

# B.T.V.T.I.- Bayesian tool for various types of inferences- マニュアル・解説集

Hiroki INOUE, M.D., Ph.D.

2020 年 9 月 6 日  
2021 年 2 月 16 日改訂

概要	…P2
環境設定	
起動方法	
操作方法	
利用規約	
A. 二項分布における生起確率の推測ツール	…P3
B. 負の二項分布における生起確率の推測ツール	…P5
C. 三項分布における各事象の生起確率の推測ツール	…P6
D. 二つの割合の差の推測（正規分布近似）ツール	…P8
E. 対数オッズ比の推測ツール	…P9
F. ピーターセン法の推測（超幾何分布の二項分布近似）ツール	…P10
G. 二群の発生率比の推測（ポアソン分布近似）ツール	…P11
H. NNT（治療必要数）の推測ツール	…P12
I. BCPNN 法によるシグナル検出の推測ツール	…P13
J. ワイブル分布の中央値の推測ツール	…P15
K. ワイブル分布の尺度母数の推測ツール	…P16
L. 感度・特異度・有病率を用いた検査的中率の推測ツール	…P17

## 概要：

本ソフト B.T.V.T.I.は種々のベイズ推測を視覚化することにより、保健医療福祉領域における意思決定へのベイズ推測の応用についての理解を助けるものである。

## 環境設定：

1. まず Python3 をダウンロード→インストールする。Python2 ではなく Python3 をダウンロード・インストールする。

2. 次に scipy、numpy、matplotlib を追加インストールする。

※Linux, MacOS はデフォルトで Python(Python2,Python3)がインストールされている場合があり、その Python3 のバージョンが古いと、同時にインストールされている scipy、numpy、matplotlib のバージョンが古い場合、または最新の scipy、numpy、matplotlib がインストールされない場合がある。必ずなるべく新しいバージョンの Python3 をインストールすることが推奨される。

※Linux で conda 系のソフトがインストールされている状態で起動すると、BTVTI の日本語文字表示が不鮮明になることがある。もし BTVTI の日本語文字表示が不鮮明になる場合は同 PC の anaconda/miniconda をアンインストールするか、または使わないようにすることが推奨される。

3. BTVTI のフォルダを解凍し、PC 内の任意の場所におく。

## 起動方法：

Linux, MacOS の場合の一例

端末を開いて、Python3+半角スペース+パス付き BTVTI.py と入力してエンターキーを押下する。

Winsows10 の場合の一例

コマンドプロンプト画面を開いて、Python+半角スペース+パス付き BTVTI.py と入力してエンターキーを押下する。または、BTVTI.py ファイルをマウスで直接ダブルクリックして起動させる。

## 操作方法：

BTVTI.py ファイルを起動すると、各ツールの選択場面が表示されるので、実行したい推測を選択する。各ツール画面右側にある、スライダー（スケール）またはアップダウンボタン（スピンボックス）を、マウスでドラッグまたはクリックして数値を入力・変化させる。事後確率の範囲もスライダー（スケール）をマウスで動かして指定する。マウスによる入力に応じて、事後確率のグラフ曲線（事前確率曲線の表示されるものも一部ある）の動きと変化を見ることができる。同時に指定された範囲の事後確率の値を見ることができる。

## 利用規約

以下お読みの上、操作を行ってください。本ソフトのダウンロード・インストール・使用・複製その他により生じた問題・不具合・事故・損害その他一切について製作者は責任を負いませんので各自の責任と判断において使用してください。本ソフトを逆アセンブルすること、解析または改造することを一切禁じます。本ソフトを販売することや営利目的、商用目的に使用することを固く禁じます。本ソフトは、大学または大学に準ずる教育研究機関または組織において保健医療・生命科学・健康科学系分野の教育・研究・学習の場面を想定して製作されておりますので、それ以外の用途・場所での使用を推奨いたしません。本ソフトは非売品であり、かつ、一般消費者向けに提供されるものではありません。製作者の承諾なく本ソフトの画面表示そのもの、または画面表示を加工したものを引用すること、本ソフトそのものに関する記事を学術文献に掲載することを禁じます。また正当な理由無くその他一般のメディアに掲載することもご遠慮ください。本ソフトの使用者が任意に行っている、全く別個の調査研究においても、本ソフトを使用し、その動作・演算・出力結果を元に判断または結論された成果を学術集会や学術文献に発表または投稿される際は、製作者にその旨かならず速やかにご一報ください。また、予告無く仕様の変更や頒布の終了など行うこともありえますので、予め上記の旨どうかご同意の上で本ソフトのご使用をお願い申し上げます。

# A.二項分布における生起確率の推測ツール

## 本ツールについて

ベルヌーイ試行をある回数繰り返した際に観測される成功回数と失敗回数から、1回あたりの成功事象の生起確率のベイズ推測を行う。成功回数と失敗回数の入力によるベータ分布曲線の変化を表示させる。事前情報に自然共役事前分布であるベータ分布  $Be(a,b)$ を用いる。事前無情報のときの事前情報は、一様分布でもある  $Be(1,1)$ を用いている。

## 使い方

1. 観測値(観測回数)の入力…画面右段上方にある、事前情報の「成功回数」、事前情報の「失敗回数」、観測値の「成功回数」、観測値の「失敗回数」、のスライダー(スケール)をそれぞれマウスで変化させる。

2. 上記入力値の変化に応じグラフ内の事前確率曲線(青)と事後確率曲線(赤)が変化する。事後確率曲線のY軸方向は適当な縮尺により最高点での高さを一定にしてある。

3. 生起確率の推測

画面右段下方にある、2つのスライダー(スケール)をマウスで動かし、成功事象  $\pi$  の生起確率を推定する範囲を指定すると、グラフ内の2本の縦線(緑色線)の間と事後確率曲線(赤線)の下部で囲まれた部分の面積が計算され表示される。

## 参考：ベイズの定理と二項分布について

二項分布において、1回のベルヌーイ試行で事象  $\pi$  が出現する(「成功」とする)確率を  $p$ 、出現しなかった(「失敗」とする)確率を  $1-p$  とする( $0 \leq p \leq 1$ )。その場合、 $N$ 回の試行で  $\pi$  の成功回数が  $x$  回である確率  $P(x | p)$ は、

$$P(x | p) = \binom{N}{x} \cdot p^x \cdot (1-p)^{N-x} \quad \dots \textcircled{1} \quad \text{となる。 (注: } \binom{N}{x} = {}_N C_x \text{ である)}$$

ここで事後分布  $P(p | x)$ を「 $N$ 回の試行で  $\pi$  の成功回数が  $x$  回だった場合の  $p$  の確率」とし、 $P(p)$ を観測を始める前の  $p$  の値についての確率とした場合、 $P(x | p)$ は、「 $N$ 回の試行で  $\pi$  の成功確率が  $p$  である場合の、 $\pi$  の成功回数が  $x$  回である確率」であることから  $P(p | x)$ の事前分布とすることができるので、ベイズの定理によると、

$$P(p | x) = \frac{P(x | p) \cdot P(p)}{P(x)} \quad \dots \textcircled{2} \text{となる。}$$

$P(x | p)$ を  $p$  の尤度とみなし、 $p$  の尤度  $L(p | x) = P(x | p)$ とすると、 $P(x)$ は  $p$  については定数項なので、 $P(p | x) \propto L(p | x) \cdot P(p) \quad \dots \textcircled{3}$ と表すことができる。

$$\text{二項分布の場合} \textcircled{1} \text{より } L(p | x) = \binom{N}{x} \cdot p^x \cdot (1-p)^{N-x} = C \cdot p^x \cdot (1-p)^{N-x}$$

