正規分布の分散における優越性と同等性のベイズ流指標および F 検定の p 値との関係

京都大学大学院医学研究科 社会健康医学系専攻 医療統計学 土居正明

生物統計において連続量の応答変数をもつ 2 群の比較を行う際，応答変数の分散が両群で異なると考えられることはよくみられる．応答変数が正規分布に従う場合，頻度論の方法による分散の比較には F 検定がよく用いられる．本研究では「過去の類似した状況のデータを利用できる」「（設定したモデルのもとで）一方の群の分散が他方より大きくなる確率，などの直感的に理解しやすい指標を計算できる」などの利点を持つベイズ流の優越性指標と同等性指標を提案し，F 検定の p 値との関係を検討する．

なお，本稿では紙面の都合上，平均既知の場合に絞って記載する．

1　分散の優越性を示すベイズ流指標

応答変数 $x_{i1},\ldots,x_{in_i}$ に対して $(i=1,2)$，$x_{i1},\ldots,x_{in_i} \overset{ind}{\sim} N(\mu_i,\sigma_i^2)$ とし，$\mathbf{x}_i := (x_{i1},\cdots,x_{in_i})'$ とおく．このとき，分散の優越性を示すベイズ流指標として，$\theta := \Pr\left(\sigma_1^2 > \sigma_2^2 \,|\, \mathbf{x}_1, \mathbf{x}_2\right)$ を考えると，以下が成り立つ．

定理 1　σ_i^2 の事後分布が逆ガンマ分布 $\textit{Inv-Ga}(a_i,b_i)$ $(a_i,b_i>0;\, i=1,2)$ のとき，θ は以下のように表せる．

$$\theta = F_{2a_1,2a_2}\left(\frac{b_1/a_1}{b_2/a_2}\right).$$

ここで，逆ガンマ分布 $\textit{Inv-Ga}(a_i,b_i)$ の確率密度関数は $f(\sigma_i^2 \mid a_i,b_i) = b_i^{a_i}(\sigma_i^2)^{-a_i-1}/\Gamma(a_i)\cdot\exp\left(-b_i/\sigma_i^2\right)$ であり，$F_{n,m}(x)$ は，自由度 (n,m) の F 分布の累積分布関数である．

いま，σ_i^2 の事前分布を逆ガンマ分布 $\textit{Inv-Ga}(\nu_i/2,\nu_i\tau_i^2/2)$ とおく $(i=1,2)$．このとき，分散の事後分布は $\textit{Inv-Ga}(a_i,b_i)$ となる (ここで，$T_i^2 := 1/n_i\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2,\ a_i := (\nu_i+n_i)/2,\ b_i := (\nu_i\tau_i^2 + n_i\cdot T_i^2)/2$)．これより，定理 1 から，ベイズ流指標 θ は以下のように表せる．

$$\theta = F_{\nu_1+n_1,\nu_2+n_2}\left(\frac{(\nu_1\tau_1^2 + n_1\cdot T_1^2)/(\nu_1+n_1)}{(\nu_2\tau_2^2 + n_2\cdot T_2^2)/(\nu_2+n_2)}\right).$$

定理 2　θ と，$H_0:\sigma_1^2=\sigma_2^2,\ H_1:\sigma_1^2>\sigma_2^2$ に対する F 検定の p 値の間に以下の関係が成り立つ．

$$\lim_{(\nu_1,\nu_2)\to(0,0)} \theta = 1-p.$$

この定理は，「事前分布の情報量を 0 に減らした極限において，ベイズ流指標と F 検定の p 値に対する $1-p$ が一致する」ことを示すものと解釈できる．次に，無情報事前分布を直接与えた場合の p 値との対応を示す．

系 1　分散の事前分布が $f(\sigma_i^2)\propto(\sigma_i^2)^{-1}$ のとき $(i=1,2)$，$\theta = 1-p$ が成り立つ．

2　分散の同等性を示すベイズ流指標

Δ を $0<\Delta<1/2$ を満たす定数とする．このとき，分散の同等性を示すベイズ流指標 $\theta_e(\Delta) := \Pr(\Delta < \sigma_1/\sigma_2 < 1/\Delta)$ を考えると，以下が成り立つ．

定理 3　分散の事後分布が $\sigma_i^2 \sim \textit{Inv-Ga}(a_i,b_i)$ のとき $(a_i,b_i>0; i=1,2)$，以下が成り立つ．

$$\theta_e(\Delta) = F_{2a_1,2a_2}\left(\frac{b_1/a_1}{b_2/a_2}\cdot\frac{1}{\Delta^2}\right) - F_{2a_1,2a_2}\left(\frac{b_1/a_1}{b_2/a_2}\cdot\Delta^2\right).$$

系 2　分散の事前分布を $f(\sigma_i^2)\propto(\sigma_i^2)^{-1}$ とおく $(i=1,2)$．ここで，$H_0^{(1)}:\sigma_1/\sigma_2=\Delta,\ H_1^{(1)}:\Delta<\sigma_1/\sigma_2$ に関する F 検定の p 値を $p^{(1)}$，$H_0^{(2)}:\sigma_1/\sigma_2=1/\Delta,\ H_1^{(2)}:\sigma_1/\sigma_2<1/\Delta$ に関する F 検定の p 値を $p^{(2)}$ とおく．このとき，以下が成り立つ．

$$\theta_e(\Delta) = 1-(p^{(1)}+p^{(2)}).$$

当日は，平均未知の場合の対応，および他の分布に関する同様の関係式との関連についても述べる．