公式集：数理統計学

A. 確率

A–1. 確率

A-1-1 [**異なる n 個のものから r 個とる順列**]

$${}_n\mathrm{P}_r = n(n-1)(n-2)\cdots(n-r+1)$$

A-1-2 [**異なる n 個のものから r とる組み合わせ**]

$${}_n\mathrm{C}_r = \frac{{}_n\mathrm{P}_r}{r!} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r(r-1)(r-2)\cdots 2\cdot 1} = \frac{n!}{r!(n-r)!}$$

ただし、${}_n\mathrm{C}_0 = 1$

A-1-3 [${}_n\mathrm{C}_r$ **の基本性質**]

$${}_n\mathrm{C}_r = {}_n\mathrm{C}_{n-r}$$

$${}_n\mathrm{C}_r = {}_{n-1}\mathrm{C}_{r-1} + {}_{n-1}\mathrm{C}_r \qquad (1 \leqq r \leqq n-1)$$

A-1-4 [**同じものがある順列**]

n 個のもののうち a が p 個、 b が q 個、 c が r 個、$\cdots$ $(p+q+r+\cdots=n)$ あるとき、それらをすべて並べてできる順列の総数は $\dfrac{n!}{p!q!r!\cdots}$

A-1-5 [**二項定理**]

$$(a+b)^n = {}_n\mathrm{C}_0 a^n + {}_n\mathrm{C}_1 a^{n-1}b + {}_n\mathrm{C}_2 a^{n-2}b^2 + \cdots + {}_n\mathrm{C}_r a^{n-r}b^r + \cdots + {}_n\mathrm{C}_{n-1} ab^{n-1} + {}_n\mathrm{C}_n b^n$$

A-1-6 [**数学的確率**]

ある試行の標本空間 Ω の根元事象の個数が N であり、それらは同じ程度の確からしさで起こるものとする。そのときの事象 A が起こる **確率** を $P(A) = \dfrac{n(A)}{N}$ で定義する。

注）ある試行の標本空間 Ω で事象 A に含まれる根元事象の個数を $n(A)$ で表す。標本空間に対して $N = n(\Omega)$ とする。

A-1-7 [**統計的確率**]

試行を N 回くり返したとき、事象 A が起こった回数が n であるとする。試行の回数 N をふやしてゆくとき、相対度数 $\dfrac{n}{N}$ が一定の値 p に近づくならば、この値 p を事象 A の **確率** という。

A-1-8 [**確率の基本性質**]

(1) 任意の事象 A に対して、 $0 \leqq P(A) \leqq 1$

(2) $P(\Omega) = 1$, $P(\phi) = 0$

(3) 事象 A と B が互いに排反ならば $P(A \cup B) = P(A) + P(B)$

A-1-9 [**一般の加法定理**]

任意の事象 $A,\ B$ に対して

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

A-1-10 [**条件付き確率**]

事象 A が起こったとき B が起こりうる条件付き確率は

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

A-1-11 [**乗法定理**]

事象 A と B の条件付き確率について

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

A-1-12 [**独立試行の確率**]

ある試行において、事象 A とその余事象 $\overline{A}$ の確率を

$$P(A) = p, \qquad P(\overline{A}) = q, \qquad p + q = 1$$

とする。その試行を独立に n 回くり返すとき A が r 回起こる確率 p_r は

$p_r = {}_n\mathrm{C}_r p^r q^{n-r}$

A–2. **確率分布**

A-2-1 [**平均（値）の定義**]

確率変数 X の確率分布が

X	x_1	x_2	$\cdots$	x_i	$\cdots$	x_r	計
確率	p_1	p_2	$\cdots$	p_i	$\cdots$	p_r	1

であたえられるとき、

$$\mu = E(X) = x_1p_1 + x_2p_2 + \cdots + x_rp_r = \sum_{i=1}^{r} x_ip_i$$

で定義される値を確率変数 X の **平均、平均値** または **期待値** といい、 μ または $E(X)$ で表す。

A-2-2 [**分散、標準偏差の定義**]

A-2-1 の確率分布表の確率変数 X の平均を μ として

$$\sigma^2 = V(X) = (x_1 - \mu)^2p_1 + (x_2 - \mu)^2p_2 + \\ + \cdots + (x_r - \mu)^2p_r = \sum_{i=1}^{r}(x_i - \mu)^2p_i$$

で定義される値 $\sigma^2 = V(X)$ を X の **分散** という。その正の平方根 $\sigma = \sqrt{V(X)}$ を X の **標準偏差** という。

A-2-3 [**分散の計算方法**]

確率変数 X の分散について

$$\sigma^2 = {x_1}^2p_1 + {x_2}^2p_2 + \cdots + {x_r}^2p_r - \mu^2$$
$$= \sum_{i=1}^{r} {x_i}^2p_i - \mu^2$$

確率変数を使って表せば

$$V(X) = E(X^2) - E(X)^2$$

A-2-4 [$aX + b$ **の平均と分散**]

