

Bonettの方法

ON BONETT'S ROBUST CONFIDENCE INTERVAL FOR A RATIO OF STANDARD DEVIATIONS (標準偏差の比に対するBONETTの頑健な信頼区間について)

Senin J. Banga and Gregory D. Fox
2013年6月18日

要旨

2つの標準偏差の比に対するBonett (2006) の信頼区間 (CI) にあるわずかな誤りを修正する代替手順を提案する。Bonettの信頼区間の基礎となっているLayard (1973) の検定統計量の併合尖度推定量は、2つの母集団の分散が等しい場合にのみ一貫性がある。2つの母集団の分散が等しい場合も等しくない場合も一貫性のある代替推定量を導き出し、この新しい推定量を使用して正しい信頼区間を計算します。シミュレーションの分析により、この新しい信頼区間は、概ねLevene/Browne-Forsytheの検定 W_{50} およびPan (1999) の検定 L_{50} に基づく信頼区間に比べてより正確で精度が高いことが明らかになりました。Panの検定と同様、Bonettの検定 W_{50} に基づく信頼区間では、小さいサンプルで精度の損失が見られ、その結果、区間の幅が無限になる場合がしばしば観測されました。Bonettの検定 L_{50} に基づく信頼区間は、対称的な分布およびほぼ対称的な分布で良好な性能を示しますが、母集団が歪んでいる分布の場合は性能が低下します。

索引用語: 等分散性、Leveneの検定、Brown-Forsytheの検定、Layardの検定、分散の比の信頼区間 (CI)

1. はじめに

古典的F検定、および関連する信頼区間 (CI) は、正規性からの逸脱に対し極端に敏感であることは広く知られています。実際、あまりにも敏感であるため、古典的F検定はほとんどの場合実際の使用には適していません。そのため、より頑健な代替方法が多数提案されてきました。そのなかでも、タイプIの誤りの特性が非常に優れていて、しかも計算と解釈が単純であることから、「検定 W_{50} 」として知られる検定がよく使用されています。(比較すべき分析については、Conover et al. (1981)、Balakrishnan and Ma (1990)、Lim and Loh (1996) を参照してください)。この検定 W_{50} はLevene (1960) によって最初に提案され、後にBrown and Forsythe (1974) で拡張された手順に基づいています。この検定 W_{50} は、広く採用されており、Minitab統計ソフトウェア、SAS、R、JMPなど、よく知られたほとんどの統計ソフトウェアパッケージに含まれています。

検定 W_{50} のタイプIIの誤りの特性は、タイプIの誤りの特性と比べてあまり注目されていません。Pan (1999) は、正規分布を含む一部の分布では、2サンプル問題での検定 W_{50} の検出力の上限は、1よりかなり下になる可能性があることを示しています。また、この上限は2つの母分散の差の大きさによる影響を受けません。この欠陥は当然、検定 W_{50} に基づく信頼区間にまで及びます。Panは、検定 W_{50} に基づく母分散の比の信頼区間が無限 $(0, +\infty)$ になるため、何も情報を得られない可能性が無視できない程度に存在することを示しています。Panの観察は、本書で後述する独自のシミュレーション結果と一致します。

Panは、 W_{50} 手順の限界を修正するため、 L_{50} と呼ばれる代替手順を提案しています。Panは、シミュレーション結果に基づき、検定 L_{50} は検定 W_{50} よりも強力であり、さらに等しく頑健で望ましい漸近特性を共有していると結論付けています。しかしながら、Panのシミュレーションのサンプルは、軽い裾から重い裾までを持つ対称分布または少し歪んだ分布から抽出されました。小さいサンプルでの L_{50} 検定の性能に対する歪みの潜在的影響については、特に説明されていません。

またPanは、 L_{50} 手順は、修正されたFligner-Killeenの順位検定やHall-Padmanabhanの適応型テストなどの、とりわけ頑健な他の手順と同様に強力であると主張しています。ただし、実際には、修正されたFligner-Killeenの順位検定とHall-Padmanabhanの適応型テストは計算に労力と時間を要し、骨の折れることであるため、検定 L_{50} および検定 W_{50} よりも有用度がいくぶん低くなります。

最近、Bonett (2006) は、Layard (1973) の等分散性検定の2サンプル版に基づく代替の信頼区間手順を提案しました。Bonettは、Layardの手順の小サンプル性能を向上するために、いくつかの調整を行っています。たとえば、BonettはLayardの併合尖度推定量と漸近的に等しいが、小サンプルの偏りがより少ない併合尖度推定量を提案しています。

残念ながら、母集団の分散が等しくない場合は、Layardの元々の併合尖度推定量とBonettが提案した代替の併合尖度推定量は、どちらも一貫性がありません。したがって、Bonett (2006) が提案する区間は、厳密な意味での信頼区間ではなく、等分散性検定の合格区間と表現するのがより適切だと言えます。したがって、Bonett (2006) で報告されたシミュレーションした被覆確率を1から引くと、等分散性検定のタイプI過誤率が得られます。このタイプI過誤率とLayardの元の検定のタイプI過誤率を比較すると、Bonettの調整によりLayardの検定の小サンプル性能が向上することが確認されます。ただし、Bonettが提案した分散の比の信頼区間は再考が必要です。

またBonettは、提案した区間を、Shoemaker (2003) の近似F検定に基づく信頼区間とも比較しています。ただし、Shoemakerの検定に関連する分散比の信頼区間は、Shoemakerの論文106ページに簡潔に記述されているように、Layardの併合尖度推定量にも基づいています。したがって、Shoemakerの論文の第7項で計算された信頼区間もまた、等分散性検定の合格区間と表現するのが最も適切です。これらの誤差にもかかわらず、Bonettのシミュレーション結果から、彼が行った調整によりLayardの等分散性検定の小サンプル性能が向上し、調整された等分散性検定はShoemakerの検定より性能が高いと結論付けることができます。

本書では、Layardの検定の2サンプル形式を拡張することでBonett (2006) のわずかな誤りを修正し、分散または標準偏差の比に関する帰無仮説を検定します。これを達成するために、任意の仮説率に対して一貫性がある併合尖度推定量を提案します。次に、検定統計量を逆変換し、比の信頼区間を得ます。最後に、シミュレーションの分析を行い、小サンプル計画における新しい信頼区間の頑健性特性を評価します。さらに、新しい信頼区間の小サンプル性能と、古典的F検定、検定 W_{50} 、および検定 L_{50} に関連する信頼区間の性能を比較します。

2. Layardの検定と拡張

$Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$ を k (個の) 独立サンプルとします。各サンプルは独立していて、まったく同じように分布され、平均は $E(Y_{ij}) = \mu_i$ 、分散は $\text{Var}(Y_{ij}) = \sigma_i^2 > 0$ です。さらに、サンプルは共通尖度 $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$ を持つ母集団から得ると仮定します。Layardが尖度過剰 $\gamma_e = \gamma - 3$ を使用している点に留意します。

\bar{Y}_i および S_i を、それぞれサンプル i の平均および標準偏差とします。また、 $\tau^2 = 2 + (1 - 1/\bar{n})\gamma_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$ とします。ここで $\bar{n} = \sum n_i / k$ です。Layard (1973) に示されているように、大きなサンプルの場合は、 $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$ です。

等分散性の帰無仮説を検定するため、Layardは帰無仮説のもとで標準正規分布として漸近的に分布されている成分 $Z_i = (n_i - 1)^{1/2} \ln S_i^2 / \tau$ のベクトルで直行変換を行います。次に、Layardは直行変換の距離維持の特性を利用して、(下で与えられる) 検定統計量 S' は等分散性の帰無仮説のもとで自由度 $k - 1$ のカイ二乗分布として漸近的に分布されることを示します。

$$S' = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \frac{\sum_{i=1}^k (n_i - 1) \ln S_i^2}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

