

古典的なノンパラメトリック検定はなにを仮定しているのか

2020 年 2 月 18 日

目次

1	読むのがめんどうな人へ	1
2	前置き	2
3	ノンパラメトリック・モデル	3
4	仮説検定一般に関する注意点	3
5	符号検定 (Sign Test)	3
5.1	R による実装例	5
6	Wilcoxon の符号付き順位検定 (Wilcoxon Signed-Rank Test)	5
6.1	シミュレーション	6
7	2 群の比較	18
7.1	シミュレーション	19

1 読むのがめんどうな人へ

たぶんシミュレーションだけ見ればいい.

2 前置き

もともと工学部出身のぼくが、医学系の研究室に行っておもしろかったことはいろいろある。

その中の一つが、医学系の人のアーチファクト（artifact）という言葉の使い方だ。

アーチファクトというのは、辞書的には人工物とか工芸品を指す。

ぼくからしたら「この結果はアーチファクトだよ」とか言われたら褒められているのかな？　みたいな感じがする。

でも違う。医学系の言葉使いではアーチファクトというのは、自然現象とか生命現象の本質じゃない人工的な混ざりもの、みたいなニュアンスだ。

ところで、ぼくは統計モデリングというのが好きだ。モデリングというのはデータを取った人からいろいろ話を聞いて、関数とか確率分布とか微分方程式とか差分方程式とかを組み合わせて、データを取った人のイメージとかを目的とかを統計の言葉に翻訳するような作業だ。モデルを作るというのは、ぼくの場合は仮定を置くことに等しい。

それはもうアーチファクトのかたまりみたいな作業だ。

そのせいかどうかは知らないが、医学系の人にはモデリングよりも検定が好きな印象がある。統計屋というと、いろんな検定を知っていて、場合に応じて正しい検定を選ぶ人、みたいな認識をされることもある。

なかでもノンパラメトリック検定が好きなようだ。

ノンパラメトリック検定とは、母集団に対して特定の確率分布を仮定しないで検定をする手法の総称とされる。

仮定が少ないほうが客観的な分析でえらいという感覚があるんだと思う。

そこがぼくの好みと相容れない。

でも好みじゃないのでノンパラメトリック検定に関する相談には乗りませんというのも大人げないので、これからちょっとノンパラメトリック検定について勉強していきたいと思っている。

この文書はそういったものだ。

この文書がこれからどういったものになるかは、ぼくもまだわかってない。

でも結びの言葉はなんとなく決めてある。

「仮定を置かないんじゃなくて仮定を明確にして、アーチファクトをもってアーチファクトを制すのだ。」

3 ノンパラメトリック・モデル

ノンパラメトリックモデルという言葉とノンパラメトリック検定という言葉はまったく別物だと思ったほうがよさそうだ.

ノンパラメトリックモデルというのは、パラメータが多すぎるモデルのことを指す.

データのサイズにほぼ比例して、モデルのパラメータが増えるようなモデルはノンパラメトリックモデルと呼ばれる.

以下では、ノンパラメトリックモデルの話はしない.

ノンパラメトリック検定の話をする.

4 仮説検定一般に関する注意点

仮説検定は「確率版背理法」と例えられることもあるが、現実のデータを扱う以上、背理法のようにすっきりとはしていない. 仮説検定では、「帰無仮説 $H_0 : \theta = 0$, 対立仮説 $H_1 : \theta \neq 0$ 」みたいな形で「仮説」を提示する. しかし、帰無仮説 H_0 の確率を評価するのに、帰無分布というのを作らなくちゃいけない.

仮説検定には、「 $H_0 : \theta = 0$ 」といった意識されやすい仮定と検定のやり方（何検定を使うか）を決めた時点で置かれている意識されにくい仮定が存在する. 例えば「〇〇検定を行った」と言った時点で、「サンプルは正規分布に従う」とか「サンプルは独立同分布とする」とかの意識されにくい仮定はクリアされたことになっている.

