

確率分布の尤度比検定は最強力検定

黒木玄

2017年10月19日

概要

2つの確率分布に関する検定において、尤度比検定が最強力であることは易しい。

目次

1	最強力検定の定義	1
2	確率分布の尤度比検定は最強力検定	2
3	Bayes 検定の例	4
3.1	指数型分布族モデルの場合	4
3.2	二項分布モデルの場合	5
3.3	正規分布モデルの場合	6

1 最強力検定の定義

簡単のため確率密度関数 $p(x)$, $q(x)$ を持つ確率分布のみを考え、確率密度関数を確率分布と呼んでしまうことにする。

確率分布 $p(x)$ に従う確率変数 X を確率分布 $p(x)$ のサンプルと呼ぶ。

確率分布 $p(x)$ のもとでの事象 A の確率を $P(A|p)$ と書くことにする。例えば、確率変数 X が確率分布 $p(x)$ に従っているという仮定のもとでの $f(X) > c$ となる確率を $P(f(X) > c|p)$ と書く。

確率分布 $p_0(x)$, $p_1(x)$ について次の2つの仮説を考える：

帰無仮説 X は確率分布 $p_0(x)$ のサンプルである。

対立仮説 X は確率分布 $p_1(x)$ のサンプルである。

関数 f と定数 c に対する (f, c) 検定の手続きを以下のように定める：

- (1) 未知の確率分布のサンプル X を採取する。
- (2) $f(X)$ の値を計算する。
- (3) もしも $f(X) > c$ ならば帰無仮説を棄却する。
(対立仮説の方がもっともらしいと判断する。)

(4) もしも $f(X) \leq c$ ならば帰無仮説を棄却しない.

(f, c) 検定の有意水準を帰無仮説が正しいのに、帰無仮説が棄却される確率

$$P(f(X) > c | p_0)$$

と定義し、 (f, c) 検定の検出力を対立仮説が正しいときに、対立仮説の方がもっともらしいと判断する確率

$$P(f(X) > c | p_1)$$

と定義する.

有意水準は低い方がよく、検出力は高い方がよい. しかし、一般に、有意水準を低くすると検出力も低くなる. 有意水準が同じならば検出力が高い検定の方が優れていると考えられる.

(f, c) 検定と (g, d) 検定の有意水準が等しいとき、 (f, c) 検定が (g, d) 検定よりも強力であるとは、 (f, c) 検定の方が (g, d) 検定よりも検出力が高いことであると定める. (f, c) 検定が同じ有意水準の検定の中で最強力なとき (f, c) 検定は最強力検定であると言う.

もしも任意の定数 c に対して、 (f, c) 検定が最強力検定であるとき、函数 f は最強力検定を与えると言うことにする.

2 確率分布の尤度比検定は最強力検定

前節の設定をそのまま引き継ぐ.

2つの確率分布 $p_0(x), p_1(x)$ のあいだの検定については次の一般的な定理が成立している.

定理 2.1. 尤度比函数 $L(x)$ を次のように定義する:

$$L(x) = \frac{p_1(x)}{p_0(x)}$$

と定義する. 尤度比函数は最強力検定を与える.

証明. 定数 a を任意に固定し、 (f, c) 検定と尤度比検定 (L, a) は同じ有意水準を持つと仮定する. すなわち

$$P(L(X) > a | p_0) = P(f(X) > c | p_0).$$

と仮定する. このとき任意の確率分布 p に関する確率 $P(\cdot) = P(\cdot | p)$ について

$$\begin{aligned} P(L(X) > a) &= P(L(X) > a, f(X) > c) + P(L(X) > a, f(X) \leq c), \\ P(f(X) > c) &= P(L(X) > a, f(X) > c) + P(L(X) \leq a, f(X) > c), \end{aligned}$$