一般に、 $Z_i = (n_i - 1)^{1/2} (\ln S_i^2 - \ln \sigma_i^2) / \tau$ は標準正規分布として漸近的に分布されます。したがって、Layardの手法を適用して、より一般化された検定統計量 T_k' を導き出すことができます。

$$T_k' = \sum_{i=1}^k (n_i - 1) \frac{(\ln S_i^2 - \ln \sigma_i^2)^2}{\tau^2} - \left(\sum_{i=1}^k (n_i - 1) \frac{\ln S_i^2 - \ln \sigma_i^2}{\tau \sqrt{\sum_{i=1}^k (n_i - 1)}} \right)^2$$

T_k' は帰無仮説と対立仮説の両方のもとで、自由度 $k - 1$ のカイ二乗分布として漸近的に分布されます。

T_k' は、 S' によく見られる形で表現できます。2次項を二重和として表現し、代数計算を行うと、次のようになります。

$$T_k' = \sum_{i=1}^k (n_i - 1) \left(\ln S_i^2 - \ln \sigma_i^2 - \frac{\sum_{i=1}^k (n_i - 1) (\ln S_i^2 - \ln \sigma_i^2)}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

すべての分散が等しい場合、 $T_k' = S'$ です。したがって、等分散性の帰無仮説を検定する場合、 S' と T_k' は同じ検定統計量です。ただし、 T_k' は分散の関数として表現される仮説を検定するために、より一般的に使用することもできます。たとえば、 T_k' を使用して、任意の $\sigma_{0i} > 0, i = 1, \dots, k$ について、 $H_0: \sigma_i = \sigma_{0i}$ の形の帰無仮説を検定できます。

$\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ は未知であるため、 S' または T_k' に基づく検定には、母集団の共通尖度の推定量 γ が必要です。たとえば、等分散性の帰無仮説を検定するために、Layardは次の共通尖度の併合推定量を提案しています。

$$\hat{\gamma} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right]^2} \sum_{i=1}^k n_i$$

ただしLayardは、分散が等しくない場合、 $\hat{\gamma}$ は必ずしも一貫性のある共通尖度推定量ではないと指摘しています。

2サンプル計画の特別なケースでは、任意の仮説率 $\rho_0 > 0$ について帰無仮説 $H_0: \sigma_1/\sigma_2 = \rho_0$ を検定することで、標準偏差の差の大きさを評価できます。ただし、この差は標準偏差の比の信頼区間を計算することで、より直接的に評価できます。

$\rho_0 = 1$ の場合、帰無仮説は等分散性の仮説と同じです。したがって、 $\hat{\tau}^2$ を求めるために、 $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$ の式の γ の代わりにLayardの2サンプルバージョンの尖度推定量を使用した後、検定を $T_2' = S'$ に基づかせることができます。

ただし、 $\rho_0 \neq 1$ の場合、検定は S' ではなく、 T_2' に基づく必要があります。さらに、 $\rho_0 \neq 1$ の場合、Layardの併合尖度推定量は必ずしも一貫性があるわけではないので、母集団の共通尖度を推定するためには使用できません。したがって、仮説率 $\rho_0 > 0$ に左右されない代替の併合尖度推定量が必要となります。

次に、そのような推定量を導き出します。これは ρ_0 の関数であるため、推定量を $\hat{\gamma}_P(\rho_0)$ として表します。また、検定統計量 $T_2 = \tau^2 T_2' / \hat{\tau}^2$ も定義します。ここで、 $\hat{\tau}^2 = 2 + (1 - 1/\bar{n})(\hat{\gamma}_P(\rho_0) - 3)$ です。Slutzkyの定理によれば、 T_2 は自由度1のカイ二乗分布として漸近的に分布されます。最後に T_2 を逆変換し、 $\rho = \sigma_1/\sigma_2$ の信頼区間を得ます。

3. 標準偏差の比の信頼区間

前のセクションでは、分散または標準偏差の比に関して記述された帰無仮説を検定する場合の、代替の尖度推定量の必要性について詳しく説明しました。次の結果から、その推定量が得られます。

結果1

$\rho = \sigma_1/\sigma_2 > 0$ の場合、2サンプルモデルの共通母集団尖度の一貫性のある併合尖度推定量は、次のように与えられます。

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

この結果の証明は、「付録A」を参照してください。

予想どおり、 $\sigma_1/\sigma_2 = 1$ は標準偏差（または分散）が等しいことを意味しているため、 $\hat{\gamma}_P(1)$ はLayardの併合尖度推定量 $\hat{\gamma}$ と同一になります。

統計量 T_2' は、2サンプルバージョンの一般統計量 T_k' であり、次のように与えられます。

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) \tau^2}$$

ここで、 $\hat{\rho} = S_1/S_2$ 、 $\rho = \sigma_1/\sigma_2$ 、および $\tau^2 = 2 + (1 - 1/\bar{n})\hat{\gamma}_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$ です。

Layard (1973) で示されたように、大きなサンプルの場合は、 $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$ です。Bonett (2006) は、代替の近似 $\text{Var}((n_i - 1)^{1/2} \ln S_i^2) \cong \gamma - (n_i - 3)/n_i$ を使用しており、これは、Shoemaker (2003) でも採用されています。大きなサンプルでは、これらの近似は同等です。ただし、Shoemakerは後者のバージョンは、小さいサンプルで自身が提案している等分散性検定を使用する場合に有利であると報告しています。この調整を使用して、統計量 T_2' は次のように修正できます。

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

ここで、 $g_i = (n_i - 3)/n_i$ です。

次に、帰無仮説 $H_0: \rho = \rho_0$ を検定するための検定統計量 $T_2 = \tau^2 T_2' / \hat{t}^2$ は、次のように与えられることになります。

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{\gamma}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho_0) - g_2}{n_2 - 1}}$$

この T_2 の式で、分母の平方根は、併合尖度の標準偏差の大きなサンプルの推定とみなすことができます。

さらに、 $\hat{\gamma}_P(1) \equiv \hat{\gamma}$ の式で、Bonett (2006) は調整比率 $1/[2(n_i - 4)^{1/2}]$ の調整サンプル平均を使用しています。これに応じて、併合尖度推定量に同じ調整を行います。

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

ここで、 m_i は、調整比率 $1/[2(n_i - 4)^{1/2}]$ を使用した、サンプル i の調整平均です。調整平均 m_i は、母集団平均 μ_i の一貫性のある推定量なので、このバージョンの併合尖度推定量と前述のバージョンは、漸近的に等しくなります。ただし、この代替のバージョンで、 T_2 に基づく検定の小サンプル性能が向上する場合があります。

これで、検定統計量 T_2 を逆変換し、分散または標準偏差の比の近似信頼区間を導出できます。ただし、その前に、標準偏差の比の信頼区間の導出でのBonett (2006) のわずかな誤りについて簡単に説明します。

3.1 Bonettの区間

信頼区間を得るのに T_2 を逆変換するかわりに、Bonett (2006) は次の統計量を逆変換しています。

$$T = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\hat{\gamma}_P(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(1) - g_2}{n_2 - 1}}$$

