目次

1 便利な定理まとめ

1.1 推定論

X をパラメータ θ を持つある確率分布に従う確率変数であるとする。Fisher 情報量 $I(\theta)$ を

$$I(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial \log f(X|\theta)}{\partial \theta}\right)^2\right] \tag{1.1}$$

と定義する。期待値 $\mathrm{E}_\theta[\cdot]$ と θ 微分の交換に関する正則条件を満たす限り、Fisher 情報量を

$$I(\theta) = -\mathrm{E}_\theta\left[\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2}\right] \tag{1.2}$$

で定義しても等価である。

なお、上記の定義において $f(X|\theta)$ となっているところを n サンプル $X_1, \ldots, X_n$ の尤度関数 $L(\theta|X_1, \ldots, X_n) = \prod_{i=1}^n f(X_i|\theta)$ に置き換えて定義したものを $I_n(\theta)$ と書くと、$X_1, \ldots, X_n$ が独立同分布である限り $I_n(\theta) = nI(\theta)$ が成り立つ。$I(\theta)$ と $I_n(\theta)$ のどちらを Fisher 情報量と呼ぶのかは分野や文献に依る。ここでは $I(\theta)$ を Fisher 情報量と呼ぶことにし、混合を避けるためにできるだけ $I(\theta)$ のみを用いることにする。

以下では n 個の独立同分布 $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ の実現値からパラメータ θ の値を推定する問題を考える。推定量は $\widehat{\theta}(\boldsymbol{X})$ のようにハットをつけて表し、引数の $\boldsymbol{X}$ はしばしば省略する。

定理 1.1. 推定量 $\widehat{\theta}$ について成り立つ以下の不等式を、Cramér-Rao の不等式という。

(1) $\widehat{\theta}$ が不偏推定量であるとき、

$$\mathrm{Var}_\theta[\widehat{\theta}] \geq \frac{1}{nI(\theta)} \tag{1.3}$$

が成り立つ。

(2) $\widehat{\theta}$ が不偏推定量でないときは、バイアスを $b(\theta) := \mathrm{E}_\theta[\widehat{\theta}] - \theta$ とすると、

$$\mathrm{Var}_\theta[\widehat{\theta}] \geq \frac{(1 + b'(\theta))^2}{nI(\theta)} \tag{1.4}$$

が成り立つ。

定理 1.2. $\widehat{\theta}$ が不偏推定量であるとする。$\widehat{\theta}$ が Cramér-Rao の不等式の下限

$$\mathrm{Var}_\theta[\widehat{\theta}] = \frac{1}{nI(\theta)} \tag{1.5}$$

を満たすならば、$\widehat{\theta}$ は一様最小分散不偏（UMVU）推定量である。

この Cramér-Rao の下限 (1.5) を用いた UMVU の判定方法は、あくまで十分条件を与えるのみであることに注意。一方、以下の完全十分統計量による判定方法は UMVU 推定量の完全な構成法を与える。

定理 1.3. T が完全十分統計量であるとき、以下の 2 つが成り立つ。

(1) T の関数である不偏推定量 $\delta(T)$ は一意に定まり、$\delta(T)$ は UMVU 推定量である。
(2) 任意の不偏推定量 $\hat{\theta}$ に対して、$\mathrm{E}[\hat{\theta}|T]$ は $\delta(T)$ に一致する。

2 正規分布

本節では、正規分布に関する推定量の基本的事項についてのまとめ・導出を行う。

実数値の確率変数 X がパラメータ μ, σ をもつ正規分布の確率密度関数

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.6}$$

に従うとき、$X \sim N(\mu, \sigma^2)$ と表記する。

2.1 1 標本問題の点推定

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, i.i.d. とする。また、$\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ とする。

命題 2.1. パラメータ μ, σ^2 の推定量に関して、以下が成り立つ。

(1) σ^2 が既知か未知かにかかわらず

$$\widehat{\mu} := \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \tag{2.7}$$

は μ の UMVU 推定量かつ最尤推定量である。

(2) μ が既知の場合、

$$\widehat{\sigma^2} := \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 \tag{2.8}$$

は σ^2 の UMVU 推定量かつ最尤推定量である。

(3) μ が未知の場合、σ^2 の UMVU 推定量および最尤推定量はそれぞれ

$$\widehat{\sigma^2_{\mathrm{UMVU}}} := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2, \qquad \widehat{\sigma^2_{\mathrm{MLE}}} := \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \tag{2.9}$$

で与えられる。

証明.

