1　記述統計

1.1　数列の基礎

総和記号 $\sum$

$$\sum_{i=1}^{n} x_i = \underbrace{x_1 + x_2 + \cdots + x_n}_{n\text{個}}$$

高校数学で重要であった（人によっては苦労した）

$$\sum_{i=1}^{n} i,\ \sum_{i=1}^{n} i^2,\ \sum_{i=1}^{n} i^3$$

など規則的な数の和は登場しない．とにかく以下で総和記号 $\sum$ に慣れること．

i に依存しない定数の n 個の和

$$\sum_{i=1}^{n} a = \underbrace{a + a + \cdots + a}_{n\text{個}} = na \tag{1}$$

は時折登場する．また以下のような関係（線形性とよばれる）

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{2}$$

$$\sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \tag{3}$$

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i \tag{4}$$

$$\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i \tag{5}$$

も具体的な和を考えると自然に理解できる．例えば，(4) は $n = 2$ として

$$ax_1 + ax_2 = a(x_1 + x_2)$$

のように先に a をかけてから足し算をするのと，足し算をしてから a をかけた結果の等号である．同様に (5) について

$$(ax_1 + by_1) + (ax_2 + by_2) = a(x_1 + x_2) + b(y_1 + y_2)$$

であり，先に $ax + by$ をしてから後で和をとっても，先に和をとってから $ax + by$ しても同じである．このような線形性とよばれる (2) – (5) の性質は，和・差・定数倍と $\sum$ についてその順番を入れ替えることが許されると解釈できる．

一方，和・差・定数倍以外の操作の代表例が二乗である．二乗と $\sum$ の操作を入れ替えたとき

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$$

であることは具体例を考えれば分かる．$n = 2$ として

$$x_1^2 + x_2^2,\ (x_1 + x_2)^2$$

が等しくないのは明らか．例えば，

$$\sum_{i=1}^{n} (x_i + y_i)^2$$

が登場した場合には，まず，項別に二乗を展開

$$= \sum_{i=1}^{n} \left(x_i^2 + 2x_i y_i + y_i^2 \right)$$

して，後に和と $\sum$ の順番を入れ替えて

$$= \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2$$

と分解できる．このような計算手順は頻出する．

最後に統計への例として，数列と平均について．$x_1, \ldots, x_n$ の平均が

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

で定義される．両辺に n をかけると

$$n\bar{x} = \sum_{i=1}^{n} x_i \tag{6}$$

であり，移項して

$$\sum_{i=1}^{n} x_i - n\bar{x} = 0$$

である．左辺について，

$$\sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} (x_i - \bar{x})$$

となる（最初の等号では (1)，２番めの等号では (3) を使う）ので，以下も成り立つ．

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

1.2 線形変換

全ての観測値に対して変換 $x \to a + bx$ を適用する．

$$x_1, x_2 \ldots, x_n \to a + bx_1, a + bx_2, \ldots, a + bx_n$$

このとき，平均と分散は

$$\frac{\sum_{i=1}^{n} (a + bx_i)}{n} = \frac{\sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i}{n} = \frac{na + b\sum_{i=1}^{n} x_i}{n} = \frac{na + bn\bar{x}}{n} = a + b\bar{x} \tag{7}$$

$$\frac{\sum_{i=1}^{n} \{(a + bx_i) - (a + b\bar{x})\}^2}{n} = \frac{\sum_{i=1}^{n} (bx_i - b\bar{x})^2}{n} = \frac{\sum_{i=1}^{n} b^2 (x_i - \bar{x})^2}{n} = b^2 \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = b^2 s^2 \tag{8}$$

1.3 標準化

連立一次方程式

$$\begin{cases} a + b\bar{x} = 0 \\ b^2 s^2 = 1 \end{cases} \tag{9}$$

を解く．解は以下の通り．

$$a = -\frac{\bar{x}}{s}, \quad b = \frac{1}{s} \tag{10}$$

1.4 偏差値

連立一次方程式

$$\begin{cases} a + b\bar{x} = 50 \\ b^2 s^2 = 10^2 \end{cases} \tag{11}$$

を解く．解は以下の通り．

$$a = 50 - 10\frac{\bar{x}}{s}, \quad b = \frac{10}{s} \tag{12}$$

1.5 分散の別表現

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) && \text{項別に 2 乗展開} \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2\bar{x}x_i + \sum_{i=1}^{n}\bar{x}^2 && \text{(2) と (3) の適用} \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 && \text{第二項に (4)，第三項に (1) の適用} \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\{n\bar{x}\} + n\bar{x}^2 && \text{第二項に (6) の適用} \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2
\end{aligned}$$

1.6　共分散の別表現

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}(x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}) \quad \text{項別に展開} \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\bar{y} - \sum_{i=1}^{n}\bar{x}y_i + \sum_{i=1}^{n}\bar{x}\bar{y} \quad (2) \text{ と } (3) \text{ の適用} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y} \quad \text{第二，三項に } (4)\text{，第四項に } (1) \text{ の適用} \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y}(n\bar{x}) - \bar{x}(n\bar{y}) + n\bar{x}\bar{y} \quad \text{第二，三項に } (6) \text{ の適用} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}
\end{aligned}$$

1.7　目的変数の単位の変化と残差平方和

残差平方和の成分 e_i の x と y による具体的表現

$$\begin{aligned}
e_i &= y_i - \hat{y}_i = y_i - (\check{a} + \check{b}x_i) \\
&= \{y_i - \bar{y}\} - \check{b}(x_i - \bar{x}) \\
&= \{y_i - \bar{y}\} - (x_i - \bar{x})\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}$$

$y_i \to cy_i + d$ とするとき

$$\begin{aligned}
\text{標本平均の変化 } \bar{y} &\to c\bar{y} + d, \qquad \text{remember } \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \\
y_i - \bar{y} &\to cy_i - c\bar{y} \qquad \Leftarrow d\text{に依存しない}
\end{aligned}$$

従って

$$\begin{aligned}
e_i &= \{cy_i - c\bar{y}\} - (x_i - \bar{x})\frac{\sum_{i=1}^{n}(x_i - \bar{x})(cy_i - c\bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= c\left(\{y_i - \bar{y}\} - (x_i - \bar{x})\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)
\end{aligned}$$

のように

$$e_i \to ce_i$$

となる．さらに残差平方和 $\sum_{i=1}^{n} e_i^2$ は c^2 倍になることも分かる．

1.8 説明変数の単位の変化と残差平方和

残差 e_i は x を線形変換しても変わらない．なぜなら $x_i \to fx_i + g$ とするとき，

$$\bar{x} \to f\bar{x} + g, \quad x_i - \bar{x} = fx_i - f\bar{x}$$

e_i の x に依存する部分

$$e_i = (fx_i - f\bar{x})\frac{\sum_{i=1}^n (fx_i - f\bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (fx_i - f\bar{x})^2} = (x_i - \bar{x})\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

