)!}(F(y)-F(x+\Delta x))^{n-2}f(x)f(y)\Delta x\Delta y. \end{aligned}$$

정리 5.5-2 밀도함수 $f(x)$에서 추출한 확률표본의 범위 $(=X_{(n)}-X_{(1)})$를 R이라 할 때, R의 밀도함수는 다음과 같다.

$$f_R(a) = \int n(n-1)(F(x+a)-F(x))^{n-2}f(x)f(x+a)dx.$$

증명 [정리 5.5-1]에서 두 확률변수 X와 Y의 결합밀도함수는

$$f_{X,Y}(x, y) = n(n-1)(F(y)-F(x))^{n-2}f(x)f(y), \quad x \le y$$

이다. 지금 a를 $a \ge 0$인 상수라고 하면 범위 R의 분포함수는 다음과 같이 얻어진다.

$$\begin{aligned} P[R \le a] &= P\{X_{(n)} - X_{(1)} \le a\} \\ &= \iint_{x_{(n)}-x_{(1)} \le a} f_{X_{(1)},X_{(n)}}(x_1, x_n)dx_1 dx_n \\ &= \int_{-\infty}^{\infty}\int_{x_1}^{x_1+a} \frac{n!}{(n-2)!}[F(x_n)-F(x_1)]^{n-2}f(x_1)f(x_n)dx_n dx_1 \end{aligned}$$

5. 순서통계량

〈위 식에서 $y = F(x_n) - F(x_1)$로 변수변환하면 $dy = f(x_n)dx_n$이므로

$$\int_{x_1}^{x_1+a} [F(x_n) - F(x_1)]^{n-2} f(x_n)dx_n = \int_0^{F(x_1+a)-F(x_1)} y^{n-2}dy$$
$$= \frac{1}{n-1}[F(x_1+a) - F(x_1)]^{n-1}$$

이 됨을 이용함.〉

따라서, 범위 R의 분포함수는

$$P[R \le a] = n\int_{-\infty}^{\infty}[F(x_1+a) - F(x_1)]^{n-1} f(x_1)dx_1$$

이며 위 식의 도함수를 구하면 밀도함수가 얻어진다.

$$f_R(a) = \int n(n-1)[F(x+a) - F(x)]^{n-2} f(x)f(x+a)dx\ .$$

예 5.5-6 확률표본 $X_1, X_2, \cdots, X_n$이 균일분포 $U(0, 1)$에서 추출되었을 때, 범위 R의 밀도함수를 구해보자.

풀이 a를 $0<a<1$인 상수라고 하자. R의 분포함수와 밀도함수는 앞의 정리에 의하여 다음과 같이 얻어진다.

$$\begin{aligned} P[R \le a] &= n\int_0^1 [F(x_1+a)-F(x_1)]^{n-1} f(x_1)dx_1 \\ &= n\int_0^{1-a} a^{n-1}dx_1 + n\int_{1-a}^1 (1-x_1)^{n-1}dx_1 \\ &= n(1-a)a^{n-1} + a^n, \end{aligned}$$

$$f_R(a) = n(n-1)a^{n-2}(1-a), \quad 0<a<1.$$

5. 순서통계량

예 5.5-7 $X_{(1)} \le X_{(2)} \le X_{(3)}$를 균일분포 $U(0, 1)$에서 추출한 3개의 확률표본의 순서 통계량이라고 하자. $X = X_{(1)}$, $Y = X_{(3)}$라 할 때, 확률 $P\left[X \le \frac{1}{3},\ Y > \frac{2}{3}\right]$을 구하라.

풀이 X와 Y의 결합밀도함수는 $f_{X,Y}(x, y) = 6(y-x)$, $0 < x \le y < 1$이므로 구하는 확률은 아래와 같다.

$$
\begin{aligned}
P\left[X \le \frac{1}{3},\ Y > \frac{2}{3}\right] &= \int_{\frac{2}{3}}^{1}\int_{0}^{\frac{1}{3}} f_{X,Y}(x, y)\,dxdy \\
&= \int_{\frac{2}{3}}^{1}\int_{0}^{\frac{1}{3}} 6(y-x)\,dxdy \\
&= \frac{4}{9}.
\end{aligned}
$$

6. 많이 쓰이는 몇 가지 분포의 결정

(1) 일차원인 경우

1) 직관적인 방법

예 5.6-1 확률변수 X의 분포가 다음과 같을 때,

X	-1	0	1
$f_X(x)$	0.3	0.4	0.3

$Y = X^2 - 1$의 분포는 다음과 같다.

