Contents

1 Maximum Likelihood Estimation (MLE, 最尤(さいゆう)法) — Review **5**

2 Bayesian Estimation (ベイズ推定) **30**

2.1 Introduction 31

2.2 Inference 36

2.2.1 Point Estimate 36

2.2.2 Interval Estimate 37

2.2.3 Marginal Likelihood (周辺尤度) 38

2.3 Example: Linear Regression 38

2.4 On Prior Distribution 42

2.4.1 Non-informative Prior 42

2.4.2 Jeffreys' Prior 43

2.5 Evaluation of Expectation 44
2.5.1 Evaluation of Expectation: Numerical Integration 45
2.5.2 Evaluation of Expectation: Monte Carlo Integration 49
2.5.3 Evaluation of Expectation: Random Number Generation 53
2.6 Sampling Method I: Random Number Generation 54
2.6.1 Uniform Distribution: $U(0, 1)$ 54
2.6.2 Transforming $U(0, 1)$: Continuous Type 63
2.7 Sampling Method II: Random Number Generation 77
2.7.1 Rejection Sampling (棄却法) 77
2.7.2 Importance Resampling (重点的リサンプリング) 84
2.7.3 Metropolis-Hastings Algorithm (メトロポリス－ハスティングス・アルゴリズム) 87
2.7.4 Gibbs Sampling 98

1 Maximum Likelihood Estimation (MLE, 最尤(さいゆう)法) — Review

1. We have random variables X_1, X_2, $\cdots$, X_n, which are assumed to be mutually independently and identically distributed.

2. The distribution function of $\{X_i\}_{i=1}^n$ is $f(x;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$ and $\theta = (\mu, \Sigma)$.

   Note that X is a vector of random variables and x is a vector of their realizations (i.e., observed data).

   Likelihood function $L(\cdot)$ is defined as $L(\theta; x) = f(x; \theta)$.

   Note that $f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$ when X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed.

The maximum likelihood estimator (MLE) of θ is θ such that:

$$\max_{\theta} \ L(\theta; X). \qquad \Longleftrightarrow \qquad \max_{\theta} \ \log L(\theta; X).$$

MLE satisfies the following two conditions:

(a) $\dfrac{\partial \log L(\theta; X)}{\partial \theta} = 0.$

(b) $\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

3. **Fisher's information matrix (フィッシャーの情報行列)** is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big),$$

where we have the following equality:

$$-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big) = \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\Big) = \mathrm{V}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big)$$

Proof of the above equality:

$$\int L(\theta; x)\mathrm{d}x = 1$$

Take a derivative with respect to θ.

$$\int \frac{\partial L(\theta; x)}{\partial \theta}\mathrm{d}x = 0$$

(We assume that (i) the domain of x does not depend on θ and (ii) the derivative $\dfrac{\partial L(\theta; x)}{\partial \theta}$ exists.)

Rewriting the above equation, we obtain:

$$\int \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x)\mathrm{d}x = 0,$$

i.e.,

$$\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0.$$

Again, differentiating the above with respect to θ, we obtain:

$$\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial L(\theta; x)}{\partial'\theta} \mathrm{d}x$$
$$= \int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial \log L(\theta; x)}{\partial\theta'} L(\theta; x)\mathrm{d}x$$
$$= \mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\Big) + \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\Big) = 0.$$

Therefore, we can derive the following equality:

$$-\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) = \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right),$$

where the second equality utilizes $\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right) = 0$.

4. **Cramer-Rao Lower Bound (クラメール・ラオの下限)**: $(I(\theta))^{-1}$

Suppose that an unbiased estimator of θ is given by $s(X)$.