1.9 R^2 の別表現

（補題１）誤差項の性質

$$\sum_{i=1}^n e_i = 0, \quad \sum_{i=1}^n x_i e_i = 0$$

Proof. 最小二乗推定量 $\check{a}, \check{b}$ は以下を満たすから

$$\frac{\partial}{\partial a} Q(a,b)|_{a=\check{a},b=\check{b}} = 0 \quad \frac{\partial}{\partial b} Q(a,b)|_{a=\check{a},b=\check{b}} = 0$$

$$\frac{\partial}{\partial a} Q(\check{a},\check{b}) = 2\sum(-1)(y_i - \check{a} - \check{b}x_i) = -2\sum e_i$$

$$\frac{\partial}{\partial b} Q(\check{a},\check{b}) = 2\sum(-x_i)(y_i - \check{a} - \check{b}x_i) = -2\sum x_i e_i$$

□

補題１を使って，補題２が示される．

（補題２）$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2$

Proof.

$$\sum_{i=1}^n e_i \hat{y}_i = \sum_{i=1}^n e_i \left(\check{a} + \check{b}x_i\right) = \check{a}\sum_{i=1}^n e_i + \check{b}\sum_{i=1}^n e_i x_i = 0$$

$$\sum_{i=1}^n e_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^n e_i \hat{y}_i - \bar{y}\sum_{i=1}^n e_i = 0$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i + e_i - \bar{y})^2 = \sum_{i=1}^n (\{\hat{y}_i - \bar{y}\} + e_i)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2$$

□

補題２を R^2 の定義式に代入すれば $R^2 = \dfrac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ を得る

1.10　決定係数と相関係数

$R^2 = \dfrac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ における分子の置き換えによる

$$\bar{y} = \frac{\sum_i y_i}{n} = \frac{\sum_i (\hat{y}_i + e_i)}{n} = \frac{\sum \hat{y}_i}{n} = \frac{\sum (\check{a} + \check{b}x_i)}{n} = \check{a} + \check{b}\bar{x}$$
$$\hat{y}_i - \bar{y} = \check{a} + \check{b}x_i - \left(\check{a} + \check{b}\bar{x}\right) = \check{b}(x_i - \bar{x})$$

よって分子 $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ は

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \check{b}^2 \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

これで分子を置き換えれば良い

1.11　最小二乗解が本当に Q を最小にしていること

以下の連立方程式の解 $\check{b}_0, \check{b}_1, \ldots, \check{b}_p$ が Q を最小化していることの証明

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}b_0}Q &= 2\sum_{i=1}^n (-1)(y_i - b_0 - b_1x_{i1} - \cdots - b_px_{ip}) = 0 \\
\frac{\mathrm{d}}{\mathrm{d}b_1}Q &= 2\sum_{i=1}^n (-x_{i1})(y_i - b_0 - b_1x_{i1} - \cdots - b_px_{ip}) = 0 \\
&\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
\frac{\mathrm{d}}{\mathrm{d}b_p}Q &= 2\sum_{i=1}^n (-x_{ip})(y_i - b_0 - b_1x_{i1} - \cdots - b_px_{ip}) = 0
\end{aligned} \tag{$\star$}$$

$\Uparrow$ Q を $b_0, \ldots, b_p$ で（偏）微分した $p+1$ 個をそれぞれ $=0$ とした連立方程式

定理 1. $c_0, \ldots, c_p$ を，$\check{b}_0, \check{b}_1, \ldots, \check{b}_p$ とは違う実数の組として

$$\sum_{i=1}^n (y_i - c_0 - c_1x_{i1} - \cdots - c_px_{ip})^2 \geqslant \sum_{i=1}^n \{y_i - \check{b}_0 - \check{b}_1x_{i1} - \cdots - \check{b}_px_{ip}\}^2$$

$\Uparrow$ $p = 1$ の場合が単回帰分析

Proof. 各項において $\check{b}_0 + \check{b}_1x_{i1} + \cdots + \check{b}_px_{ip}$ を引いて足す（差し引きゼロ）

$$\begin{aligned}
&\sum_{i=1}^n (y_i - c_0 - c_1x_{i1} - \cdots - c_px_{ip})^2 \\
&= \sum_{i=1}^n \left\{y_i - \check{b}_0 - \check{b}_1x_{i1} - \cdots - \check{b}_px_{ip} + (\check{b}_0 - c_0) + (\check{b}_1 - c_1)x_{i1} + \cdots + (\check{b}_p - c_p)x_{ip}\right\}^2
\end{aligned}$$

A_i と B_i を以下のように定める．

$$\begin{aligned}
A_i &= y_i - \check{b}_0 - \check{b}_1x_{i1} - \cdots - \check{b}_px_{ip} \\
B_i &= (\check{b}_0 - c_0) + (\check{b}_1 - c_1)x_{i1} + \cdots + (\check{b}_p - c_p)x_{ip}
\end{aligned}$$

このとき 2 乗の展開

$$\sum_{i=1}^{n}(A_i+B_i)^2=\sum_{i=1}^{n}A_i^2+\sum_{i=1}^{n}B_i^2+2\sum_{i=1}^{n}A_iB_i$$

に対応して以下を得る．

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-c_0-c_1x_{i1}-\cdots-c_px_{ip})^2 &= \sum_{i=1}^{n}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\}^2 \\
&\quad+\sum_{i=1}^{n}\{(\check{b}_0-c_0)+(\check{b}_1-c_1)x_{i1}+\cdots+(\check{b}_p-c_p)x_{ip}\}^2 \\
&\quad+2(\check{b}_0-c_0)\sum_{i=1}^{n}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\} \\
&\quad+2(\check{b}_1-c_1)\sum_{i=1}^{n}x_{i1}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\} \\
&\quad+\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad+2(\check{b}_p-c_p)\sum_{i=1}^{n}x_{ip}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\}
\end{aligned}$$

ここで，$\sum_{i=1}^{n}A_iB_i$ に由来する以下の具体的な表現

$$\begin{cases}
\sum_{i=1}^{n}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\} \\
\sum_{i=1}^{n}x_{i1}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\sum_{i=1}^{n}x_{ip}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\}
\end{cases}$$

は連立方程式 ($\star$) の左辺に $\check{b}_0,\check{b}_1,\ldots,\check{b}_p$ をプラグインした形である．$\check{b}_0,\check{b}_1,\ldots,\check{b}_p$ は連立方程式の解なので全て 0．従って，

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-c_0-c_1x_{i1}-\cdots-c_px_{ip})^2 &= \sum_{i=1}^{n}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\}^2 \\
&\quad+\sum_{i=1}^{n}\{(\check{b}_0-c_0)+(\check{b}_1-c_1)x_{i1}+\cdots+(\check{b}_p-c_p)x_{ip}\}^2
\end{aligned}$$

である．右辺 2 項目は非負なので

$$\sum_{i=1}^{n}(y_i-c_0-c_1x_{i1}-\cdots-c_px_{ip})^2 \geqslant \sum_{i=1}^{n}\{y_i-\check{b}_0-\check{b}_1x_{i1}-\cdots-\check{b}_px_{ip}\}^2$$