$a,\ b$ を定数、 $a \neq 0$ として、確率変数 X の 1 次式 $aX + b$ の平均と分散は

$$E(aX + b) = aE(X) + b, \qquad V(aX + b) = a^2V(X)$$

A-2-5 [**二項分布の平均と分散**]

二項分布 $B(n,\ p)$ の平均と分散は

平均：$\mu = E(X) = np$

分散：$\sigma^2 = V(X) = npq = np(1-p)$

A-2-6 [**確率変数の和の平均**]

確率変数の平均について次の公式が成り立つ。

$E(X+Y) = E(X) + E(Y)$

A-2-7 [**互いに独立な確率変数の積の平均**]

互いに独立な確率変数 $X,\ Y$ に対して

$E(XY) = E(X)E(Y)$

A-2-8 [**互いに独立な確率変数の和の分散**]

互いに独立な確率変数 $X,\ Y$ に対して

$V(X+Y) = V(X) + V(Y)$

B. 統計

B–3. 資料の整理

B-3-1 [**代表値**]

資料の集団的特性を１つの数値で表すものに、総計・平均・メジアン・モードなどがある。このような数値を **代表値** という。

B-3-2 [**資料の形**]

(1) n 個の測定値 $x_1,\ x_2,\ \cdots, x_n$ がそのまま与えられている。

(2) 測定値 $x_1,\ x_2,\ \cdots,\ x_k$ とその度数 $f_1,\ f_2,\ \cdots,\ f_k$ が与えられている。

(3) 階級値 $m_1,\ m_2,\ \cdots,\ m_k$ とその度数 $f_1,\ f_2,\ \cdots,\ f_k$ が与えられている。この場合、個々の測定値は階級値と一般に異なるが、１つの階級に属する測定値はその階級値を取るものとして扱う。

☆注意：[B-3-3] から [B-3-6] の (1) (2) (3) は [B-3-2] の (1) (2) (3) に対応している。

B-3-3 [**資料の総計**]

資料の測定値の総和 T を**総計**という。

(1) $T = x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$

(2) $T = x_1 f_1 + x_2 f_2 + \cdots + x_k f_k = \sum_{i=1}^{k} x_i f_i$

(3) $T = m_1 f_1 + m_2 f_2 + \cdots + m_k f_k = \sum_{i=1}^{k} m_i f_i$

B-3-4 [**資料の平均**]

資料の総計 T を資料の個体の総数 n で割った値を **平均** といい、 $\overline{x}$ で表す。

$\overline{x}=\dfrac{T}{n}$

(1) $\overline{x}=\dfrac{1}{n}(x_1+x_2+\cdots+x_n)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i$

(2) $\overline{x}=\dfrac{1}{n}(x_1f_1+x_2f_2+\cdots+x_kf_k)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{k}x_if_i$

(3) $\overline{x}=\dfrac{1}{n}(m_1f_1+m_2f_2+\cdots+m_kf_k)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{k}m_if_i$

B-3-5 [**資料の分散・標準偏差**]

資料の各個体のとる値 x_i と平均 $\overline{x}$ との差の平方の平均を **分散** といい、 s^2 で表す。

(1) $s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2$

(2) $s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+\cdots+(x_k-\overline{x})^2f_k=\dfrac{1}{n}\displaystyle\sum_{i=1}^{k}(x_i-\overline{x})^2f_i$

(3) $s^2=\dfrac{1}{n}\{(m_1-\overline{x})^2f_1+(m_2-\overline{x})^2f_2+\cdots+(m_k-\overline{x})^2f_k=\dfrac{1}{n}\displaystyle\sum_{i=1}^{k}(m_i-\overline{x})^2f_i$

分散 s^2 の正または 0 の平方根 s を **標準偏差** という。

B-3-6 [**分散の公式**]

(1) $s^2=\dfrac{1}{n}({x_1}^2+{x_2}^2+\cdots+{x_n}^2)-\overline{x}^2$

(2) $s^2=\dfrac{1}{n}({x_1}^2f_1+{x_2}^2f_2+\cdots+{x_k}^2f_k)-\overline{x}^2$

(3) $s^2=\dfrac{1}{n}({m_1}^2f_1+{m_2}^2f_2+\cdots+{m_k}^2f_k)-\overline{x}^2$

B-3-7 [**積和・共分散**]

大きさ n の 2 変量の標本値 $\{(1_x,\ y_1),\ (x_2,\ y_2),\ \cdots(x_n,\ y_n)\}$ に対して

積和：$T_{xy}=x_1y_1+x_2y_2+\cdots x_ny_n=\displaystyle\sum_{i=1}^{n}x_iy_i$

共分散：$s_{xy}=Cov(x,\ y)$

$$\begin{aligned}&=\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots+(x_n-\overline{x})(y_n-\overline{y})\}\\&=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})\\&=\frac{1}{n}T_{xy}-\overline{xy}\end{aligned}$$

B-3-8 [**相関係数**]

