指数分布モデルにおける
同等性の統計解析法

2011SE079 石黒弘隆
2011SE186 長坂侑貴
指導教員：白石高章

1　はじめに

現在, 統計学はさまざまな分野で使われており, 特に医学の分野で発展してきている. 今回は, 臨床試験の一例を上げる. 臨床試験とはヒトを対象にし, 効果的か安全かを調べる試験である. 特に必要以上に被験者を増やしてはならないという倫理学的要請がある (角間・服部 [1]). そこでどのように統計学を用いて臨床試験を行うかを統計的仮説を用いて考える. 本論では, 統計的仮説の中でも特に同等性の仮説検定について焦点をあてて考えてみる. 角間・服部 [1] では、主に正規分布における, 統計解析が示されている. また指数分布における片側検定は加藤・丹羽 [2] の卒業論文で示されている. そこで本論では, それら 2 つの文献を参考に指数分布モデルにおける同等性の両側仮説検定を論じる.

2　指数分布モデル

2.1　指数分布

指数分布について補題 1 として, 知られている結果を述べる.

補題 1　(加藤・丹羽 [2])　$X_1, ..., X_n$は互いに独立で同一の密度関数 $f_z(x) = (1/\mu)e^{-x/\mu}I_{(0,\infty)}$をもつ指数分布 $EX(1/\mu)$ に従うと仮定する.
このとき,

$$T \equiv \left(\frac{2n}{\mu}\right)\overline{X}_n \sim \chi^2_{2n}$$

が成り立つ. ここで, χ^2_{2n}は自由度 $2n$ のχ^2分布である.
ただし

$$\overline{X}_n \equiv \left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i$$

とする. □

次に, 帰無仮説 $H_0 : \mu \le \mu_0$ vs. 対立仮説 $K_0 : \mu > \mu_0$ に対する検定統計量は, 白石 [3] より

$$T_0 = 2\sum_{i=1}^{n}\frac{X_i}{\mu_0}$$

で与えられる.

定理 1　(加藤・丹羽 [2])　$\chi^2_{2n}(\alpha)$ は自由度 $2n$ のカイ二乗分布の上側 $100\alpha\%$ 点とする. $\beta(\mu) = P_\mu(T_0 \ge \chi^2_{2n}(\alpha))$ とおくと $\mu \le \mu_0$ ならば $\beta(\mu) \le \alpha$となる. □

2.2　両側検定

帰無仮説 $H : \mu = \mu_0$ vs. 対立仮説 $K : \mu \ne \mu_0$ に対する検定について考える.
検定統計量は T_0 とする. T_0 は H の下で, 補題 1 より自由度 2n のカイ二乗分布に従う.

これにより棄却域は

$$\left\{T_0 \le \chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right\} \cup \left\{\chi^2_{2n}\left(\frac{\alpha}{2}\right) \le T_0\right\} \tag{2.1}$$

で与えられる.

定理 2　(2.1) の検出力関数は

$$\beta(\mu) = \int_0^{\frac{\mu_0}{\mu}\chi^2_{2n}(1-\frac{\alpha}{2})} f_\chi(t|2n)dt + \int_{\frac{\mu_0}{\mu}\chi^2_{2n}(\frac{\alpha}{2})}^{\infty} f_\chi(t|2n)dt \tag{2.2}$$

で与えられる.
定理 2 の証明は本論によって示されている.

3　同等性の仮説検定

同等性（equivalence）とは, 新たに開発された医薬品の効果 μ が既存の医薬品の効果 μ_0 と比較して, 劣ってなくかつ優越でもないことを意味する (角間・服部 [1]).