□

1.12 単回帰分析における最小化

$$
\begin{aligned}
Q(a,b) &= \sum_{i=1}^{n} |y_i - (a + bx_i)|^2 \\
&= \sum_{i=1}^{n} \{y_i - (a + bx_i)\}^2 \\
&= \sum_{i=1}^{n} \{y_i - \bar{y} + \bar{y} - a - b\bar{x} + b\bar{x} - bx_i)\}^2 \\
&= \sum_{i=1}^{n} \{(y_i - \bar{y}) + (\bar{y} - a - b\bar{x}) + (b\bar{x} - bx_i)\}^2 \\
&= \sum_{i=1}^{n} \{(y_i - \bar{y}) + (\bar{y} - a - b\bar{x}) - b(x_i - \bar{x})\}^2 \\
&= \sum_{i=1}^{n} \{(y_i - \bar{y})^2 + (\bar{y} - a - b\bar{x})^2 + b^2(x_i - \bar{x})^2 \\
&\quad +2(y_i - \bar{y})(\bar{y} - a - b\bar{x}) - 2b(y_i - \bar{y})(x_i - \bar{x}) - 2b(\bar{y} - a - b\bar{x})(x_i - \bar{x})\} \\
&= ns_{yy} + n(\bar{y} - a - b\bar{x})^2 + b^2 ns_{xx} - 2bs_{xy} \\
&= ns_{yy} + n(\bar{y} - a - b\bar{x})^2 + ns_{xx}\left(b - \frac{s_{xy}}{s_{xx}}\right)^2 - n\frac{s_{xy}^2}{s_{xx}} \\
&= n\left\{s_{yy}\left(1 - \frac{s_{xy}^2}{s_{xx}s_{yy}}\right) + (\bar{y} - a - b\bar{x})^2 + s_{xx}\left(b - \frac{s_{xy}}{s_{xx}}\right)^2\right\}
\end{aligned}
$$

1.13 重回帰分析の R^2

最小二乗解 $\check{b}_0, \ldots, \check{b}_p$ は，以下を満たす

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}b_0}Q &= 2\sum_{i=1}^{n}(-1)(y_i - b_0 - b_1x_{i1} - \cdots - b_px_{ip}) = 0 \\
\frac{\mathrm{d}}{\mathrm{d}b_1}Q &= 2\sum_{i=1}^{n}(-x_{i1})(y_i - b_0 - b_1x_{i1} - \cdots - b_px_{ip}) = 0 \\
&\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
\frac{\mathrm{d}}{\mathrm{d}b_p}Q &= 2\sum_{i=1}^{n}(-x_{ip})(y_i - b_0 - b_1x_{i1} - \cdots - b_px_{ip}) = 0
\end{aligned}
$$

書きなおすと

$$
\sum e_i = 0, \sum x_{i1}e_i = 0, \ldots, \sum x_{ip}e_i = 0
$$

誘導される性質

$$\sum_{i=1}^n e_i\hat{y}_i = \sum_{i=1}^n e_i\left(\check{b}_0 + \check{b}_1 x_{i1} + \cdots + \check{b}_p x_{ip}\right) = \check{b}_0\sum_{i=1}^n e_i + \check{b}_1\sum_{i=1}^n e_i x_{i1} + \cdots + \check{b}_p\sum_{i=1}^n e_i x_{ip} = 0$$

$$\sum_{i=1}^n e_i(\hat{y}_i - \bar{y}) = \sum_{i=1}^n e_i\hat{y}_i - \bar{y}\sum_{i=1}^n e_i = 0$$

$$\begin{aligned}\sum_{i=1}^n (y_i-\bar{y})^2 &= \sum_{i=1}^n(\hat{y}_i + e_i - \bar{y})^2 = \sum_{i=1}^n(\hat{y}_i - \bar{y} + e_i)^2 = \sum_{i=1}^n(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^n e_i^2 + 2\sum_{i=1}^n(\hat{y}_i-\bar{y})e_i \\ &= \sum_{i=1}^n(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^n e_i^2\end{aligned}$$

R^2 の定義式に代入すれば $R^2 = \dfrac{\sum_{i=1}^n(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^n(y_i-\bar{y})^2}$ を得る

2　確率

2.1　同時確率密度関数

$a < b$，$c < d$ に対し，連続確率変数の場合に

$$P(a < X \leqslant b, c < Y \leqslant d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

が従う．なぜなら

$$\begin{aligned}&P(a < X \leqslant b, c < Y \leqslant d) \\ &= P(a < X \leqslant b, Y \leqslant d) - P(a < X \leqslant b, Y \leqslant c) \\ &= \{P(X \leqslant b, Y \leqslant d) - P(X \leqslant a, Y \leqslant d)\} \\ &\quad - \{P(X \leqslant b, Y \leqslant c) - P(X \leqslant a, Y \leqslant c)\} \\ &= F(b,d) - F(a,d) - F(b,c) + F(a,c)\end{aligned}$$

- , は and つまり積集合 $\cap$
- 最初の等号
  - 事象 $G_1 : a < X \leqslant b$, $G_2 : Y \leqslant c$, $G_3 : c < Y \leqslant d$,
  - $G_2 \cap G_3 = \varnothing$ なので $P(G_1 \cap G_3) + P(G_1 \cap G_2) = P(G_1 \cap (G_2 \cup G_3))$
- ２番目も同様

$b = a + \Delta a$, $d = c + \Delta c$ として，

$$\begin{aligned}&P(a < X \leqslant a + \Delta a, c < Y \leqslant c + \Delta c) \\ &= F(a+\Delta a, c+\Delta c) - F(a, c+\Delta c) - F(a+\Delta a, c) + F(a,c) \\ &= \Delta a\left\{\frac{F(a+\Delta a, c+\Delta c) - F(a, c+\Delta c)}{\Delta a} - \frac{F(a+\Delta a, c) - F(a,c)}{\Delta a}\right\} \\ &= \underbrace{\Delta a \Delta c}_{\text{底面積（長方形）}}\left\{\underbrace{\frac{\tilde{F}(c+\Delta c; a, \Delta a) - \tilde{F}(c; a, \Delta a)}{\Delta c}}_{\to \frac{\partial^2}{\partial x \partial y}F(x,y)|_{x=a,y=c} \text{ as } \Delta a, \Delta c \to 0}\right\}\end{aligned}$$

where

$$\tilde{F}(c; a, \Delta a) = \frac{F(a + \Delta a, c) - F(a, c)}{\Delta a}$$

2.2 和の期待値と期待値の和

$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$ の確認（多次元離散確率分布の場合）．

$$\begin{aligned}
\mathrm{E}[X+Y] &= \sum_{(x,y)} (x+y) \underbrace{p(x,y)}_{\text{同時確率関数}} = \sum_{(x,y)} xp(x,y) + \sum_{(x,y)} yp(x,y) \\
&= \sum_x x\{\sum_y p(x,y)\} + \sum_x y\{\sum_x p(x,y)\} \\
&= \sum_x xp(x) + \sum_y yp(y) = \mathrm{E}[X] + \mathrm{E}[Y]
\end{aligned}$$

第一の等号は定義．第二の等号は和の性質．第四の等号は周辺確率関数の定義．

2.3 期待値の線形和と線形和の期待値

X を離散確率変数とする

$$\begin{aligned}
\mathrm{E}[a+bX] &= \sum_x (a+bx)p(x) = \sum_x ap(x) + \sum_x bxp(x) = a \underbrace{\sum_x p(x)}_{1} + b \underbrace{\sum_x xp(x)}_{\mathrm{E}[X]} \\
&= a + b\mathrm{E}[X]
\end{aligned}$$