であり、これを③に代入すると、事後分布  $P(p | x)$ は  $P(p | x) = C' \cdot p^x \cdot (1-p)^{N-x} \cdot P(p)$  となる( $C, C'$ は定数)。...④

$$\text{今、事前分布 } P(p) \text{ をベータ分布 ( } B(a,b) \text{ はベータ関数) } \quad Be(a,b) = \frac{1}{B(a,b)} \cdot p^{a-1} \cdot (1-p)^{b-1}$$

で与えると、「 $N$ 回の試行で  $\pi$  の成功回数が  $x$  回」という観測値が得られた時の事後分布  $P(p | x)$ は、

$$P(p | x) = Be(a+x, b+N-x)$$

となる。二項分布の場合、ベータ分布は自然共役事前分布である(すなわち事前分布にベータ分布を用いると、事後分布もベータ分布になる)。

ちなみに  $Be(a,b)$ の平均(期待値)、分散、モードは各々、

$$\text{期待値} = \frac{a}{a+b}, \quad \text{分散} = \frac{ab}{(a+b)^2(a+b+1)}, \quad \text{モード} = \frac{a-1}{a+b-2} \quad (\text{ただし } a > 1, b > 1 \text{ の場合}) \quad \text{である。}$$

例題：ある治療法の有効性を推定するとき、観測されたデータは 20 回中有効 12 回、無効 8 回であった。私的な事前情報は無情報であった。なおこの治療法は独立して行われていたとする。この想定で本治療法の有効性  $p$  が 50%以上になる確率を計算する。例解:私的な事前情報は無情報より事前分布  $\text{Be}(1,1)$  であるので、事後分布は  $\text{Be}(13,9)$  となる。本ツールは  $p$  が 0.5 以上であると設定することにより、 $p$  が 0.5 から 1 の間の値を取る事後確率は 80.8%となる。

#### 文献

東京大学教養学部統計学教室編：自然科学の統計学，東京大学出版会，第 1 版，251-263，東京，1992.  
他多数

## B.負の二項分布における生起確率の推測ツール

### 本ツールについて

ベルヌーイ試行を繰り返し、 $m$  回目の成功の前に  $y$  回の失敗を観測したとする。 $m-1$  回の成功と  $y$  回の失敗のすべての組み合わせの後に  $m$  回目の成功を得たとする、 $y$  は負の二項分布に従う。観測される失敗回数と成功回数から、ベルヌーイ試行 1 回あたりの成功事象の生起確率のベイズ推測を行う。自然共役事前分布は二項分布と同じベータ分布  $Be(a,b)$  になるが、そのままだと前述 A. のツールと同じになるので事前情報は、無情報事前分布である  $Be(1,1)$  の代わりに Jeffrey's rule による事前情報を本ツールでは用いている。

### 使い方

1. 観測値の入力…画面右段上方にある、観測値の「成功回数」、観測値の「失敗回数」、のスライダー(スケール)をそれぞれマウスで変化させる。
2. 上記入力値の変化に応じグラフ内の事後確率曲線(赤)が変化する。事後確率曲線の Y 軸方向は縮尺により最高点での高さを一定にしてある。
3. 生起確率の推測…画面右段下方にある、2 つのスライダー(スケール)をマウスで動かし、成功事象  $\pi$  の生起確率を推定する範囲を指定すると、グラフ内の 2 本の縦線(緑色線)の間の事後確率が計算表示される。

### 参考：ベイズの定理と負の二項分布について

負の二項分布は文献によって定義が異なるので、Lee(2004)を引用して本ツールでは以下のように定義する。ベルヌーイ試行を繰り返し、 $m$  回目の成功の前に  $y$  回の失敗を観測したとする。 $M-1$  回の成功と  $y$  回の失敗のすべての組み合わせの後に  $m$  回目の成功を得たとする、 $y$  は負の二項分布に従う。1 回のベルヌーイ試行で成功事象  $\pi$  が出現する確率を  $p$  とすると、 $\pi$  の失敗回数が  $y$  である確率  $P(y | p)$  は、

$$P(y|p) = \binom{m+y-1}{y} p^m (1-p)^y \cdots \textcircled{1} \quad \text{となる。 (注: } \binom{m+y-1}{y} = {}_{m+y-1}C_y \text{ である) }$$

$$\text{二項分布と同様に、} \textcircled{1} \text{より } L(p|y) = \binom{m+y-1}{y} p^m (1-p)^y = C \cdot p^m (1-p)^y \cdots \textcircled{2}$$

$$\text{事後分布 } P(p | y) \text{ は } P(p|y) = C' \cdot p^m (1-p)^y \cdot P(p) \text{ となる}(C, C' \text{ は定数})。$$

今、事前分布  $P(p)$  を Jeffrey's rule による事前情報  $Be\left(0, \frac{1}{2}\right)$  ( $Be(a,b)$  は母数  $a, b$  のベータ分布) で与えると (Lee(2004) による)、「 $m$  回めの成功が生じるまでに  $y$  回失敗」という観測値が得られた時の事後分布  $P(p | y)$  は、 $P(p|y) = Be\left(m, \frac{1}{2} + y\right)$  となる。

例題：ある新薬の治験を行うとき、20 回の有効を観測するまでに 12 回の無効を観測した。この治療法は各回独立して行われていたとする。以上の想定で本治療法の有効率  $p$  が 50% 以上になる確率を計算する。なお観測者に事前のデータは無く、事前情報を Jeffrey's rule により定めるものとする。例解：ベイズ推測によると  $p$  の事後分布は  $Be(20, 12.5)$  となる。本ツールによると  $p$  が 0.5~1 の範囲を取る事後確率は 91% となる。

### 文献

Peter M. Lee : Bayesian Statistics: An Introduction, Hodder Arnold, 第 3 版, 202-204, London, 2004.

## C.三項分布における各事象の生起確率の推測ツール

本ツールについて

意思決定においてしばしば3つ以上に設定されるアウトカムに当てはめられる、多項分布モデルに対するベイズ推測を視覚的に理解する。本ツールで用いた多項分布は、使用者の理解のしやすさを考慮し三項分布としている。観測された各事象の生起度数を入力し、多項分布のベイズ推測において事前確率分布に自然共役事前分布であるディリクレ分布を用い、各事象の生起確率の周辺事後確率分布(ベータ分布)の変化をグラフ曲線表示し、多項分布の周辺事後確率分布の変化をそれぞれの分布のグラフを表示する事により視覚的に理解することができる。

使い方：

### 1. 事前情報の入力

画面右段上方の、3つのスライダー(スケール)をマウスで動かして各事象の生起度数を入力する。

### 2. 各事象の観測値(観測回数)の入力

画面右段中ほどの、3つのスライダー(スケール)をマウスで動かして各事象の観測値(観測回数)を入力する。

### 3. 上記入力値の変化に応じグラフ内の周辺事後確率曲線が変化する。各事象は事象a：赤、事象b：青、事象c：紫である。周辺事後確率曲線のY軸方向は適当な縮尺により最高点での高さを一定にしてある。

### 4. 生起確率の推測

画面右下段にある、2つのスライダー(スケール)をマウスで動かして、a、b、c各事象の生起確率を推定する範囲を指定すると、グラフ内の2本の縦線(緑色線)の間とそれぞれの周辺事後確率曲線の下部で囲まれた部分の面積が計算され表示される。