なので、これらの差を取ると共通部分の確率はキャンセルする (これが証明のポイント!). ゆえに、 $L(x) > a$ と $p_1(x) > p_0(x)$ が同値であることに注意すると、

$$\begin{aligned} & P(L(X) > a|p_1) - P(f(X) > c|p_1) \\ &= P(L(X) > a, f(X) \leq c|p_1) - P(L(X) \leq a, f(X) > c|p_1) \\ &= \int_{p_1(x) > ap_0(x), f(x) \leq c} p_1(x) dx - \int_{p_1(x) \leq ap_0(x), f(x) > c} p_1(x) dx \\ &\geq \int_{p_1(x) > ap_0(x), f(x) \leq c} ap_0(x) dx - \int_{p_1(x) \leq ap_0(x), f(x) > c} ap_0(x) dx \\ &= aP(L(X) > a, f(X) \leq c|p_0) - aP(L(X) \leq a, f(X) > c|p_0) = 0. \end{aligned}$$

これで尤度比検定 (L, a) 検定の方が任意の (f, c) 検定よりも検出力が高いこと

$$P(L(X) > a|p_1) \geq P(f(X) > c|p_1)$$

が示された. □

例 2.2 (NeymanPearson の補題). 独立同分布な試行で生成されたサンプル (X_1, \dots, X_n) が従う確率分布は $\prod_{k=1}^n p(x_k)$ の形になる. その形の確率分布に定理を適用すると、尤度比関数

$$L(x_1, \dots, x_n) = \frac{\prod_{k=1}^n p_1(x_k)}{\prod_{k=1}^n p_0(x_k)}$$

が最強力検定を与える. 確率分布がパラメーター w によって $p(x|w)$ の形で与えられている場合には 2 つのパラメーターの値 w_0, w_1 に関する最強力検定が尤度比関数

$$L(x_1, \dots, x_n) = \frac{\prod_{k=1}^n p(x_k|w_1)}{\prod_{k=1}^n p(x_k|w_0)}$$

によって得られる. □

例 2.3 (Bayes 検定). パラメーター w に関する確率分布 $\varphi(w)$ とパラメーター w 付きの x に関する確率分布 $p(x|w)$ に対して、 (x_1, \dots, x_n) に関する確率分布 $Z(x_1, \dots, x_n)$ が

$$Z(x_1, \dots, x_n) = \int dw \varphi(w) \prod_{k=1}^n p(x_k|w)$$

によって定義される. この形の確率分布に定理を適用すると、パラメーターの確率分布 $\varphi_0(w), \varphi_1(w)$ に関する最強力検定が尤度比関数

$$L(x_1, \dots, x_n) = \frac{\int dw \varphi_1(w) \prod_{k=1}^n p(x_k|w)}{\int dw \varphi_0(w) \prod_{k=1}^n p(x_k|w)}$$

によって得られる. この検定を **Bayes 検定** と呼ぶ. $\varphi_\nu(w) = \delta_{w_\nu}(w)$ (デルタ分布) の場合がちょうど NeymanPearson の補題の場合になっている.

Bayes 検定に関するより詳しい説明については、渡辺澄夫著『ベイズ統計の理論と方法』(2012) の第 6.4 節を参照せよ. □

注意 2.4 (最尤法の尤度比検定は要注意). パラメーター空間 W_ν における最尤法の解を \hat{w}_ν と書いたときの尤度比関数

$$L(x_1, \dots, x_n) = \frac{\prod_{k=1}^n p(x_k | \hat{w}_1)}{\prod_{k=1}^n p(x_k | \hat{w}_0)}$$

は 2 つの確定したパラメーター値 \hat{w}_0, \hat{w}_1 のあいだの最強力検定を与えるが, 上の定理を用いても, 2 つのパラメーター空間 W_0, W_1 のあいだの最強力検定を与えるとは言えない. 特に, 最尤法に関する Wilks の定理の文脈における対数尤度比検定は最強力検定を与えるとは言えない. この点に関しては誤解し易いところなので注意した方がよい.