第一の等号は定義．第二の等号は和の性質．以上は離散なので $\sum$ が登場．連続ならば $\int$ を使った証明となる．

分散

$$\begin{aligned}
\mathrm{Var}[a+bX] &= \mathrm{E}\left[\{a+bX-\mathrm{E}[a+bX]\}^2\right] = \mathrm{E}\left[\{a+bX-(a+b\mathrm{E}[X])\}^2\right] \\
&= \mathrm{E}\left[(bX-b\mathrm{E}[X])^2\right] = \mathrm{E}\left[b^2(X-\mathrm{E}[X])^2\right] = b^2\mathrm{E}\left[(X-\mathrm{E}[X])^2\right] = b^2\mathrm{Var}[X]
\end{aligned}$$

第一の等号は定義．第ニの等号は上で求めた $\mathrm{E}[a+bX] = a + b\mathrm{E}[X]$ より．

2.4 多項分布

講義では，一般論として「離散」「二次元」確率変数を学んだ．しかし，少し一般化して「離散」「多次元」確率変数をすることはそれほど難しくない．

将棋の駒（７面体）を投げると，wikipedia のまわり将棋のページの真ん中付近の５枚の写真（横向きと下向きはそれぞれ２通りある）のようなイベントが観測できる．「まわり将棋」は，４枚の「金将」を投げるが，ここでは n 回１枚の金将を投げて，７通りのイベント（表，横 1，横 2，上，下 1，下 2，裏）がそれぞれ何回起こるかを観察すると多項分布（７項分布）になる．例えば，それぞれの確率を

1. p_1 上
2. p_2 表
3. p_3 横 1

4. p_4 下 1
5. p_5 裏
6. p_6 横 2
7. p_7 下 2

（ただし，$\sum_{i=1}^{7} p_i = 1$）として，それぞれ $x_1, x_2, \ldots, x_7$ 回イベントが起こるとすると，その同時確率関数は

$$p(x_1, \ldots, x_7) = \frac{n!}{x_1! \ldots x_7!} p_1^{x_1} \ldots p_7^{x_7}, \text{ただし} \sum_{i=1}^{7} x_i = n \tag{13}$$

となる．$p(x_1, \ldots, x_7)$ と 7 変数の同時確率関数となっているが，実際には $\sum_{i=1}^{7} x_i = n$ の制約があるので，6 変数の同時確率関数である．また，$\sum_{i=1}^{7} p_i = 1$ の関係もある．従って，$p_1, \ldots, p_6$ を 6 つのパラメータとして

$$p(x_1, \ldots, x_6) = \frac{n!}{x_1! \ldots (n - \sum_{i=1}^{6} x_i)!} p_1^{x_1} \ldots p_6^{x_6} \left(1 - \sum_{i=1}^{6}\right)^{n - \sum_{i=1}^{6} x_i} \tag{14}$$

と書くのが正確かもしれない．しかし対称性を重視した (13) の方がよく用いられる．

(13)，(14) が分かりにくければ，二項分布を考えれば良い．コインの表と裏の 2 つのイベント，表が出る確率が p．そのようなコインを n 回投げる．表の出た回数が二項分布に従う．対称性を重視した記述であれば，表の出た回数 x，裏の出た回数 y に対して，

$$p(x, y) = \frac{n!}{x!y!} p^x q^y, \text{ただし} x + y = n, \quad p + q = 1 \tag{15}$$

であるが，二項分布は表が出た回数のみに注目して（裏の出た回数は【$n-$ 表の回数】なので）

$$p(x) = \frac{n!}{x!(n-x)!} p^x \ldots (1-p)^{n-x} \tag{16}$$

を確率関数とする．

2.4.1　より単純化したモデル

見た目による考察から，

$$p_2 = p_5,\ p_3 = p_6,\ p_4 = p_7$$

として，より単純化したモデル（p_4 は制約から決まるので，3 つのパラメータ p_1, p_2, p_3 による統計モデル）を考えることも出来る．またデータに基づいてどちらの方が尤もらしいか判断することも可能である．

2.5　ポアソン分布の期待値と分散

$\mathrm{E}[X] = \lambda$

$$\begin{aligned}\mathrm{E}[X] &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda\end{aligned}$$

$\mathrm{Var}[X] = \lambda$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \{\mathrm{E}[X]\}^2 = \mathrm{E}[X(X-1)] + \mathrm{E}[X] - \{\mathrm{E}[X]\}^2$$

に注意．以下の $\mathrm{E}[X(X-1)] = \lambda^2$ から $\mathrm{Var}[X] = \lambda$ が従う

$$\begin{aligned}
\mathrm{E}[X(X-1)] &= \textstyle\sum_{x=0}^{\infty} x(x-1)\dfrac{e^{-\lambda}\lambda^x}{x!} = \textstyle\sum_{x=2}^{\infty} x(x-1)\dfrac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda^2 e^{-\lambda} \textstyle\sum_{x=2}^{\infty} \dfrac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda} \textstyle\sum_{y=0}^{\infty} \dfrac{\lambda^y}{y!} = \lambda^2
\end{aligned}$$

2.6 標準正規分布の分散

$\phi'(x) = -x\phi(x)$

- （数学 III の内容）$\{d/dx\}e^x = e^x$
- chain rule $\phi'(x) = \dfrac{1}{(2\pi)^{1/2}}\left\{\exp\left(-\dfrac{x^2}{2}\right)\right\} \times \left\{-\dfrac{x^2}{2}\right\}'$
  $\Rightarrow \phi'(x) = -x\phi(x)$

部分積分を用いると

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x-0)^2\phi(x)\mathrm{d}x = \int_{-\infty}^{\infty} x^2\phi(x)\mathrm{d}x \\
&= -\int_{-\infty}^{\infty} x\phi'(x)\mathrm{d}x \\
&= [-x\phi(x)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \phi(x)\mathrm{d}x = 1
\end{aligned}$$

- $\displaystyle\lim_{x\to\pm\infty} x\phi(x) = \lim_{x\to\pm\infty} \frac{\sqrt{2\pi}x}{e^{x^2/2}} = 0$
- $\displaystyle\int_{-\infty}^{\infty} \phi(x)\mathrm{d}x = 1$，$\phi(x)$ は確率密度関数

2.7 たたみ込みの確率密度関数の導出と具体的な計算例

- 定理
  - X と Y は独立な連続確率変数
  - 確率密度関数はそれぞれ $f(x)$ と $g(y)$
    $\Rightarrow X+Y$ の確率密度関数 $\displaystyle\int_{-\infty}^{\infty} f(x)g(z-x)\mathrm{d}x$
- 証明
  - 累積分布関数 $P(X+Y \leqslant z)$ の z に関する微分