参考：三項分布のベイズ推測について

三項分布において、1回のベルヌーイ試行でそれぞれ排反な事象 $\pi_1$ ,  $\pi_2$ ,  $\pi_3$ が出現する確率をそれぞれ $p_1$ ,  $p_2$ ,  $p_3$  ( $p_1+p_2+p_3=1$ )とする ( $0 \leq p_1 \leq 1$ ,  $0 \leq p_2 \leq 1$ ,  $0 \leq p_3 \leq 1$ )。その場合、N回の試行で $\pi_1$ ,  $\pi_2$ ,  $\pi_3$ の出現度数が各々 $n_1$ ,  $n_2$ ,  $n_3$ 回である確率 $P(n_1, n_2, n_3 | p_1, p_2, p_3)$ は、

$$P(n_1, n_2, n_3 | p_1, p_2, p_3) = \frac{(n_1 + n_2 + n_3)!}{n_1! n_2! n_3!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3}$$

$$\begin{aligned} \text{ベイズの定理によると、} P(p_1, p_2, p_3 | n_1, n_2, n_3) &= \frac{P(n_1, n_2, n_3 | p_1, p_2, p_3)}{P(n_1, n_2, n_3)} \cdot P(p_1, p_2, p_3) \\ &\propto L(p_1, p_2, p_3 | n_1, n_2, n_3) \cdot P(p_1, p_2, p_3) \end{aligned}$$

$$L(p_1, p_2, p_3 | n_1, n_2, n_3) \propto C \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3}$$

今、事前分布 $P(p_1, p_2, p_3)$ をDirichlet分布 ( $\Gamma(n)$ はガンマ関数)

$$P(p_1, p_2, p_3) = \frac{\Gamma(n_{01} + n_{02} + n_{03})}{\Gamma(n_{01})\Gamma(n_{02})\Gamma(n_{03})} \cdot p_1^{n_{01}-1} \cdot p_2^{n_{02}-1} \cdot p_3^{n_{03}-1} \quad \text{とすると、}$$

$$P(p_1, p_2, p_3 | n_1, n_2, n_3) \propto L(p_1, p_2, p_3 | n_1, n_2, n_3) \cdot P(p_1, p_2, p_3)$$

$$\propto C \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot \frac{\Gamma(n_{01} + n_{02} + n_{03})}{\Gamma(n_{01})\Gamma(n_{02})\Gamma(n_{03})} \cdot p_1^{n_{01}-1} \cdot p_2^{n_{02}-1} \cdot p_3^{n_{03}-1}$$

$$\propto C \cdot \frac{\Gamma(n_{01} + n_{02} + n_{03})}{\Gamma(n_{01})\Gamma(n_{02})\Gamma(n_{03})} \cdot p_1^{n_1+n_{01}-1} \cdot p_2^{n_2+n_{02}-1} \cdot p_3^{n_3+n_{03}-1}$$

$$\dots$$

$$= \frac{\Gamma(n_1 + n_{01} + n_2 + n_{02} + n_3 + n_{03})}{\Gamma(n_1 + n_{01})\Gamma(n_2 + n_{02})\Gamma(n_3 + n_{03})} \cdot p_1^{n_1+n_{01}-1} \cdot p_2^{n_2+n_{02}-1} \cdot p_3^{n_3+n_{03}-1}$$

であり、事前分布  $P(p_1, p_2, p_3)$  が Dirichlet 分布のとき、事後分布  $P(p_1, p_2, p_3 | n_1, n_2, n_3)$  も Dirichlet 分布となる。

各事象の周辺事後確率分布について、 $P(p_1)$  の周辺事後確率分布  $P(p_1 | n_1, n_2, n_3)$  を考える。

$$P(p_1 | n_1, n_2, n_3) = \int \frac{\Gamma(n_1 + n_2 + n_3)}{\Gamma(n_1)\Gamma(n_2)\Gamma(n_3)} \cdot p_1^{n_1-1} \cdot p_2^{n_2-1} \cdot (1-p_1-p_2)^{n_3-1} dp_2$$

$$= \frac{\Gamma(n_1 + n_2 + n_3)}{\Gamma(n_1)\Gamma(n_2)\Gamma(n_3)} \cdot p_1^{n_1-1} \cdot \int p_2^{n_2-1} \cdot (1-p_1-p_2)^{n_3-1} dp_2$$

$$= \frac{\Gamma(n_1 + n_2 + n_3)}{\Gamma(n_1)\Gamma(n_2)\Gamma(n_3)} \cdot p_1^{n_1-1} \cdot (1-p_1)^{n_2+n_3-1} \int \left(\frac{p_2}{1-p_1}\right)^{n_2-1} \cdot \left(1-\frac{p_2}{1-p_1}\right)^{n_3-1} dp_2 \quad \theta_2 = \frac{p_2}{1-p_1} \text{ とすると、}$$

$$= \frac{\Gamma(n_1 + n_2 + n_3)}{\Gamma(n_1)\Gamma(n_2)\Gamma(n_3)} \cdot p_1^{n_1-1} \cdot (1-p_1)^{n_2+n_3-1} \int \theta_2^{n_2-1} \cdot (1-\theta_2)^{n_3-1} d\theta_2$$

$$= \frac{\Gamma(n_1 + n_2 + n_3)}{\Gamma(n_1)\Gamma(n_2)\Gamma(n_3)} \cdot p_1^{n_1-1} \cdot (1-p_1)^{n_2+n_3-1} \frac{\Gamma(n_2)\Gamma(n_3)}{\Gamma(n_2 + n_3)}$$

$$= \frac{\Gamma(n_1 + (n_2 + n_3))}{\Gamma(n_1)\Gamma(n_2 + n_3)} \cdot p_1^{n_1-1} \cdot (1-p_1)^{n_2+n_3-1}$$

よって周辺事後確率分布  $P(p_1 | n_1, n_2, n_3)$  は母数  $n_1, n_2+n_3$  のベータ分布  $Be(n_1, n_2+n_3)$  となる。 $(P(p_2 | n_1, n_2, n_3))$  と  $P(p_3 | n_1, n_2, n_3)$  も同様

ちなみに  $Be(a, b)$  の平均(期待値)、分散、モードは各々、

$$\text{期待値} = \frac{a}{a+b}, \quad \text{分散} = \frac{ab}{(a+b)^2(a+b+1)}, \quad \text{モード} = \frac{a-1}{a+b-2} \quad (\text{ただし } a > 1, b > 1 \text{ の場合})$$

である。

例題：ある疾患に対する、ある薬物治療のアウトカムを、奏効( $P_1$ )、不変( $P_2$ )、増悪( $P_3$ )とした。各アウトカムが 50%以上の確率で生起する周辺事後確信度( $P_1(p_1 \geq 0.5)$ ,  $P_2(p_2 \geq 0.5)$ ,  $P_3(p_3 \geq 0.5)$ )をそれぞれ求める。観測症例数 25 で集計を行い、それぞれの度数 8, 10, 7 を得た。事前情報は得られていなかったとする。本ソフトでは、 $P_1(p_1 \geq 0.5)$ ,  $P_2(p_2 \geq 0.5)$ ,  $P_3(p_3 \geq 0.5)$  は各々 2.6%、12.4%、1.0% であると表示される。

文献

Lesaffre E, Lawson A: Bayesian Biostatistics (Statistics in Practice), Wiley, 第 1 版, 90-92, West Sussex, 2012.