その結果として得た区間は、等分散性検定の単なる合格領域です。これは、併合尖度推定量 $\hat{\rho}_p(1)$ は、分散が等しい場合、または仮説率が1の場合にのみ一貫性があるからです。結果として得た区間は、Bonett (2006) で次のように報告されています。

$$\exp[\ln(c S_1^2/S_2^2) \pm z_{\alpha/2} se]$$

ここで

$$se^2 = \frac{\hat{y}(1) - g_1}{n_1 - 1} + \frac{\hat{y}(1) - g_2}{n_2 - 1}$$

アンバランス型計画で等しくない裾部による誤差の確率の影響を削減するために、定数 c が小サンプル調整として含まれています。この定数は、次によって与えられます。

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

計画がバランス型で、サンプルサイズの増加とともにその影響を無視できるようになると、定数は0とみなせます。

表1に、上記の区間を信頼区間と誤って解釈した場合の結果を示します。これらの結果は、Bonett (2006) の区間に基づいてシミュレートした被覆確率を計算する小規模なシミュレーションの分析に基づいています。等分散のケース（左列）では、標準正規分布から2つの独立サンプルを抽出します。不等分散のケース（右列）では、2番目のサンプルの観測値を定数係数4でスケールリングします。推定される被覆確率は100,000反復に基づきます。目標とする名目被覆は、0.95です。

表1 Bonett (2006) の信頼区間 ($\alpha = 0.05$) に対する母集団の分散が等しくない場合の影響

n_1, n_2	シミュレートした被覆確率	
	等分散	不等分散
10, 10	0.963	0.972
50, 50	0.952	0.991
100, 100	0.952	0.994

区間が一貫性のある併合尖度推定量に基づいていた場合、2つのケースの被覆確率は同一になることが期待されます。しかしながら、分散が等しくない場合は、区間は一貫してより控え目になることに注目してください。さらに、被覆確率はサンプルサイズが増加するにつれて1に近づきます。Shoemaker (2003) の近似信頼区間でも同様の結果が得られます。

3.2 信頼区間の計算

対立仮説 $H_A: \rho \neq \rho_0$ に対する帰無仮説 $H_0: \rho = \rho_0$ の検定の問題について検討してみます。前述の検定統計量 T_2 に基づき、 $\rho = \sigma_1/\sigma_2$ および $\rho_0 > 0$ です。帰無仮説では、検定統計量

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{y}_P(\rho_0) - g_2}{n_2 - 1}}$$

は自由度1のカイ二乗分布として漸近的に分布されます。そのため、次の場合にのみ、検定では有意水準 α で帰無仮説が棄却されます。

$$(\ln \hat{\rho}^2 - \ln \rho_0^2)^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

ここで、 z_{α} は標準正規分布の $\alpha \times 100$ 番目の上位百分位数点を表します。自由度1のカイ二乗分布 $\chi_{1,\alpha}^2$ の $\alpha \times 100$ 番目の上位百分位数点は、 $\chi_{1,\alpha}^2 = z_{\alpha/2}^2$ という条件を満たしています。

Bonett (2006) のシミュレーション結果は、アンバランス型計画で等しくない裾部による誤差の影響を削減する小サンプル調整が有効であったことを示しています。したがって、 T_2 に基づいて、検定に同様の調整を行います。この調整を行うと、検定では次の場合にのみ帰無仮説が棄却されます。

$$(\ln \rho_0^2 - \ln(c\hat{\rho}^2))^2 > z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

ここで、 c は次のように与えられるBonettの調整定数です。

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

同等に、 T_2 に基づく $\rho = \sigma_1/\sigma_2$ の約 $(1 - \alpha)100$ パーセントの信頼セットは、次のように与えられます。

$$\left\{ \rho \in (0, \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{Y}_P(\rho) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho) - g_2}{n_2 - 1} \right) \leq 0 \right\}$$

ただし、 c はバランス型計画にはまったく影響せず、アンバランス型計画の大サンプルに無視できる程度の影響しか与えません。

次の結果で、その性質を説明する便利な形で信頼セットの代替式を示します。この式では、併合尖度推定量は次のように個々のサンプル尖度として書き換えられます。

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

結果2

T_2 に基づく $\rho = \sigma_1/\sigma_2$ の約 $(1 - \alpha)100$ パーセントの信頼セットは、次のように表現できます。

$$\hat{\rho}\sqrt{c} \{r \in (0, \infty): H(r^2) \leq 0\}$$

または、 $\rho^2 = \sigma_1^2/\sigma_2^2$ の信頼セットを次のように表現できます。

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

ここで

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{g_1}{n_1 - 1} + \frac{g_2}{n_2 - 1}, K = \frac{n_1 - 1}{n_2 - 1}$$

この結果の証明は、「付録B」を参照してください。

$H(0) = H(+\infty) = +\infty$ と $H(1) < 0$ を使用すると、関数 $H(x)$ は正の実数直線上で連続していることが簡単に検証されます。したがって、中間値の定理によって、関数 $H(x)$ では区間 $(0, 1)$ で少なくとも1つの根、区間 $(0, +\infty)$ で少なくとも1つの根が認められます。

次の結果で、信頼セットを区間または互いに素な区間の和集合として説明します。

結果3

$H(x)$ に正確に2つの根 x_L と x_U がある場合、 $0 < x_L < 1 < x_U$ および $\rho^2 = \sigma_1^2/\sigma_2^2$ の信頼セットは、次のように与えられる区間です。

$$[c\hat{\rho}^2x_L, c\hat{\rho}^2x_U]$$

したがって、 $\rho = \sigma_1/\sigma_2$ の信頼区間は次の区間になります。

$$[\hat{\rho}\sqrt{cx_L}, \hat{\rho}\sqrt{cx_U}]$$

一方、関数 $H(x)$ に3つ以上の根がある場合、 $\rho^2 = \sigma_1^2/\sigma_2^2$ の信頼セットは重なり合わない区間の和集合です。各区間のエンドポイントは関数が上向きに開く連続する根です。

この結果の証明は、「付録C」を参照してください。

注釈

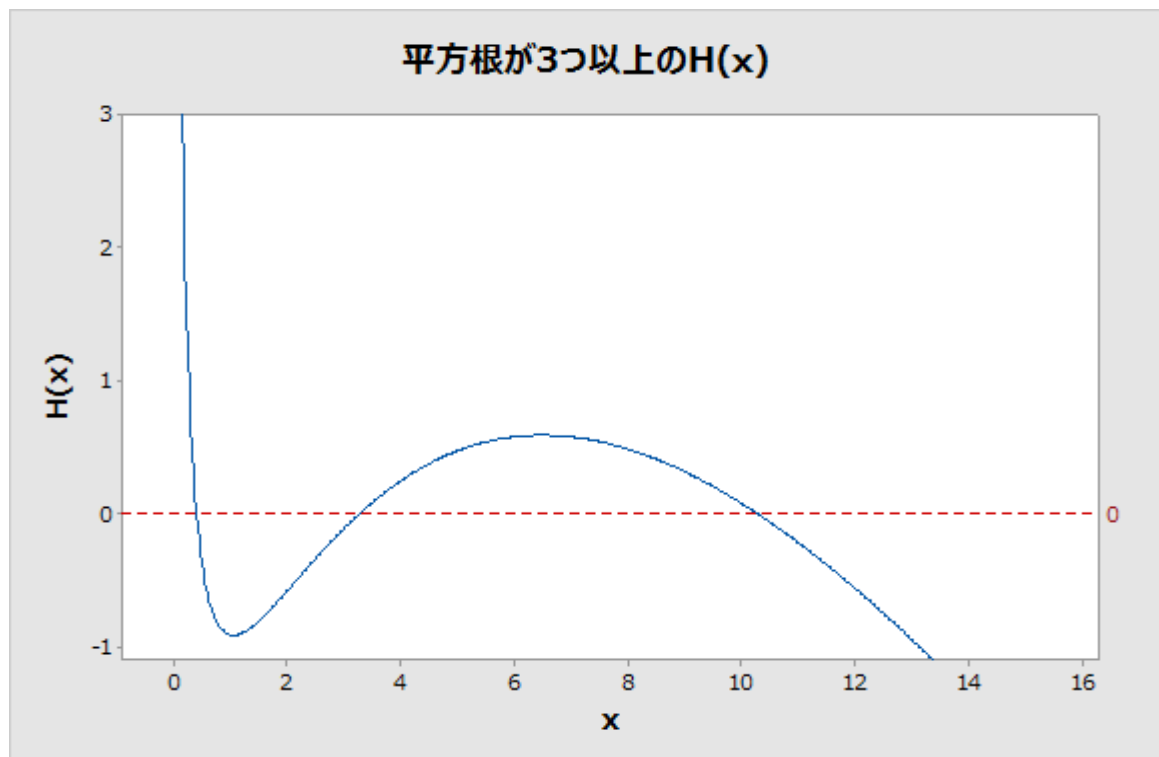
関数 $H(x)$ で3つ以上の根が認められることは数学的に可能ですが、これは片方または両方のサンプルが小さすぎるか、ひどくアンバランスである極端に異常で実質的に無意味な計画でのみ起こることを観測しました。 $H(x)$ には、2つまたは4つの根があると推測します。