– X と Y の同時密度関数を $x+y \leqslant z$ の領域で積分

$$\begin{aligned} P(X+Y \leqslant z) &= \iint_{\{(x,y):x+y\leqslant z\}} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \iint_{\{(x,y):x+y\leqslant z\}} f(x)g(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} g(y)\mathrm{d}y\right) f(x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} g(u-x)\mathrm{d}u\right) f(x)\mathrm{d}x \\ &= \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f(x)g(u-x)\mathrm{d}x\right) \mathrm{d}u \end{aligned}$$

1. 定義
2. X と Y の独立性より $f_{X,Y}(x,y) = f(x)g(y)$
3. 以下の関係及び累次積分「y の積分」の後「x の積分」

$$\{(x,y): x+y \leqslant z\} \;\Leftrightarrow\; \left\{(x,y): \begin{matrix} -\infty < x < \infty \\ -\infty < y \leqslant z-x \end{matrix}\right\}$$

4. 置換積分 $y = u - x$
5. 累次積分で順序を交換．「x の積分」の後「u の積分」

– 累積分布関数 $P(X+Y \leqslant z)$ の z に関する微分は，

$$f_Z(z) = \int_{-\infty}^{\infty} f(x)g(z-x)\mathrm{d}x$$

2.7.1 正規分布 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$

$$f(x;\mu_1,\sigma_1^2) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x-\mu_1)^2}{\sigma_1^2}\right),\; g(z-x;\mu_2,\sigma_2^2) = \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(z-x-\mu_1)^2}{\sigma_2^2}\right)$$

exp の中身に -2 をかけた量

$$\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(z-x-\mu_2)^2}{\sigma_2^2}$$

を x について平方完成

$$
\begin{aligned}
&\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(z-x-\mu_2)^2}{\sigma_2^2}\\
&=\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(x-z+\mu_2)^2}{\sigma_2^2}\\
&=\left(1/\sigma_1^2+1/\sigma_2^2\right)(x-\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\})^2\\
&\quad-\left(1/\sigma_1^2+1/\sigma_2^2\right)\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\}^2+\frac{\mu_1^2}{\sigma_1^2}+\frac{(z-\mu_2)^2}{\sigma_2^2}\\
&=\left(1/\sigma_1^2+1/\sigma_2^2\right)(x-\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\})^2\\
&\quad-\left(1/\sigma_1^2+1/\sigma_2^2\right)\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\}^2+\left(1/\sigma_1^2+1/\sigma_2^2\right)\{\alpha\mu_1^2+(1-\alpha)(z-\mu_2)^2\}\\
&=\left(1/\sigma_1^2+1/\sigma_2^2\right)(x-\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\})^2+\left(1/\sigma_1^2+1/\sigma_2^2\right)\alpha(1-\alpha)\{\mu_1-(z-\mu_2)\}^2\\
&=\left(1/\sigma_1^2+1/\sigma_2^2\right)(x-\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\})^2+\frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}
\end{aligned}
$$

ただし，

$$\alpha=\frac{1/\sigma_1^2}{1/\sigma_1^2+1/\sigma_2^2}$$

である．従って，

$$
\begin{aligned}
f(x;\mu_1,\sigma_1^2)g(z-x;\mu_2,\sigma_2^2)=\frac{1}{\sqrt{2\pi\sigma_1^2}}\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(x-\{\alpha\mu_1+(1-\alpha)(z-\mu_2)\})^2}{2\left(1/\sigma_1^2+1/\sigma_2^2\right)^{-1}}\right)\\
\times\exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)
\end{aligned}
$$

ところで，ガウス積分より，μ_3 が何であれ，

$$\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu_3)^2}{2\left(1/\sigma_1^2+1/\sigma_2^2\right)^{-1}}\right)\mathrm{d}x=\sqrt{2\pi}\sqrt{\left(1/\sigma_1^2+1/\sigma_2^2\right)^{-1}}=\sqrt{2\pi}\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}$$

よって，

$$\int_{-\infty}^{\infty}f(x;\mu_1,\sigma_1^2)g(z-x;\mu_2,\sigma_2^2)\mathrm{d}x=\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)$$

であり，これは $N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$ の確率密度関数

2.7.2 ガンマ分布 $X\sim\mathrm{Ga}(a,\tau)$，$Y\sim\mathrm{Ga}(b,\tau)$

$$f(x;a,\tau)=\frac{x^{a-1}e^{-x/\tau}}{\Gamma(a)\tau^a}I_{\{x:x>0\}}(x),\ g(y;b,\tau)=\frac{y^{b-1}e^{-y/\tau}}{\Gamma(b)\tau^b}I_{\{y:y>0\}}(y)$$

たたみこみの公式より，

$$\int_{-\infty}^{\infty}f(x;a,\tau)g(z-x;b,\tau)\mathrm{d}x$$

を考えるが，$x \notin (0,z)$ では，$f(x;a,\tau)g(z-x;b,\tau)=0$ であるので，積分範囲は実質的に $0<x<z$ であることに注意

$$\int_0^z f(x;a,\tau)g(z-x;b,\tau)\mathrm{d}x$$

ここで，

$$f(x;a,\tau)g(z-x;b,\tau) = \frac{x^{a-1}e^{-x/\tau}}{\Gamma(a)\tau^a}\frac{(z-x)^{b-1}e^{-(z-x)/\tau}}{\Gamma(b)\tau^b} = \frac{x^{a-1}(z-x)^{b-1}e^{-z/\tau}}{\Gamma(a)\tau^a\Gamma(b)\tau^b}$$

である．従って，

$$\int_0^z x^{a-1}(z-x)^{b-1}\mathrm{d}x = z^{a+b-1}\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t = z^{a+b-1}B(a,b) = z^{a+b-1}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

（最初の等号は変数変換 $x=zt$，二番目はベータ関数の定義，三番目はガンマ関数とベータ関数の有名な関係）であるので，

$$\int_{-\infty}^{\infty} f(x;a,\tau)g(z-x;b,\tau)\mathrm{d}x = \frac{z^{a+b-1}e^{-z/\tau}}{\Gamma(a+b)\tau^{a+b}}I_{\{z:z>0\}}(z)$$

これは $\mathrm{Ga}(a+b,\tau)$ の密度関数

2.8 $X \sim N(0,1)$ のときの X^2 の分布が $\mathrm{Ga}(1/2,2)$

累積分布関数 $P(X^2 \leqslant v)$

$$\begin{aligned} P(X^2 \leqslant v) &= P(-\sqrt{v} \leqslant X \leqslant \sqrt{v}) \\ &= \Phi(\sqrt{v}) - \Phi(-\sqrt{v}) \\ &= \Phi(\sqrt{v}) - \{1-\Phi(\sqrt{v})\} \\ &= 2\Phi(\sqrt{v}) - 1 \end{aligned}$$

X^2 の確率密度関数は，$\dfrac{\mathrm{d}}{\mathrm{d}v}P(X^2 \leqslant v)$ で与えられる．

$$2\phi(v^{1/2}) \times \frac{1}{2}v^{1/2-1} = \frac{1}{\sqrt{2\pi}}\exp(-v/2)v^{1/2-1} = \frac{v^{1/2-1}\exp(-v/2)}{\Gamma(1/2)2^{1/2}}$$