Then, we have the following:

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1}$$

Proof:

The expectation of $s(X)$ is:

$$\mathrm{E}(s(X)) = \int s(x)L(\theta; x)\mathrm{d}x.$$

Differentiating the above with respect to θ,

$$\begin{aligned}
\frac{\partial \mathrm{E}(s(X))}{\partial \theta'} &= \int s(x)\frac{\partial L(\theta; x)}{\partial \theta'}\mathrm{d}x = \int s(x)\frac{\partial \log L(\theta; x)}{\partial \theta'}L(\theta; x)\mathrm{d}x \\
&= \mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)
\end{aligned}$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial\theta}\right)^2 = \left(\mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta;X)}{\partial\theta}\right)\right)^2 = \rho^2 \mathrm{V}\,(s(X))\,\mathrm{V}\left(\frac{\partial \log L(\theta;X)}{\partial\theta}\right)$$
$$\leq \mathrm{V}\,(s(X))\,\mathrm{V}\left(\frac{\partial \log L(\theta;X)}{\partial\theta}\right),$$

where ρ denotes the correlation coefficient between $s(X)$ and $\dfrac{\partial \log L(\theta;X)}{\partial\theta}$, i.e.,

$$\rho = \frac{\mathrm{Cov}\left(s(X), \dfrac{\partial \log L(\theta;X)}{\partial\theta}\right)}{\sqrt{\mathrm{V}\,(s(X))}\sqrt{\mathrm{V}\left(\dfrac{\partial \log L(\theta;X)}{\partial\theta}\right)}}.$$

Note that $|\rho| \leq 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 \leq \mathrm{V}(s(X))\ \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$\mathrm{V}(s(X)) \geq \frac{\left(\dfrac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2}{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $\mathrm{E}(s(X)) = \theta$,

$$\mathrm{V}(s(X)) \geq \frac{1}{-\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$I(\theta) = -\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right)$$
$$= \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right).$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

5. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ.

As n goes to infinity, we have the following result:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right),$$

where it is assumed that $\lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)$ converges.

That is, when n is large, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $\mathrm{V}(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

Practically, we utilize the following approximated distribution:

$$\tilde{\theta} \sim N\left(\theta, (I(\tilde{\theta}))^{-1}\right).$$

Then, we can obtain the significance test and the confidence interval for θ

6. **Central Limit Theorem:** Let $X_1, X_2, \cdots, X_n$ be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for $i = 1, 2, \cdots, n$.

   Define $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$.

   Then, the central limit theorem is given by:

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1).$$

   Note that $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{V}(\overline{X}) = \sigma^2/n$.

That is,

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \sigma^2).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance $\Sigma < \infty$, the central limit theorem is given by:

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \Sigma).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \Sigma$.

7. **Central Limit Theorem II:** Let X_1, X_2, $\cdots$, X_n be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma_i^2$ for $i = 1, 2, \cdots, n$.

Assume:

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 < \infty.$$

Define $\overline{X} = (1/n) \sum_{i=1}^{n} X_i$.

Then, the central limit theorem is given by:

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1),$$

i.e.,

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \sigma^2).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) \longrightarrow \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance Σ_i, the central limit theorem is given by:

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \Sigma),$$

where $\Sigma = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Sigma_i < \infty$.

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) \longrightarrow \Sigma$.

[Review of Asymptotic Theories]

- **Convergence in Probability (確率収束)** $X_n \longrightarrow a$, i.e., X converges in probability to a, where a is a fixed number.

- **Convergence in Distribution (分布収束)** $X_n \longrightarrow X$, i.e., X converges in distribution to X. The distribution of X_n converges to the distribution of X as n goes to infinity.

Some Formulas

X_n and Y_n : Convergence in Probability

Z_n : Convergence in Distribution

- If $X_n \longrightarrow a$, then $f(X_n) \longrightarrow f(a)$.
- If $X_n \longrightarrow a$ and $Y_n \longrightarrow b$, then $f(X_n Y_n) \longrightarrow f(ab)$.
- If $X_n \longrightarrow a$ and $Z_n \longrightarrow Z$, then $X_n Z_n \longrightarrow aZ$, i.e., aZ is distributed with mean $\mathrm{E}(aZ) = a\mathrm{E}(Z)$ and variance $\mathrm{V}(aZ) = a^2\mathrm{V}(Z)$.

