

Rstan を利用するための R 関数

作成 2017,03,20 : 早稲田大学 豊田秀樹

目次

1 Rstan を利用するための R 関数	1
1.1 G1mean: 1 群の正規分布に関する推測	1
1.2 G2Ind: 独立した 2 群の差の推測	2
1.3 G2pair: 対応ある 2 群の差の推測	2
1.4 E1Ind: 1 要因実験の推測	3
1.5 printIJ: 連言命題が正しい確率	3
1.6 E1betw_level: 特に興味のある 2 水準間の比較	4
1.7 E2Ind: 2 要因実験の推測	4
1.8 E2betw_level: 特に興味のある 2 セル間の比較	5
1.9 Bi01: 1 つの 2 項分布に関する推測	5
1.10 Bi02: 2 つの 2 項分布に関する推測	5
1.11 Bi03: g 個の 2 項分布に関する推測	6
1.12 Mu01: カテゴリ数が k の比率の推測	6
1.13 Mu02: 対応ある a × b のクロス表の分析	7
1.14 Reg: 回帰分析	8
2 Stan スクリプト	8

1 Rstan を利用するための R 関数

MCMC を利用して母数の事後分布を求める関数と、それを印刷するメソッド関数の、2 種類の関数を実行することで分析を行う。

生成量や研究仮説が正しい確率は、結果を見ながら再計算を繰り返しながら分析を進めることが一般的である。再計算の際には、印刷のためのメソッド関数の

引数を渡して計算する。時間のかかる MCMC の計算からやり直すのではないから、素早い再計算・分析が可能になる。

セル番号や水準番号を指定し、連言命題が正しい確率を計算するメソッド関数も用意してある。

1.1 G1mean: 1 群の正規分布に関する推測

```
G1mean(x,prior=F, mL=-1000, mH=1000, sL=0, sH=100,
       prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
       see=1234,cha=5,war=1000,ite=21000, fi=NA)
```

引数

x ベクトル形式のデータ

prior 論理値。T なら事前分布の範囲を指定。F なら指定せず (stan のデフォルト)。

mL, mH, sL, sH 事前分布のパラメタ/一様分布の平均下限・上限/標準偏差下限・上限

prob 事後分布で報告する確率点

see, cha, war, ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

mu 平均の事後分布

sigma 標準偏差の事後分布

xaste 予測分布

log_lik 対数尤度

印刷のメソッド

```
print.G1mean(x,degits=3,cr1=F,cr2a=F,cr2b=F,cr3=F,cr4=F,cr5=F,pr1=F,pr2=F)
```

x クラス 'G1mean' のオブジェクト

degits=3 小数点以下の数字

cr1 効果量を定義するための基準点 (測定値で指定)
cr2a 予測分布の特定区間の確率を求めるための分布関数上限 (測定値で指定)
cr2b 予測分布の特定区間の確率を求めるための分布関数下限 (測定値で指定)
cr3 測定値との比を計算する際の基準点 (測定値で指定)
cr4 母平均値が cr4 より小さい確率 (測定値で指定)
cr5 cr1 による効果量が cr5 より小さい確率 (効果量の単位で指定)
pr1 予測分布の%点の推測 (下からの確率で指定)
pr2 cr2 の領域が観測される確率が、pr2 より大きい確率 (確率で指定)

1.2 G2Ind: 独立した 2 群の差の推測

```
G2Ind(x1,x2,EQU=1,prior=F,mL=-1000, mH=1000, sL=0, sH=100,
      prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

x1 ベクトル形式の第 1 群のデータ (平均値の大きな群)
x2 ベクトル形式の第 2 群のデータ (平均値の小さな群)
EQU 論理値。1 なら等分散。0 なら異なる分散。
prior 論理値。T なら事前分布の範囲を指定。F なら指定せず (stan のデフォルト)。
mL, mH, sL, sH 事前分布のパラメタ/一様分布の平均下限・上限/標準偏差下限・上限
prob 事後分布で報告する確率点
see, cha, war,ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/
fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力
par 母数リスト
prob 事後分布で報告する確率点ベクトル,
mu1,mu2 第 1 群平均, 第 2 群平均,
sigma1, sigma2 第 1 群標準偏差, 第 2 群標準偏差
xaste 予測分布
log.lik 対数尤度

印刷のメソッド

