データ解析 (第3回)

静岡大学システム工学科
安藤 和敏

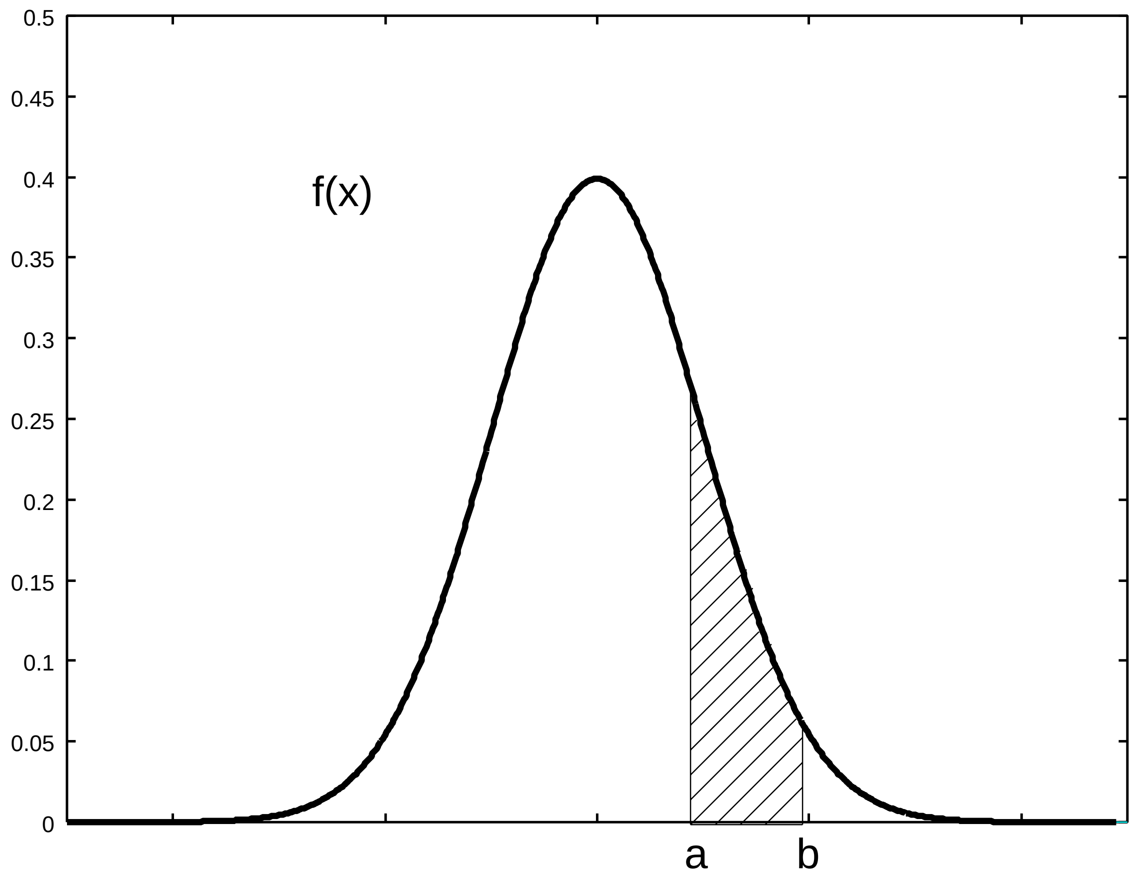
(1) 確率密度関数

$$f(x) \geq 0 \quad (-\infty < x < \infty)$$

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

を満たす関数 f を確率密度関数と呼ぶ.

$$\Pr(a \leq x \leq b) = \int_a^b f(x)\mathrm{d}x \qquad (2.22)$$



(2) 期待値と分散

期待値の定義

$$E(x) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \qquad (2.23)$$

期待値の性質

$$E(ax + b) = aE(x) + b \qquad (2.24)$$

$y = ax + b$の分布

xが確率変数のとき, $y = ax + b$で定義されるyも確率変数となる. さらに, xの確率密度関数が$f(x)$のとき, yの確率密度関数$g(y)$は, $g(y) = \frac{1}{a} f(\frac{y-b}{a})$で与えられる.