[End of Review]

8. **Weak Law of Large Numbers (大数の弱法則) — Review:**

Suppose that $X_1, X_2, \cdots, X_n$ are distributed.

As $n \longrightarrow \infty$, $\overline{X} \longrightarrow \lim_{n\to\infty} \mathrm{E}(\overline{X})$ under $\lim_{n\to\infty} n\mathrm{V}(\overline{X}) < \infty$, which is called the **weak law of large numbers**.

$\longrightarrow$ Convergence in probability

$\longrightarrow$ Proved by Chebyshev's inequality

(i) Suppose that $X_1, X_2, \cdots, X_n$ are assumed to be mutually independently and identically distributed with $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2 < \infty$.

Consider $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$.

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \sigma^2$.

(ii) Suppoose that $X_1, X_2, \cdots, X_n$ are assumed to be mutually independently distributed with $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{V}(X_i) = \sigma_i^2$.

Assume that

(a) $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i \longrightarrow \mu$, i.e., $\lim_{n\to\infty} \mathrm{E}(\overline{X}) = \mu$, and

(b) $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \longrightarrow \sigma^2 < \infty$, ie., $\lim_{n\to\infty} n\mathrm{V}(\overline{X}) = \sigma^2 < \infty$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$,

Note that $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i$ and $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sigma_i^2$.

(iii) Suppose that $X_1, X_2, \cdots, X_n$ are assumed to be serially correlated with $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{Cov}(X_i, X_j) = \sigma_{ij}$.

Assume that

(a) $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i \longrightarrow \mu$, i.e., $\lim_{n\to\infty}\mathrm{E}(\overline{X}) = \mu$, and

(b) $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij} \longrightarrow \sigma^2 < \infty$, ie., $\lim_{n\to\infty} n\mathrm{V}(\overline{X}) = \sigma^2 < \infty$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$,

Note that $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i$ and $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}$.

9. **Some Formulas of Expectaion and Variance in Multivariate Cases**

— Review:

A vector of randam variavle X: $\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) \equiv \mathrm{E}((X-\mu)(X-\mu)') = \Sigma$

Then, $\mathrm{E}(AX) = A\mu$ and $\mathrm{V}(AX) = A\Sigma A'$.

Proof:

$\mathrm{E}(AX) = A\mathrm{E}(X) = A\mu$

$$\begin{aligned}\mathrm{V}(AX) &= \mathrm{E}((AX - A\mu)(AX - A\mu)') = \mathrm{E}(A(X-\mu)(A(X-\mu))') \\ &= \mathrm{E}(A(X-\mu)(X-\mu)'A') = A\mathrm{E}((X-\mu)(X-\mu)')A' = A\mathrm{V}(X)A' = A\Sigma A'\end{aligned}$$

10. **Asymptotic Normality of MLE — Proof:**

The density (or probability) function of X_i is given by $f(x_i;\theta)$.

The likelihood function is: $L(\theta; x) \equiv f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$.

MLE of θ results in the following maximization problem:

$$\max_{\theta} \quad \log L(\theta; x).$$

A solution of the above problem is given by MLE of θ, denoted by $\tilde{\theta}$.

That is, $\tilde{\theta}$ is given by the θ which satisfies the following equation:

$$\frac{\partial \log L(\theta; x)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log f(x_i; \theta)}{\partial \theta} = 0.$$

Replacing x_i by the underlying random variable X_i, $\dfrac{\partial \log f(X_i; \theta)}{\partial \theta}$ is taken as the ith random variable, i.e., X_i in the **Central Limit Theorem II**.