仮説検定を行った結果、帰無仮説が棄却されたということ意識されやすい仮定が棄却された感じがするが、実は意識されにくい仮定のほうがもっとずれていたのかもしれない.

「 H_0 でないから H_1 だ」という背理法ほど、仮説検定はすっきりしていない.

5 符号検定 (Sign Test)

検定というのは帰無分布を一つ決めて、そこからのずれを測っている.

帰無分布を決めなきゃいけないのに、「特定の確率分布を仮定しない」なんてことができるのか.

どうやらできるようだ.

データ X_1, X_2, \dots, X_n を生成した分布が連続な分布関数 F を持つということだけを仮定しよう. これはどちらかというと意識されにくい仮定だ. θ を F の中央値とする.

その上で母集団の中央値が 0 であるという帰無仮説を検定したい。まず、帰無仮説 $H_0 : \theta = 0$, 対立仮説 $H_1 : \theta > 0$ の片側検定を考える。

そして,

$$S_n = \sum_{i=1}^n I_{\{X_i > 0\}} \quad (1)$$

という統計量を作ることにする。ここで I_A は指示関数（つまり $X_i > 0$ なら 1, さもなくば 0 を返す）である。

S_n は符号が + になった回数を単に数えただけだ。

帰無仮説のもとで, S_n は二項分布 $\text{Binomial}(n, 1/2)$ に従う。

中央値（メジアン）の定義を確認しておく, 分布 F に従う確率変数を X と書くとき,

$$\Pr(X \geq m) \geq \frac{1}{2}, \quad \Pr(X \leq m) \geq \frac{1}{2} \quad (2)$$

を満たす実数 m が中央値である。

分布関数が連続であれば,

$$\Pr(X \leq m) = \frac{1}{2} \quad (3)$$

である。いま, X が連続な分布関数を持つことを仮定したので, 帰無分布が二項分布 $\text{Binomial}(n, 1/2)$ になるのは, ほぼ定義そのものである。

帰無仮説のもとで A という事象の起こる確率を $\Pr_{H_0}(A)$ と書くことにすると, この検定の有意水準 α での棄却域は, $\Pr_{H_0}(S_n > k) \leq \alpha$ を満たす k に対して $S_n > k$ となる範囲である。

対立仮説 $H_1 : \theta < 0$ の片側検定をやりたい場合, 棄却域は, $\Pr_{H_0}(S_n < k) \leq \alpha$ を満たす k に対して $S_n < k$ となる範囲である。

対立仮説 $H_1 : \theta \neq 0$ の両側検定をやりたい場合, 棄却域は, $\Pr_{H_0}(\min(S_n, n - S_n) < k) \leq \alpha/2$ を満たす k に対して $S_n < k$ となる範囲である。

中央値が 0 である場合だけでなく, 一般の θ_0 の検定を考えたいときは, X_i それぞれから θ_0 を引いてやれば, 同じ問題に帰着される。

2 群の比較ならわかるけど, 1 群の検定なんて使う場面あるのかと思うかもしれないので, こんな例を与えておく: 「患者さんたちにある睡眠薬を飲ませたところ, 睡眠時間がそれぞれ増えたり減ったりしました。睡眠時間の増減の中央値が 0 より大きければ, 睡眠薬は効果ありと言えます。」

■0 がある場合. 連続型の分布を仮定したので, 理想的にはちょうど 0 が出る確率は 0 である. 現実的には 0 がある場合はそれをのぞいて検定統計量を出す.

5.1 R による実装例

統計ソフト R で上記の符号検定の p 値を計算するには, 次のように書けばよい. まず, 対立仮説 $H_1: \theta > 0$ の片側検定の場合,

```
theta0 <- 0
n <- length(X)
Sn <- sum(X>theta0)
pbinom(Sn-1,n,0.5,lower.tail = FALSE)
```

対立仮説 $H_1: \theta > 0$ の片側検定の場合,

```
pbinom(Sn-1,n,0.5,lower.tail = FALSE)
```

対立仮説 $H_1: \neq 0$ の両側検定の場合,

```
2*pbinom(min(Sn,n-Sn),n,0.5)
```

こう書いても同じ.

```
2*min(pbinom(Sn-1,n,0.5,lower.tail = FALSE),pbinom(Sn,n,0.5))
```

6 Wilcoxon の符号付き順位検定 (Wilcoxon Signed-Rank Test)

符号検定は符号の情報しか使っていないので, 検出力はあまり高くないことが予想される (本当にそうなのかは, 6.1 節で検討する予定).