次の例は、関数 $H(x)$ が強制的に3つ以上の根を持つように作成したデータに基づいています。データを要約すると、 $n_1 = 169$ 、 $n_2 = 7$ 、 $S_1 = 301.855$ 、 $S_2 = 4606.170$ 、 $\hat{\gamma}_1 = 1.877$ 、 $\hat{\gamma}_2 = 6.761$ 、 $c = 0.728$ 、 $A = 30.381$ 、 $B = 0.101$ 、 $K = 28.000$ となります。

$\alpha = 0.05$ の場合、関数 $H(x)$ は次のようになります。

$$H(x) = (\ln x)^2 - 1.960^2 \left(30.381 \frac{1.877 \times 28^2/169 + 6.761 \times (.728x)^2/7}{(28.000 + 0.728x)^2} - 0.101 \right)$$

このケースの関数 $H(x)$ には4つの根があります。関数のグラフを下に示します。4番目の根は大きすぎるため、グラフには表示されていません。ただし、 $H(+\infty) = +\infty$ であるため、4番目の根が存在することがわかります。



4つの根は $x_1 = 0.389$ 、 $x_2 = 3.282$ 、 $x_3 = 10.194$ 、 $x_4 = 39685.0$ として数値的に計算されます。標準偏差の推定比は $\hat{\rho} = S_1/S_2 = 0.066$ です。 $\rho^2 = \sigma_1^2/\sigma_2^2$ の信頼セットを次のように表現できます。

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] \cup [c \hat{\rho}^2 x_3, c \hat{\rho}^2 x_4] = [0.001, 0.010] \cup [0.032, 124.072]$$

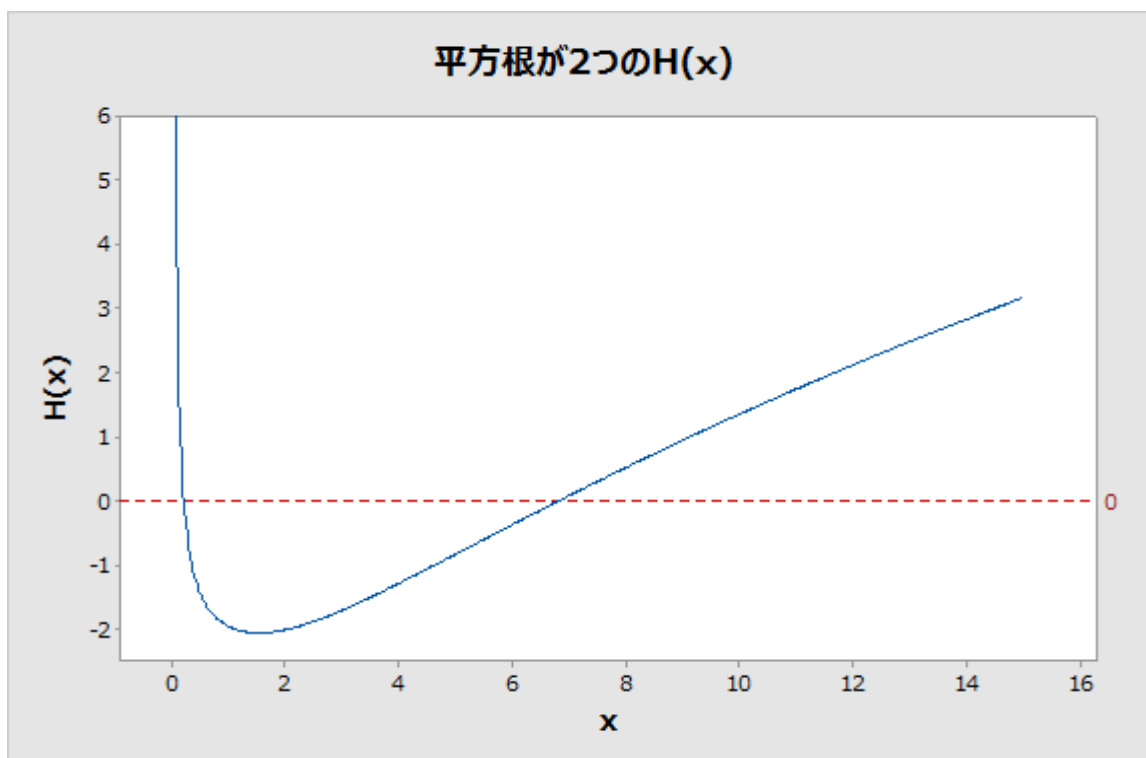
標準偏差の比 ρ の信頼セットは、区間のエンドポイントの平方根を取ることで得られます。

サンプルが小さすぎず ($n_i \geq 10$)、サイズの違いが大きくない場合、関数 $H(x)$ では一般に2つの根が認められます。結果2で説明したように、一方の根は1未満、他方の根は1より大きくなります。次に、ランダムに生成されたデータに基づく例を示します。データを要約すると、 $n_1 = 10$ 、 $n_2 = 12$ 、 $S_1 = 1.150$ 、 $S_2 = 1.043$ 、 $\hat{\gamma}_1 = 2.704$ 、 $\hat{\gamma}_2 = 3.671$ 、 $c = 1.041$ 、 $A = 4.444$ 、 $B = 0.146$ 、 $K = 0.818$ となります。

$\alpha = 0.05$ の場合、このケースでは関数 $H(x)$ は次のようになります。

$$H(x) = (\ln x)^2 - 1.960^2 \left(4.444 - \frac{2.704 \times 0.818^2/10 + 3.671 \times (1.041x)^2/12}{(0.818 + 1.041x)^2} - 0.146 \right)$$

次に示すように、関数 $H(x)$ には2つの根があります。



2つの根は、 $x_1 = 0.200$ および $x_2 = 6.824$ として数値的に計算されます。標準偏差の推定比は $\hat{\rho} = S_1/S_2 = 1.102$ です。 $\rho^2 = \sigma_1^2/\sigma_2^2$ の信頼セットは、次のように与えられる区間です。