Consider applying **Central Limit Theorem II** as follows:

$$\frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta} - \mathrm{E}\Big(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)}{\sqrt{\mathrm{V}\Big(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)}} = \frac{\dfrac{1}{n}\dfrac{\partial \log L(\theta;X)}{\partial\theta} - \mathrm{E}\Big(\dfrac{1}{n}\dfrac{\partial \log L(\theta;X)}{\partial\theta}\Big)}{\sqrt{\mathrm{V}\Big(\dfrac{1}{n}\dfrac{\partial \log L(\theta;X)}{\partial\theta}\Big)}}.$$

Note that

$$\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta} = \frac{\partial \log L(\theta;X)}{\partial\theta}$$

In this case, we need the following expectation and variance:

$$\mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \mathrm{E}\Big(\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big) = 0,$$

and

$$\mathrm{V}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \mathrm{V}\Big(\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big) = \frac{1}{n^2}I(\theta).$$

Note that $\mathrm{E}\Big(\dfrac{\partial \log L(\theta;X)}{\partial\theta}\Big) = 0$ and $\mathrm{V}\Big(\dfrac{\partial \log L(\theta;X)}{\partial\theta}\Big) = I(\theta)$.

Thus, the asymptotic distribution of

$$\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}$$

is given by:

$$\begin{aligned}
&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta} - \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)\right)\\
&= \sqrt{n}\left(\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta} - \mathrm{E}\Big(\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)\right)\\
&= \frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta} \longrightarrow N(0,\Sigma)
\end{aligned}$$

where

$$n\mathrm{V}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \frac{1}{n}\mathrm{V}\Big(\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \frac{1}{n}\mathrm{V}\Big(\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$$

$$= \frac{1}{n}I(\theta) \longrightarrow \Sigma.$$

That is,

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta; X)}{\partial \theta} \longrightarrow N(0, \Sigma),$$

where $X = (X_1, X_2, \cdots, X_n)$.

Now, replacing θ by $\tilde{\theta}$, consider the asymptotic distribution of

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\tilde{\theta}; X)}{\partial \theta},$$

which is expanded around $\tilde{\theta} = \theta$ as follows:

$$0 = \frac{1}{\sqrt{n}}\frac{\partial \log L(\tilde{\theta}; X)}{\partial \theta} \approx \frac{1}{\sqrt{n}}\frac{\partial \log L(\theta; X)}{\partial \theta} + \frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}(\tilde{\theta} - \theta).$$

Therefore,

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}(\tilde{\theta} - \theta) \approx \frac{1}{\sqrt{n}}\frac{\partial \log L(\theta; X)}{\partial \theta} \longrightarrow N(0, \Sigma).$$

The left-hand side is rewritten as:

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}(\tilde{\theta}-\theta) = \sqrt{n}\left(-\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\right)(\tilde{\theta}-\theta).$$

Then,

$$\sqrt{n}(\tilde{\theta}-\theta) \approx \Big(-\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$$

$$\longrightarrow N(0,\Sigma^{-1}\Sigma\Sigma^{-1}) = N(0,\Sigma^{-1}).$$

Using the law of large number, note that

$$-\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'} \longrightarrow \lim_{n\to\infty}\frac{1}{n}\left(-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big)\right)$$

$$= \lim_{n\to\infty}\frac{1}{n}\left(\mathrm{V}\Big(\frac{\partial \log L(\theta;X)}{\partial}\Big)\right) = \lim_{n\to\infty}\frac{1}{n}I(\theta) = \Sigma,$$

and $\Big(\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$ has the same asymptotic distribution as

$$\Sigma^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta; X)}{\partial \theta}\Big).$$

11. **Optimization (最適化):**

MLE of θ results in the following maximization problem:

$$\max_{\theta} \quad \log L(\theta; x).$$

We often have the case where the solution of θ is not derived in closed form.