## D.二つの割合の差の推測(正規分布近似)ツール

本ツールについて

2つの事象のそれぞれ割合である、P1とP2の差の事後確率を、直接正規分布に近似して推測する。画面上方の表にある通り、P1の事前情報  $a_0/b_0$ 、P2の事前情報  $c_0/d_0$ 、P1の観測値  $a/b$ 、P2の観測値  $c/d$  である。観測値(度数、自然数)  $a_0$ 、 $b_0$ 、 $c_0$ 、 $d_0$ 、 $a$ 、 $b$ 、 $c$ 、 $d$  を画面右段のアップダウンボックス(スピンボックス)8か所にそれぞれ入力する。

すると、正規分布近似された  $P_1 - P_2$  の事後確率分布のグラフ曲線が表示され、入力値8つの値を変化させると同時にグラフ曲線の変化も観察することができる。

画面右段下のスライダー(スケール)を動かして、指定した範囲の信用区間の事前確率と事後確率の値がそれぞれ算出表示される。

### 参考：2つの割合の差の推測(正規分布近似)について

2つのベータ分布の差の分布は通常母数が4つになり、このままでは解析が困難である。そこで直接正規分布に近似して推測する

独立に割合  $p_1$  の事後分布  $\sim \text{Be}(a+1, b+1)$ 、割合  $p_2$  の事後分布  $\sim \text{Be}(c+1, d+1)$  となるとき、 $\text{Be}(a+1, b+1)$  の平均  $\mu$  (期待値)、分散  $\sigma^2$  は各々、

$$\mu = \frac{a+1}{a+1+b+1}, \quad \sigma^2 = \frac{(a+1)(b+1)}{(a+1+b+1)^2(a+1+b+1+1)}$$
 であるので、 $a$ 、 $b$  が十分大きい時、(ともに  $\geq 10$ )、

$\text{Be}(a+1, b+1)$  を、正規分布  $N(\mu, \sigma^2)$  に近似することができる。このとき、近似的に

$$p_1 - p_2 \text{ の事後分布} \sim N\left(\frac{a+1}{a+1+b+1} - \frac{c+1}{c+1+d+1}, \frac{(a+1)(b+1)}{(a+1+b+1)^2(a+1+b+1+1)} + \frac{(c+1)(d+1)}{(c+1+d+1)^2(c+1+d+1+1)}\right)$$

となる。また、 $p_1$  の事前分布  $\sim B(a_0, b_0)$ 、 $p_2$  の事前分布  $\sim B(c_0, d_0)$  のとき、 $p_1 - p_2$  の事後分布は近似的に

$$N\left(\frac{a+a_0}{a+a_0+b+b_0} - \frac{c+c_0}{c+c_0+d+d_0}, \frac{(a+a_0)(b+b_0)}{(a+a_0+b+b_0)^2(a+a_0+b+b_0+1)} + \frac{(c+c_0)(d+d_0)}{(c+c_0+d+d_0)^2(c+c_0+d+d_0+1)}\right) \text{ に従う}$$

ただし、 $a_0$ 、 $b_0$ 、 $c_0$ 、 $d_0$ 、 $a$ 、 $b$ 、 $c$ 、 $d$  は十分大きい(ともに  $\geq 10$ )ことが望ましい。

文献

繁枅算男：ベイズ統計入門，東京大学出版会，第1版，136-141，東京，1985.

## E.対数オッズ比の推測ツール

本ツールについて

2つの事象を表示した2×2表において、対数オッズ比の事後確率を、正規分布に近似して推測する。画面上方の表にある通り、事前情報から得た事前対数オッズ比は  $\text{Ln}(a_0d_0/b_0c_0)$ 、事前情報と観測値から得た事後対数オッズ比は  $\text{Ln}((a+a_0)(d+d_0)/(b+b_0)(c+c_0)) - \text{Ln}(a_0d_0/b_0c_0)$  である。

観測値(度数。自然数)  $a_0, b_0, c_0, d_0, a, b, c, d$  を画面右段のアップダウンボックス(スピンボックス)8か所にマウスでそれぞれ入力する。

すると、正規分布近似された対数オッズ比の事後確率分布のグラフ曲線が表示され、入力値8つの値を変化させると同時にグラフ曲線の変化も観察することができる。

画面右段下のスライダー(スケール)をマウスで動かして範囲を指定すると、指定した範囲の信用区間の事前確率と事後確率の値がそれぞれ算出表示される。

### 参考：対数オッズ比の推測(正規分布近似)について

2×2表において対数オッズ比を推測する。

事前無情報のとき、群1、2が独立に、群1において  $p_1$  の観測値： $\pi$  が出現する回数  $a$ /しない回数  $b$ 、群2において  $p_2$  の観測値： $\pi$  が出現する回数  $c$ /しない回数  $d$ 、とすると、割合  $p_1$  の事後分布  $\sim \text{Be}(a+1, b+1)$ 、割合  $p_2$  の事後分布  $\sim \text{Be}(c+1, d+1)$  となる。このとき、 $a, b, c, d$  が独立または交換可能性を満たし観測されたならば、 $p_1$  のオッズ  $= \frac{p_1}{1-p_1}$ 、 $p_2$  のオッズ  $= \frac{p_2}{1-p_2}$  とすると、 $p_1, p_2$  の事

後分布の対数オッズ比は Woolf の方法と同様に、近似的に 
$$N \left( \log \frac{a + \frac{1}{2}}{c + \frac{1}{2}} \frac{b + \frac{1}{2}}{d + \frac{1}{2}}, \frac{1}{a+1} + \frac{1}{b+1} + \frac{1}{c+1} + \frac{1}{d+1} \right)$$
 に従う。

また、 $p_1$  の事前分布  $\sim \text{Be}(a_0, b_0)$ 、 $p_2$  の事前分布  $\sim \text{Be}(c_0, d_0)$  のとき、

$p_1, p_2$  の事後分布の対数オッズ比は、近似的に 
$$N \left( \log \frac{a + a_0 - \frac{1}{2}}{c + c_0 - \frac{1}{2}} \frac{b + b_0 - \frac{1}{2}}{d + d_0 - \frac{1}{2}}, \frac{1}{a+a_0} + \frac{1}{b+b_0} + \frac{1}{c+c_0} + \frac{1}{d+d_0} \right)$$
 に従う。

ただし近似の条件として、事前無情報の場合、 $a, b, c, d$  いずれも  $>5$ 、事前情報ありの場合、 $a_0+a, b_0+b, c_0+c, d_0+d$  いずれも  $>5$  となることが望ましい。

文献

渡部洋：ベイズ統計学入門，福村出版，第1版，94-96，東京，1999.