ここで許容できる効果の比を $\mu > 0$ を用いて行う.

$$K : 1-\delta_1 < \frac{\mu}{\mu_0} < 1+\delta_2$$

$$\delta_1 > 0, \quad \delta_2 > 0$$

と表すと, 同等性の帰無仮説はこれを否定するので

$$H : \frac{\mu}{\mu_0} \le 1-\delta_1 \quad \text{または} \quad \frac{\mu}{\mu_0} \ge 1+\delta_2$$

と表される.
$\theta = \mu/\mu_0$とおくと, 同等性の仮説は,

$$H : \theta \le 1-\delta_1 \quad \text{または} \quad \theta \ge 1+\delta_2$$

$$K : 1-\delta_1 < \theta < 1+\delta_2$$

さらにこの 2 つの仮説は,

$$H_1 : \theta \leq 1-\delta_1 \text{ vs. } K_1 : \theta > 1-\delta_1$$
$$H_2 : \theta \geq 1+\delta_2 \text{ vs. } K_2 : \theta < 1+\delta_2$$

の 2 つの片側検定として表すことができる.

H_1 vs. K_1 の場合

$$T_1(X) = 2\sum_{i=1}^{n} \frac{X_i}{\mu_0(1-\delta_1)} \tag{3.1}$$

を用いる. $T_1(X) \geq \chi^2_{2n}(\alpha/2)$ のとき H_1を棄却する検定がサイズ$\alpha/2$ 検定となる.

H_2 vs. K_2 の場合

$$T_2(X) = 2\sum_{i=1}^{n} \frac{X_i}{\mu_0(1+\delta_2)} \tag{3.2}$$

を用いる. $T_2(X) \geq \chi^2_{2n}(1-\alpha/2)$ のとき H_2を棄却する検定がサイズ $1-(\alpha/2)$ 検定となる.

2 つの帰無仮説 H_1と H_2を共に棄却する棄却域は以下のようになる.

$$\left\{T_1(X) \geq \chi^2_{2n}\left(\frac{\alpha}{2}\right)\right\} \cap \left\{T_2(X) \leq \chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right\} \tag{3.3}$$

定理 3 (3.3) の検出力関数は,

$$\beta(\mu) = \int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)} f_\chi(t|2n)dt - \int_{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right)}^{\infty} f_\chi(t|2n)dt \tag{3.4}$$

で与えられる.

証明

$$\begin{aligned}
\beta(\mu) &= P_\mu\left(\left\{T_1(X) \geq \chi^2_{2n}\left(\frac{\alpha}{2}\right)\right\} \cap \left\{T_2(X) \leq \chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right\}\right) \\
&= P_\mu\left(\left\{2\sum_{i=1}^{n}\frac{X_i}{\mu_0(1-\delta_1)} \geq \chi^2_{2n}\left(\frac{\alpha}{2}\right)\right\} \cap \left\{2\sum_{i=1}^{n}\frac{X_i}{\mu_0(1+\delta_2)} \leq \chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right\}\right) \\
&= P_\mu\left(\left\{2\sum_{i=1}^{n}\frac{X_i}{\mu} \geq \frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right)\right\} \cap \left\{2\sum_{i=1}^{n}\frac{X_i}{\mu} \leq \frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right\}\right) \\
&= P_\mu\left(\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right) \leq 2\sum_{i=1}^{n}\frac{X_i}{\mu} \leq \frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right) \\
&= P_\mu\left(2\sum_{i=1}^{n}\frac{X_i}{\mu} \leq \frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\right) - P_\mu\left(\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right) \leq 2\sum_{i=1}^{n}\frac{X_i}{\mu}\right) \\
&= F_\chi\left(\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)\middle|2n\right) - 1 + F_\chi\left(\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right)\middle|2n\right) \\
&= \int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)} f_\chi(t|2n)dt - 1 + \int_0^{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right)} f_\chi(t|2n)dt \\
&= \int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(1-\frac{\alpha}{2}\right)} f_\chi(t|2n)dt - \int_{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right)}^{\infty} f_\chi(t|2n)dt
\end{aligned}$$

ただし, $F_\chi(t|2n)$ は自由度 $2n$ のカイ二乗分布の分布関数である.また, $f_\chi(t|2n)$ は, 自由度 $2n$ のカイ二乗分布の密度関数である. □

$\mu > \mu_0$ が真の場合と $\mu < \mu_0$ が真の場合に分けて考えてみる.