$\Longrightarrow$ Optimization procedure

$$0 = \frac{\partial \log L(\theta; x)}{\partial \theta} = \frac{\partial \log L(\theta^*; x)}{\partial \theta} + \frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}(\theta - \theta^*).$$

Solving the above equation with respect to θ, we obtain the following:

$$\theta = \theta^* - \left(\frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log L(\theta^*; x)}{\partial \theta}.$$

Replace the variables as follows:

$$\theta \longrightarrow \theta^{(i+1)}, \qquad \theta^* \longrightarrow \theta^{(i)}.$$

Then, we have:

$$\theta^{(i+1)} = \theta^{(i)} - \left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta}.$$

$\Longrightarrow$ **Newton-Raphson method (ニュートン・ラプソン法)**

Replacing $\dfrac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}$ by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}\right)$, we obtain the following optimization algorithm:

$$\begin{aligned}\theta^{(i+1)} &= \theta^{(i)} - \left(\mathrm{E}\left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}\right)\right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta} \\ &= \theta^{(i)} + \left(I(\theta^{(i)})\right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta}\end{aligned}$$

$\Longrightarrow$ **Method of Scoring (スコア法)**

2 Bayesian Estimation (ベイズ推定)

Greenberg, E. (2013) *Introduction to Bayesian Econometrics* (2nd ed.)

安藤知寛 (2010) 『ベイズ統計モデリング』 (朝倉書店)

豊田秀樹編 (2008) 『マルコフ連鎖モンテカルロ法』 (朝倉書店)

Dey, D.K. and Rao, C.R., (2005) *Handbook of Statistics, Vol.25: Bayesian Thinking: Modeling and Computation*

繁桝・岸野・大森監訳 (2011) 『ベイズ統計分析ハンドブック』 (朝倉書店)

2.1 Introduction

Two Events: A and B

Conditional Probability:

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

Posterior Distribution (事後分布): $f_{\theta|y}(\theta|y)$:

$$f_{\theta|y}(\theta|y) = \frac{f_{y|\theta}(y|\theta)f_\theta(\theta)}{f_y(y)} = \frac{f_{y|\theta}(y|\theta)f_\theta(\theta)}{\int f_{y|\theta}(y|\theta)f_\theta(\theta)\mathrm{d}\theta} \propto f_{y|\theta}(y|\theta)f_\theta(\theta),$$

where $f_\theta(\theta)$ is called the prior distribution (事前分布).

Example 1: Let x be the number of successes in a series of n trials with probability θ of success in each.

That is, x has the binomial probability function, given θ,

$$f_{x|\theta}(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \qquad x = 0, 1, \cdots, n.$$

θ is assumed to be the beta distribution:

$$f_\theta(\theta) = \frac{1}{B(p,q)} \theta^{p-1} (1-\theta)^{q-1},$$

for $\leq \theta \leq 1$, which corresponds to a prior distribution.

Before applying Bayes' theorem, $f_x(x)$ is given by:

$$\begin{aligned} f_x(x) &= \int f_{x|\theta}(x|\theta) f_\theta(\theta) \mathrm{d}\theta \\ &= \binom{n}{r} \frac{1}{B(p,q)} \int_0^1 \theta^{p+x-1} (1-\theta)^{q+n-x-1} \mathrm{d}\theta \\ &= \binom{n}{r} \frac{B(p+x, q+n-x)}{B(p,q)}. \end{aligned}$$

The posterior distribution of θ is:

$$f_{\theta|x}(\theta|x) = \frac{1}{B(p+x, q+n-x)}\theta^{p+x-1}(1-\theta)^{q+n-x-1},$$

which is also a beta distribution with prameters $p+x$ and $q+n-x$.

The posterior mean and variance are:

$$\mathrm{E}(\theta|x) = \frac{p+x}{p+q+n}, \qquad \mathrm{V}(\theta|x) = \frac{(p+x)(q+n-x)}{(p+q+n)^2(p+q+n+1)}.$$

Example 2: $x|\theta \sim N(\theta, v)$, where v is known.

$\theta \sim N(m, w)$, where m and w are known. $\Longrightarrow$ prior dist.

Then, the posterior distribution of θ is:

$$\theta|x \sim N\left(\frac{wx+vm}{w+v}, \frac{vw}{w+v}\right).$$