(1) $\mathrm{E}[\widehat{\mu}]=\mu$ が成り立つから、$\widehat{\mu}$ は μ の不偏推定量。

$$\log f(X|\mu,\sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(X-\mu)^2}{2\sigma^2} \tag{2.10}$$

より、μ に関する Fisher 情報量は

$$I(\mu) = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^2}\,. \tag{2.11}$$

$X_1,\ldots,X_n$ は互いに独立だから、

$$\mathrm{Var}[\widehat{\mu}] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \sigma^2/n\,. \tag{2.12}$$

よって定理 1.2 より $\widehat{\mu}$ は UMVU 推定量である。
次に最尤推定量が $\widehat{\mu}$ になることを示す。対数尤度は、

$$\ell(\mu,\sigma^2|\boldsymbol{X}) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{2\sigma^2} \tag{2.13}$$

これを μ について最大化すると、

$$\widehat{\mu}_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n}X_i \tag{2.14}$$

となり、これは $\widehat{\mu}$ に一致。

(2) $X_1,\ldots,X_n$ は互いに独立であることから $\mathrm{E}[\widehat{\sigma^2}]=\sigma^2$ が成り立つため、$\widehat{\sigma^2}$ は σ^2 の不偏推定量。
$\tau=\sigma^2$ とおくと

$$\frac{\partial^2}{\partial\tau^2}\log f(X|\mu,\tau) = \frac{1}{2\tau^2} - \frac{(X-\mu)^2}{\tau^3} \tag{2.15}$$

となり、τ に関する Fisher 情報量は

$$I(\tau) = -\frac{1}{2\tau^2} + \frac{1}{\tau^2} = \frac{1}{2\tau^2}\,. \tag{2.16}$$

一方、$\widehat{\sigma^2}$ の分散は

$$\mathrm{Var}[\widehat{\sigma^2}] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[(X_i-\mu)^2] = \frac{2\sigma^4}{n} \tag{2.17}$$

である。最後の等号は $((X_i-\mu)/\sigma)^2$ が自由度 1 の χ^2 分布に従うことから（χ^2 分布の分散を知っていれば）求めることができる（当然定義通りに計算してもよいが、4 次モーメントを求め

る必要があるためやや面倒)。これらの結果と定理 1.2 より $\widehat{\sigma^2}$ は UMVU 推定量である。次に最尤推定量を求める。(2.13) より

$$\frac{\partial}{\partial \tau}\ell(\mu, \tau|\boldsymbol{X}) = -\frac{n}{2\tau} + \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\tau^2} . \tag{2.18}$$

よって最尤推定量は

$$\widehat{\sigma^2_{\mathrm{MLE}}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \tag{2.19}$$

となり、$\widehat{\sigma^2}$ に一致する。

(3) 分散の定義が平均値 μ に陽に依存するため、μ が既知か未知かによって推定量が変わり得る。

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= \sum_{i=1}^{n}\mathrm{E}\left[(X_i - \mu + \mu - \bar{X})^2\right] \\
&= \sum_{i=1}^{n}\mathrm{E}[(X_i - \mu)^2] - 2\,\mathrm{E}\left[\sum_{i=1}^{n}(X_i - \mu)(\bar{X} - \mu)\right] + \sum_{i=1}^{n}\mathrm{E}\left[(\bar{X} - \mu)^2\right] \\
&= n\sigma^2 - 2n\,\mathrm{E}[(\bar{X} - \mu)^2] + \mathrm{E}[(\bar{X} - \mu)^2] \\
&= n\sigma^2 - n\,\mathrm{Var}(\bar{X}) \\
&= (n-1)\sigma^2
\end{aligned} \tag{2.20}$$

より、σ^2 の不偏推定量は

$$\widehat{\sigma^2_{\mathrm{UMVU}}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{2.21}$$

のように、n の代わりに $n-1$ で割ったものとなる。ここでは実際に確かめることはしないが、これは Cramér-Rao の不等式の下限を満たさない。よってこれが実際に UMVU 推定量であることを示すためには定理 1.3 を使う必要がある。$X_1, \ldots, X_n$ の同時密度分布が

$$\begin{aligned}
f(\boldsymbol{x}) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\prod_{i=1}^{n}\exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)
\end{aligned}$$

と書けることから、$T = (\bar{X}, \sum_{i=1}^{n}(X_i - \bar{X})^2)$ は十分統計量であり、証明は省略するがこれは完備である。以上のことから、$\widehat{\sigma^2_{\mathrm{UMVU}}}$ は不偏推定量かつ完備十分統計量（の関数）であるから、定理 1.3 より UMVU 推定量である。

次に最尤推定量を求める。この場合は対数尤度を μ と σ で同時に最大化する必要がある。しかし、(1) で求めた μ の最尤推定量 (2.14) は σ に依らないため今の場合も同様に成り立ち、その $\widehat{\mu}_{\mathrm{MLE}}$ を μ に代入して σ について最大化すればよい。当然その結果は (2) の結果 (2.19) の μ を

$\widehat{\mu}_{\mathrm{MLE}}$ で置き換えたものになるため、

$$\widehat{\sigma^2_{\mathrm{MLE}}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{2.22}$$

となる。

□