期待値の性質 (2.24) の証明

$$
\begin{aligned}
E(y) &= \int_{-\infty}^{\infty} y g(y) \mathrm{d}y \\
&= \int_{-\infty}^{\infty} y \frac{1}{a} f\left(\frac{y-b}{a}\right) \mathrm{d}y \\
&= \int_{-\infty}^{\infty} (ax+b) \frac{1}{a} f(x) a \mathrm{d}x \\
&= \int_{-\infty}^{\infty} (ax+b) f(x) \mathrm{d}x
\end{aligned}
$$

期待値の性質(2.24)の証明(続き)

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} (ax+b) f(x) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} ax f(x) \mathrm{d}x + \int_{-\infty}^{\infty} b f(x) \mathrm{d}x \\
&= a \int_{-\infty}^{\infty} x f(x) \mathrm{d}x + b \int_{-\infty}^{\infty} f(x) \mathrm{d}x \\
&= aE(x) + b.
\end{aligned}
$$

分散の定義

$$V(x) = E\{(x - E(x))^2\} \tag{2.25}$$

分散の性質

$$V(x) = E(x^2) - \{E(x)\}^2 \tag{2.26}$$

$$V(ax + b) = a^2 V(x) \tag{2.27}$$

分散の性質(2. 26)の証明

$$
\begin{aligned}
V(x) &= E((x - E(x))^2) \\
&= E(x^2 - 2xE(x) + E(x)^2) \\
&= E(x^2) - 2E(x)E(x) + E(x)^2 \\
&\qquad\qquad\qquad ((2.24)\text{を用いた}) \\
&= E(x^2) - E(x)^2.
\end{aligned}
$$

分散の性質(2. 27)の証明

$$
\begin{aligned}
V(ax+b) &= E((ax+b-E(ax+b))^2) \\
&= E((ax+b-aE(x)-b)^2) \\
&= E((a(x-E(x)))^2) \\
&= E(a^2(x-E(x))^2)) \\
&= a^2E((x-E(x))^2)) \\
&\qquad\qquad ((2.24)\text{を用いた}) \\
&= a^2V(x).
\end{aligned}
$$

(3) 正規分布

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{2.28}$$

で定義される f を確率密度関数にもつ分布は,
正規分布と呼ばれる.
確率変数 x が期待値 μ, 分散 σ^2 の正規分布に従うということを $x \sim N(\mu, \sigma^2)$ で表す.
μ 母平均, σ 母標準偏差, σ^2 母分散.

標準正規分布

期待値 0, 分散 1 の正規分布 $N(0,1)$ は, 標準正規分布と呼ばれる.

