第 2 章

いろいろな分布と統計量

2.1　母集団と統計量

母集団の確率分布をμとしたとき, μ母集団と呼ぶ。
μ母集団から n 個の標本とは $X_1 \sim X_2 \sim \cdots \sim X_n \sim \mu$で独立とする。
(注意　$X_1 \sim X_2$とは, X_1の分布 $=$ X_2の分布　$X_n \sim \mu$とは, X_nの分布 $= \mu$)

統計量θとは $\theta = \theta(X_1, X_2, ..., X_n)$ とn 個の標本の関数

$$\text{最も重要な統計量} = \text{標本平均} = \overline{X} = \frac{X_1 + X_2 + \cdots X_n}{n}$$

すると母集団の平均 $=$ 母平均 $= m$, 母集団の分散 $=$ 母分散 $= \sigma^2$とすると

$$E(\overline{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{E(X_1) + \cdots + E(X_n)}{n} = \frac{nm}{n} = m$$

◎この性質を標本平均$\overline{X}$は母平均 m の不偏推定量という

$$\begin{aligned} V(\overline{X}) &= V\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(V(X_1) + \cdots + V(X_n)\right) \qquad (\because \text{独立}) \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

(注意　$\lim_{n\to\infty} V(\overline{X}) = 0$ より, $\overline{X}$は $n \to \infty$ で定数 $(= m)$ に近づくことが分かる)

$$\text{次に重要な統計量} = \text{不偏標本分散} = \hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

◎$\hat{S}^2$は母分散σ^2の不偏推定量 つまり $E(\hat{S}^2) = \sigma^2$

(証明)

$$
\begin{aligned}
E(\hat{S}^2) &= \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + (\overline{X})^2)\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}E(X_i^2) - 2E\left(\sum_{i=1}^{n}X_i\overline{X}\right) + nE((\overline{X})^2)\right) \\
&= \frac{1}{n-1}(nE(X_1^2) - nE((\overline{X})^2)) = \frac{n}{n-1}(\ (V(X_1) + E(X_1)^2) - (V(\overline{X}) + E(\overline{X})^2)\ ) \\
&= \frac{n}{n-1}\left((\sigma^2 + m^2) - \left(\frac{\sigma^2}{n} + m^2\right)\right) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2
\end{aligned}
$$

2.2　最尤推定量（値）

μ母集団からの n 個の標本とその実現値を $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ とすると,

$$
\text{尤度関数 } L(x_1, x_2, ..., x_n|\Theta) = \begin{cases} P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_n = x_n) & \cdots \text{離散} \\ f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) & \cdots \text{連続} \end{cases}
$$

$\boxed{\text{最尤推定値 } \hat{\theta}}$

尤度関数が最大 (確率最大) となるθを $x_1, x_2, ..., x_n$ で表したもの $\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n)$ を母パラメータΘの最尤推定値という。意味は, 実現値 $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ を得たのには, それが確率最大だからという理由で出たものであり, その関係式に基づいて母パラメータの推定値$\hat{\theta}$を $x_1, x_2, ..., x_n$ で表す。
実際には対数尤度 $\log L$ をθで偏微分して, $\boxed{\frac{\partial}{\partial\theta}\log L = 0}$ を解いて求めることが多い。

2.3　有効推定量とクラメール・ラオの不等式

大数の法則より, 標本数を大きくしていけば不偏推定量は母パラメータに近づく。分散が小さければ小さいほどこの収束が早くなるので望ましい統計量と言える。最小のものがあれば最も望ましいが, ここで述べるクラメール・ラオの不等式により, 不偏推定量の中での下限が分かり, その下限を達成する推定量が最小分散推定量となる。
T_nを母パラメータθの不偏推定量とするとき, $L = f_X(x_1)\cdots f_X(x_n)$ として,

$$
V(T_n) \geq \frac{1}{E\left(\left(\frac{\partial}{\partial\theta}\log L\right)^2\right)} = \frac{1}{V\left(\frac{\partial}{\partial\theta}\log L\right)} = \frac{1}{I_n(\theta)} \ (= \text{フィッシャー情報量})
$$

が成立し, 等号が成立するような$\hat{T}_n$を有効推定量(= 最小分散推定量) と呼ぶ。

(証明) 母集団が連続のときのみ考えるが, 離散のときもほとんど同じ

$$
1 = \int_{\mathbb{R}} f_X(x_1)dx_1 \cdots \int_{\mathbb{R}} f_X(x_n)dx_n = \int \cdots \int_{\mathbb{R}^n} L(x_1, x_2, ..., x_n|\theta)dx_1 \cdots dx_n
$$

