第8講：確率変数の関数

平成16年12月8日

確率変数の関数：離散の場合

- X: 離散型確率変数
- $Y = g(X)$: 離散型確率変数
- $g^{-1}(y) = \{x | g(x) = y\}$
- Y の分布：

$$\begin{aligned} P(Y = y) &= P[g(X) = y] \\ &= P[X \in g^{-1}(y)] \\ &= \sum_{x \in g^{-1}(y)} P(X = x) \end{aligned}$$

例

- X が一様分布

$$P(X=x)=\frac{1}{101}, \quad (x=0,1,2,\cdots,100)$$

-

$$Y=g(X)=\begin{cases} 0 & \text{if} \ \ 0 \le X < 60 \\ 1 & \text{if} \ \ 60 \le X < 80 \\ 2 & \text{if} \ \ 80 \le X \le 100 \end{cases}$$

- $Y=g(X)$ の分布:

$$\begin{aligned} P(Y=0) &= P(0 \le X < 60) \\ &= \sum_{x=0}^{59} P(X=x) = \frac{60}{101} \\ P(Y=1) &= \sum_{x=60}^{79} P(X=x) = \frac{20}{101} \\ P(Y=2) &= \sum_{x=80}^{100} P(X=x) = \frac{21}{101} \end{aligned}$$

確率変数の関数：連続の場合

- X: 連続型確率変数で、密度関数 $f(x)$
- $Y = g(X)$: 微分可能な単調関数
- Y の密度関数 $h(y)$:

$$h(y) = \begin{cases} f[g^{-1}(y)]\left[g^{-1}(y)\right]' & g(x): \text{単調増加} \\ -f[g^{-1}(y)]\left[g^{-1}(y)\right]' & g(x): \text{単調減少} \end{cases}$$

ただし、$x = g^{-1}(y)$ は $y = g(x)$ の逆関数

証明: 任意の $a, b \in R$ に対し

$$\begin{aligned} P[a \le Y \le b] &= P[a \le g(X) \le b] \\ &= P[g^{-1}(a) \le X \le g^{-1}(b)] \\ &= \int_{g^{-1}(a)}^{g^{-1}(b)} f(x)\,dx \\ &\overset{x=g^{-1}(y)}{=} \int_a^b f[g^{-1}(y)]\left[g^{-1}(y)\right]'\,dy \end{aligned}$$

例

- $X \sim N(\mu,\, \sigma^2)$
- $Y = aX + b\,, \quad (a \neq 0)$

$$\Longrightarrow \quad Y \sim N(a\mu + b,\, a^2\sigma^2)$$

解: $y = g(x) = ax + b$ から

$$x = \frac{y-b}{a}, \quad \frac{dx}{dy} = \frac{1}{a}$$

したがって、

$$\begin{aligned} f[g^{-1}(y)]\frac{dx}{dy} &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{[(y-b)/a-\mu]^2}{2\sigma^2}\right\}\frac{1}{|a|} \\ &= \frac{1}{\sqrt{2\pi}\sigma|a|}\exp\left\{-\frac{[y-(a\mu+b)]^2}{2\sigma^2a^2}\right\} \\ &= N(a\mu+b,\, a^2\sigma^2) \end{aligned}$$

例

- X の密度関数: $f(x)$
- $Y = X^2$ の密度関数: $g(y)$

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left(f(\sqrt{y}) + f(-\sqrt{y})\right) & y > 0 \\ 0 & y \leq 0 \end{cases}$$