そこで, データの順位の情報も使うことを考える.

R_i を X_i の順位とする. フォーマルに書くと,

$$R_i = \#\{j; X_j \leq X_i\} \quad (4)$$

である. ここで, $\#A$ は集合 A の要素の個数を返す関数である.

次の統計量を考える.

$$T = \sum_{i=1}^n R_i I_{\{X_i > 0\}} \quad (5)$$

ここで分布 F の密度関数が 0 を中心に左右対称であるという仮定をする. 「0 を中心に」というのは意識されやすい仮定, 「左右対称である」というのは意識されにくい仮定だ. この仮定のもとで, $I_{\{X_i > 0\}}$ はベルヌーイ分布 Bernoulli(1/2) に従う.

$\{R_i, I_{\{X_i > 0\}}\}$ は n 個の順位それぞれと, 0 か 1 なので, 取りうる組み合わせは 2^n 通りある. その中で $T = k$ となる場合の数の合計を $c_{n,k}$ とすると, $\Pr\{T = k\} = c_{n,k}/2^n$ と書ける. これが帰無分布である.

データから求めた T の実現値を T_0 とするとき,

- 帰無仮説 $H_0 : \theta = 0$, 対立仮説 $H_1 : \theta > 0$ の片側検定の p 値は $\Pr\{T \geq T_0\}$,
- 帰無仮説 $H_0 : \theta = 0$, 対立仮説 $H_1 : \theta < 0$ の片側検定の p 値は $\Pr\{T \leq T_0\}$,
- 帰無仮説 $H_0 : \theta = 0$, 対立仮説 $H_1 : \theta \neq 0$ の両側検定の p 値は $T > n(n+1)/4$ のとき, $\Pr\{T \geq T_0\}$, さもなくば $\Pr\{T \leq T_0\}$

となる. ここで $n(n+1)/4$ というのは, T の取りうる値の範囲 0 から $\sum_{i=1}^n i = n(n+1)/2$ のまんなかである.

中央値が 0 である場合だけでなく, 一般の θ_0 の検定を考えたいときは, X_i それぞれから θ_0 を引いてやれば, 同じ問題に帰着される.

6.1 シミュレーション

よい検定の第 1 条件はまずなによりも実際の α エラーが, 設定した有意水準以下であることだ. これが成り立っていないとオオカミ少年になってしまい, どんなに検出力が高くても評価されない.

なのでここではまず α エラーのシミュレーションを行う.

正規分布でやるとノンパラメトリック検定のありがたみがわからないので, まずはロジスティック分布 Logistic(0, 1) からサンプリングしたデータで, 帰無仮説 $H_0 : \theta = 0$, 対立仮説 $H_1 : \theta \neq 0$ の両側検定を行い p 値の分布を調べる.

ロジスティック分布は正規分布と似た形をしており, 左右対称だが正規分布より裾が重い.

サンプルサイズ $n = 5$ のときのシミュレーション結果を, 図 1-3 に示す.