```
print.G2Ind(x,degits=3,cr1=F,cr2=F,cr3=F,pr1=F,pr2=F,pr3=F)
```

x クラス 'G2Ind' のオブジェクト
degits=3 小数点以下の数字
cr1 平均値の差の基準点
cr2 閾上率の基準点
cr3 効果量の基準点
pr1 非重複度の基準確率
pr2 優越率の基準確率
pr3 閾上率の基準確率

1.3 G2pair: 対応ある 2 群の差の推測

```
G2pair(x,EQU=1,prior=F,mL=-1000, mH=1000, sL=0, sH=100,
      prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

x n 行 2 列の行列形式のデータ (平均値の大きな群は 1 列目に指定)
EQU 論理値。1 なら等分散。0 なら異なる分散。
prior 論理値。T なら事前分布の範囲を指定。F なら指定せず。
mL, mH, sL, sH 事前分布のパラメタ/一様分布の平均下限・上限/標準偏差下限・上限
prob 事後分布で報告する確率点
see, cha, war,ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/
fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力
par 母数リスト
prob 事後分布で報告する確率点ベクトル,
mu1,mu2 第 1 群平均, 第 2 群平均,

1. Rstan を利用するための R 関数

sigma1, sigma2 第 1 群標準偏差, 第 2 群標準偏差

rho 相関係数

xaste1, xaste2 予測分布 1, 予測分布 2

log_lik 対数尤度

印刷のメソッド

```
print.G2pair(x, onlydiff=T, degits=3, cr1=F, cr2=F, cr3=F, cr4=F, ra= 1.0,
  rb= -1.0, pr1=F, pr2=F, pr3=F, pr4=F)
```

x クラス'G2pair' のオブジェクト

onlydiff 論理値。T なら差得点に関する出力のみ。F なら群差の出力もする

degits=3 小数点以下の数字

cr1 平均値の差の基準点

cr2 閾上率の基準点

cr3 効果量の基準点

cr4 差得点の標準偏差の基準点

ra 相関上限

rb 相関下限

pr1 非重複度の基準確率

pr2 優越率の基準確率

pr3 閾上率の基準確率

pr4 同順率の基準確率

1.4 E1Ind: 1 要因実験の推測

```
E1Ind(y,A,prior=F,mL=-1000, mH=1000, sL=0, sH=100,
  prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
  see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

y ベクトル形式の特性値

A ベクトル形式の水準 (1 から a まで、抜けなしで、整数で指定)

prior 論理値。T なら事前分布の範囲を指定。F なら指定せず。

mL, mH, sL, sH 事前分布のパラメタ/一様分布の平均下限・上限/標準偏差下限・上限

prob 事後分布で報告する確率点

see, cha, war, ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

muA 水準の平均

sigmaE 水準内 sd

sigmaA 水準間 sd

eta2 説明率

delta 効果量

mu 全平均

aj 水準の効果

Ubig 水準の効果 0 以上

Usma 水準の効果 0 以下

U2 i 行が j 列より大きい

log_lik 対数尤度

印刷のメソッド

```
print.E1Ind(x,degits=3)
```

x クラス'E1Ind' のオブジェクト

degits=3 小数点以下の数字

出力

事後分布 水準の平均・誤差標準偏差・要因の標準偏差・説明率・効果量・全平均・水準の効果

確率 水準の効果が 0 以上 (以下) である確率・水準 i が水準 j より大きい確率

WAIC

1.5 printIJ: 連言命題が正しい確率