よって, θで偏微分して

$$\begin{aligned}0 &= \int \cdots \int \frac{\partial}{\partial\theta} L(x_1, x_2, ..., x_n|\theta) dx_1 \cdots dx_n \\ &= \int \cdots \int \left(\frac{\partial}{\partial\theta} \log L(x_1, x_2, ..., x_n|\theta)\right) L(x_1, x_2, ..., x_n|\theta) dx_1 \cdots dx_n \\ &= E\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right) \quad \cdots ①\end{aligned}$$

また, T_nはθの不偏推定量より,

$$\theta = E(T_n(X_1, X_2, .., X_n)) = \int \cdots \int T_n(x_1, x_2, .., x_n) L(x_1, x_2, ..., x_n|\theta) dx_1 \cdots dx_n$$

θで偏微分して

$$\begin{aligned}1 &= \int \cdots \int T_n(x_1, x_2, .., x_n) \frac{\partial}{\partial\theta} L(x_1, x_2, ..., x_n|\theta) dx_1 \cdots dx_n \\ &= \int \cdots \int T_n(x_1, x_2, .., x_n) \left(\frac{\partial}{\partial\theta} \log L(x_1, x_2, ..., x_n|\theta)\right) L(x_1, x_2, ..., x_n|\theta) dx_1 \cdots dx_n \\ &= E\left(T_n(X_1, X_2, .., X_n) \frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right) \quad \cdots ②\end{aligned}$$

$\therefore$ ①, ②より, $1 = E\left((T_n(X_1, X_2, .., X_n) - \theta) \frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right)$

$\therefore$ コーシーシュワルツの不等式 $E(XY) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$より,

$$1^2 \leq E\left((T_n - \theta)^2\right) E\left(\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right)^2\right)$$

$\theta = E(T_n)$ ($\because$ 不偏性) より,

$$V(T_n) \geq \frac{1}{E\left(\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right)^2\right)}$$

また, $E\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right) = 0$ より,

$$\begin{aligned}E\left(\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right)^2\right) &= V\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right) \\ &= V\left(\frac{\partial}{\partial\theta} \log f(X_1) + \cdots + \frac{\partial}{\partial\theta} \log f(X_n)\right) \\ &= nV\left(\frac{\partial}{\partial\theta} \log f(X_1)\right) = nV\left(\frac{\frac{\partial f}{\partial\theta}}{f}(X_1)\right) \quad (= I_n(\theta) = nI_1(\theta))\end{aligned}$$

注意

等号成立を調べると, $T_n - \hat{\theta} = c\left(\frac{\partial}{\partial\theta} \log L(X_1, X_2, ..., X_n)\right)$ となる定数 c が存在して, ここに実現値 $x_1, ..., x_n$ を考え, 最尤推定値とすると, $0 = \frac{\partial}{\partial\theta} \log L(x_1, x_2, ..., x_n)$ より, $T_n = \hat{\theta} =$ 最尤推定値となる

2.4 色々な分布

2.4.1 ベルヌーイ分布 Be(p) と 2 項分布 B(n,p)

$X \sim \mathrm{Be}(p)$ とは

X	0	1
確率	$1-p$	p

$X=1$: 成功　$X=0$: 失敗

このとき,

$$E(X)=1\times p=p \qquad V(X)=E(X^2)-(E(X))^2=p-p^2 \qquad g_X(t)=E(t^X)=1-p+pt$$

n 個の独立な Be(p) の和, つまり $Y=X_1+X_2+\cdots+X_n$　$X_i\sim \mathrm{Be}(p)$ で独立のとき,
$Y\sim \mathrm{B}(n,p)$ (パラメータ n,p の 2 項分布)　　(n 回の独立なベルヌーイ試行での成功回数)
すると,

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k} \qquad (k=0,1,2,...,n)$$

ここで,

$$\binom{n}{k}={}_n\mathrm{C}_k=\frac{n(n-1)\cdots(n-k+1)}{k!}$$

$$\begin{aligned}
E(Y)&=E(X_1+X_2+\cdots+X_n)=nE(X_1)=np\\
V(Y)&=V(X_1+X_2+\cdots+X_n)=nV(X_1)=np(1-p) \qquad (\because \text{独立})\\
g_Y(t)&=E(t^Y)=E(t^{X_1+\cdots+X_n})=E(t^{X_1})\cdots E(t^{X_n})=(1-p+pt)^n \qquad (\because \text{独立})
\end{aligned}$$