## F. ピーターセン法の推測（超幾何分布の二項分布近似） ツール

本ツールについて

総数  $N$  の集団から  $M$  個体を捕獲し、ラベル後放流し、その後再捕獲した  $n$  個体のうち  $r$  個体がラベルされている場合、 $r$  は超幾何分布に従う。

$n, r \ll N, M$  の場合、 $r$  は二項分布に近似される。このとき、 $p=M/N$  は観測値  $n, r$  により事後確率はベータ分布に従う。

本ツールでは上記の想定で  $p$  の推測を行う。

使い方：

1. 画面右段中ほどのスライダー(スケール)で、 $r$  と  $n \cdot r$  をマウスを使って値を入力する。
2.  $P$  の事後分布のグラフ(赤)の表示が変化する。
3. 画面右段下方の、スライダー(スケール)をマウスで変化させることにより、指定した範囲(信用区間)で  $p$  の取りうる確信度を求めることができる。
4. グラフ曲線の  $Y$  軸方向は、縮尺により最高点(モード)での高さを常に一定に表示されるようになっている。
5. 事前情報は無情報であるベータ分布  $Be(1, 1)$ としている。

参考：超幾何分布のベイズ推測と二項分布近似について

ピーターセン法ではサンプルサイズ  $N$  の有限母集団から  $M$  捕獲してラベルし放流し、 $n$  個体再捕獲したうちの  $r$  個体にラベルがあった場合に  $N$  を推定する。このとき  $r$  は超幾何分布に従う。

超幾何分布の確率密度関数  $P(r)$  は、

$$P(r) = \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}} = \binom{n}{r} \frac{M^r (N-M)^{n-r}}{N^n} \quad \text{である。}$$

ここで  $n, r \ll N, M$  の場合、 $r$  は二項分布に近似される。 $M/N = p$  とすると、

$$P(r, n, p) = \binom{n}{r} p^r (1-p)^{n-r}$$

ベイズの推測では  $p$  の事後確率は事前無情報の場合、母数  $r+1, n-r+1$  のベータ分布  $Be(r+1, n-r+1)$  に従う。

例題：  $M=200$  体の個体数にラベルした生物を放流して 30 体を再捕獲した。そのうちラベルされていた個体は 2 体であった。このとき  $p=M/N$  は 90% の確信度で 0.03 から 0.20 の間の値を取ることがツールに表示された。この  $p$  の範囲だと  $N=p/M$  より  $N$  は 90% の確信度で 1000 から 6666 の間の値を取ることが推測された。

文献

赤嶺達郎：枠どり法と Petersen 法の区間推定における伝統的統計学とベイズ統計学との比較，水研センター研報，（ 2 ）： 25 - 34 ， 2002.

## G. 二群の発生率比の推測（ポアソン分布近似） ツール

本ツールについて

二つの事象の発生回数をそれぞれポアソン分布近似して、二つの発生率比をベイズ推測する。  
事象の発生がまれな場合に用いる。群 1、群 2 における事象  $\pi$  の(回数×単位時間)あたりの平均発生数  $\theta_1 = a/(N \times b)$ ,  $\theta_2 = c/(N \times d)$  について発生数の観測値(自然数)  $a, c, b, d$  を、画面右段のアップダウンボックス(スピンボックス)4 箇所それぞれ入力すると、単位時間  $N$  あたりの発生回数  $\theta_1/N$  と  $\theta_2/N$  の比、すなわち  $\theta_1$  と  $\theta_2$  の率比は  $F$  分布の形で表され、 $(\theta_1/\theta_2)(n_2/n_1)/(t_2/t_1)$  の事後分布である  $F$  分布のグラフ曲線が表示され、観測値 4 つの値を変化させると同時にグラフ曲線の変化も観察することができる。画面右段下のスライダー(スケール)を動かして、指定した範囲の換算された信用区間の率比( $\theta_1/\theta_2$ )の事後確率の値が算出表示される。

**参考：** 事象の発生回数をポアソン分布近似して、二群の発生率比をベイズ推測する

事象  $\pi$  の発生がまれなものとして、 $\pi$  の発生回数  $x$  がポアソン分布に従うとする。単位回数  $m$  回あたりの  $\pi$  の平均発生回数  $\theta$  とすると、平均発生回数は  $\theta/m$  で表される。

$$\text{ポアソン分布の確率密度関数： } P(x) = \frac{e^{-\theta} \theta^x}{x!}$$

$\pi$  の平均発生回数  $\theta$  は、 $\theta > 0$  であることから、ジェフリーズの法則より事前無情報での  $\theta$  の事前分布を  $\propto \theta^{-1}$  とする。

$\pi$  の発生回数の観測値が、 $N \times t$  回あたり ( $N$  は単位時間)、 $n$  回であるとする、 $\theta$  の事後分布は、 $G(n, t)$  に従う。 $G(n, t)$  はガンマ分布である。ガンマ分布の確率密度関数： $G(n, t): P(\theta) = \frac{t^n}{\Gamma(n)} \theta^{n-1} e^{-t\theta}$

$\theta$  の事前分布が  $G(n_0, t_0)$  のとき、 $\theta$  の事後分布は  $G(n+n_0, t+t_0)$  となる。

独立に、群 1 で  $\theta_1$  の事後分布  $\sim G(n_1, t_1)$ 、群 2 で、 $\theta_2$  の事後分布  $\sim G(n_2, t_2)$ 、とすると、

$$\frac{t_1 \times \frac{\theta_1}{n_1}}{t_2 \times \frac{\theta_2}{n_2}} = (\theta_1/\theta_2)(n_2/n_1)/(t_2/t_1) \sim F(2n_1, 2n_2) \quad \text{となり、自由度 } 2n_1, 2n_2 \text{ の } F \text{ 分布に従う。これより、}$$
$$\frac{\theta_1}{\theta_2} = \frac{\frac{\theta_1}{N}}{\frac{\theta_2}{N}} \quad \text{であることから、群 1 と群 2 の平均発生回数の率比の事後確率を推測することができるように}$$

なる。

※ポアソン分布の性格から  $N$  が単位時間となることが多く、この方法は発生率比の場合に適用されることの方が多いと考えられる。

文献

- ・ Gelman A : Bayesian Data Analysis, Chapman & Hall/CRC, 第 2 版, 53-55, Boca Raton, 2003.
- ・ 養谷千風彦：統計分布ハンドブック, 朝倉書店, 第 1 版, 240-260, 東京, 2003.

# H.NNT（治療必要数）の推測ツール

本ツールについて

介入群と対照群において観測された、二値データで表される事象の度数をマウスで入力することによって、算出される NNT(治療必要数)についてベイズ推測を行い、事後確率の変化をグラフ曲線表示する事により視覚的に理解することができる。推測の方法は Thabane (2003)によった。

使い方：

1. NNT の式は、画面左段上方にある  $2 \times 2$  表にある通りである。この表によると介入群と対照群の事象の割合はそれぞれ介入群： $P_i = a/(a+b)$ 、対象群： $P_e = c/(c+d)$ となる。画面右段上方から縦並びにある 4 つのアップダウンボックス(スピンボックス)に、マウスでそれぞれ観測値  $a$ 、 $b$ 、 $c$ 、 $d$  の値を入力すると、NNT の事後確率のグラフ曲線表示が変化し、 $a$ 、 $b$ 、 $c$ 、 $d$  の値の変化と同時にグラフ曲線の変化も観察することができる
2. 同時に、画面右下のスライダー(スケール)をマウスで左右に動かして、指定した範囲の信用区間の NNT の事後確率の値を算出表示する。
3. 事後確率曲線の縦軸方向は縮尺により最高点での高さを一定にしてある。

参考：NNT(治療必要数)のベイズ推測について

Thabane (2003)によると、2つの治療手段においてイベント発生数をそれぞれ $x_1$ ,  $x_2$ 、全症例数をそれぞれ $n_1$ ,  $n_2$ としたとき、 $p_1 = x_1/n_1$ ,  $p_2 = x_2/n_2$ であることから  $NNT = 1/|p_1 - p_2|$ となる。 $Be(a, b)$ を母数 $a, b$ のベータ分布とすると、 $p_1$ の事後分布 $\sim Be(x_1+1, n_1 - x_1+1)$ 、 $p_2$ の事後分布 $\sim Be(x_2+1, n_2 - x_2+1)$ であり、 $p_1 - p_2$ の事後確率の期待値 $\mu_p$ 、分散 $\sigma_p^2$ は、 $\mu_p = \mu_1 - \mu_2$ ,  $\sigma_p^2 = \sum_{i=1}^2 \frac{\mu_i(1-\mu_i)}{n_i+1+1+1}$ となる。 $\mu_i = (x_i+1)/(n_i+1+1)$ である。