ただし，$\Gamma(1/2)=\sqrt{\pi}$ を使った

2.9 t 分布の確率密度関数の導出

- 定義

$$T = \frac{X}{\sqrt{V/m}}$$

（ここで $X \sim N(0,1)$ 標準正規分布，$V \sim \chi^2_m$ 自由度 m のカイ二乗分布，X と V は独立）の従う分布

- X と V は独立なのでその同時確率密度関数は

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \times \frac{1}{\Gamma(m/2)2^{m/2}}v^{m/2-1}e^{-v/2}$$

- 累積分布関数 $P(T \leqslant t)$

$$
\begin{aligned}
P(T \leqslant t) &= P\left(\frac{X}{\sqrt{V/m}} \leqslant t\right) \\
&= \iint_{\{(x,v):x/\sqrt{v/m}\leqslant t\}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\Gamma(m/2)2^{m/2}} v^{m/2-1} e^{-v/2} \mathrm{d}x\mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(m/2)2^{m/2}} \int_0^\infty \left(\int_{-\infty}^{\sqrt{v/m}t} e^{-x^2/2}\mathrm{d}x\right) v^{m/2-1} e^{-v/2}\mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(m/2)2^{m/2}} \int_0^\infty \frac{\sqrt{v}}{\sqrt{m}} \left(\int_{-\infty}^{t} e^{-vw^2/(2m)}\mathrm{d}w\right) v^{m/2-1} e^{-v/2}\mathrm{d}v \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(m/2)2^{m/2}} \frac{1}{\sqrt{m}} \int_{-\infty}^{t} \left(\int_0^\infty v^{(m+1)/2-1} e^{-v(1+w^2/m)/2}\mathrm{d}v\right)\mathrm{d}w \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(m/2)2^{m/2}} \frac{1}{\sqrt{m}} \int_{-\infty}^{t} \Gamma((m+1)/2)2^{(m+1)/2}\left(1+w^2/m\right)^{-(m+1)/2}\mathrm{d}w \\
&= \int_{-\infty}^{t} \left(\frac{\Gamma((m+1)/2)}{\sqrt{\pi m}\Gamma(m/2)}\left(1+w^2/m\right)^{-(m+1)/2}\right)\mathrm{d}w
\end{aligned}
$$

  等号
  1. 定義
  2. 同時確率密度関数を確率を計算したい領域で積分
  3. 累次積分「x で積分」の後「v で積分」
  4. 置換積分：$x = (\sqrt{v}/\sqrt{m})w$ と変数変換．ヤコビアン $\sqrt{v}/\sqrt{m}$．$w \in (-\infty, t)$
  5. 積分の順序を交換．「v で積分」の後「w で積分」
  6. ガンマ分布やガンマ関数を思い出す

$$
\int_0^\infty x^{\alpha-1} e^{-x/\beta}\mathrm{d}x = \Gamma(\alpha)\beta^\alpha
$$

  7. 約分して整理

- 累積分布関数 $P(T \leqslant t)$ の t に関する微分が確率密度関数

$$
f_T(t) = \frac{\Gamma((m+1)/2)}{\sqrt{\pi m}\Gamma(m/2)}\left(1+t^2/m\right)^{-(m+1)/2}
$$

2.10 大数の法則の証明

$X_1, \ldots, X_n, \ldots,$ を独立同一分布に従う確率変数（証明の簡単のため連続確率変数とするが，離散でも同じ）の無限列とし，また $\mu = \mathrm{E}[X_i]$，$\sigma^2 = \mathrm{Var}[X_i]$ とする．

まずチェビシェフの不等式より

$$
P(|Z - \mathrm{E}[Z]| \geqslant c) \leqslant \frac{\mathrm{Var}[Z]}{c^2}
$$

であることを思い出す．

特に $Z = \bar{X}_n$ の場合を考える．このとき $\mathrm{E}[\bar{X}_n] = \mu$，$\mathrm{Var}[\bar{X}_n] = \sigma^2/n$ である．

任意の $\epsilon > 0$ に対し，チェビシェフの不等式（2.11 節）より以下が成り立つ

$$\begin{aligned} P(|\bar{X}_n - \mu| \geqslant \epsilon) \leqslant \frac{\sigma^2}{n\epsilon^2} &\Leftrightarrow 1 - P(|\bar{X}_n - \mu| < \epsilon) \leqslant \frac{\sigma^2}{n\epsilon^2} \\ &\Leftrightarrow P(|\bar{X}_n - \mu| \leqslant \epsilon) \geqslant 1 - \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

ところで $0 < \forall\eta < 1$ に対し，$N(\epsilon, \eta) = \left[\frac{\sigma^2}{\epsilon^2\eta}\right] + 1$（[] はガウス記号）と定める．このとき $n \geqslant n(\epsilon, \eta)$ を満たす全ての n に対し，

$$\begin{aligned} P(|\bar{X}_n - \mu| \leqslant \epsilon) &\geqslant 1 - \frac{\sigma^2}{n\epsilon^2} \geqslant 1 - \frac{\sigma^2}{N(\epsilon, \eta)\epsilon^2} \\ &\geqslant 1 - \frac{\sigma^2}{\frac{\sigma^2}{\epsilon^2\eta}\epsilon^2} = 1 - \eta \end{aligned}$$

2.11 チェビシェフの不等式の証明

$$\begin{aligned} \mathrm{Var}[Z] &= \int_{-\infty}^{\infty} (z-\mu)^2 f(z)\mathrm{d}z \quad \mu = \mathrm{E}[Z] \\ &= \left(\int_{\mu-c}^{\mu+c} + \int_{z \notin (\mu-c, \mu+c)}\right)(z-\mu)^2 f(z)\mathrm{d}z \\ &\geqslant \int_{z-\mu \notin (-c,c)} (z-\mu)^2 f(z)\mathrm{d}z \\ &\geqslant \int_{z-\mu \notin (-c,c)} c^2 f(z)\mathrm{d}z \\ &= c^2 \int_{z-\mu \notin (-c,c)} f(z)\mathrm{d}z = c^2 P(|Z-\mu| \geqslant c) \end{aligned}$$

3 推測統計

3.1 標本不偏分散

- $X_1, \ldots, X_n$ の関数で

$$\mathrm{E}[T(X_1, \ldots, X_n)] = \sigma^2$$

  となるような $T(X_1, \ldots, X_n)$ はあるのか？
- 記述統計の標本分散 $\sum_{i=1}^n (x_i - \bar{x})^2/n$ の形が参考になる

$$\mathrm{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = (n-1)\sigma^2$$

  - 復習 $\mathrm{E}[aZ+b] = a\mathrm{E}[Z] + b$ 期待値の線形性
  - 復習 $\mathrm{Var}[aZ+b] = a^2\mathrm{Var}[Z]$，$\mathrm{Var}[Z] = \mathrm{E}[Z^2] - \{\mathrm{E}[Z]\}^2$
  - 仮定の再掲 $\mathrm{E}[X_i] = \mu$，$\mathrm{Var}[X_i] = \sigma^2$

