演習「統計的処理について」3

蛭川　潤一

新潟大学・自然科学系

2018 年 5 月 21 日

正規分布

正規分布

次の確率密度関数を持つ分布を
正規分布（normal distribution）と呼び
$N(\mu, \sigma^2)$ で表す

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \tag{1}$$

ただし、$-\infty < \mu < \infty$, $0 < \sigma < \infty$ である

特に、$\mu = 0$, $\sigma = 1$ のとき
標準正規分布（standard normal distribution）と呼び
$N(0, 1)$ で表す

正規分布

$Z \cdots$ 標準正規分布に従う確率変数

$\phi(x) \cdots$ 標準正規分布の密度関数

$$\Phi(x) = P(Z \le x) = \int_{-\infty}^{x} \phi(y)dy \tag{2}$$

を標準正規分布の分布関数（distribution function）と呼ぶ

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty \tag{3}$$

正規分布

Remark

観測値や測定値には近似的に正規分布に従う場合が多くあり
正規分布は統計学におけるもっとも重要な分布である

Theorem

確率変数 X が正規分布 $N(\mu, \sigma^2)$ に従うとき

平均と分散は

$$E(X) = \mu, \quad V(X) = \sigma^2 \tag{4}$$

となる

正規分布

確率変数 Z が標準正規分布 $N(0,1)$ に従うとき

任意の $0 < \alpha < 1$ に対して

(5) $$P(Z > z) = 1 - \Phi(z) = \int_z^\infty \phi(x)dx = \alpha$$

をみたす z を

標準正規分布の上側 α 点
（あるいは $100\alpha\%$ 点）と呼び
z_α で表す

α を上側確率と呼ぶ

z_α は標準正規分布表（付表１）から求めることができる

標準正規分布表

z	.00	.01	.02	$\cdots$
.0	$\Phi(0.00)$	$\Phi(0.01)$	$\Phi(0.02)$	$\cdots$
.1	$\Phi(0.10)$	$\Phi(0.11)$	$\Phi(0.12)$	$\cdots$
.2	$\Phi(0.20)$	$\Phi(0.21)$	$\Phi(0.22)$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

中心極限定理

$X_1, \ldots, X_n$：独立な確率変数
共通の平均 $\mu = E(X_i)$ と，共通の有限な正の分散 $\sigma^2 = V(X_i)$
$(i = 1, \ldots, n)$ をもつ

$$\overline{X}_n := \frac{Y_n}{n}$$

Theorem (中心極限定理)

n が限りなく大きくなれば

$$Z_n := \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$$

の分布は，限りなく標準正規分布 $N(0,1)$ に近づく
$Z_n \xrightarrow{\mathcal{L}} Z$ （分布収束），$Z \sim N(0,1)$ と表す

中心極限定理

Remark

- 正の有限な分散 σ^2 をもつすべての分布について
  n が十分大きいとき
  $Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ の分布は
  標準正規分布 $N(0,1)$ で近似できる
- 分布が何であっても
  極限分布として正規分布が現れる

スラツキーの定理

Theorem (スラツキー（Slutsky）の定理)

確率変数列 $\{X_n\}$ $\{Y_n\}$ と
確率変数 X と定数 c に対して
$X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{P} c$ であれば次が成り立つ

$$(i)\ X_n + Y_n \xrightarrow{\mathcal{L}} X + c$$
$$(ii)\ X_n Y_n \xrightarrow{\mathcal{L}} cX$$

中心極限定理

標準偏差 σ を
（不偏）標本標準偏差

$$S_n = \sqrt{S_n^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

で置き換えると
$\frac{\sigma}{S_n} \xrightarrow{P} 1 \Rightarrow$

$$Z_n^* = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{S_n} = Z_n \times \frac{\sigma}{S_n} \xrightarrow{\mathcal{L}} Z \times 1 \sim N(0,1)$$

母平均の区間推定

信頼度 $1-\alpha$ に対して
$\Downarrow$ 標準正規分布表から

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$$

となる $N(0,1)$ の上側 $\alpha/2$ 点 $z_{\alpha/2}$ を求める
$\Downarrow$

$$\begin{aligned} 1-\alpha &\approx P\left(-z_{\alpha/2} \leq Z_n^* \leq z_{\alpha/2}\right) \\ &= P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{S_n} \leq z_{\alpha/2}\right) \\ &= P\left(\overline{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}\right) \end{aligned}$$

一般母集団の母平均の区間推定

Theorem

大標本のとき、母分散 σ^2 が未知ならば
母平均の信頼度 $1-\alpha$ の信頼区間は近似的に

$$\left[\overline{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}}, \overline{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}\right]$$

となる
95% 近似信頼区間は $z_{\alpha/2}$ を 1.96 とすればよい

混合正規分布

$X_1 \sim N\left(\mu_1, \sigma_1^2\right)$, $X_2 \sim N\left(\mu_1, \sigma_2^2\right)$, $M \sim Ber\left(p\right)$
X_1, X_2, M は独立

$$Y := M \times X_1 + (1 - M) \times X_2$$

$\Rightarrow$

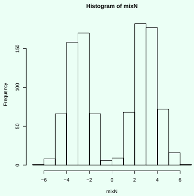
$$\begin{aligned} \mu :&= E\left(Y\right) = E\left(M\right) \times E\left(X_1\right) + \left\{1 - E\left(M\right)\right\} \times E\left(X_2\right) \\ &= p\mu_1 + (1 - p)\,\mu_2 \end{aligned}$$

$\Downarrow p = \frac{1}{2}, \mu_2 = -\mu_1$

$$\mu = \frac{\mu_1}{2} - \frac{\mu_1}{2} = 0$$

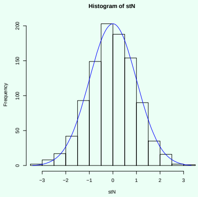
混合正規分布

混合正規分布のヒストグラム：



中心極限定理

標準化した変数のヒストグラムと標準正規分布の密度関数：



中心極限定理

1000 個の 95% 信頼区間：