それに対して, 上の Bayes 検定は W_0, W_1 のそれぞれに台を持つ確率分布 φ_0, φ_1 (例えば W_0, W_1 のそれぞれに台を持つ一様分布) のあいだの最強力検定を与える. \square

3 Bayes 検定の例

ベイズ検定とは, 確率モデル $p(x|w)$ と二つの事前分布 $\varphi_0(w), \varphi_1(w)$ に関する
 帰無仮説 サンプル X は確率分布 $p_0(x) = \int p(x|w)\varphi_0(w) dw$ によって生成された.
 対立仮説 サンプル X は確率分布 $p_1(x) = \int p(x|w)\varphi_1(w) dw$ によって生成された.
 について,

$$L(x) := \frac{\int p(x|w)\varphi_1(w) dw}{\int p(x|w)\varphi_0(w) dw} > a$$

という条件が満たされたら帰無仮説を棄却するという方法で行う検定のことである. この検定は有意水準が

$$\alpha = \int_{L(x) > a} p_0(x) dx$$

に等しい検定の中で最強力である. 以下では Bayes 検定の簡単な例について説明する.

3.1 指数型分布族モデルの場合

$w = (w_1, \dots, w_r)$ と $g = (g_1, \dots, g_r)$ に対して,

$$\langle w, g \rangle = \sum_{i=1}^r w_i g_i$$

と書く. $f(x) = (f_1(x), \dots, f_r(x))$ についても同様である.

パラメーター w を持つ確率密度関数 $p(x|w)$ が指数型分布族であるとは, それが以下の形をしていることだと定義される:

$$p(x|w) = Z(w)^{-1} \exp(-\langle w, f(x) \rangle) q(x),$$

$$Z(w) = \int \exp(-\langle w, f(x) \rangle) q(x) dx.$$

この指数型分布族の共役事前分布族 $\varphi(w|\nu, g)$ は次のように定義される:

$$\begin{aligned}\varphi(w|\nu, g) &= W(\nu, g)^{-1} Z(w)^{-\nu} \exp(-\langle w, g \rangle), \\ W(\nu, g) &= \int Z(w)^{-\nu} \exp(-\langle w, g \rangle) dw.\end{aligned}$$

このとき, パラメーター ν, g 付きの確率密度関数 $p(x|\nu, g)$ を

$$\begin{aligned}p(x|\nu, g) &= \int p(x|w)\varphi(w|\nu, g) dw \\ &= \frac{q(x)}{W(\nu, g)} \int Z(w)^{-(\nu+1)} \exp(-\langle w, g + f(x) \rangle) dw \\ &= q(x) \frac{W(\nu + 1, g + f(x))}{W(\nu, g)}\end{aligned}$$

と定義する. このとき, 2つの事前分布 $\varphi_0(w) = \varphi(w|\nu_0, g_0)$, $\varphi_1(w) = \varphi(w|\nu_1, g_1)$ のあいだの Bayes 検定は次の尤度比によって行われる:

$$L(x) = \frac{p(x|\nu_1, g_1)}{p(x|\nu_0, g_0)} = \frac{W(\nu_1 + 1, g_1 + f(x))}{W(\nu_1, g_1)} \frac{W(\nu_0, g_0)}{W(\nu_0 + 1, g_0 + f(x))}.$$

このように $W(\nu, g)$ の形さえ決定できれば, 指数型分布族の共役事前分布のあいだの Bayes 検定の条件は具体的に書き下せる.

3.2 二項分布モデルの場合

二項分布は $k = 0, 1, \dots, n$ に関する次の形の離散確率分布 $p(k|\theta)$ として定義される:

$$\begin{aligned}p(k|\theta) &= \binom{n}{k} \theta^k (1 - \theta)^{n-k} = (1 - \theta)^n \left(\frac{\theta}{1 - \theta} \right)^k \binom{n}{k} = Z(\beta)^{-1} \exp(-\beta k) q(k), \\ Z(\beta)^{-1} &= (1 - \theta)^n = (1 - e^{-\beta})^n, \quad q(k) = \binom{n}{k}.\end{aligned}$$

ここで $e^{-\beta} = \theta/(1 - \theta)$, すなわち $\theta = e^{-\beta}/(1 + e^{-\beta})$ とおいた. これより二項分布は指数型分布族であることがわかる. 以下では座標系 β ではなく, θ の方を使う.