Y	0	-1
$f_Y(y)$	0.6	0.4

예 5.6-2 확률변수 X의 분포가 이항분포 $B(n, p)$를 따를 때, $Y = X^2$의 분포를 구하라.

풀이 $P[Y = y] = P[X^2 = y] = P[X = \sqrt{y}]$

$$= f_X(\sqrt{y}) = \binom{n}{\sqrt{y}} p^{\sqrt{y}}(1-p)^{n-\sqrt{y}}, \quad y = 0, 1, 4, \cdots, n^2.$$

6. 많이 쓰이는 몇 가지 분포의 결정

2) 누적분포함수를 이용하는 방법

예 5.6-3 확률변수 X의 분포가 표준지수분포(모수가 1인 지수분포)를 따를 때, $Y=\sqrt{X}$의 분포를 구하라.

풀이 확률변수 X의 밀도함수와 분포함수는 각각 $f_X(x)=e^{-x}$, $F_X(x)=1-e^{-x}$, $x>0$이므로 구하는 분포함수는 다음과 같다.

$$
\begin{aligned}
F_Y(y) &= P[Y\le y] = P[\sqrt{X}\le y] = P[X\le y^2] = F_X(y^2) \\
&= 1-e^{-y^2}, \quad y>0.
\end{aligned}
$$

밀도함수는 위의 분포함수의 도함수를 구하면 되고 아래와 같다.

$$f_Y(y) = 2ye^{-y^2}, \qquad y>0.$$

참고로 이 분포는 **웨이블 분포** (Weibull distribution)이다.

6. 많이 쓰이는 몇 가지 분포의 결정

예 5.6-4 확률변수 X가 균일분포 $U(-1, 1)$를 따를 때, $Y = X^2$의 분포를 구하라.

풀이 X의 밀도함수는 $f_X(x) = \frac{1}{2}$, $-1 < x < 1$이므로 구하고자 하는 분포함수는 다음과 같다.

$$F_Y(y) = P[Y \le y] = P[X^2 \le y] = P[-\sqrt{y} \le X \le \sqrt{y}]$$
$$= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\,dx = \sqrt{y}, \quad 0 \le y < 1 .$$

밀도함수는 위의 분포함수를 미분하면 얻어지며 다음과 같다.

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}, & 0 < y < 1 \\ 0, & \text{그 외.} \end{cases}$$

6. 많이 쓰이는 몇 가지 분포의 결정

3) 변환에 의한 방법

정리 5.6-1 확률변수 X의 밀도함수를 $f_X(x)$라 하고, 함수 $g(x)$가 $f_X(x)$의 정의구간에서 단조증가함수이거나 단조감소함수이면 $Y=g(X)$의 밀도함수는 다음과 같이 주어진다.

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x), \quad x = g^{-1}(y).$$

증명 (1) $g(x)$가 단조증가함수인 경우를 먼저 증명해 보도록 한다.
$[g(X) \le y] \Leftrightarrow [X \le g^{-1}(y)]$이므로 Y의 분포함수는 다음과 같이 얻어진다.

$$F_Y(y) = P[X \le g^{-1}(y)] = \int_{-\infty}^{g^{-1}(y)} f_X(x)\, dx.$$

위 분포함수를 미분하면 다음의 밀도함수를 얻게 된다.

$$f_Y(y) = \left\{\frac{d}{dy} g^{-1}(y)\right\} f_X(g^{-1}(y)).$$

(2) $g(x)$가 단조감소함수인 경우에는 $[g(X) \le y] \Leftrightarrow [X \ge g^{-1}(y)]$이므로 Y의 분포함수는 다음과 같이 얻어진다.

$$F_Y(y) = P[X \ge g^{-1}(y)] = 1 - P[X \le g^{-1}(y)].$$

따라서 밀도함수는 $f_Y(y) = \left(-\frac{d}{dy} g^{-1}(y)\right) f_X(g^{-1}(y))$이 된다.