– 復習 $\mathrm{E}[\bar{X}] = \mu$

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i] = \frac{n\mu}{n} = \mu$$

２番目３番目の等号は期待値の線形性，４番目の等号は独立「同一」分布

– 復習 $\mathrm{Var}[\bar{X}] = \sigma^2/n$

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{\mathrm{Var}\left[\sum_{i=1}^n X_i\right]}{n^2} = \frac{\sum_{i=1}^n \mathrm{Var}\left[X_i\right]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

２番目の等号は公式 $\mathrm{Var}[aZ+b] = a^2\mathrm{Var}[Z]$，３番目の等号は「独立」同一分布なので「和の分散 = 分散の和」，４番目は独立「同一」分布

– $\mathrm{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \mathrm{E}\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] = \sum_{i=1}^n \mathrm{E}\left[X_i^2\right] - n\mathrm{E}\left[\bar{X}^2\right]$

$$= n\mathrm{E}[X_i^2] - n\mathrm{E}[\bar{X}^2] = n\left(\sigma^2 + \mu^2 - \left\{\frac{\sigma^2}{n} + \mu^2\right\}\right) = (n-1)\sigma^2$$

２番目の等号は期待値の線形性，３番目の等号は独立「同一」分布であること，４番目の等号は以下より従う

$$\mathrm{E}[\bar{X}^2] = \mathrm{Var}[\bar{X}] + \{\mathrm{E}[\bar{X}]\}^2 = \sigma^2/n + \mu^2$$
$$\mathrm{E}[X_i^2] = \mathrm{Var}[X_i] + \{\mathrm{E}[X_i]\}^2 = \sigma^2 + \mu^2$$

- 従って $\mathrm{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] = \sigma^2$
- $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ は不偏分散と呼ばれる

3.2 推定論：情報量不等式を使った具体例

3.2.1 二項分布

ベルヌーイ分布から考える．$X_1, \ldots, X_n$ が独立同一にベルヌーイ分布に従うときの $X = Y_1 + \cdots + Y_n$ の分布が二項分布．$\mathrm{E}[Y_i] = p$, $\mathrm{Var}[Y_i] = \mathrm{E}[(Y_i - p)^2] = p(1-p)$ より，

$$\mathrm{E}[\frac{\sum_{i=1}^n Y_i}{n}] = p,$$
$$\mathrm{Var}[\frac{\sum_{i=1}^n Y_i}{n}] = \frac{\mathrm{Var}[\sum_{i=1}^n Y_i]}{n^2} = \frac{\sum_{i=1}^n \mathrm{Var}[X_i]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

ところで，ベルヌーイ分布の確率関数は

$$f(y,p) = p^y(1-p)^{1-y},\ y = 0, 1$$

である．$\log f(y,p) = y\log p + (1-y)\log(1-p)$ であり，また

$$\frac{\mathrm{d}}{\mathrm{d}p}\log f(y,p) = \frac{y}{p} - \frac{1-y}{1-p} = \frac{y-p}{p(1-p)}$$

である．ただし，合成関数の微分を使って

$$\frac{\mathrm{d}}{\mathrm{d}p}\log(1-p) = \frac{1}{1-p} \times \frac{\mathrm{d}}{\mathrm{d}p}(1-p) = -\frac{1}{1-p}$$

を得た．従って，フィッシャー情報量

$$I(p) = \mathrm{E}\left[\left(\frac{Y-p}{p(1-p)}\right)^2\right] = \frac{\mathrm{E}[(Y-p)^2]}{p^2(1-p)^2} = \frac{1}{p(1-p)}$$

であり，不偏推定量の分散の下限は $\dfrac{1}{nI(p)} = \dfrac{p(1-p)}{n}$．これは $\dfrac{X}{n}$ の分散に等しい．

3.2.2 正規分布

$X_1, \ldots, X_n$ を正規分布 $N(\mu, 1)$ からの $i.i.d.$ 標本とする．μ の推定量 $\bar{X}$ が UMVUE であることを示したい．$\mathrm{E}[X] = \mu$，$\mathrm{Var}[X] = \sigma^2$ なる母集団からの i.i.d. 標本の特殊ケースである．分布が「正規分布」，$\sigma^2 = 1$ である．まず $\mathrm{Var}[\bar{X}] = \dfrac{\sigma^2}{n} = \dfrac{1}{n}$ に注意する．

正規分布 $N(\mu, 1)$ の確率密度関数の対数は

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

であり，その μ に関する微分は

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\log\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2}\right)\right) = x - \mu$$

で得られる．フィッシャー情報量は二乗の期待値であり，

$$I(\mu) = \mathrm{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\mu}\log\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2}\right)\right)\right)^2\right] = \mathrm{E}[(X-\mu)^2] = 1$$

で得られる．よって，μ の不偏推定量の分散の下限は $\dfrac{1}{nI(\mu)} = \dfrac{1}{n}$ で与えられる．これは，$\bar{X}$ の分散に等しい．

3.3 最尤推定量の具体例

3.3.1 二項分布 or ベルヌーイ分布

ベルヌーイ分布の

$$\text{確率関数} \quad f(y,p) = p^y(1-p)^{1-y},\ y = 0, 1$$

は $Y_1, \ldots, Y_n$ の同時確率関数は $x = \sum_{i=1}^n y_i$ としたとき

$$\prod_{i=1}^n p^{y_i}(1-p)^{1-y_i} = p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i} = p^x(1-p)^{n-x}$$

で与えられる．尤度関数 $L(p, y_1, \ldots, y_n) = L(p, x) = p^x(1-p)^{n-x}$，対数尤度関数 $\log L(p,x) = x\log p + (n-x)\log(1-p)$ と得られる．またその微分は

$$\frac{\mathrm{d}}{\mathrm{d}p}\log L(p,x) = \frac{x}{p} - \frac{n-x}{1-p}$$

であり，高校レベルの数学で実際に x/n で対数尤度関数が最大化されていることが確認できる．

3.3.2 正規分布

$X_1, \dots, X_n$ を正規分布 $N(\mu, 1)$ からの $i.i.d.$ 標本とする．尤度関数

$$L(\mu, \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{(x_i-\mu)^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\mu)^2 &= n\left(\mu - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} + \sum_{i=1}^{n} x_i^2 = n\left(\mu-\bar{x}\right)^2 - n\bar{x}^2 + \sum_{i=1}^{n} x_i^2 \\ &= n\left(\mu-\bar{x}\right)^2 + \sum_{i=1}^{n}(x_i-\bar{x})^2\end{aligned}$$

3.4 平均二乗誤差の分解

推定量 $\hat{\theta}(X)$ に対して，平均二乗誤差の分解 $\mathrm{E}[(\hat{\theta}(X)-\theta)^2]$ を与える．ただし，$b(\theta)$ はバイアス

$$b(\theta) = E(\hat{\theta}) - \theta$$

である．このとき，

$$\begin{aligned}\mathrm{E}[(\hat{\theta}(X)-\theta)^2] &= \mathrm{E}[(\hat{\theta}(X) - \mathrm{E}(\hat{\theta}) + \mathrm{E}(\hat{\theta}) - \theta)^2] \\ &= \mathrm{E}[\{\hat{\theta}(X) - E(\hat{\theta})\}^2] + 2\mathrm{E}[(\hat{\theta}(X) - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)] + \{E(\hat{\theta}) - \theta\}^2 \\ &= \mathrm{Var}(\hat{\theta}(X)) + b(\theta)^2\end{aligned}$$