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] = [0.253, 8.634]$$

標準偏差の比 ρ の信頼区間は、上記の区間のエンドポイントの平方根を取ることで得られます。

つぎに、信頼限界を見つける2つのアルゴリズムについて説明します。

1番目のアルゴリズムでは、関数 $H(x)$ の根を見つける数値的求根手順を使用します。分散の比の下側信頼限界に対応する根は、区間 $(0, 1)$ に制限されます。この根を x_L で表すと、結果3により、分散の比の下側信頼限界は $c\hat{\rho}^2 x_L$ で計算され、標準偏差の比の下側信頼限界は $\hat{\rho}\sqrt{cx_L}$ で得られます。同様に、分散の比の上側信頼限界は $c\hat{\rho}^2 x_U$ 、標準偏差の比の上側信頼限界は $\hat{\rho}\sqrt{cx_U}$ となります。ここで、 $x_U > 1$ は $H(x)$ のもう一方の根です。上側信頼区間を見つける単純な手法は、 $1/\rho^2$ の下側限界は ρ^2 の上側限界であるという事実を利用することです。最初に、比 $1/\rho^2 = \sigma_2^2/\sigma_1^2$ の信頼限界を計算するかのように、1番目のサンプルと2番目のサンプルの役割が関数 $H(x)$ の式で交換されます。次に、下限を見つけるアルゴリズムが新しい関数 $H(x)$ に適用されます。最後に、結果の限界を逆変換して、目的の上側信頼限界を得ます。

代替の手法では、次の漸化式を使用して、分散の比の下側信頼限界を再帰的に計算します。

$$\rho_0^2 = 1$$

$$\rho_{i+1}^2 = \exp \left[\ln(c \hat{\rho}^2) - z_{\alpha/2} \sqrt{\frac{\hat{Y}_P(\rho_i) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_i) - g_2}{n_2 - 1}} \right], i = 0, 1, 2, \dots$$

分散の比の下側信頼限界は ρ_{j+1}^2 で、 $|\rho_{j+1}^2 - \rho_j^2| < \varepsilon$ となります。ここで、 $j > 0$ と ε は小さい値が選ばれます（たとえば、 $\varepsilon = 10^{-6}$ ）。上側信頼限界を見つけるには、単純に上の式で $-z_{\alpha/2}$ を $+z_{\alpha/2}$ に置き換えます。

再帰の手順は基本的に、 ρ^2 について方程式 $H(\rho^2/(c\hat{\rho}^2)) = 0$ を解く反復手順であるため、信頼限界を計算する2つのアルゴリズムは明らかに同等です。再帰アルゴリズムは実行しやすいため、求根手順を使用できないときに、代わりに役立ちます。

4. シミュレーションの分析と結果

本書では、分散の比を検定するため、Layardの2サンプルの等分散性検定を拡張する手順を導き出しました。この手順を、拡張されたLayardの比検定（ELTR: Extended Layard's Test for the Ratio）と呼びます。このセクションでは、ELTR手順に基づく信頼区間の小サンプル特性を調べます。Bonett（2006）が使用した一般的な手法に従います。

ELTR手順に基づく信頼区間と、検定 L_{50} （Pan 1999）および検定 W_{50} （Levene/Brown-Forsytheの検定）に基づく信頼区間を比較します。分析1では、比較のために古典的F検定に基づく信頼区間も含めました。データが正規分布に従っている場合、古典的F検定が最適であることはよく知られています。検定 W_{50} と検定 L_{50} に基づく信頼区間の計算は、Pan（1999）により示されています。F検定に基づく信頼区間の計算は、多くの統計学入門の教科書に記載されています。またこれらの計算は、Bonett（2006）でも示されています。

3つのシミュレーションの分析を行い、それぞれ100,000個のサンプル反復を使用しました。各反復は、小規模から中規模サイズの独立サンプル2個で構成されます。各サンプルは、対称性、非対称性、重い裾、軽い裾など、既知の特性を持つ親母集団から抽出されました。各シミュレーションに関連付けられた標準誤差は、90%、95%、99%の名目信頼水準で、それぞれ約0.0009、0.0007、0.0003です。

各手順の性能を評価するため、分散の比のシミュレートした区間の達成被覆確率と平均幅を報告します。検定 W_{50} に関連する一部の区間は、幅が無限になりました（Pan（1999）によって明らかにされた可能性）。このようなケースでは、有限区間の平均幅、および無限な幅を持つ区間の割合の両方を報告します。すべてのシミュレーションは、Mathematicaソフトウェアパッケージのバージョン8を使用して行われました。

分析1：正規データの被覆確率の比較

最初の分析では、正規分布からさまざまなサイズのランダムサンプルを生成します。表2に結果を示します。

表2 被覆確率と平均区間幅の比較

$1 - \alpha$	n_1, n_2	測定	手順			
			F	ELTR	L_{50}	W_{50}
0.90	10, 10	被覆	0.898	0.918	0.913	0.921
		幅	3.72	5.06	4.72	8.03 (0.01%)
	30, 10	被覆	0.900	0.909	0.897	0.911

$1 - \alpha$	n_1, n_2	測定	手順			
			F	ELTR	L_{50}	W_{50}
		幅	2.42	3.01	3.58	3.17
	25, 25	被覆	0.902	0.907	0.914	0.916
		幅	1.61	1.73	1.85	1.938
	50, 50	被覆	0.900	0.901	0.906	0.907
		幅	1.03	1.06	1.13	1.15
0.95	10, 10	被覆	0.949	0.963	0.958	0.964
		幅	4.90	7.72	6.52	497.24 (0.20%)
	30, 10	被覆	0.950	0.957	0.945	0.959
		幅	2.98	4.91	4.67	4.07
	25, 25	被覆	0.951	0.955	0.958	0.961
		幅	1.99	2.24	2.31	2.49
	50, 50	被覆	0.951	0.952	0.953	0.954
		幅	1.25	1.31	1.38	1.41
0.99	10, 10	被覆	0.989	0.993	0.992	0.994
		幅	8.29	17.76	12.52	$> 10^4$ (8.8%)
	30, 10	被覆	0.990	0.992	0.986	0.994
		幅	4.26	15.76	8.26	6.77
	25, 25	被覆	0.990	0.992	0.992	0.993
		幅	2.86	3.66	3.43	4.03
	50, 50	被覆	0.990	0.991	0.991	0.991
		幅	1.71	1.89	1.92	2.02

陰影の付いた行は、各信頼水準 ($1 - \alpha$) とサンプルサイズ (n_1, n_2) の各組み合わせで、各手順の達成被覆確率 (被覆) を示しています。区間幅の平均 (幅) は、各被覆確率の下に示されています。ある条件の区間が無限だった場合、有限区間の平均、および無限だった区間の割合の両方を報告します。

予想どおり、結果はF手順に関連する信頼区間が最も正確で精度が高いことを示しています。F手順の達成被覆確率は、他の手順に関連するものよりも目標被覆に近くなります。F手順に関連する区間の平均幅は、他の手順に関連するものより小さくなります。ただし、この

表では、ELTR手順および L_{50} 手順を使用して構築される信頼区間は、F手順に基づくものと同程度に正確で精度が高いことも示されています。

検定 W_{50} に基づく区間もかなり正確です。ただし、 W_{50} 区間は、サンプルのサイズによっては、幅が非常に広くなったり、無限になったりすることもあります。両方のサンプルの観測値が10個しかない場合、 W_{50} 手順で生じる区間の少なくとも0.01%で幅が無限になるという点に留意してください。また、目標被覆が増加すると、無限区間の割合が増加します。ほとんどの条件下で、ELTR区間と L_{50} 区間の幅は、 W_{50} 区間の平均幅より小さくなります。

分析2: 非正規データの被覆確率の比較

2つ目の分析は、親分布が非正規の場合、ELTR手順、 L_{50} 手順、 W_{50} 手順の性能を評価および比較するために設計されています。また、手順の性能に対する外れ値の影響を評価するために、混合正規分布も含めています。観測値の90%は標準正規分布から抽出されるのに対し、残りの10%は平均が0、標準偏差が3の正規分布から抽出されることを示すため、この混合分布をCN(0.1, 3)と表します。結果を表3に示します。