(1)과 (2)를 종합하면 $f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| f_X(g^{-1}(y))$로 표현된다.

3) 변환에 의한 방법

따름정리 $g(x) = ax + b$이면, $Y = g(X) = aX + b$의 밀도함수는

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), \quad a \neq 0$$

이며, 특히 Y가 이산확률변수이면 밀도함수는 $f_Y(y) = f_X\left(\frac{y-b}{b}\right)$이다.

증명 $x = g^{-1}(y) = \frac{y-b}{a}$, $\frac{dx}{dy} = \frac{1}{a}$이므로 앞의 정리로부터 따름정리가 나온다. Y가 이산확률변수이면

$$P[Y=y] = P[aX+b=y] = P\left[X=\frac{y-b}{a}\right] = f_X\left(\frac{y-b}{a}\right)$$

임을 알 수 있다.

예 5.6-6 확률변수 X가 정규분포 $N(\mu, \sigma^2)$를 따를 때, $Y = \frac{X-\mu}{\sigma}$의 분포를 구하라.

풀이 $X = \sigma Y + \mu$이므로 앞의 정리로부터 다음의 결과를 얻는다.

6. 많이 쓰이는 몇 가지 분포의 결정

3) 변환에 의한 방법

따름정리 $g(x) = ax + b$이면, $Y = g(X) = aX + b$의 밀도함수는

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), \quad a \neq 0$$

이며, 특히 Y가 이산확률변수이면 밀도함수는 $f_Y(y) = f_X\left(\frac{y-b}{b}\right)$이다.

증명 $x = g^{-1}(y) = \frac{y-b}{a}$, $\frac{dx}{dy} = \frac{1}{a}$이므로 앞의 정리로부터 따름정리가 나온다. Y가 이산확률변수이면

$$P[Y=y] = P[aX+b=y] = P\left[X=\frac{y-b}{a}\right] = f_X\left(\frac{y-b}{a}\right)$$

임을 알 수 있다.

예 5.6-6 확률변수 X가 정규분포 $N(\mu, \sigma^2)$를 따를 때, $Y = \frac{X-\mu}{\sigma}$의 분포를 구하라.

풀이 $X = \sigma Y + \mu$이므로 앞의 정리로부터 다음의 결과를 얻는다.

$$\begin{aligned} f_Y(y) &= |\sigma| f_X(\sigma y + \mu) \\ &= \sigma \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(\sigma y+\mu-\mu)^2}{\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}, \quad -\infty < y < \infty. \end{aligned}$$

이는 표준정규분포의 밀도함수임을 알 수 있다.

6. 많이 쓰이는 몇 가지 분포의 결정

4) 적률생성함수를 이용하는 방법

예 5.6-7 확률변수 X의 분포가 표준정규분포 $N(0, 1)$일 때, $Y = X^2$의 분포를 구하라.

풀이 $M_Y(t) = E[e^{tY}] = E[e^{tX^2}]$

$$= \int_{-\infty}^{\infty} e^{tx^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(1-2t)x^2} dx$$

〈 $u = (1-2t)^{\frac{1}{2}} x$, $du = (1-2t)^{\frac{1}{2}} dx$로 치환을 한다.〉

$$= (1-2t)^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du$$

$$= (1-2t)^{-\frac{1}{2}}, \quad 1-2t > 0.$$

이는 감마분포 $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$의 적률생성함수이며 또한 자유도가 1인 카이제곱분포 $\chi^2(1)$의 적률생성함수이기도 하다.

6. 많이 쓰이는 몇 가지 분포의 결정

예 5.6-8 확률변수 X의 분포가 정규분포 $N(\mu, \sigma^2)$일 때, $Z = \dfrac{X-\mu}{\sigma}$의 분포를 구하라.