(別解)

$$\begin{aligned}
E(Y)&=\sum_{k=0}^{n}kP(Y=k)=\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n}n\binom{n-1}{k-1}p^k(1-p)^{n-k}\\
&=n\sum_{l=0}^{n-1}\binom{n-1}{l}p^{l+1}(1-p)^{n-1-l} \qquad (l=k-1)\\
&=np(p+(1-p))^{n-1}=np
\end{aligned}$$

$$\begin{aligned}
E(Y(Y-1))&=\sum_{k=0}^{n}k(k-1)\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=2}^{n}n(n-1)\binom{n-2}{k-2}p^k(1-p)^{n-k}\\
&=n(n-1)\sum_{l=0}^{n-2}\binom{n-2}{l}p^{l+2}(1-p)^{n-2-l} \qquad (l=k-2)\\
&=n(n-1)p^2(p+(1-p))^{n-2}=n(n-1)p^2
\end{aligned}$$

$$\begin{aligned}
\therefore V(Y)&=E(Y(Y-1))+E(Y)-(E(Y))^2\\
&=n(n-1)p^2+np-(np)^2\\
&=np(1-p)
\end{aligned}$$

$$g_Y(t) = E(t^Y) = \sum_{k=0}^{n} t^k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pt)^k (1-p)^{n-k} = (1-p+pt)^n$$

これより, 2 項分布の再生性が分かる。

$X \sim \mathrm{B}(n,p)$ $Y \sim \mathrm{B}(m,p)$ で独立なら $X+Y \sim \mathrm{B}(n+m,p)$

$$\because g_{X+Y}(t) = g_X(t)g_Y(t) = (1-p+pt)^n(1-p+pt)^m = (1-p+pt)^{n+m} = g_{\mathrm{B}(n+m,p)}(t)$$

○ $\mathrm{Be}(p)$ 母集団の最尤推定値

$X_1 \sim X_2 \sim \cdots \sim X_n \sim \mathrm{Be}(p)$ で独立とする。

$\mathrm{Be}(p)$ の母集団からの n 個の標本 $X_1, X_2, ..., X_n$の実現値を $x_1, x_2, ..., x_n$ $x_i \in \{0,1\}$ とする。

$$P(X_1 = x_1) = \begin{cases} p & \cdots x_1 = 1 \\ 1-p & \cdots x_1 = 0 \end{cases} \text{より,} \quad P(X_1 = x_1) = p^{x_1}(1-p)^{1-x_1}$$

$$\begin{aligned} \therefore L(x_1, x_2, ..., x_n|\theta) &= P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{x_1+x_2+...+x_n}(1-p)^{n-(x_1+x_2+...+x_n)} \end{aligned}$$

$$\therefore \log L = (x_1+x_2+...+x_n)\log p + (n-(x_1+x_2+...+x_n))\log(1-p)$$

$$0 = \frac{\partial}{\partial p}\log L = (x_1+x_2+...+x_n)\frac{1}{\hat{p}} + (n-(x_1+x_2+...+x_n))\frac{-1}{1-\hat{p}}$$

$$\therefore \hat{p} = \frac{x_1+x_2+...+x_n}{n} \quad (= \text{標本平均})$$

フィッシャー情報量

$$\begin{aligned} \log P(X=x) &= x\log p + (1-x)\log(1-p) \\ \frac{\partial}{\partial p}\log P(X=x) &= \frac{x}{p} - \frac{1-x}{1-p} = \frac{x(1-p)-p(1-x)}{p(1-p)} = \frac{x-p}{p(1-p)} \\ \therefore V\left(\frac{\partial}{\partial p}\log P(X)\right) &= V\left(\frac{X-p}{p(1-p)}\right) = \frac{V(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)} \\ \therefore \frac{1}{I_n(p)} &= \frac{1}{nI_1(p)} = \frac{p(1-p)}{n} \end{aligned}$$

また, $V(\hat{p}) = V\left(\dfrac{X_1+X_2+...+X_n}{n}\right) = \dfrac{V(X_1)}{n} = \dfrac{p(1-p)}{n}$ と $\hat{p}$はクラメール・ラオの不等式の下限を実現する。

$\therefore \hat{p} = \dfrac{X_1+X_2+...+X_n}{n}$ は有効推定量

練習問題 2.1　120 回サイコロを投げて 6 の目が出た回数 $= X$ とする。

(1)$P(X=k)$　(2)$E(X)$　(3)$V(X)$　(4)$g_X(t) = E(t^X)$　(5)$E(X(X-1)(X-2))$

(6)$P(X=k)$ を最大にする k (モードという)