表3 一部の非正規モデルの正規信頼水準が $1 - \alpha = 0.95$ の場合の被覆確率と平均区間幅の比較

分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}	分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}
一様				$\chi^2(5)$			
[1.8]				[5.4]			
10, 10	0.971	0.971	0.966	10, 10	0.956	0.938	0.956
	5.27	4.87	42.08 (0.1%)		11.61	8.78	$> 10^4$ (2.6%)
10, 30	0.964	0.961	0.957	10, 30	0.959	0.923	0.956
	2.51	2.4	2.89		6.25	4.14	190.645 (0.3%)
25, 25	0.967	0.972	0.968	25, 25	0.956	0.944	0.954
	1.43	1.79	1.88		3.66	2.92	3.26
50, 50	0.959	0.962	0.959	50, 50	0.959	0.946	0.952
	0.83	1.06	1.08		2.07	1.7	1.77
Beta(3, 3)				指数			
[2.5]				[9]			
10, 10	0.968	0.966	0.966	10, 10	0.947	0.916	0.950
	6.26	5.59	254.62 (0.1%)		20.99	14.47	$> 10^4$ (9.1%)

分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}	分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}
10, 30	0.960	0.954	0.960	10, 30	0.954	0.896	0.953
	3.14	2.76	3.71		10.46	6.19	$> 10^4$ (4.1%)
25, 25	0.959	0.966	0.965	25, 25	0.956	0.931	0.951
	1.81	2.06	2.18		6.09	4.13	5.48 (0.008%)
50, 50	0.957	0.959	0.958	50, 50	0.962	0.942	0.952
	1.06	1.23	1.26		3.18	2.24	2.38
ラプラス				$\chi^2(1)$			
[6]				[15]			
10, 10	0.946	0.935	0.961	10, 10	0.928	0.889	0.947
	13.47	10.45	$> 10^4$ (3.0%)		55.09	37.4	$> 10^5$ (25.1%)
10, 30	0.947	0.919	0.957	10, 30	0.943	0.882	0.956
	6.78	4.82	$> 10^4$ (0.4%)		18.71	11.14	$> 10^6$ (25.7%)
25, 25	0.945	0.940	0.952	25, 25	0.952	0.925	0.954
	4.00	3.372	3.86		10.97	6.84	$> 10^4$ (0.4%)
50, 50	0.952	0.949	0.955	50, 50	0.958	0.936	0.951
	2.19	1.91	1.99		5.08	3.31	3.75 (0.001%)
$t(5)$				対数正規			
[9]				[113.9]			
10, 10	0.957	0.946	0.965	10, 10	0.923	0.876	0.955
	11.07	8.81	$> 10^3$ (2.0%)		59.22	46.15	$> 10^5$ (23.0%)
10, 30	0.957	0.930	0.959	10, 30	0.949	0.866	0.958
	6.06	4.24	$> 10^3$ (0.7%)		29.13	17.67	$> 10^6$ (31.6%)

分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}	分布 [γ] n_1, n_2	ELTR	L_{50}	W_{50}
25, 25	0.954	0.948	0.960	25, 25	0.947	0.917	0.965
	3.54	2.93	4.86 (0.01%)		16.21	8.73	$> 10^4$ (2.4%)
50, 50	0.954	0.947	0.954	50, 50	0.955	0.928	0.960
	2.10	1.71	1.77 (0.003%)		8.62	4.11	164.38 (0.2%)
半正規				CN(0.1, 3)			
[3.9]				[8.3]			
10, 10	0.956	0.942	0.954	10, 10	0.977	0.965	0.979
	10.41	7.89	$> 10^4$ (1.5%)		12.64	9.52	$> 10^4$ (4.9%)
10, 30	0.959	0.930	0.954	10, 30	0.981	0.952	0.979
	5.18	3.64	13.00 (0.02%)		7.82	4.71	944.68 (1.1%)
25, 25	0.959	0.952	0.959	25, 25	0.982	0.972	0.981
	3.01	2.62	2.88		4.63	3.22	3.71
50, 50	0.960	0.951	0.954	50, 50	0.983	0.972	0.978
	1.69	1.54	1.59		2.64	1.83	1.91

陰影の付いた行は、各手順、親分布、およびサンプルサイズの組み合わせで達成被覆確率を示しています。区間幅の平均は、各被覆確率の下に示されています。ある条件の区間が無限だった場合、有限区間の平均、および無限だった区間の割合の両方を報告します。各親分布の尖度 (γ) は、角括弧 [] で示されています。

対称的で裾の軽い分布の場合、結果は3つすべての方法で同様に控え目な被覆確率が出ることを示しています。ただし、ELTR区間と L_{50} 区間は、小さいサンプルでは、 W_{50} 区間よりも精度が高くなっています。たとえば、サンプルが(3, 3)のパラメータのベータ分布から抽出される場合、ELTR区間と L_{50} 区間の達成被覆確率は、少なくとも W_{50} 区間と同じぐらい正確ですが、 W_{50} 区間は一貫して幅がより広くなります。

また、ELTR区間と W_{50} 区間は、対称的で裾の重い分布では少し控え目であるのに対し、 L_{50} 区間は大き目な値を示します。 L_{50} 区間は、計画がアンバランス型だと、さらに大き目な値を示します。たとえば、サイズ10と30のサンプルがラプラス分布から抽出される場合、 L_{50} 区間の達成被覆確率は0.919です。同じサイズのサンプルが自由度5のt分布から抽出される場合、 L_{50} 区間の達成被覆確率は0.930です。

また、 L_{50} 区間は、小さいサンプルが大きく歪んだ裾の重い分布から抽出されると、かなり大き目な値を示します。たとえば、サンプルが対数正規分布から抽出されると、達成被覆確

率は0.866まで低くなる場合があります。これらの分布では、 W_{50} の方法が、3つの方法の中で最も大きな値を示す方法ではありません。ただし、 W_{50} 区間は、幅が無限になる区間が多すぎます。たとえば、サンプルが自由度1のカイ二乗分布 ($\chi^2(1)$) から抽出される場合、 W_{50} 区間の25%超で幅が無限になる可能性があります。ELTR区間は、やや正確さに欠けますが、幅がかなり狭いので、 W_{50} 区間より参考になります。

最後に、3つすべての手順は、外れ値による悪影響を受けることに気付きました。影響が最小なのは L_{50} の方法です。 L_{50} の方法は検定 W_{50} (Pan 1999) に対する外れ値の影響を削減するために導出されたので、これは予想されたことです。サンプルが混合正規分布CN(0.1, 3) から抽出される場合、ELTR手順および W_{50} 手順の達成被覆確率の最小値は0.977です。ここに記載されていないその他のシミュレーション結果は、これらの区間はサンプルサイズの増加によって、緩やかに向上することを示しています。

分析3: 等尖度仮定に対する感度

最後の分析では、尖度が等しいという仮定に対するELTR手順の感度を調べます。親母集団の尖度が等しくない、つまり $\gamma_1 \neq \gamma_2$ のときの、ELTR手順の性能を調べます。また、母集団が類似しているという仮定のもとで導出されるため、 L_{50} 手順と W_{50} 手順も含めます。親母集団の尖度が等しくない場合、この類似性の仮定は損なわれます。表4に結果を示します。