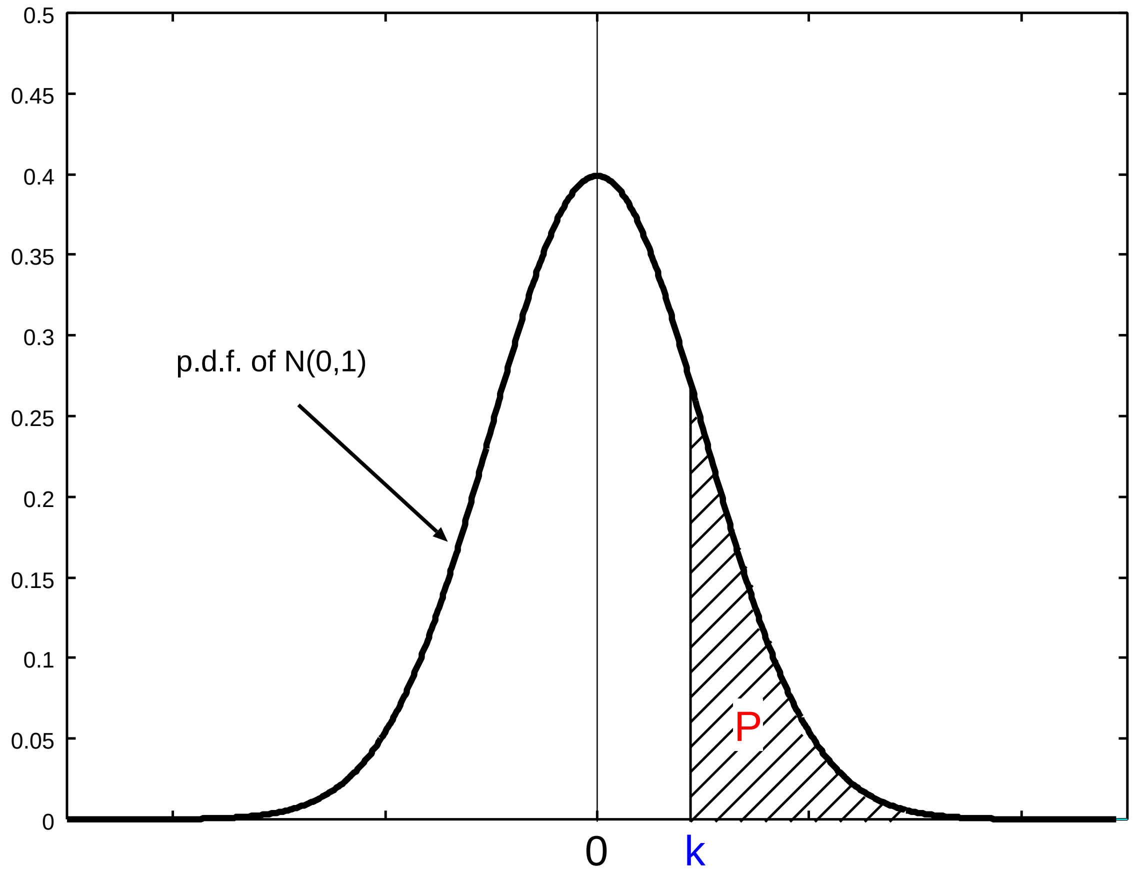
確率変数の標準化

$$x \Longrightarrow u = \frac{x-u}{\sigma}$$

正規分布に従う変数を正規化すると, それは標準正規分布に従う.

$$x \sim N(\mu, \sigma^2) \Longrightarrow u \sim N(0,1)$$

正規分布表(p.234-235)の読み方



$$P = \int_k^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \mathrm{d}u$$

k から P を求める方法

k	*=0	*=1	*=2	*=3	*=4	*=5
0.0*						
0.1*						
0.2*						
0.3*						
0.4*						
0.5*						
0.6*				.2643		
0.7*						
0.8*						
0.9*						
1.0*						
1.1*						
1.2*						

k=0.63 ---> P=0.2643

正規分布表を使った応用

$x \sim N(\mu, \sigma^2)$ とする.

$$Pr\left(\mu - k\sigma \leq x \leq \mu + k\sigma\right) = 0.90$$

となるような k を求めてみよう.

x 正規化すると $\frac{x-\mu}{\sigma} \sim N(0,1)$ であるから,
p.235 の表から,

$$Pr\left(\frac{x-\mu}{\sigma} \geq 1.645\right) = 0.05.$$

$N(0,1)$ の密度関数は y 軸対称であるから,

$$Pr\left(\frac{x-\mu}{\sigma} \leq -1.645\right) = 0.05.$$

ゆえに,

$$Pr\left(\frac{x-\mu}{\sigma} \leq -1.645, \text{または}, \frac{x-\mu}{\sigma} \geq 1.645\right) =$$

よって,

$$Pr\left(-1.645 \leq \frac{x-\mu}{\sigma} \leq 1.645\right) = 0.90.$$

書き換えると

$$Pr\left(\mu - 1.645\sigma \leq x \leq \mu + 1.645\sigma\right) = 0.90. \tag{2.31}$$