以上より、 $x_1, x_2, n_1, n_2$ が与えられたときのNNTの事後確率の分布密度関数 $P(y | x_1, x_2, n_1, n_2)$ は以下のよう

に漸近的に逆正規分布となる(註)。 $P(y | x_1, x_2, n_1, n_2) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left\{-\frac{(1/y - \mu_p)^2}{2\sigma_p^2}\right\}$  この事後分布の確率

密度関数は、左右の高さが異なる2峰性の曲線であり、期待値と分散が存在しない。モードは2つあり、

$$\text{モード(1)} = -\frac{\mu_p + \sqrt{\mu_p^2 + 8\sigma_p^2}}{4\sigma_p^2}, \text{モード(2)} = \frac{\sqrt{\mu_p^2 + 8\sigma_p^2} - \mu_p}{4\sigma_p^2} \text{となる。}$$

$P(y | x_1, x_2, n_1, n_2)$ の確率密度関数は事後の NNT が負の値をとる場合の事後確率も表示されているが、その部分についての解釈のしかたは一定していない。Thabane は扱われる治療手段が対照よりも優位であることが十分期待でき、NNT を構成するリスク差が正の値であることが確定している場合に限り推測を行うことを提唱している。

例題：ある新しい治療手段の有効症例数/全実施症例数=11/20、対して従来の治療手段の有効症例数/全実施症例数=5/20であった( $a=11$ ,  $b=9$ ,  $c=5$ ,  $d=15$ と入力)。この場合 NNT は 3.3 となるが、本ソフト上では NNT 値の事後確率は、95%の確信度で信用区間 1 から 23 の間を取ることが即座に表示される。

文献：詳細は以下を参照のこと

THABANE L : A closer look at the distribution of number needed to treat (NNT): a Bayesian approach, Biostatistics, 4 ( 3 ) : 365-370, 2003.

註：逆正規分布(inverse normal distribution)は逆ガウス分布(ワルド分布)とは異なる。詳細は「統計分布ハンドブック(朝倉書店)」初版 P282 と P382 を参照のこと

# I.BCPNN 法によるシグナル検出の視覚化ツール

本ツールについて

薬剤有害事象のベイズ的なシグナル検出手法の一つである BCPNN 法(Bayesian Confidence Propagation Neural Network Method)を、有害事象に関する報告の観測度数をマウスで入力することによって、IC<sub>11</sub>(後述)についてのベイズ推測を行い、観測値の入力に伴う IC<sub>11</sub> の事後確率の変化をグラフ曲線表示するとともに、シグナル検出有無の表示を視覚的に理解することができる。

使い方：

1. 画面左段上方にある、注目薬剤とその他の薬剤、注目イベント数とその他イベント数で、報告集団を 4 グループに分割した 2×2 表において、それぞれの観測値(a、b、c、d)を画面右段にある、アップダウンボックス(スピンボックス)にそれぞれマウスで入力すると、薬剤の注目するイベントに関する IC<sub>11</sub> (IC ; information component)がベイズ推測され、IC<sub>11</sub> の事後確率のグラフ曲線(正規近似)が表示される。
2. 2×2 表の観測値を変化させると同時にグラフ曲線の変化も観察することができる。同時に、画面右下のスライダー(スケール)を動かして、指定した範囲の信用区間の IC<sub>11</sub> の事後確率の値が算出表示される。
3. 計算の結果、「シグナルあり」か「シグナルなし」の判定結果が画面右下に表示される。
4. 事後確率曲線の縦軸方向は縮尺により最高点での高さを一定にしてある。

参考：BCPNN(Bayesian Confidence Propagation Neural Network)について

薬剤の有害事象に関する報告は、以下の 2 表のように 2×2 表に分割される。

各セルは報告件数

	注 目 イ ベ ン ト	そ の 他 イ ベ ン ト	計
注目薬剤	n <sub>11</sub>	n <sub>12</sub>	n <sub>1+</sub>
その他薬剤	n <sub>21</sub>	n <sub>22</sub>	n <sub>2+</sub>
計	n <sub>+1</sub>	n <sub>+2</sub>	n <sub>++</sub>

各セルは確率( $\alpha_1, \alpha_2, \beta_1, \beta_2$  はベータ分布のパラメータ。 $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$  はディリクレ分布のパラメータ。)

	注 目 イ ベ ン ト	そ の 他 イ ベ ン ト	計
注目薬剤	P <sub>11</sub> ( $\gamma_{11}$ )	P <sub>12</sub> ( $\gamma_{12}$ )	P <sub>1+</sub> ( $\alpha_1$ )
その他薬剤	P <sub>21</sub> ( $\gamma_{21}$ )	P <sub>22</sub> ( $\gamma_{22}$ )	P <sub>2+</sub> ( $\alpha_2$ )
計	P <sub>+1</sub> ( $\beta_1$ )	P <sub>+2</sub> ( $\beta_2$ )	P <sub>++</sub> = 1

$$\begin{aligned} \alpha_1 + \alpha_2 &= \alpha & P_{1+} + P_{2+} &= 1 \\ \beta_1 + \beta_2 &= \beta & P_{+1} + P_{+2} &= 1 \end{aligned}$$

P<sub>1+</sub>、P<sub>+1</sub>、はそれぞれベータ分布に従うとする。すなわち P<sub>1+</sub> ~ Be( $\alpha_1, \alpha - \alpha_1$ )、P<sub>+1</sub> ~ Be( $\beta_1, \beta - \beta_1$ )

BCPNNの場合、P<sub>11</sub>の期待値(理論値)は、P<sub>1+</sub>・P<sub>+1</sub>

$$\log_2 \frac{\text{実測値}}{\text{期待値}} \text{である、} IC_{ij} = \log_2 \frac{P_{ij}}{P_{i+} P_{+j}} \text{を考えると、このとき、興味があるのは} IC_{11} = \log_2 \frac{P_{11}}{P_{1+} P_{+1}}$$

P<sub>ij</sub>=P(1, 1)P(1, 2)P(2, 1)P(2, 2) の同時分布はディリクレ分布

$$P_{ij} = \frac{1}{Z} \cdot P_{11}^{\gamma_{11}-1} \cdot P_{12}^{\gamma_{12}-1} \cdot P_{21}^{\gamma_{21}-1} \cdot P_{22}^{\gamma_{22}-1} = \frac{\Gamma(\gamma_{11} + \gamma_{12} + \gamma_{21} + \gamma_{22})}{\Gamma(\gamma_{11})\Gamma(\gamma_{12})\Gamma(\gamma_{21})\Gamma(\gamma_{22})} \cdot P_{11}^{\gamma_{11}-1} P_{12}^{\gamma_{12}-1} P_{21}^{\gamma_{21}-1} P_{22}^{\gamma_{22}-1}$$