表4 正規信頼水準が $1 - \alpha = 0.95$ の場合の、等尖度仮定に対するELTR手順の感度

分布1, 分布2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	分布1, 分布2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
Beta (3, 3)、 正規 [2.5, 3]				正規、 CN (0.9, 3) [3, 8.3]			
10, 10	0.964	0.961	0.964	10, 10	0.955	0.948	0.951
	0.27	0.23	204.50 (0.20%)		6.88	5.16	$> 10^4$ (4.89%)
30, 10	0.946	0.939	0.946	30, 10	0.941	0.910	0.942
	0.16	0.17	0.15		5.26	3.77	3.20
10, 30	0.966	0.956	0.967	10, 30	0.961	0.950	0.958
	0.14	0.11	0.17		4.26	2.40	630.42 (1.10%)
50, 50	0.951	0.950	0.949	50, 50	0.936	0.910	0.907
	0.04	0.05	0.05		1.27	1.11	1.19
正規、 ラプラス [3, 6]				半正規、 $\chi^2(5)$ [3.9, 5.4]			
10, 10	0.941	0.935	0.947	10, 10	0.956	0.940	0.954

分布1, 分布2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}	分布1, 分布2 [γ_1, γ_2] n_1, n_2	ELTR	L_{50}	W_{50}
	6.67	5.17	$> 10^6$ (2.90%)		0.42	0.32	304.41 (2.60%)
30, 10	0.912	0.888	0.914	30, 10	0.954	0.918	0.949
	5.06	3.85	3.21		0.33	0.22	0.20
10, 30	0.963	0.943	0.955	10, 30	0.962	0.934	0.958
	3.33	2.25	$> 10^3$ (0.40%)		0.23	0.15	3.28 (0.30%)
50, 50	0.935	0.894	0.889	50, 50	0.955	0.941	0.945
	0.98	1.04	1.12		0.07	0.06	0.07
正規、 半正規 [3, 3.9]				$\chi^2(5)$ 、 指数 [5.4, 9]			
10, 10	0.956	0.948	0.957	10, 10	0.938	0.914	0.940
	28.16	20.65	$> 10^4$ (1.50%)		211.17	137.88	$> 10^6$ (9.10%)
30, 10	0.946	0.924	0.947	30, 10	0.928	0.875	0.929
	20.59	14.83	12.78		194.70	93.02	83.02
10, 30	0.961	0.946	0.962	10, 30	0.968	0.930	0.954
	14.06	9.37	49.11 (0.02%)		102.35	55.29	$> 10^5$ (3.90%)
50, 50	0.953	0.950	0.952	50, 50	0.950	0.920	0.923
	4.32	4.16	4.33		29.64	23.37	25.54

陰影の付いた行は、各手順、親分布の組み合わせ（分布1, 分布2）、およびサンプルサイズの組み合わせで達成被覆確率を示しています。区間幅の平均は、各被覆確率の下に示されています。ある条件区間が無限である場合、有限区間の平均、および無限だった区間の割合の両方を報告します。各親分布の尖度 (γ_1, γ_2) は、角括弧 [] で示されています。

一般に、ELTR手順の性能は、サンプルが十分に大きい場合、等しくない尖度による悪影響は受けないように見えます。ただし、計画がアンバランス型で、より小さいサンプルがより裾の重い分布から抽出される場合、達成被覆確率は大き目な値を示します。より大きなサンプルがより裾の重い分布から抽出されると、達成被覆確率は良くなります。

また、サンプルサイズが十分に大きい場合、 L_{50} と W_{50} 区間は、等しくない尖度に起因する分布の相違性に対して一般に頑健であるように見えます。ただし、サンプルが正規分布およ

びラプラス分布、または正規分布および混合正規分布から抽出される場合、 L_{50} 区間と W_{50} 区間の被覆確率はサンプルが50と大きくても安定していません。

L_{50} 区間は、一般にELTR区間と W_{50} よりも大きな値を示します。3つのケースで、 L_{50} 区間の達成被覆確率は0.90未満です。対照的に、 W_{50} 区間の達成被覆確率が0.90未満なのは1つのみです。ELTR区間の最低の達成被覆確率は0.912です。

前述の分析（表3）は、両方のサンプルが混合正規分布CN(0.1, 3)から抽出される場合、3つすべての手順から、明らかにより保守的な区間が生じることを示しています。現在の分析は、1つのサンプルのみがCN(0.1, 3)から抽出される場合、3つすべての手順の性能が明らかに良くなることを示しています。ただし、 L_{50} 区間と W_{50} 区間の性能は、サンプルサイズが50まで増加すると、かなり低下するよう見えます。

5. 例

このセクションでは、F、ELTR、 L_{50} 、 W_{50} という4つすべての手順をPan (1999) から取得したデータセットに適用します。Ott (1993, 352ページ) は、このデータについて次のように説明しています。

鉄鉱山の化学者が、鉱石1ポンドあたりの酸化鉄の量（オンス単位の重量）は、1ポンドあたりの酸化鉄の平均量が増加するにつれ、増加する傾向にあるのではないかと考えています。この理論を検定するため、1ポンドの鉱石のサンプルを10個ずつ、2か所から抽出します。場所1の酸化鉄の平均含有量は、場所2よりもかなり多くなっています。鉱石サンプルに含まれている酸化鉄の量は次のとおりです。

場所1	8.1	7.4	9.3	7.5	7.1	8.7	9.1	7.9	8.4	8.8
場所2	3.9	4.4	4.7	3.6	4.1	3.9	4.6	3.5	4.0	4.2

次の表に、4つの方法で計算された $\sigma_2/\sigma_1 = 1/\rho$ の95%信頼区間を示します。

手順	95%信頼区間
F	(0.262, 1.055)
ELTR	(0.277, 0.924)
L_{50} (Pan)	(0.295, 0.938)
W_{50} (Levene/Brown-Forsythe)	(0.237, 0.908)

6. 結論

今回のシミュレーションは、一般に、ELTR手順に基づく信頼区間は、検定 L_{50} と検定 W_{50} から導出される信頼区間と同じくらい正確であることを示しています。ただし、ELTR区間と L_{50} 区間は、ほとんどの分布で W_{50} 区間より高い精度を示しています。大きく歪んだ裾の重い分布から小さいサンプルが抽出される場合、 W_{50} 区間はELTR区間と L_{50} 区間より正確である傾向があります。しかしながら、この利点は通常、精度の著しい損失によって相殺されます。その結果、 W_{50} 区間は一般的に広くなりすぎ、幅が無限になる可能性が高くなります。

設計どおり、 L_{50} 区間によって W_{50} 区間の精度が向上します。けれども、歪んだ母集団の場合、 L_{50} 区間は短すぎて、被覆確率が過度に大き目の値を示します。対照的に、ELTR区間は一般に、より安定しています。ELTR区間は、通常長すぎたり、短すぎたりしないため、被覆確率も控え目すぎたり、過度に大き目の値を示したりしません。したがってELTR手順は、ほとんどの実用目的に最適な手順であるように見えます。

ELTR区間は、検定 L_{50} または検定 W_{50} に基づく区間よりも計算に少し手間がかかります。けれども、一般に (W_{50} 区間と比較した) 精度の向上と (L_{50} と比較した) 正確さの向上は、余分な計算にかかる労力を補って余りあります。ELTR手順は、Minitab統計ソフトウェアのリリース17で、2サンプル分散分析の一部として実装済みで、Bonettの手順と呼ばれています。

今後の研究のために、次のように、Layardの併合尖度推定量がBonettのより頑健なバージョンに置き換えられる場合に、多重サンプル計画でのLayardの検定の小サンプル特性を調べることを検討してみることができます。

$$\hat{\gamma}_A = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{\left[\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^2 \right]^2} \sum_{i=1}^k n_i$$