二項分布の共役事前分布は次の形になる:

$$\begin{aligned}\varphi(\theta|\nu, g) &= W(\nu, g)^{-1} (1 - \theta)^{n\nu} \left(\frac{\theta}{1 - \theta} \right)^g = W(\nu, g)^{-1} \theta^g (1 - \theta)^{n\nu - g}, \\ W(\nu, g) &= B(g + 1, n\nu - g + 1).\end{aligned}$$

ゆえに,

$$\begin{aligned}p(k|\nu, g) &= \binom{n}{k} \frac{B(g + k + 1, n\nu - (g + k) + 1)}{B(g + 1, n\nu - g + 1)} \\ &= \binom{n}{k} \frac{(\alpha + 1) \cdots (\alpha + k) \cdot (\beta + 1) \cdots (\beta + (n - k))}{(\alpha + \beta + 1 + 1) \cdots (\alpha + \beta + 1 + n)}.\end{aligned}$$

ここで $\alpha = g$, $\beta = n\nu - g$ とおいた.

$\alpha = \beta = 0$ すなわち $\nu = g = 0$ のとき

$$p(k|0,0) = \frac{1}{n+1}.$$

これは $k = 0, 1, \dots, n$ に関する離散一様分布である.

$0 < \theta < 1$ であるとする. このとき, $\alpha = N\theta$, $\beta = N(1 - \theta)$ すなわち $g = N\theta$, $n\nu = N$ とおき, $N \rightarrow \infty$ とすると,

$$\lim_{N \rightarrow \infty} p(k|N/n, N\theta) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} = p(k|\theta).$$

これは, 確率 θ の二項分布である.

このように $p(k|\nu, g)$ は極端な場合として離散一様分布と通常二項分布を含んでいると考えてよい.

Bayes 検定を与える尤度比は次のように表わされる:

$$L(x) = \frac{(\alpha_1 + 1) \cdots (\alpha_1 + k) \cdot (\beta_1 + 1) \cdots (\beta_1 + (n - k))}{(\alpha_1 + \beta_1 + 1 + 1) \cdots (\alpha_1 + \beta_1 + 1 + n)} \\ \times \frac{(\alpha_0 + \beta_0 + 1 + 1) \cdots (\alpha_0 + \beta_0 + 1 + n)}{(\alpha_0 + 1) \cdots (\alpha_0 + k) \cdot (\beta_0 + 1) \cdots (\beta_0 + (n - k))}.$$

たとえば $\alpha_1 = \beta_1 = 0$, $\alpha_0 = N\theta_0$, $\beta_0 = N(1 - \theta_0)$ とおいて $N \rightarrow \infty$ とすると

$$L(x) = \frac{1}{(n+1)p(k|\theta_0)}.$$

これは $\varphi_0(\theta) = \delta(\theta - \theta_0)$ (デルタ分布) と $\varphi_1(\theta) = 1$ (一様分布) のあいだの尤度比

$$L(x) = \frac{p(k|0,0)}{p(k|\theta_0)}$$

に一致する. そして, この尤度比による検定の条件 $L(x) > a$ は

$$p(k|\theta_0) < \frac{1}{a(n+1)}$$

と同値である. そしてこの条件は二項分布に関する通常両側仮説検定で使われる条件に等しい.

以上によって, これで二項分布の事前分布に関する Bayes 検定は通常両側仮説検定を特別な場合として含んでいることがわかった.