$P_{11}$ の周辺分布は、 $\gamma_{11} + \gamma_{12} + \gamma_{21} + \gamma_{22} = \gamma$  とすると、 $P_{11} = \frac{\Gamma(\gamma)}{\Gamma(\gamma_{11})\Gamma(\gamma - \gamma_{11})} P_{11}^{\gamma_{11}-1} (1 - P_{11})^{\gamma - \gamma_{11}-1} \sim \text{Be}(\gamma_{11}, \gamma - \gamma_{11})$

事前情報  $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_{11}$  が所与のとき、Orre (2000) によると  $IC_{11}$  の事後分布の期待値  $E(IC_{11})$  と分散  $V(IC_{11})$  は、以下のように表される(※ツール作成者注：近似と考えられる)。

$$E(IC_{11}) = \log_2 \frac{(n_{11} + \gamma_{11})(n_{++} + \alpha)(n_{++} + \beta)}{(n_{++} + \gamma)(n_{1+} + \alpha_1)(n_{+1} + \beta_1)}$$

$$V(IC_{11}) = \left( \frac{1}{\log 2} \right)^2 \left[ \frac{n_{++} - n_{11} + \gamma - \gamma_{11}}{(n_{11} + \gamma_{11})(1 + n_{++} + \gamma)} + \frac{n_{++} - n_{1+} + \alpha - \alpha_1}{(n_{1+} + \alpha_1)(1 + n_{++} + \alpha)} + \frac{n_{++} - n_{+1} + \beta - \beta_1}{(n_{+1} + \beta_1)(1 + n_{++} + \beta)} \right]$$

$$\text{ただし、} \gamma = \gamma_{11} \frac{(n_{++} + \alpha)(n_{++} + \beta)}{(n_{1+} + \alpha_1)(n_{+1} + \beta_1)}$$

$$\text{事前無情報の場合、} \gamma_{11} = 1, \quad \alpha_1 = \beta_1 = 1, \quad \alpha = \beta = 2 \quad \text{において、} \quad \gamma = 1 \cdot \frac{(n_{++} + 2)^2}{(n_{1+} + 1)(n_{+1} + 1)}$$

$$\text{このとき、} E(IC_{11}) = \log_2 \frac{(n_{11} + 1)(n_{++} + 2)^2}{(n_{++} + \gamma)(n_{1+} + 1)(n_{+1} + 1)}$$

$$V(IC_{11}) = \left( \frac{1}{\log 2} \right)^2 \left[ \frac{n_{++} - n_{11} + \gamma - 1}{(n_{11} + 1)(1 + n_{++} + \gamma)} + \frac{n_{++} - n_{1+} + 1}{(n_{1+} + 1)(n_{++} + 3)} + \frac{n_{++} - n_{+1} + 1}{(n_{+1} + 1)(n_{++} + 3)} \right]$$

「おおよその」95%信頼区間(信用区間)の下限が0より大きい場合、すなわち

$$E(IC_{11}) - 2SE_{11} = E(IC_{11}) - 2\sqrt{V(IC_{11})} > 0 \quad \text{のとき「シグナルあり」と判断する。}$$

※ツールでは  $n_{11} : a, n_{12} : b, n_{21} : c, n_{22} : d$  にそれぞれ対応している。

例題：ある新薬に関する有害事象についての観測値がそれぞれ  $a = 5, b = 240, c = 160, d = 60000$  であった。入力に応じて操作画面上に  $IC_{11}$  の事後分布のグラフ曲線が表示される。 $IC_{11}$  の95%信頼区間の下限  $> 0$  となり、シグナルありと判定される。また、データセットの値を変化させることにより、シグナル検出の状態をシミュレートすることができる。例として  $a = 6, b = 240, c = 610, d = 60000$  の場合は  $IC_{11}$  の95%信頼区間の下限  $< 0$  となり、シグナルなしと判定される。

文献：BCPNN 法とその他のシグナル検出手法についての詳細は、以下を参照のこと

・渡邊裕之，松下泰之，渡辺篤ら：重要な安全性情報を早期に検出する仕組み—シグナル検出の最近の手法について—，計量生物学，25（1）：37-60，2004.

・久保田潔：自発報告からのシグナル検出—英国 MCA，米国 FDA，WHO の新しい方法—，薬剤疫学，6（2）：101-108，2001.

・Orre R, Lansner A, Bate A, et al. : Bayesian neural networks with confidence estimations applied to data mining, Computational Statistics & Data Analysis, 34（4）：473-493, 2000.

## J. ワイブル分布の中央値の推測ツール

本ツールについて

母集団の生存時間が 2 母数のワイブル分布に従う場合において、得られた観測時間データとイベント回数、形状母数(値所与)の入力により、Achcar(1984)による近似を用いて生存時間中央値の事後分布をベイズ推測し、生存時間中央値の事後確率分布を動的視覚的に理解することができる。

使い方：

1. 画面右段上の縦並びのアップダウンボックス(スピンドルボックス)に、マウスで観測時間のデータを入力する。データ要素の個数  $n$  は最大 10 個である。入力する観測時間のデータはソートされていなくても良い。  
※観測値 0 は観測なしとして扱われる。
2. その下のイベント回数  $r$  を入力する。 $r > n$  にならないようにする( $n - r$  は右側打ち切り回数になる)。
3. ワイブル分布の形状母数の値を設定する。画面右段やや下のスライダー(スケール)を左右に動かして設定する。本ツールではワイブル分布の形状母数  $\theta_2$  の値は 0.01~2 で設定できる。

### 4. 事後確率の測定

上記の操作により、生存時間中央値の事後確率のグラフ曲線の変化を見ることができる。表出された事後確率のグラフ曲線に対し、画面右下方の、スライダー (スケール) をマウスで変化させることにより、指定した範囲で生存時間中央値の事後確信度を概算で求めることができる。

### 5. 注意

グラフ曲線の Y 軸方向は、縮尺させ最高点(モード)での高さを常に一定に表示されるようになっている。各グラフにおける中央値の区間推定は指定した範囲の、グラフ内の 2 本の縦線(緑色線)の間とグラフ曲線の下部で囲まれた部分の面積が計算され表示される(注：本数値は概算値である)。

参考：ワイブル分布の生存時間の中央値のベイズ推測について

観測時間データ  $x$ (ベクトルデータ)、イベント回数  $r$ 、ワイブル分布の形状母数  $\theta_2$  から母集団の生存時間中央値  $m$  の事後分布を推測する際に、Achcar(1984)による手法(近似法)を用いると、 $n$  個の観測時間データ  $x_1, \dots, x_r, x_{r+1}, \dots, x_n$  において時間  $x_r$  でイベントが  $r$  回発生し、右側打ち切りが  $n - r$  回あるモデルを表す、2 母数のワイブル分布(尺度母数： $\theta_1$ 、形状母数： $\theta_2$ )では、

$\theta_1$  が事前無情報で  $\theta_1 = \text{定数}$  とすると、

$$m \text{ の条件付事後確率は、 } P(m|\theta_2, x) = S^{r+1} \frac{\theta_2}{m} \exp(-S)/r! \quad \text{と近似される。ただし、} S = \text{Ln} 2 \sum_{i=1}^n \left( \frac{x_i}{m} \right)^{\theta_2}$$

例題：生存曲線がワイブル分布に従うとされる、ある集団の生存時間データ(年)を観測したところ、 $\{1, 2, 3, 4, 6, 8, 9, 11\}$ であり、最初の 6 例が死亡、後ろの 2 例は打ち切りであった。また、この集団の生存モデルの形状母数は 1.2 であった。観測データとモデルの形状母数を入力すると、本ソフト上では集団の生存時間中央値は、95%の確信度で 2.24 年から 8.96 年であることが即座に表示された。

文献

Dellaportas P, Wright DE : Numerical prediction for the two-parameter Weibull distribution, The Statistician, 40 ( 4 ) : 365-372, 1991.