(a) $\mu > \mu_0$ が真のとき
ここで (3.4) の右辺の 2 つの積分を分けて考える.

$$\int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}(1-\frac{\alpha}{2})} f_\chi(t|2n)dt < \frac{\alpha}{2} \tag{3.5}$$

$$\int_{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})}^{\infty} f_\chi(t|2n)dt > \frac{\alpha}{2} \tag{3.6}$$

(3.5) の左辺は μ を大きくすると無視できるほど小さくなるため, 検出力は次のように近似する.

$$\beta(\mu) \approx \int_{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})}^{\infty} f_\chi(t|2n)dt$$

(b) $\mu < \mu_0$ が真のとき
(a) と同様に (3.4) の右辺の 2 つの積分を分けて考える.

$$\int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}(1-\frac{\alpha}{2})} f_\chi(t|2n)dt > \frac{\alpha}{2} \tag{3.7}$$

$$\int_{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})}^{\infty} f_\chi(t|2n)dt < \frac{\alpha}{2} \tag{3.8}$$

(3.8) の左辺は μ を大きくすると無視できるほど小さくなるため, 検出力は次のように近似する.

$$\beta(\mu) \approx \int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}(1-\frac{\alpha}{2})} f_\chi(t|2n)dt$$

3.1 症例数の算出

$\mu > \mu_0$ が真のとき

$$\beta(\mu) \approx \int_{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})}^{\infty} f_\chi(t|2n)dt \geq \beta_0$$

$$1 - \int_0^{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})} f_\chi(t|2n)dt \geq \beta_0$$

$$\int_0^{\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})} f_\chi(t|2n)dt \leq 1-\beta_0 = \int_0^{\chi^2_{2n}(\beta_0)} f_\chi(t|2n)dt$$

$$\frac{\mu_0(1-\delta_1)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right) \leq \chi^2_{2n}(\beta_0)$$

帰無仮説 $H_1 : \theta \leq 1-\delta_1$ vs. 対立仮説 $K_1 : \theta > 1-\delta_1$ に対する有意水準がα, 検出力がβ_0以上の検定でΔ 以上の効果の比 $\Delta \leq \mu/\mu_0$を検出するためには少なくとも

$$\frac{1-\delta_1}{\Delta}\chi^2_{2n}\left(\frac{\alpha}{2}\right) \leq \chi^2_{2n}(\beta_0)$$

$$(1-\delta_1)\frac{\chi^2_{2n}(\frac{\alpha}{2})}{\chi^2_{2n}(\beta_0)} \leq \Delta \tag{3.9}$$

となる症例数 n が必要となる.特に（3.9）を満たす最小の n を求めることが重要である.

$\mu < \mu_0$ が真のとき

$$\beta(\mu) \approx \int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}(1-\frac{\alpha}{2})} f_\chi(t|2n)dt$$

$$1 - \int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})} f_\chi(t|2n)dt \geq \beta_0$$

$$\int_0^{\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}(\frac{\alpha}{2})} f_\chi(t|2n)dt \leq 1-\beta_0 = \int_0^{\chi^2_{2n}(\beta_0)} f_\chi(t|2n)dt$$

$$\frac{\mu_0(1+\delta_2)}{\mu}\chi^2_{2n}\left(\frac{\alpha}{2}\right) \leq \chi^2_{2n}(\beta_0)$$

帰無仮説 $H_2 : \theta \geq 1+\delta_2$ vs. 対立仮説 $K_2 : \theta < 1+\delta_2$ に対する有意水準がα, 検出力がβ_0以上の検定でΔ 以上の効果の比 $\Delta \leq \mu/\mu_0$を検出するためには少なくとも

$$\frac{1+\delta_2}{\Delta}\chi^2_{2n}\left(\frac{\alpha}{2}\right) \leq \chi^2_{2n}(\beta_0)$$

$$(1+\delta_2)\frac{\chi^2_{2n}(\frac{\alpha}{2})}{\chi^2_{2n}(\beta_0)} \leq \Delta \tag{3.10}$$

となる症例数 n が必要となる.