3.3 正規分布モデルの場合

$p(x|\mu)$ を次のように定める:

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

これは次のように表わされる:

$$p(x|\mu) = \exp\left(-\frac{\mu^2}{2}\right) \exp(\mu x) \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

ゆえにこれの共役事前分布は次のように表わされる:

$$\varphi(\mu|\nu, g) = W(\nu, g)^{-1} \exp(-\nu \frac{\mu^2}{2}) \exp(\mu g).$$

ここで $W(\nu, g)$ の形は以下のようにして決定される:

$$\begin{aligned} W(\nu, g) &= \int_{\mathbb{R}} \exp\left(-\nu \frac{\mu^2}{2} + \mu g\right) d\mu = \int_{\mathbb{R}} \exp\left(-\frac{\nu}{2} \left(\mu - \frac{g}{\nu}\right)^2 + \frac{g^2}{2\nu}\right) d\mu \\ &= \sqrt{\frac{2\pi}{\nu}} \exp\left(\frac{g^2}{2\nu}\right). \end{aligned}$$

ゆえに

$$p(x|\nu, g) = \frac{e^{-x^2/2} W(\nu+1, g+x)}{\sqrt{2\pi} W(\nu, g)} = \sqrt{\frac{\nu}{2\pi(\nu+1)}} \exp\left(-\frac{1}{2} \frac{\nu}{\nu+1} \left(x - \frac{g}{\nu}\right)^2\right).$$

これは平均 g/ν , 分散 $(\nu+1)/\nu = 1+1/\nu > 1$ の正規分布である. $g = \nu\mu$, $\nu = 1/(\rho^2 - 1)$, $\rho^2 > 1$ とおくと

$$\lim_{\rho^2 \rightarrow 1} p(x|\mu/(\rho^2 - 1), 1/(\rho^2 - 1)) = \frac{1}{\sqrt{2\pi\rho^2}} \exp\left(-\frac{(x - \mu)^2}{2\rho^2}\right) =: p_{\text{Normal}}(x|\mu, \rho).$$

ゆえに, 以上のケース内の Bayes 検定で使われる尤度比は分散が 1 より大きい正規分布の比になる.

$\rho^2 \rightarrow 1$ の極限で

$$\lim_{\rho^2 \rightarrow 1} p(x|\mu/(\rho^2 - 1), 1/(\rho^2 - 1)) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2}\right) = p(x|\mu).$$

ゆえに $p(x|\nu, g)$ は極限として, $p(x|\mu)$ を含んでいる. これの μ を μ_0 に置き換えたものはデルタ事前分布 $\varphi_0(\mu) = \delta(\mu - \mu_0)$ に対する $p_0(x)$ に一致する:

$$p_0(x) = \int_{\mathbb{R}} p(x|\mu) \varphi_0(\mu) d\mu = p(x|\mu_0) = p_{\text{Normal}}(x|\mu_0, 1).$$

これと $\varphi_1(x) = \varphi(x|\mu_1/(\rho_1^2 - 1), 1/(\rho_1^2 - 1))$ に対する

$$p_1(x) = \int_{\mathbb{R}} p(x|\mu) \varphi_1(\mu) d\mu = \frac{1}{\sqrt{2\pi\rho_1^2}} \exp\left(-\frac{(x - \mu_1)^2}{2\rho_1^2}\right) = p_{\text{Normal}}(x|\mu_1, \rho_1)$$

のあいだの Bayes 検定を与える尤度比は

$$L(x) = \frac{p_1(x)}{p_0(x)} = \frac{p_{\text{Normal}}(x|\mu_1, \rho_1)}{p_{\text{Normal}}(x|\mu_0, 1)} \quad (\rho_1 > 1)$$

である. Bayes 検定の条件 $L(x) > a$ は

$$p_{\text{Normal}}(x|\mu_0, 1) < a^{-1} p_{\text{Normal}}(x|\mu_1, \rho_1) = \frac{a^{-1}}{\sqrt{2\pi\rho_1^2}} \exp\left(-\frac{(x - \mu_1)^2}{2\rho_1^2}\right)$$

と書き直せる. $a^{-1} = c\sqrt{2\pi\rho_1^2}$ とおき, $\rho_1 \rightarrow \infty$ の極限をとると,

$$p_{\text{Normal}}(x|\mu_0, 1) < c.$$

これは通常の両側仮説検定の条件と同じである.

このように, 分散が 1 に固定されていてパラメーターが平均 μ のみの正規分布族に関する Bayes 検定は通常の両側仮説検定を極限として含んでいると考えられる.