```
printIJ(x,degits=3,IJ)
```

1. Rstan を利用するための R 関数

引数

x :E1Ind の出力の U2[,i,j]、あるいは E2Ind の出力の U2A,U2B、Mu01 の出力の U2、Bi03 の出力の U2 (E1Ind と E2Ind と Mu01 と Bi03 で共通に利用する関数である)

degits=3 小数点以下の数字

IJ 2 列の行列 (1 列目は行、2 列目は列)

1.6 E1betw_level: 特に興味のある 2 水準間の比較

```
E1betw_level(x,degits=3,I,J,cr1=F)
```

引数

x クラス'E1Ind' のオブジェクト

degits=3 小数点以下の数字

I 整数、平均値の大きい水準

J 整数、平均値の小さい水準

cr1 閾上率の基準値 (測定値で指定)

1.7 E2Ind: 2 要因実験の推測

```
E2Ind(y,A,B,prior=F,mL=-1000, mH=1000, sL=0, sH=100,
      prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

y ベクトル形式の特性値

A ベクトル形式の水準 (1 から a まで、抜けなしで、整数で指定)

B ベクトル形式の水準 (1 から b まで、抜けなしで、整数で指定)

prior 論理値。T なら事前分布の範囲を指定。F なら指定せず。

mL, mH, sL, sH 事前分布のパラメタ/一様分布の平均下限・上限/標準偏差下限・上限

prob 事後分布で報告する確率点

see, cha, war,ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

mu 全平均の事後分布、ここから下は全て事後分布であり、第 1 次元は乱数数 t 。

muA 要因 A の水準の効果の事後分布, サイズは [,a]

muB 要因 B の水準の効果の事後分布, サイズは [,b]

muAB 交互作用の効果の事後分布, サイズは [,a,b]

cellmean セルごとの平均の事後分布, サイズは [,a,b]

sigmaA 要因 A の標準偏差の事後分布

sigmaB 要因 B の標準偏差の事後分布

sigmaAB 交互作用 AB の標準偏差の事後分布

sigmaE 誤差 E の標準偏差の事後分布

eta2A 要因 A の説明率の事後分布

eta2B 要因 B の説明率の事後分布

eta2AB 交互作用 AB の説明率の事後分布

eta2T モデル全体での説明率の事後分布

deltaA 要因 A の効果量の事後分布, サイズは [,a, サイズは [,a,b]

deltaB 要因 B の効果量の事後分布, サイズは [,b]

deltaAB 交互作用 AB の効果量の事後分布, サイズは [,a,b]

UbigA 要因 A の水準の効果が 0 以上のとき 1 が立つ, サイズは [,a]

UsmaA 要因 A の水準の効果が 0 以下のとき 1 が立つ, サイズは [,a]

UbigB 要因 B の水準の効果が 0 以上のとき 1 が立つ, サイズは [,b]

UsmaB 要因 B の水準の効果が 0 以下のとき 1 が立つ, サイズは [,b]

UbigAB 交互作用 AB の効果が 0 以上のとき 1 が立つ, サイズは [,a,b]

UsmaAB 交互作用 AB の効果が 0 以下のとき 1 が立つ, サイズは [,a,b]

U2A 要因 A の行の水準より列の水準が大きい場合に 1 が立つ, サイズは [,a,a]

U2B 要因 B の行の水準より列の水準が大きい場合に 1 が立つ, サイズは [,b,b]

log_lik 対数尤度

印刷のメソッド

```
print.E2Ind(x,degits=3)
```

x クラス'E2Ind' のオブジェクト

degits=3 小数点以下の数字

1. Rstan を利用するための R 関数

出力

事後分布 全平均・水準の効果・交互作用効果・セル平均・要因の主効果 (2 つ)・交互作用効果・誤差標準偏差・説明率 (4 つ)・効果量 (3 つ)