풀이 $M_Z(t) = E[e^{tZ}] = E\left[\exp\left(\dfrac{X-\mu}{\sigma}t\right)\right] = E\left[\exp\left(\dfrac{t}{\sigma}X - \dfrac{\mu}{\sigma}t\right)\right]$

$$= \exp\left(-\frac{\mu}{\sigma}t\right) \cdot E\left[\exp\left(\frac{t}{\sigma}X\right)\right] = \exp\left(-\frac{\mu}{\sigma}t\right) \cdot M_X\left(\frac{t}{\sigma}\right).$$

$M_X\left(\dfrac{t}{\sigma}\right) = \exp\left(\dfrac{\mu}{\sigma}t + \dfrac{1}{2}t^2\right)$이므로

$$M_Z(t) = \exp\left(-\frac{\mu}{\sigma}t\right) \cdot \exp\left(\frac{\mu}{\sigma}t + \frac{1}{2}t^2\right) = \exp\left(\frac{1}{2}t^2\right)$$

이다. 이는 표준정규분포 $N(0, 1)$의 적률생성함수이다.

6. 많이 쓰이는 몇 가지 분포의 결정

(1) 이차원인 경우

1) 기본이론

정리 5.6-2 확률변수 X가 고정되어 있을 때, $W=t(X, Y)$가 Y의 단조증가함수이거나 단조감소함수이면 W의 밀도함수는 아래와 같이 주어진다.

$$f_W(w) = \int \left| \frac{\partial y}{\partial w} \right| f_{X,Y}(x, y)\, dx.$$

증명 조건부 분포를 이용하여 위의 정리를 증명하도록 한다. 일차원의 경우에서와 같이 고정된 $X=x$에 대하여

$$f_{W|x}(w) = \left| \frac{\partial y}{\partial w} \right| f_{Y|x}(y)$$

이고, $f_{X,Y}(x, y) = f_X(x) f_{Y|x}(y)$, $f_{X,W}(x, w) = f_X(x) f_{W|x}(w)$이므로

$$f_{X,W}(x, w) = \left| \frac{\partial y}{\partial w} \right| f_{X,Y}(x, y)$$

가 됨을 알 수 있다.

6. 많이 쓰이는 몇 가지 분포의 결정

SOOKHEE KWON

2) 합의 이론

따름정리 확률변수 X와 Y의 결합확률밀도함수를 $f_{X,Y}(x, y)$라고 하면 $W = X + Y$의 밀도함수는 아래와 같다.

$$f_W(w) = \int f_{X,Y}(w, w-x)dx.$$

증명 $W = X + Y$이므로 $Y = W - X$이며 $\left|\frac{\partial y}{\partial w}\right| = 1$이다. 따라서 앞의 정리에 의하여 따름정리가 증명된다.

예 5.6-9 확률변수 X와 Y가 서로 독립이고 표준균일분포 $U(0, 1)$을 따를 때, $W=X+Y$의 분포를 구하라.

풀이 확률변수 X와 Y의 결합밀도함수는

$$f_{X,Y}(x, y) = 1, \quad 0 < x < 1, \quad 0 < y < 1$$

이며 앞의 따름정리에 의하면 W의 밀도함수는 다음과 같다.

$$f_W(w) = \int f_{X,Y}(x, w-x)dx = \int 1\, dx.$$

그런데 위의 적분에서 x의 적분구간이 결정되어야 $f_W(w)$의 형태가 결정된다. $0 < w < 2$, $0 < y = w-x < 1$이므로 x의 적분구간은 다음의 세 가지 조건을 만족해야 한다.

$$w-1 < x < w, \quad 0 < x < 1, \quad 0 < w < 2.$$

즉, $0 < w \le 1$이면 x의 범위는 $0 < x < w$이고 $1 < w < 2$이면 x의 범위는 $w-1 < x < 1$이다. 따라서 W의 밀도함수는 다음과 같다.

$$f_W(w) = \begin{cases} \int_0^w dx = w, & 0 < w < 1 \\ \int_{w-1}^1 dx = 2-w, & 1 < w < 2. \end{cases}$$

6. 많이 쓰이는 몇 가지 분포의 결정

3) 차의 이론

따름정리 $W = Y - X$의 밀도함수는 다음과 같이 주어진다.

$$f_W(w) = \int f_{X,Y}(x,\ w+x)dx.$$

증명 앞의 따름정리의 증명과 유사하므로 여기서는 생략하기로 한다.