2 変量の標本値 $\{(1_x,\ y_1),\ (x_2,\ y_2),\ \cdots(x_n,\ y_n)\}$ に対して

$$r=\frac{s_{xy}}{s_xs_y}=\frac{\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\displaystyle\sum_{i=1}^{n}(y_i-\overline{y})^2}}$$

を **相関係数** という。

B-3-9 [**相関係数の性質**]

相関係数について不等式 $-1 \leqq r \leqq 1$ が成り立つ。 $r = \pm 1$ のときはすべての点 $P_i(x_i,\ y_i)$ が 1 直線上にある。

B-3-10 [**回帰直線**]

y の x への回帰直線の方程式は

$$y - \overline{y} = \frac{s_{xy}}{{s_x}^2}(x - \overline{x})$$

B–4. 正規分布

B-4-1 [**正規分布の平均と分散**]

標準正規分布 $N(0,\ 1)$ の確率変数 Z の平均 $E(Z)$ と分散 $V(Z)$ は

$$E(Z) = 0, \quad V(Z) = 1$$

正規分布 $N(\mu,\ \sigma^2)$ に従う確率変数 X の平均と分散は

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

B-4-2 [**二項分布の正規分布による近似**]

n が十分大きいとき、二項分布 $B(n,\ p)$ は正規分布 $N(np,\ npq)$ で近似される。

B-4-3 [**ポアソン分布の総計、平均、分散**]

ポアソン分布 $P(\mu)$ の確率の総計、平均、分散は

総計： $\displaystyle\sum_{r=0}^{\infty} p_r = 1$

平均： $\displaystyle E(X) = \sum_{r=0}^{\infty} r p_r = \mu$

分散： $\displaystyle V(X) = \sum_{r=0}^{\infty} (r - \mu)^2 p_r = \mu$

B–5. 統計的推定・検定

B-5-1 [**標本平均の平均と分散**]

母集団 Ω の母平均が μ、母分散が σ^2 であるとする。Ω から大きさ n の標本を復元抽出するとき、標本平均 $\overline{X}$ の平均と分散は

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

B-5-2 [**中心極限定理**]

母集団 Ω の母平均が μ 、母分散が σ^2 であるとする。復元抽出による大きさ n の標本平均 $\overline{X}$ の確率分布は、n を十分大きくすれば、正規分布 $N(\mu,\ \dfrac{\sigma^2}{n})$ によって近似される。

B-5-3 [**母平均の推定**]

母集団 Ω が正規分布 $N(\mu,\ \sigma^2)$ に従いその母分散 σ^2 がわかっているとする。Ω から抽出された大きさ n の標本の平均（の実現値）が $\overline{x}$ であるとき、母平均 μ の信頼度 $1 - \alpha$ の信頼区間は

$$\overline{x} - A_\alpha \frac{\sigma}{\sqrt{n}} \leqq \mu \leqq \overline{x} + A_\alpha \frac{\sigma}{\sqrt{n}}$$

である。

B-5-4 [**母平均の検定**]

母集団 Ω が分散 σ^2 の正規分布 $N(\mu,\ \sigma^2)$ に従うことがわかっているとき、母平均 μ について

$$\text{仮説}\quad H_0\ :\ \mu = \mu_0$$

をたてる。Ω からの大きさ n の標本の平均（の実現値）が $\overline{x}$ であったとする。有意水準を α とするとき、$\overline{x}$ が不等式

$$|z| = \frac{\sqrt{n}}{\sigma}|\overline{x} - \mu_0| > A_\alpha$$

を満たすならば、仮説 H_0 を棄却する。そうでなければ仮説 H_0 を棄却しない。

B-5-5 [**標本平均と t 分布**]

正規母集団 $N(\mu,\ \sigma^2)$ から大きさ n $(n > 2)$ の独立な確率変数を $X_1,\ X_2,\ \cdots,\ X_n$、その標本平均を $\overline{X}$、標本分散を S^2 とするとき

$$T_{n-1} = \frac{\sqrt{n-1}(\overline{X} - \mu)}{S}\quad \left(= \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n-1}}}\right)$$

は自由度 $n-1$ の t 分布に従う。

B-5-6 [**母平均の t 推定（母分散がに未知の場合）**]

正規母集団 Ω から抽出された大きさ n の標本の平均（の実現値）が $\overline{x}$ 、分散が s^2 であるとき、母平均 μ の信頼度 $1-\alpha$ の信頼区間は

$$\overline{x} - t_{n-1}(\alpha)\frac{s}{\sqrt{n-1}} \leqq \mu \leqq \overline{x} + t_{n-1}(\alpha)\frac{s}{\sqrt{n-1}}$$

B-5-7 [**母平均の t 検定（母分散が未知の場合）**]

正規母集団 Ω の母平均 μ について

$$\text{仮説}\quad H_0\ :\ \mu = \mu_0$$

をたてる。Ω から大きさ n の標本の平均（の実現値）が $\overline{x}$、分散が s^2 であるとする。有意水準を α として

$$|t| = \frac{\sqrt{n-1}|\overline{x} - \mu_0|}{s} > t_{n-1}(\alpha)$$

ならば、仮説 H_0 は棄却する。