2016/06/03
Hajime Okano

対数正規分布

● 対数正規分布の誘導

$y=\ln x$ として、y が正規分布に従うとする。

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\,\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

ここに、

$$\mu_Y=E(\ln X) \qquad \mu=\sqrt{Var(\ln X)}$$

これを、X に関する確率密度関数に書き換える。

確率密度関数の変数変換より、

$$f_Y(y)=\left|\frac{dx}{dy}\right|f_X(x)$$

上式は、変数 x で表しても、変数 y で表しても、dx 区間と dy 区間の確率は等しくなければならないことを表している。すなわち、次のように書いた方が、元々の意味は分かり易い。

$$f_Y(y)|dy|=f_X(x)|dx|$$

ここで絶対値を用いているのは、y が x の減少関数である場合にも成立するようにするためである。ここで、

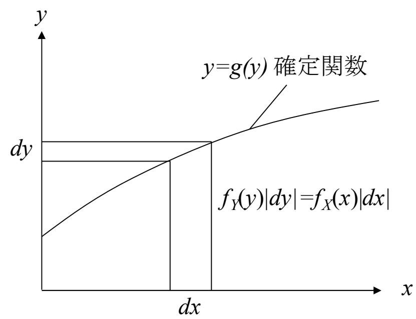
$$\frac{dy}{dx}=\frac{1}{x}$$

より、

$$f_X(x)=\frac{1}{x}\cdot f_Y(y)$$

よって、

$$f_X(x)=\frac{1}{\sqrt{2\pi}\,\sigma_Y\cdot x}\exp\left[-\frac{1}{2}\left(\frac{\ln x-\mu_Y}{\sigma_Y}\right)^2\right]$$



確率変数の変数変換

対数正規分布は、通常、次式のように表記する。

$$f_X(x)=\frac{1}{\varsigma\cdot x\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln x-\lambda}{\varsigma}\right)^2\right]$$

ここに、

$$\lambda=E(\ln X) \qquad \varsigma=\sqrt{Var(\ln X)}$$

であり、それぞれ対数平均，対数標準偏差と呼ばれる。

● 対数正規分布の平均値と分散

ここで、対数正規分布の平均値、分散などを求めておこう。X の平均は、

$$\mu_X=\int_0^\infty\frac{1}{\varsigma\cdot\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln x-\lambda}{\varsigma}\right)^2\right]dx$$

$dy/dx=1/x \rightarrow dx=x\,dy=e^y dy$ より、

$$\mu_X = \int_{-\infty}^{\infty} \frac{1}{\varsigma \cdot \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\lambda}{\varsigma} \right)^2 \right] e^y dy$$

$$= \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[y - \frac{1}{2}\left(\frac{y-\lambda}{\varsigma} \right)^2 \right] dy$$

指数部を正規分布の形になるように変形していくと、

$$\mu_X = \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[y - \frac{1}{2}\left(\frac{y-\lambda}{\varsigma} \right)^2 \right] dy$$

$$= \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \frac{-2y\varsigma^2 + y^2 - 2\lambda y + \lambda^2}{\varsigma^2} \right] dy$$

$$= \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \frac{\left\{ y - \left(\varsigma^2 + \lambda \right) \right\}^2 - \varsigma^4 - 2\varsigma^2\lambda - \lambda^2 + \lambda^2}{\varsigma^2} \right] dy$$

$$= \left[\frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\left(\frac{y - \left(\varsigma^2 + \lambda \right)}{\varsigma} \right)^2 \right\} dy \right] \exp\left(\lambda + \frac{1}{2}\varsigma^2 \right)$$

上式で、大括弧内は、平均$\left(\varsigma^2 + \lambda\right)$、標準偏差$\varsigma$の正規分布を表しているので、積分は１になる。

以上より、次式を得る。

$$\mu_X = \exp\left(\lambda + \frac{1}{2}\varsigma^2 \right)$$

次に、Xの分散を求めるために、Xの２次モーメントを求める。

$$E\left(X^2\right) = \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{2y} \exp\left[-\frac{1}{2}\left(\frac{y-\lambda}{\varsigma} \right)^2 \right] dy$$

$$= \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \frac{-4y\varsigma^2 + y^2 - 2\lambda y + \lambda^2}{\varsigma^2} \right] dy$$

$$= \frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \frac{\left\{ y - \left(2\varsigma^2 + \lambda \right) \right\}^2 - 4\varsigma^4 - 4\varsigma^2\lambda - \lambda^2 + \lambda^2}{\varsigma^2} \right] dy$$

$$= \left[\frac{1}{\varsigma \cdot \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\left(\frac{y - \left(2\varsigma^2 + \lambda \right)}{\varsigma} \right)^2 \right\} dy \right] \exp\left\{ 2\left(\lambda + \varsigma^2 \right) \right\}$$

$$= \exp\left\{ 2\left(\lambda + \varsigma^2 \right) \right\}$$

$$Var(X) = E\left(X^2\right) - {\mu_X}^2$$

$$= \exp\left\{ 2\left(\lambda + \varsigma^2 \right) \right\} - \exp\left\{ 2\left(\lambda + \frac{1}{2}\varsigma^2 \right) \right\}$$

$$= \exp\left\{ 2\left(\lambda + \frac{1}{2}\varsigma^2 \right) \right\} \left(\exp \varsigma^2 - 1 \right)$$

$$= {\mu_X}^2 \cdot \left(\exp \varsigma^2 - 1 \right)$$

これより、

$$\frac{{\sigma_X}^2}{{\mu_X}^2} + 1 = \exp \varsigma^2$$

$$\varsigma^2 = \ln\left(1 + \frac{{\sigma_X}^2}{{\mu_X}^2} \right)$$

１の周りのテイラー展開より、

$$\ln(1+z) \approx \ln 1 + \ln'(1) z = z \qquad z \prec 1$$

が成り立つ。よって、

$$\varsigma \approx \frac{\sigma_X}{\mu_X}$$

が成り立つ。これより、対数標準偏差ςは変動係数に近い値となることが分かる。

次に、対数平均λの持つ意味について考えてみる。$y = \ln x$は正規分布に従い、λはその平均$E[\ln x]$である。正規分布では平均値を越えない確率も超える確率も 50%、つまり中央値である。したがって、λに対応するxは中央値となる。

$$x_m = e^{\lambda}$$

ここで下添え字mは中央値を表す。

最後に、次式で示すようなXが区間$[a, b]$に含まれる確率を計算する方法を示しておく。

$$P(a < X \le b) = \int_a^b \frac{1}{\varsigma \cdot x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\varsigma}\right)^2\right] dx$$

ここで、

$$s = \frac{\ln x - \lambda}{\varsigma}$$

とおくと（標準正規化）、

$$P(a < X \le b) = \int_{(\ln a - \lambda)/\varsigma}^{(\ln b - \lambda)/\varsigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}s^2\right] dx$$

$$= \Phi\left(\frac{\ln b - \lambda}{\varsigma}\right) - \Phi\left(\frac{\ln a - \lambda}{\varsigma}\right)$$

ここで、Φは標準正規確率分布（平均０、分散１の正規分布の累積確率分布）である。

$a \to -\infty$とすると、

$$P(-\infty < X \le b) = \int_0^{(\ln b - \lambda)/\varsigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}s^2\right] dx$$

$$= \Phi\left(\frac{\ln b - \lambda}{\varsigma}\right)$$