確率 (1) 水準の効果と交互作用効果が 0 以上 (以下) である確率、(2) 要因 A と B に関して行の水準より列の水準が大きい確率 (2 × 2 の要因計画の場合は (1)(2) は一致する)

WAIC

1.8 E2betw_level: 特に興味のある 2 セル間の比較

```
E2betw_level(x,degits=3,H="A",F=1,I=1,J=2,cr1=F)
```

引数

x クラス'E2Ind' のオブジェクト

degits=3 小数点以下の数字

H 固定する因子 "A" or "B"

F 整数、固定する因子の水準

I 整数、比較する因子の平均値の大きい水準

J 整数、比較する因子の平均値の小さい

cr1 閾上率の基準値

1.9 Bi01: 1 つの 2 項分布に関する推測

```
Bi01(x,n,prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

x 正反応数

n ベルヌイ試行数

prob 事後分布で報告する確率点

see, cha, war,ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル、

theta 母比率

xaste 予測分布

Odds オッズ

log_lik 対数尤度

印刷のメソッド

```
print.Bi01(x,degits=3,pr1=F,cr1=F,cr2=F)
```

x クラス'Bi01' のオブジェクト

degits=3 小数点以下の数字

pr1 区間 (0,1) の実数、比率の最小値、その値より大きい確率を求める

cr1 x*の最小値、測定値がその値より大きくなる確率を求める

cr2 オッズの最小値、その値より大きい確率を求める

出力

事後分布 比率・オッズ

予測分布

確率 pr1,cr1,cr2 で指定した確率

WAIC

1.10 Bi02: 2 つの 2 項分布に関する推測

```
Bi02(x,n,prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

x 正反応数 (長さ 2 のベクトル、標本比率の大きいものを先に指定する)

n ベルヌイ試行数 (長さ 2 のベクトル)

prob 事後分布で報告する確率点

1. Rstan を利用するための R 関数

see, cha, war,ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた **stanfit** のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

p 母比率 (2)

xaste 予測分布の $x^{*(2)}$

p_sa 比率の差

p_hi 比率の比

Odds オッズ (2)

Odds_hi オッズ比

log_lik 対数尤度

印刷のメソッド

```
print.Bi02(x,degits=3,cr1=F,cr2=F,cr3=F)
```

x クラス 'Bi02' のオブジェクト

degits=3 小数点以下の数字

cr1 比率の差の最小値、その値より大きい確率を求める

cr2 比率の比の最小値、その値より大きい確率を求める

cr3 オッズ比の最小値、その値より大きい確率を求める

出力

事後分布 2つの比率・比率の差・比率の比・2つのオッズ・オッズ比

確率 引数で指定した確率

WAIC

1.11 Bi03: g 個の 2 項分布に関する推測

```
Bi03(x,n,prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

x 正反応数 (長さ g のベクトル)

n ベルヌイ試行数 (長さ g のベクトル)

prob 事後分布で報告する確率点

see, cha, war,ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた **stanfit** のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

p 母比率 (2)

xaste 予測分布の $x^{*(2)}$

U2 行の母比率が列の母比率より大きいとき 1 を、それ以外は 0 ($g \times g$)

log_lik 対数尤度

印刷のメソッド

```
print.Bi03(x,degits=3)
```

x クラス 'Bi03' のオブジェクト

degits=3 小数点以下の数字

出力

事後分布 g 個の比率

予測分布

確率表 行の比率が列の比率より大きい確率

WAIC

1.12 Mu01: カテゴリ数が k の比率の推測

```
Mu01(x,prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

1. Rstan を利用するための R 関数

x 正反応数 (長さ k のベクトル)

prob 事後分布で報告する確率点

see, cha, war, ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M /

fi 以前の計算から導かれた `stanfit` のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

pi 母比率 (k)