# K. ワイブル分布の尺度母数の推測ツール

## 本ツールについて

母集団の生存時間が 2 母数のワイブル分布に従う場合において、値所与の形状母数のもと、得られた観測時間データと打ち切りデータの入力により尺度母数をベイズ推測する。尺度母数・形状母数(所与)から変数変換して得た  $\lambda$  の事後確率分布を動的視覚的に理解することができる。尺度母数・形状母数から変数変換して得た  $\lambda$  を用いて尺度母数を推測する方法は貝瀬ら(2000)の方法によった。

## 使い方：

1. 画面右段上の縦並びのアップダウンボックス(スピンボックス)に、マウスで観測時間のデータを入力する。データ要素の個数  $n$  は最大 10 個である。入力する観測時間のデータはソートされていなくても良い。  
※観測値 0 は観測なしとして扱われる。
2. その下のイベント回数  $r$  を入力する。 $r > n$  にならないようにする( $n - r$  は右側打ち切り回数になる)。
3. ワイブル分布の形状母数の値を設定する。画面右段やや下のスライダー(スケール)を左右に動かして設定する。本ツールではワイブル分布の形状母数  $m$  の値は 0.1~2 で設定できる。

## 4. 事後確率の測定

上記の操作により、ワイブル母集団に関する変数  $\lambda$  の事後確率分布の曲線変化を見ることができる。表出された事後確率のグラフ曲線に対し、画面右下方の、スライダー(スケール)をマウスで変化させることにより、 $\lambda$  が指定した範囲を取る(このとき変換前の尺度母数  $\eta$  が  $\lambda$  により指定された範囲を取る)事後確信度を求めることができる。

## 5. 注意

※変数変換が  $m$  のべき乗を伴うため  $m$  の値が極端に小さい場合、やイベント回数  $r$  が少ない場合は  $\eta$  の推測が漠然としたものになる。

グラフ曲線の Y 軸方向は、縮尺により最高点(モード)での高さを常に一定に表示されるようになっている。

## 参考：ワイブル生存時間データのハザード比のベイズ推測について

母集団の寿命  $t$  がワイブル分布に従う場合、

$$f(t|m, \eta) = \frac{m}{\eta} \left(\frac{t}{\eta}\right)^{m-1} \exp\left[-\left(\frac{t}{\eta}\right)^m\right] \quad m: \text{形状母数} \quad \eta: \text{尺度母数} \quad \text{のとき、}$$

$$\text{信頼度関数は、} R(t|m, \eta) = \exp\left[-\left(\frac{t}{\eta}\right)^m\right] \quad \text{故障率関数 (ハザード関数) は、} H(t|m, \eta) = \frac{m}{\eta} \left(\frac{t}{\eta}\right)^{m-1}$$

ここで、貝瀬らの方法により  $\lambda = \left(\frac{1}{\eta}\right)^m$  とおくと、 $f(t|m, \lambda) = \lambda m t^{m-1} \exp[-\lambda t^m]$  となり、 $R(t) = \exp[-\lambda t^m]$  となることから、 $H(t|m, \eta) = \lambda m t^{m-1} = \lambda(t)$  となる。

このとき、 $n$  個の観測時間データ  $x_1, \dots, x_n$  において時間  $t_i$  ( $i = 1, 2, \dots, n_1$ ) でイベントが発生または、打ち切りとなった(打ち切りの種類は問わない)場合、ワイブル分布の所与の形状母数値から導き出される母集団

の  $\lambda$  の事後分布は、 $p(\lambda|x) = S^m \frac{1}{\Gamma(n_1)} \lambda^{n_1-1} \exp[-S\lambda]$  となり、形状母数  $n_1$ 、尺度母数  $1/S$  を持つガンマ分布と

なる。ただし、 $S = \sum_{i=1}^n t_i^m$ 。(  $\lambda$  の推測から、平均寿命(ワイブル分布の期待値)  $\eta \Gamma(1+1/m)$  も推測される。)

例題：ワイブル分布型の寿命モデルを有する、集団の生存時間データ(年)を観測し、{3, 5, 9, 1, 11, 8}であり、4 例は死亡、2 例が打ち切りであった。母集団の形状母数は 1.2 であった。本ツールではこの集団で 90% の確信度で  $\lambda$  の値が 0.03~0.190 を取る、すなわち尺度母数の値が 4.0~18.6 を取ることが即座に表示される。

## 文献

貝瀬徹, 堀籠教夫: ワイブル分布を用いたベイズアプローチによる信頼性解析, REAJ誌, 22 (3) : 197 - 202 , 2000.

# L. 感度・特異度・有病率を用いた検査的中率の推測ツール

## —ベイズの公式を用いた—

本ツールについて

ベイズの公式を用いて検査の感度・特異度と、検査前確率である有病率を指定することにより、検査後確率である陽性検査的中率、陰性検査的中率を測定する。

画面右段の感度、特異度、有病率(検査前確率)、の各スライダー(スケール)の数値をマウスで変化させると、各々の数値の変化に合わせて陽性検査的中率(赤線)と陰性検査的中率(青線)のグラフ曲線が変化する。同時に陽性検査的中率と陰性検査的中率の値が計算・表示される。それぞれの用語については成書を参考にする事。

参考：ベイズの定理と二項分布について

一般に、陽性予測値(陽性検査的中率。検査が陽性に出てしかも病気である確率)は以下の式で表される。

$$\text{陽性予測値} = \frac{\text{有病率} \times \text{感度}}{(\text{有病率} \times \text{感度}) + \{(1 - \text{有病率}) \times (\text{偽陽性率})\}}$$

すなわち

$$P(D | +) = \frac{P(D) \times P(+ | D)}{P(D) \times P(+ | D) + \{(1 - P(D)) \times P(+ | \text{not } D)\}}$$

また、陰性予測値(陰性検査的中率。検査が陰性に出てしかも病気でない確率)は以下の式で表される。

$$\text{陰性予測値} = \frac{(1 - \text{有病率}) \times \text{特異度}}{(\text{有病率} \times \text{偽陰性率}) + \{(1 - \text{有病率}) \times (\text{特異度})\}}$$

すなわち

$$P(\text{not } D | -) = \frac{(1 - P(D)) \times P(- | \text{not } D)}{P(D) \times (1 - P(+ | D)) + \{(1 - P(D)) \times P(- | \text{not } D)\}}$$

※ここで、 $P(D)$  = 有病率、 $P(- | \text{not } D)$  = 特異度、 $P(+ | D)$  = 感度、 $P(+ | \text{not } D)$  = 偽陽性率、感度 + 偽陰性率 = 1、特異度 + 偽陽性率 = 1 である。

例題：ある疾患の検査の感度が 80%、特異度が 99%であった。有病率 5%で、ある患者の検査が陽性であったとき本ツールではその患者の陽性検査的中率(赤線)と陰性検査的中率(青線)の曲線が表示され、陽性検査的中率 80.8%、陰性検査的中率 98.9%と計算表示される。

文献

松原望：入門ベイズ統計、東京図書、第 1 版、135-153、東京、2008.

大橋靖雄：診断・治療の評価、Surgery Frontier、11 (2) : 196 - 203, 2004.

他多数