(3.9) の左辺のグラフを図 1 に, (3.10) の左辺のグラフを図 2 に書いた. いずれも単調減少となっている.

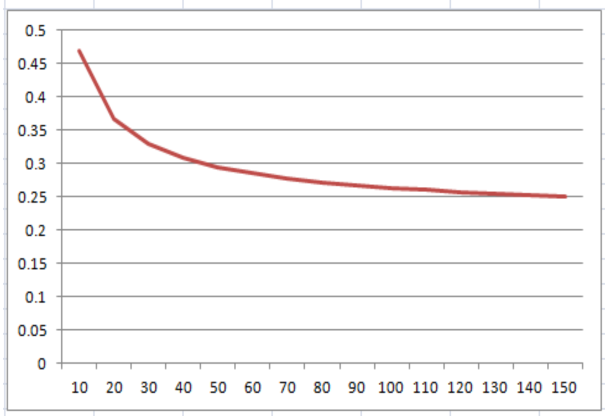

図 1 $\alpha = 0.05$, $\beta_0 = 0.8$, $\delta_1 = 0.8$ の場合の $y = (1-\delta_1)\chi^2_{2n}(\alpha/2)/\chi^2_{2n}(\beta_0)$ のグラフ (横軸は n)

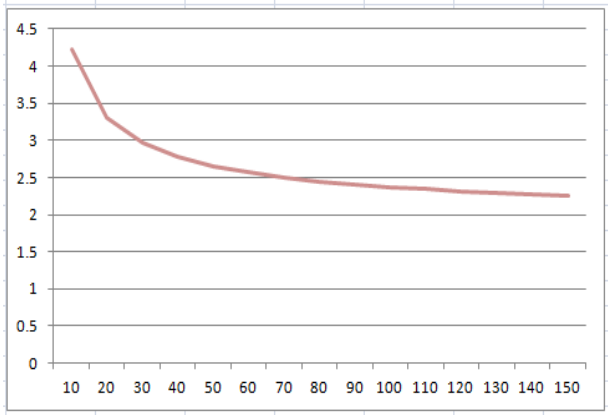

図 2 $\alpha = 0.05$, $\beta_0 = 0.8$, $\delta_2 = 0.8$ の場合の $y = (1+\delta_2)\chi^2_{2n}(\alpha/2)/\chi^2_{2n}(\beta_0)$ のグラフ (横軸は n)

4 最小の症例数を計算するプログラムの解説

(3.9) を満たす最小の n を n_1とし, (3.10) を満たす最小の n を n_2とする. 早川 [5] を参考にし n_1, n_2を求めるプログラムを C 言語で作成した.

4.1 プログラム詳細

以下のプログラムは main プログラムである.

float solution1, solution2;
input();/*自由度, 有意水準, 許容できる効果の比を入力する関数*/
for(NU=1; NU<=200; NU++)/*自由度を 1 から 200 まで繰り返す*/
solution1=KAI1(ALPHA/2);/*$\chi^2_{2n}(\alpha)$ の値の置き換え*/
solution2=KAI2(BETA);/*$\chi^2_{2n}(\beta)$ の値の置き換え*/
if(kekka<DELTA)
break;
/*kekka の値が DELTA の値を超えたら止まる*/
output1(NU, ALPHA, solution1);/*$\chi^2_{2n}(\alpha)$ の値を出力した関数*/
output2(NU, BETA, solution2);/*$\chi^2_{2n}(\beta)$ の値を出力した関数*/
result(NU, delta, solution1, solution2);/*検定結果を出力する関数*/

以下は引数と関数の詳細である.