これらの図では横軸にシミュレーションした α エラー, 縦軸に名目上の有意水準を取っ

てプロットした。折れ線（実線）が対角線より下に来ていれば、 α エラーが名目上の有意水がより下に保たれていることを示す。

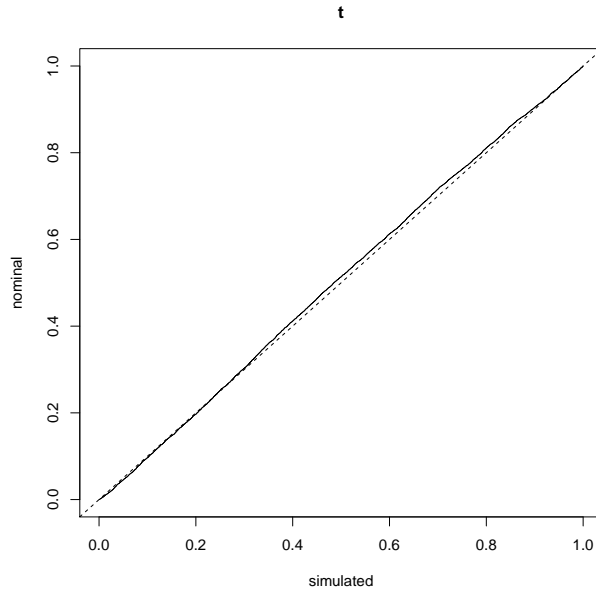


図1 α エラーのプロット. ロジスティック分布 $\text{Logistic}(0, 1)$. $n = 5$. t 検定

t 検定はサンプルサイズが小さくても、割と名目通りの p 値を与えている。

ノンパラメトリック検定は離散的な p 値を与える。特に符号検定でその傾向が顕著である。

正規分布に似た形だと t 検定も使えそうなので、一様分布 $\text{Uniform}(-1, 1)$ で同様の検定をやってみよう（図 4-6）。

t 検定は有意水準が小さいときに少しだけ、 α エラーが有意水準を上回っている。

サンプルサイズを 10 まで増やすと（図 7-9）， t 検定でもかなり正確な p 値を与える。

次は、左右対称の仮定を崩してみる。

ワイブル分布 $\text{Weibull}(m, \eta)$ は $m > 1$ のとき、モードの周りにやや集中した形をとり、 $m < 1$ のとき、指数分布よりも裾の重い分布となる。ワイブル分布は左右対称でないので、 t 検定では平均 μ に対しての、帰無仮説 $H_0 : \mu = \mu_0$ 、対立仮説 $H_1 : \mu \neq \mu_0$ 、符号検定と Wilcoxon の符号付き順位検定では中央値 θ に関しての、帰無仮説 $H_0 : \theta = \theta_0$ 、対立仮説 $H_1 : \theta \neq \theta_0$ の検定を行った。

まず、 $\text{Weibull}(2, 1)$ からサンプリングして $n = 5$ のときを図 10-12 に示す。

Wilcoxon の符号付き順位検定、もっと崩れるかと思ったけど意外といい。

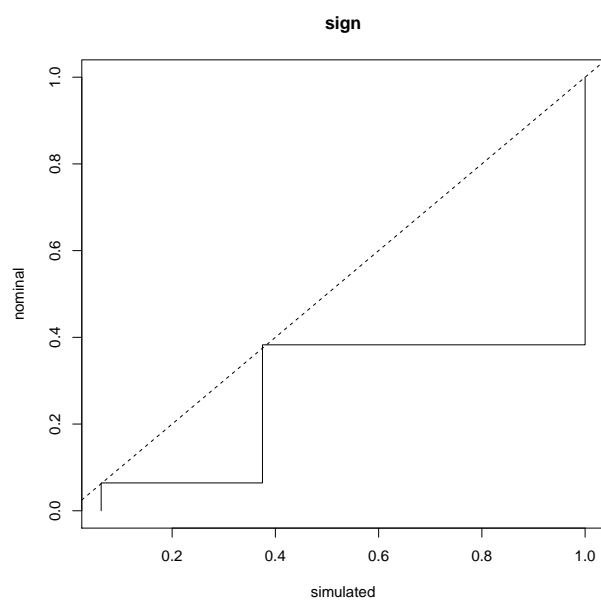


図 2 α エラーのプロット. ロジスティック分布 $\text{Logistic}(0, 1)$. $n = 5$. 符号検定.