U2 行の母比率が列の母比率より大きいとき 1 を、それ以外は 0 ($k \times k$)

log.lik 対数尤度

印刷のメソッド

```
print.Mu01(x,degits=3)
```

x クラス 'Mu01' のオブジェクト

degits=3 小数点以下の数字

出力

事後分布 各カテゴリの比率

確率表 行のカテゴリ比率が列のカテゴリ比率より大きい確率

WAIC

1.13 Mu02: 対応ある $a \times b$ のクロス表の分析

```
Mu02(x,prob=c(0.025, 0.05, 0.5, 0.95, 0.975),
      see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

x クロス表 ($a \times b$) 行列形式で指定)

prob 事後分布で報告する確率点

see, cha, war, ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M /

fi 以前の計算から導かれた `stanfit` のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

prob 事後分布で報告する確率点ベクトル,

pim i 行 j 列の母比率 ($a \times b$)

xaste 予測分布の x^* ($a \times b$)

res ピアソン残差 ($a \times b$)

V クラメル の 連関係数

pa 第 i 行の周辺確率 (a)

pb 第 j 列の周辺確率 (b)

Up ピアソン残差が 0 より大きい確率 ($a \times b$)

Um ピアソン残差が 0 以下の確率 ($a \times b$)

log.lik 対数尤度

印刷のメソッド

```
print.Mu02(x,degits=3)
```

x クラス 'Mu02' のオブジェクト

degits=3 小数点以下の数字

出力

事後分布 同時確率・残差・クラメル の 連関係数・行周辺確率・列周辺確率

予測分布

確率表 1 ピアソン残差が 0 より大きい確率 ($a \times b$)

確率表 2 ピアソン残差が 0 以下の確率 ($a \times b$)

WAIC

注意 `Up<-out$Up; Um<-out$Um; round(mean(Up[,1,1]*Up[,3,3]*Um[,3,1]),3)` のように、ピアソン残差に関する連言命題が正しい確率は、関数がないので、自分で書く

1.14 Reg: 回帰分析

```
Reg(y,X,see=1234, cha=5, war=1000, ite=21000, fi=NA)
```

引数

y 基準変数 (n)

X 予測変数行列 (n*p)

see, cha, war, ite 乱数の種/チェーン数/ウォームアップ期間/生成数 M/

fi 以前の計算から導かれた stanfit のインスタンス。同じ関数のインスタンスを指定することにより、コンパイル済みのモデルが再利用でき、コンパイルに費やされる時間を省略できる。デフォルトは NA。セッションの初めは NA を指定する。NA の場合はコンパイルされる。

戻り値

fit stan の出力

par 母数リスト

b 回帰係数ベクトル (p)

sigma 誤差標準偏差

r2 決定係数

vyhat yhat の分散

log_lik 対数尤度

印刷のメソッド

```
print.Reg(x,degits=3, prob=c(0.025, 0.05, 0.5, 0.95, 0.975),Xnew=F)
```

x クラス 'Reg' のオブジェクト

degits 小数点以下の数字

prob 事後分布で報告する確率点

Xnew (n*P) の行列、単回帰の場合はサイズ n のベクトル 予測分布用データ

出力

fit stan の出力

waic,y,X waic,y,X

yhat 予測値

resi 残差 (X による)

Ga Gc 標準編回帰係数

yhata yhatc 回帰式の事後分布 (Xnew による)

yasta yastc 予測分布 (Xnew による)

2 Stan スクリプト

rstan 関数によって呼ばれる stan スクリプトは以下である。ユーザーが直接触れることはないので、解説は省略する。引数と戻り値の意味は、スクリプト内のコメント文に記述した。

G1meanPF.stan

G1meanPT.stan

G2IndPF.stan

G2IndPT.stan

G2pairPF.stan

G2pairPT.stan

E1IndPF.stan

E1IndPT.stan

E2IndPF.stan

E2IndPT.stan

Bi01.stan

Bi02.stan

Bi03.stan

Mu01.stan

Mu02.stan

Reg.stan