である．式展開における注意

- 1 つ目の等号で差し引きした $\mathrm{E}(\hat{\theta}) = \mathrm{E}(\hat{\theta}(X))$ は定数．
- 交差項 $\mathrm{E}[(\hat{\theta}(X) - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)]$ で $\mathrm{E}(\hat{\theta}) - \theta$ は定数

$$\begin{aligned}\mathrm{E}[(\hat{\theta}(X) - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)] &= (\mathrm{E}(\hat{\theta}) - \theta)\mathrm{E}[\hat{\theta}(X) - \mathrm{E}(\hat{\theta})] \\ &= (\mathrm{E}(\hat{\theta}) - \theta)\left(\mathrm{E}[\hat{\theta}(X)] - \mathrm{E}[\hat{\theta}]\right) = 0\end{aligned}$$

3.5 「不偏標本共分散」と「不偏標本相関係数？」

- 設定
  - 母集団：壺からの 2 つの玉が刺さった団子の抽出．
  - 母集団 $\mathrm{E}[X] = \mu_x$，$\mathrm{E}[Y] = \mu_Y$，$\mathrm{Var}[X] = \mathrm{E}[(X-\mu_X)^2] = \sigma_X^2$，$\mathrm{Var}[Y] = \mathrm{E}[(Y-\mu_Y)^2] = \sigma_Y^2$，$\mathrm{COV}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \rho\sigma_X\sigma_Y$
  - 母集団からの独立同一な n 個の標本（確率変数ベクトル）$(X_1, Y_1), \dots, (X_n, Y_n)$ があるとする
- 定理：不偏標本共分散は $\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$ つまり，

$$\mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right] = \mathrm{COV}(X,Y) \qquad (*)$$

- 証明
  - $\mathrm{E}[(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)] = \mathrm{COV}(X,Y)/n$ なぜなら

$$
\begin{aligned}
\mathrm{E}[(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)] &= \mathrm{E}\left[\frac{\sum_{i=1}^n (X_i-\mu_X)}{n}\frac{\sum_{i=1}^n (Y_i-\mu_Y)}{n}\right] \\
&= \frac{1}{n^2}\mathrm{E}\left[\sum_{i=1}^n (X_i-\mu_X)\sum_{i=1}^n (Y_i-\mu_Y)\right] \\
&= \frac{1}{n^2}\mathrm{E}\left[\sum_{i=1}^n (X_i-\mu_X)(Y_i-\mu_Y) + \sum_{i\neq j}(X_i-\mu_X)(Y_j-\mu_Y)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathrm{E}\left[(X_i-\mu_X)(Y_i-\mu_Y)\right] + \sum_{i\neq j}\mathrm{E}\left[(X_i-\mu_X)(Y_j-\mu_Y)\right] \\
&= \frac{n\mathrm{COV}(X,Y)}{n^2} + 0 \\
&= \frac{\mathrm{COV}(X,Y)}{n}
\end{aligned}
$$

  二番目の等号：期待値の線形性 $a\mathrm{E}[Z] = \mathrm{E}[aZ]$，三番目の等号：$n$ 項の和同士の積を添え字が同じ n 項と違う $n(n-1)$ 項に分離，四番目の等号：期待値の線形性 $\mathrm{E}[Z_1+Z_2] = \mathrm{E}[Z_1]+\mathrm{E}[Z_2]$，五番目の等号（第 1 項）：独立「同一」分布，五番目の等号（第 2 項）：「独立」同一分布なので共分散は 0，$\mathrm{E}\left[(X_i-\mu_X)(Y_j-\mu_Y)\right] = \mathrm{E}\left[X_i-\mu_X\right]\mathrm{E}\left[Y_j-\mu_Y\right] = 0$，
  - (*) の分子

$$
\begin{aligned}
&\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y}) = \sum_{i=1}^n (\{X_i-\mu_X\}-\{\bar{X}-\mu_X\})(\{Y_i-\mu_Y\}-\{\bar{Y}-\mu_Y\}) \\
&= \sum_{i=1}^n (X_i-\mu_X)(Y_i-\mu_Y) - (\bar{Y}-\mu_Y)\sum_{i=1}^n (X_i-\mu_X) - (\bar{X}-\mu_X)\sum_{i=1}^n (Y_i-\mu_Y) + n(\bar{X}-\mu_X)(\bar{Y}-\mu_Y) \\
&= \sum_{i=1}^n (X_i-\mu_X)(Y_i-\mu_Y) - n(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)
\end{aligned}
$$

  - $\mathrm{E}\left[\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})\right]$ について

$$
\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})\right] &= \mathrm{E}\left[\sum_{i=1}^n (X_i-\mu_X)(Y_i-\mu_Y)\right] - n\mathrm{E}\left[(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)\right] \\
&= n\mathrm{COV}(X,Y) - \mathrm{COV}(X,Y) = (n-1)\mathrm{COV}(X,Y)
\end{aligned}
$$

- 母相関係数は，

$$
\rho = \frac{\mathrm{COV}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}
$$

  であり，$\mathrm{COV}(X,Y)$，$\mathrm{Var}[X]$，$\mathrm{Var}[Y]$ それぞれについて，不偏推定量が

$$
\frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{n-1},\ \frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1},\ \frac{\sum_{i=1}^n (Y_i-\bar{Y})^2}{n-1}
$$

  と得られている

- 相関係数のモーメント推定量は，標本相関係数

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

  しかし，不偏性は持たない

$$\rho = \frac{\mathrm{COV}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \neq \mathrm{E}\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}\right]$$

- 不偏性を持つ相関係数の推定量の理論は難しいので略.

3.6 正規分布と P 値

まず，

$$W \sim N(0, \sigma^2/n)$$

は原点対称の分布であることに注意する．従って，

$$\Pr(|W| \geqslant |\bar{x}|) = \Pr(W \leqslant -|\bar{x}|) + \Pr(W \geqslant |\bar{x}|) = 2\Pr(W \geqslant |\bar{x}|)$$

が従う．ところで正規分布の線形変換も正規分布であること

$$Y \sim N(\theta, \eta^2) \;\Rightarrow\; a + bY \sim N(a + b\theta, b^2\eta^2)$$

を思い出すと，

$$Z = \frac{\sqrt{n}}{\sigma}W \sim N(0,1)$$

のように Z の分布は標準正規分布であることが分かる．これらから，

$$\begin{aligned}\Pr(|W| \geqslant |\bar{x}|) &= 2\Pr(W \geqslant |\bar{x}|) = 2\Pr(\frac{\sqrt{n}}{\sigma}W \geqslant \frac{\sqrt{n}}{\sigma}|\bar{x}|) \\ &= 2\Pr(Z \geqslant \frac{\sqrt{n}}{\sigma}|\bar{x}|) = 2\int_{\sqrt{n}|\bar{x}|/\sigma}^{\infty} \frac{\exp(-z^2/2)}{\sqrt{2\pi}}\mathrm{d}z\end{aligned}$$