ここで、 m_i は、調整比率 $1/[2(n_i - 4)^{1/2}]$ と $i = 1, \dots, k$ を使用した、サンプル i の調整平均です。

さらに、対数変換したサンプル分散の漸近分散のShoemakerの近似を使用すると、有益である可能性があります。

最後に、Bonett (2006) によって提案された区間は、信頼区間としては適していませんが、2サンプルの等分散性検定で合格領域と解釈される場合には、ほとんどの分布で際立って正確で高い精度を示します。この合格領域は、複数の分散を比較するグラフを用いた手順の基本として適切です。Hochberg, Weiss, and Hart (1982) は、平均が等しいことを検定する同様の手順を提案しました。このような手順は、Minitab統計ソフトウェアのリリース17で、等分散性分析の一部として実装済みで、多重比較手順と呼ばれています。

7. 付録

付録A: 結果1の証明

セクション2の表記法と仮定を使用して、任意の $\rho = \sigma_1/\sigma_2$ に対して、 $X_j = \rho Y_{2j}$ とします。そうすると、次のようになります。

$$\text{Var}(X_j) = \rho^2 \text{Var}(Y_{2j}) = \rho^2 \sigma_2^2 = \sigma_1^2 = \text{Var}(Y_{1j})$$

および

$$E\left(\frac{X_j - \mu_{X_j}}{\sigma_{X_j}^4}\right)^4 = \rho^4 E\left(\frac{Y_{2j} - \mu_2}{\rho^4 \sigma_2^4}\right)^4 = E\left(\frac{Y_{2j} - \mu_2}{\sigma_2^4}\right)^4 = \gamma$$

仮定により $E(Y_{1j} - \mu_1)^4 / \sigma_1^4 = \gamma$ なので、2つのサンプル Y_{1j} と $X_j = \rho Y_{2j}$ の親母集団には、同じ分散 σ_1^2 と同じ尖度 γ があるということになります。Layard (1973) により、2つのサンプル Y_{1j} と X_j に基づく一貫した尖度推定量の γ は、

$$\begin{aligned}\hat{\gamma}' &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \sum_{j=1}^{n_2} (X_j - \bar{X})^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_X^2]^2} \\ &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = \hat{\gamma}_P(\rho)\end{aligned}$$

として、要求に応じて与えられます。

付録B: 結果2の証明

T_2 に基づく $\rho = \sigma_1/\sigma_2$ の約 $(1 - \alpha)100$ パーセントの信頼セットは、次のように与えられるということです。すでに立証済みです。

$$\left\{ \rho \in (0, \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \leq 0 \right\}$$

併合尖度推定量は、個々のサンプルの尖度推定量として表現でき、次のように与えられます。

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1, 2$$

より具体的に言えば、 $t = \rho/\hat{\rho}$ とすると、次のようになります。

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = (n_1 + n_2) \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2}$$

ここで、 $K = (n_1 - 1)/(n_2 - 1)$ です。

その結果、2次標準誤差項は、次のように表現できます。

$$\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2} - B$$

ここで

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{k_1}{n_1 - 1} + \frac{k_2}{n_2 - 1}$$

したがって、 $r^2 = \rho^2/(c\hat{\rho}^2)$ とすると、次のようになることがわかりやすくなります。

$$\begin{aligned}(\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left(\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \\ = (\ln r^2)^2 - z_{\alpha/2}^2 \left(A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 c^2 r^4/n_2}{(K + c r^2)^2} - B \right)\end{aligned}$$

その結果、 T_2 に基づく $\rho = \sigma_1/\sigma_2$ の約 $(1 - \alpha)100$ パーセントの信頼セットは、次のように与えられます。

$$\hat{\rho}\sqrt{c} \{ r \in (0, \infty): H(r^2) \leq 0 \}$$

または、 $\rho^2 = \sigma_1^2/\sigma_2^2$ の信頼セットを次のように表現できます。

$$c\hat{\rho}^2 \{r \in (0, \infty): H(r) \leq 0\}$$

ここで

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

および

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K+x)^2} - B$$

付録C: 結果3の証明

$H(0) = H(+\infty) = +\infty$ と $H(1) < 0$ を使用すると、 $H(x)$ は正の実数直線上で連続していることが簡単に検証されます。中間値の定理によって、関数 $H(x)$ では区間 $(0, 1)$ で少なくとも1つの根、区間 $(0, +\infty)$ で少なくとも1つの根が認められます。したがって、関数 $H(x)$ には正確に2つの根があり、一方は1未満、他方は1より大きくなります。この関数は上向きに開くので、不等式 $H(r) \leq 0$ は、 r が根の間にある場合に満たされます。これらの根によって、 $\rho^2/(c\hat{\rho}^2)$ の信頼区間のエンドポイントが定義されます。したがって、 $x_L < 1 < x_U$ を2つの根とすると、結果2により、分散の比の下側信頼限界 ρ^2 は $c\hat{\rho}^2 x_L$ で計算され、標準偏差の比の下側信頼限界は $\hat{\rho}\sqrt{cx_L}$ で得られます。同様に、分散の比の上側信頼限界は $c\hat{\rho}^2 x_U$ 、標準偏差の比の上側信頼限界は $\hat{\rho}\sqrt{cx_U}$ となります。

一方、関数 $H(x)$ に3つ以上の根がある場合、不等式 $H(r) \leq 0$ は、関数が上向きに開く連続する根の間に r がある場合に満たされます。したがって、信頼セットは、重なり合わない区間の和集合です。

8. 参考文献

- Balakrishnan, N. and Ma, C. W. (1990). A Comparative Study of Various Tests for the Equality of Two Population Variances. *Journal of Statistical Computation and Simulation*, 35, 41-89.
- Bonett D. G. (2006). Robust Confidence Interval for a Ratio of Standard Deviations. *Applied Psychological Measurements*, 30, 432-439.
- Boos, D. D. and Brownie, C. (1989). Bootstrap Methods for Testing Homogeneity of Variances. *Technometrics*, 31, 69-82.
- Brown, M. B., and Forsythe A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, 69, 364-367.
- Conover, W. J., Johnson, M. E. and Johnson, M. M. (1981). A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics*, 23, 351-361.
- Fligner, M. A. and Killeen, T. J. (1976). Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association*, 71, 210-213.
- Hall, P. and Padmanabhan, A. R. (1997). Adaptive Inference for the Two-Sample Scale Problem. *Technometrics*, 39, 412-422.

- Hochberg, Y., Weiss, G., and Hart S., (1982). On Graphical Procedures for Multiple Comparisons. *Journal of the American Statistical Association*, 77, 767-772.
- Layard, M. W. J. (1973). Robust Large-Sample Tests for Homogeneity of Variances. *Journal of the American Statistical Association*, 68, 195-198.
- Levene, H. (1960). "Robust Tests for Equality of Variances," in I. Olkin, ed., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278-292.
- Lim, T.-S. and Loh, W.-Y. (1996). A Comparison of Tests of Equality of Variances. *Computational Statistics and Data Analysis*, 22, 287-301.
- Ott, L. (1993). *An Introduction to Statistical Methods and Data Analysis*, Belmont, CA: Duxbury Press.
- Pan, G. (1999). On a Levene Type Test for Equality of Two Variances. *Journal of Statistical Computation and Simulation*, 63, 59-71.
- Shoemaker, L. H. (2003). Fixing the F Test for Equal Variances. *The American Statistician*, 57, 105-114.
- Wolfram, S. (1999). *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press.

© 2015, 2017 Minitab Inc. All rights reserved.

Minitab®, Quality. Analysis. Results.® and the Minitab® logo are all registered trademarks of Minitab, Inc., in the United States and other countries. See minitab.com/legal/trademarks for more information.