NU: 自由度 n
ALPHA, BETA: 有意水準, 検出力
delta1, delta2: 許容できる効果の比
KAI1 関数: 有意水準 α における χ^2 分布の上側 $100\alpha\%$ 点を求める関数
KAI2 関数: 有意水準 β における χ^2 分布の上側 $100\beta\%$ 点を求める関数
NORMP 関数: 正規分布の上側確率を求める関数
ppow1, ppow2 関数: 上側確率を求める際に使用する積の計算
CHI2PA 関数: 有意水準 α での上側確率を求める関数
CHI2PB 関数: 有意水準 β での上側確率を求める関数

4.2 症例数 n_1を求めるプログラムの実行例

一例として, $\Delta = 0.5, \Delta = 0.3$ のときの実行結果を表示する.

$\Delta = 0.5$ のとき
ラージデルタ, アルファ, ベータ, デルタ 1 を入力してください
0.5, 0.05, 0.8, 0.8
誤差 0.001 以下の自由度 18.000 のカイ二乗分布の上側 2.500 パーセント点は 31.526
誤差 0.001 以下の自由度 18.000 のカイ二乗分布の上側 80.000 パーセント点は 12.857
ラージデルタの値が 0.5 である症例数は 9 となる

$\Delta = 0.3$ のとき
ラージデルタ, アルファ, ベータ, デルタ 1 を入力してください
0.3, 0.05, 0.8, 0.8
誤差 0.001 以下の自由度 92.000 のカイ二乗分布の上側 2.500 パーセント点は 120.427
誤差 0.001 以下の自由度 92.000 のカイ二乗分布の上側 80.000 パーセント点は 80.433
ラージデルタの値が 0.3 である症例数は 46 となる

4.3 症例数 n_2を求めるプログラムの実行例

一例として, $\Delta = 3.0, \Delta = 2.5$ のときの実行結果を表示する.

$\Delta = 3.0$ のとき
ラージデルタ, アルファ, ベータ, デルタ 2 を入力してください
3.0, 0.05, 0.8, 0.8
誤差 0.001 以下の自由度 56.000 のカイ二乗分布の上側 2.500 パーセント点は 78.567
誤差 0.001 以下の自由度 56.000 のカイ二乗分布の上側 80.000 パーセント点は 46.955
ラージデルタの値が 3.0 である症例数は 28 となる

$\Delta = 2.5$ のとき
ラージデルタ, アルファ, ベータ, デルタ 2 を入力してください
2.5, 0.05, 0.8, 0.8
誤差 0.001 以下の自由度 140.000 のカイ二乗分布の上側 2.500 パーセント点は 174.647
誤差 0.001 以下の自由度 140.000 のカイ二乗分布の上側 80.000 パーセント点は 125.758
ラージデルタの値が 2.5 である症例数は 70 となる

5 おわりに

指数分布モデルの同等性の仮説検定で, 必要とされる症例数を導くことができた. また C 言語によってプログラムを作成し症例数の設定を行うことで, より理解を深められた.

参考文献

[1] 角間辰之・服部聡:『臨床試験のデザインと解析』. 近代科学者, 東京, 2012.

[2] 加藤駿介・丹羽雄士:『指数分布モデルにおける非劣性の統計解析法』. 2013 年度南山大学情報理工学部情報システム数理学科卒業論文.

[3] 白石高章:『 統計科学の基礎　データと確率の結びつきがよくなる数理』. 日本評論社, 東京,2012.

[4] 芳賀敏郎・永田靖: 『データ解析に役立つ Excel 関数』. 日科技連出版社.

[5] 早川由宏:『Mathematica と C 言語による統計プログラムの基礎』. 2012 年度南山大学情報理工学部情報システム数理学科卒業論文.