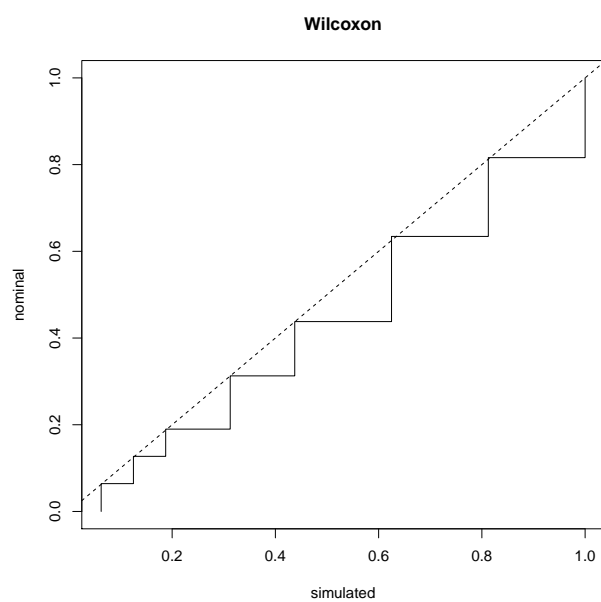


図 3 α エラーのプロット. ロジスティック分布 $\text{Logistic}(0, 1)$. $n = 5$. Wilcoxon の符号付き順位検定.

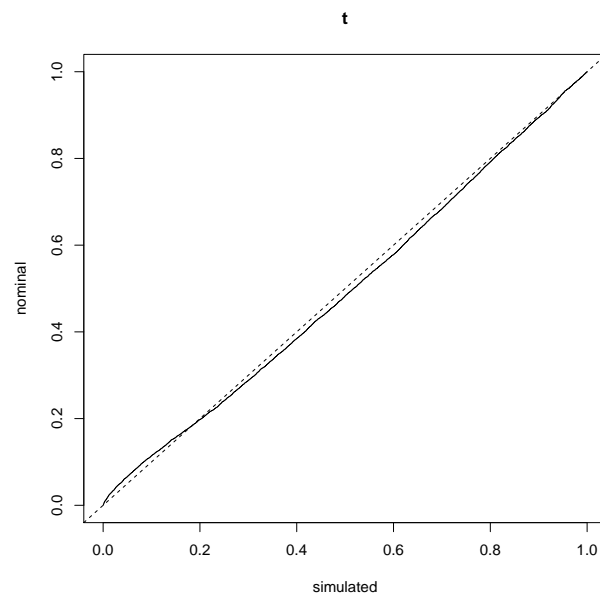


図4 α エラーのプロット. 一様分布 $\text{Uniform}(0, 1)$. $n = 5$. t 検定

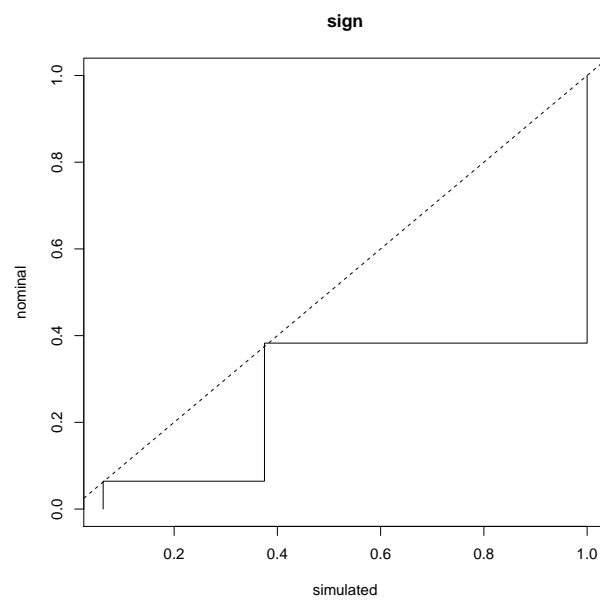


図5 α エラーのプロット. 一様分布 $\text{Uniform}(0, 1)$. $n = 5$. 符号検定.