예 5.6-10 두 확률변수 X와 Y는 서로 독립이고 모수가 1인 표준지수분포를 따른다고 할 때, $W=Y-X$의 분포를 구하라.

풀이 X와 Y의 결합밀도함수는

$$f_{X,Y}(x, y) = e^{-(x+y)}, \quad x>0, \quad y>0$$

이며 W의 밀도함수는 위의 따름정리에 의하여

$$f_W(w) = \int e^{-(x+w+x)}dx = \int e^{-2x-w}dx$$

이다. 여기서도 적분구간을 살펴보아야 한다. $x>0$, $y>0$, $w=y-x$이므로 $-\infty<w<\infty$이고 $y=x+w$, $y>0$이므로 $x+w>0$이다. 따라서 x는 $x>-w$, $x>0$, $-\infty<w<\infty$를 만족하여야 한다. 이를 정리하면 w에 따른 x의 범위는 다음과 같다.

w의 범위가 $-\infty<w\le 0$일 때, x의 범위는 $-\infty<x<\infty$,
w의 범위가 $w>0$일 때, x의 범위는 $x>0$ 이다.

따라서, W의 밀도함수는 아래와 같이 주어진다.

$$f_W(w) = \begin{cases} \int_{-w}^{\infty} e^{-2x-w}dx = \frac{1}{2}e^{w}, & w<0 \\ \int_{0}^{\infty} e^{-2x-w}dx = \frac{1}{2}e^{-w}, & w>0. \end{cases}$$

이 결과를 결합하면 $f_W(w) = \frac{1}{2}e^{-|w|}$, $-\infty<w<\infty$가 되며 이는 이중지수분포(Double exponential distribution)의 밀도함수이다.

6. 많이 쓰이는 몇 가지 분포의 결정

3) 곱과 몫의 이론

따름정리　두 확률변수 X와 Y의 곱 $W = YX$의 밀도함수는 다음과 같이 주어진다.

$$f_W(w) = \int \left|\frac{1}{x}\right| f_{X,Y}\left(x, \frac{w}{x}\right) dx.$$

증명　$\frac{\partial Y}{\partial W} = \frac{1}{X}$, $Y = \frac{W}{X}$이므로 앞의 정리에 의하여 증명된다.

따름정리　두 확률변수 X와 Y의 몫 $W = Y/X$의 밀도함수는 다음과 같이 주어진다.

$$f_W(w) = \int |x| f_{X,Y}(x, wx) dx.$$

증명　$\frac{\partial Y}{\partial W} = X$, $Y = WX$이므로 앞의 정리에 의하여 증명된다.

예 5.6-11 확률변수 Z는 표준정규분포 $N(0, 1)$을 따르고 확률변수 U는 자유도가 n인 카이제곱분포 $\chi^2(n)$을 따르며 Z와 U가 서로 독립일 때, $W = \dfrac{\sqrt{n}Z}{\sqrt{U}}$의 분포를 구하라.

풀이 $X = \sqrt{U}$라 두고 $Y = \sqrt{n}Z$라 두자. Y의 분포는 정규분포 $N(0, n)$임을 바로 알 수 있다. 한편 X의 밀도함수는

$$\begin{aligned} f_X(x) &= \left|\frac{\partial u}{\partial x}\right| f_U(x^2) \\ &= \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}-1}} x^{n-1} e^{-x^2/2}, \quad x > 0 \end{aligned}$$

이고 X와 Y는 서로 독립이므로 X와 Y의 결합 pdf는 다음과 같다.

