

公式集 (統計学)

離散型の確率

値 $x_1, x_2, \dots$ を取る離散型確率変数 X について, 確率 $P(a \leq X \leq b)$ は

$$P(a \leq X \leq b) = \sum_{a \leq x_i \leq b} P(X = x_i)$$

期待値 $E(X)$, 分散 $V(X)$, 標準偏差 $\sigma(X)$ は

$$E(X) = \sum_i x_i P(X = x_i)$$

$$\begin{aligned} V(X) &= \sum_i (x_i - E(X))^2 P(X = x_i) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

$$\sigma(X) = \sqrt{V(X)}$$

連続型の確率

確率密度関数 $f(x)$ の連続型確率変数 X について, 確率 $P(a \leq X \leq b)$ は

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

期待値 $E(X)$, 分散 $V(X)$, 標準偏差 $\sigma(X)$ は

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$\begin{aligned} V(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx \\ &= E(X^2) - E(X)^2 \end{aligned}$$

$$\sigma(X) = \sqrt{V(X)}$$

二項分布

二項分布 $\text{Bin}(n, p)$ に従う離散型確率変数 X について, 確率 $P(X = k)$ は

$$P(X = k) = {}_n C_k p^k (1-p)^{n-k}$$

期待値 $E(X)$, 分散 $V(X)$ は

$$E(X) = np \quad V(X) = np(1-p)$$

正規分布

正規分布 $N(\mu, \sigma^2)$ に従う連続型確率変数 X について, 確率密度関数 $f(x)$ は

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

期待値 $E(X)$, 分散 $V(X)$ は

$$E(X) = \mu \quad V(X) = \sigma^2$$

標準正規分布 $N(0, 1)$ と変数変換

X が $N(\mu, \sigma^2)$ に従い, Z が $N(0, 1)$ に従うとき, 積分の変数変換により

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

$$P(a \leq Z \leq b) = P(\mu + a\sigma \leq X \leq \mu + b\sigma)$$

標本平均 $\bar{X}$, 標本分散 S^2 , 不偏分散 U^2

n 個の無作為標本 $X_1, X_2, \dots, X_n$ について, 標本平均 $\bar{X}$, 標本分散 S^2 , 不偏分散 U^2 は

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - (\bar{X})^2$$

$$U^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

母平均 μ , 母分散 σ^2 とすれば

$$E(\bar{X}) = \mu \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

$$E(S^2) = \frac{n-1}{n} \sigma^2$$

$$E(U^2) = \sigma^2$$

区間推定

無作為標本の数 n が十分大きいとき, 信頼度 $(1 - \alpha)$ の区間は

$$\bar{X} - z \left(\frac{\alpha}{2} \right) \frac{U}{\sqrt{n}} \leq \mu \leq \bar{X} + z \left(\frac{\alpha}{2} \right) \frac{U}{\sqrt{n}}$$