証明　$0 < a < b$ に対して

$$\begin{aligned}
P(a < Y < b) &= P(a < X^2 < b) \\
&= P(\sqrt{a} < X < \sqrt{b}) + P(-\sqrt{b} < X < -\sqrt{a}) \\
&= \int_{\sqrt{a}}^{\sqrt{b}} f(x)\,dx + \int_{-\sqrt{b}}^{-\sqrt{a}} f(x)\,dx \\
&\overset{y=x^2}{=} \int_a^b f(\sqrt{y})\frac{1}{2\sqrt{y}}\,dy + \int_b^a f(-\sqrt{y})\frac{1}{-2\sqrt{y}}\,dy \\
&= \int_a^b \frac{1}{2\sqrt{y}}\left(f(\sqrt{y}) + f(-\sqrt{y})\right)dy
\end{aligned}$$

2つの確率変数の関数：離散の場合

- X, Y: 離散型確率変数
- (X, Y): 同時分布 $p(x, y)$

$$p(x,y) = P(X = x, Y = y)$$

- $Z = \psi(X, Y)$ の分布：

$$\begin{aligned} P(Z = z) &= P[\psi(X,Y) = z] \\ &= \sum_{(x,y)\in D} p(x,y) \end{aligned}$$

ただし、

$$D = \{(x,y) | \psi(x,y) = z\}$$

例

- $X \sim \mathrm{Bi}(m,p), \quad Y \sim \mathrm{Bi}(n,p)$
- X, Y: 独立
- $Z = X + Y \sim \mathrm{Bi}(m+n, p)$

証明 恒等式 ${}_{m+n}C_z = \sum_{x+y=z} {}_mC_x \cdot {}_nC_y$ に注意すると、

$$
\begin{aligned}
P(Z=z) &= P(X+Y=z) \\
&= \sum_{x+y=z} P(X=x)\,P(Y=y) \\
&= \sum_{x+y=z} {}_mC_x \cdot {}_nC_y\, p^{x+y}(1-p)^{m+n-(x+y)} \\
&= p^z(1-p)^{m+n-z} \sum_{x+y=z} {}_mC_x \cdot {}_nC_y \\
&= {}_{m+n}C_z\, p^z(1-p)^{m+n-z}
\end{aligned}
$$

積率母関数の性質：

定理 1 $M(t)$ がある確率分布の積率母関数とするとき、$M(t)$ を積率母関数とするような確率分布はただ１つしかない。

例：二項分布の再生性

$X \sim \mathrm{Bi}(m,p)$, $Y \sim \mathrm{Bi}(n,p)$ で、また X と Y が独立ならば、$X+Y$ も二項分布に従う。

証明　X, Y が独立のとき

$$\begin{aligned} M_{X+Y}(t) &= E(e^{(X+Y)t}) \\ &= E(e^{Xt}e^{Yt}) \\ &= E(e^{Xt})E(e^{Yt}) \\ &= M_X(t)\,M_Y(t) \end{aligned}$$

に注意すると、

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)\,M_Y(t) \\ &= (p\,e^t + 1 - p)^m\,(p\,e^t + 1 - p)^n \\ &= (p\,e^t + 1 - p)^{m+n} \end{aligned}$$

従って、$X+Y \sim \mathrm{Bi}(m+n,p)$

例：ポアソン分布の再生性

$X \sim \mathrm{Po}(\lambda_1)$, $Y \sim \mathrm{Po}(\lambda_2)$ で、また X と Y が独立ならば、$X+Y$ もポアソン分布に従う。

証明 X と Y の独立性から

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)\,M_Y(t) \\ &= e^{\lambda_1(e^t-1))}\,e^{\lambda_2(e^t-1))} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1))} \end{aligned}$$

従って、

$$X+Y \sim \mathrm{Po}(\lambda_1+\lambda_2)$$

例：正規分布の再生性

$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ で、また X と Y が独立ならば、$X+Y$ も正規分布に従う。

証明 X と Y の独立性から

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)\, M_Y(t) \\ &= \exp\left\{\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right\} \exp\left\{\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right\} \\ &= \exp\left\{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)}{2} t^2\right\} \end{aligned}$$

従って、

$$X + Y \sim N(\mu_1+\mu_2,\, \sigma_1^2+\sigma_2^2)$$