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) \cdot f_Y(y) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}-1}} x^{n-1} e^{-x^2} \cdot \frac{1}{\sqrt{2\pi n}} e^{-\frac{1}{2}\frac{y^2}{n}} \\ &= \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n-1}{2}} \sqrt{n\pi}} x^{n-1} e^{-\frac{1}{2}\left(x^2 + \frac{y^2}{n}\right)}, \quad x > 0, \quad -\infty < y < \infty. \end{aligned}$$

6. 많이 쓰이는 몇 가지 분포의 결정

한편 $W = \dfrac{\sqrt{n}Z}{\sqrt{U}} = \dfrac{Y}{X}$ 이므로 앞의 따름정리에 의하여 W의 밀도함수는 다음과 같이 얻어진다.

$$
\begin{aligned}
f_W(w) &= \int_0^\infty x \cdot f_{X,Y}(x, wx)dx \\
&= \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n-1}{2}}\sqrt{n\pi}} x^n \cdot e^{-\frac{1}{2}\left(x^2+\frac{w^2x^2}{n}\right)}dx
\end{aligned}
$$

$\langle v = x^2\left(1+\dfrac{w^2}{n}\right)/2$로 치환하면 $x = \left(\dfrac{2v}{1+w^2/n}\right)^{1/2}$이며,

$dx = x(1+w^2/n)dx = \sqrt{2v(1+w^2/n)}\,dx$ 이므로$\rangle$

$$
\begin{aligned}
&= \frac{(1+w^2/n)^{-\frac{1}{2}(n+1)}}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \int_0^\infty v^{\frac{1}{2}(n-1)} e^{-v}dv \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{w^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad -\infty < w < \infty.
\end{aligned}
$$

6. 많이 쓰이는 몇 가지 분포의 결정

[참고] 위의 등식에서 $\int_0^\infty v^{\frac{1}{2}(n-1)} e^{-v} dv = \int_0^\infty v^{\frac{1}{2}(n+1)-1} e^{-v} dv = \Gamma\left(\frac{n+1}{2}\right)$임을 사용하였고, 위의 밀도함수는 자유도가 n인 t-분포(t-distribution)의 밀도함수이다.

예 5.6-12 확률변수 U는 자유도가 m인 카이제곱분포 $\chi^2(m)$을 따르고 확률변수 V는 자유도가 n인 카이제곱분포 $\chi^2(n)$을 따르며 U와 V가 서로 독립일 때, $W=\frac{U/m}{V/n}$의 분포를 구하라.

풀이 $Y=U/m$, $X=V/n$으로 두면 Y와 X의 pdf는 아래와 같다.

$$f_Y(y) = \left|\frac{\partial u}{\partial y}\right| f_U(my) = \frac{m^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} y^{\frac{m}{2}-1} e^{-\frac{1}{2}my}, \quad y>0,$$

$$f_X(x) = \left|\frac{\partial v}{\partial x}\right| f_V(nx) = \frac{n^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}nx}, \quad x>0.$$

U와 V는 서로 독립이므로 W의 pdf는 따름정리에 의하여 다음과 같이 얻어진다.

U와 V는 서로 독립이므로 W의 pdf는 따름정리에 의하여 다음과 같이 얻어진다.

$$\begin{aligned}
f_W(w) &= \int_0^\infty x \cdot f_X(x) f_Y(wx)\, dx \\
&= \frac{m^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \frac{n^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} \int_0^\infty x^{\frac{n}{2}} e^{-\frac{1}{2}nx} (wx)^{\frac{m}{2}-1} e^{-\frac{1}{2}mwx} dx \\
&= \frac{m^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \frac{n^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} w^{\frac{m}{2}-1} \int_0^\infty x^{\frac{m+n}{2}-1} e^{-\frac{1}{2}x(n+mw)} dx
\end{aligned}$$

〈 $v=\frac{1}{2}x(n+mw)$로 치환하면 $dv=\frac{1}{2}(n+mw)dx$ 이므로 〉

$$\begin{aligned}
&= \frac{m^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}} \frac{n^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} \frac{w^{\frac{m}{2}-1} 2^{\frac{1}{2}(m+n)}}{(n+mw)^{\frac{1}{2}(m+n)}} \int_0^\infty v^{\frac{1}{2}(m+n)-1} e^{-v} dv \\
&= \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{w^{\frac{m}{2}-1}}{(n+mw)^{\frac{1}{2}(m+n)}}.
\end{aligned}$$

[참고] 위의 W의 pdf는 분자의 자유도가 m, 분모의 자유도가 n인 F-분포(F-distribution), $F(m, n)$의 밀도함수이다.