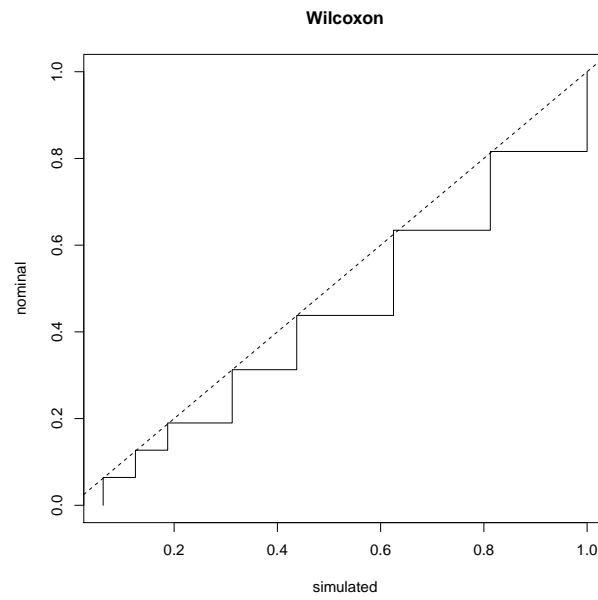


図 6 α エラーのプロット. 一様分布 $\text{Uniform}(0, 1)$. $n = 5$. Wilcoxon の符号付き順位検定.

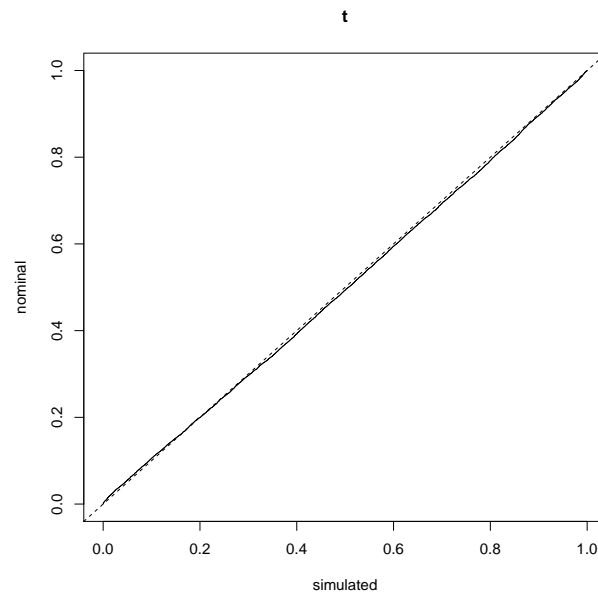


図 7 α エラーのプロット. 一様分布 $\text{Uniform}(0, 1)$. $n = 10$. t 検定

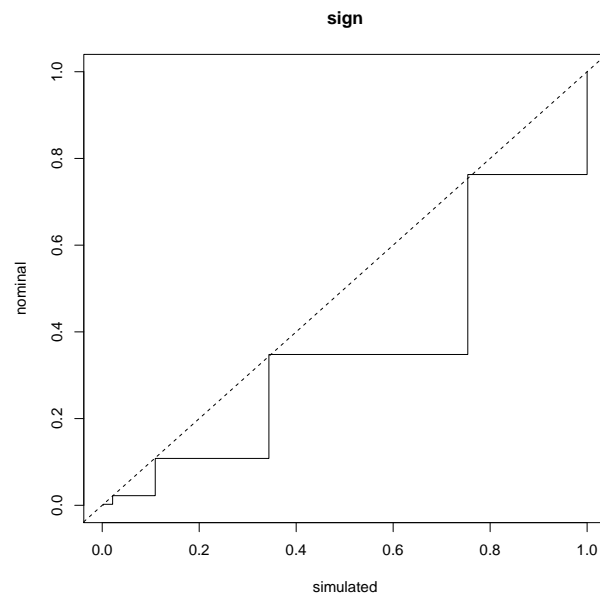


図 8 α エラーのプロット. 一様分布 $\text{Uniform}(0, 1)$. $n = 10$. 符号検定.

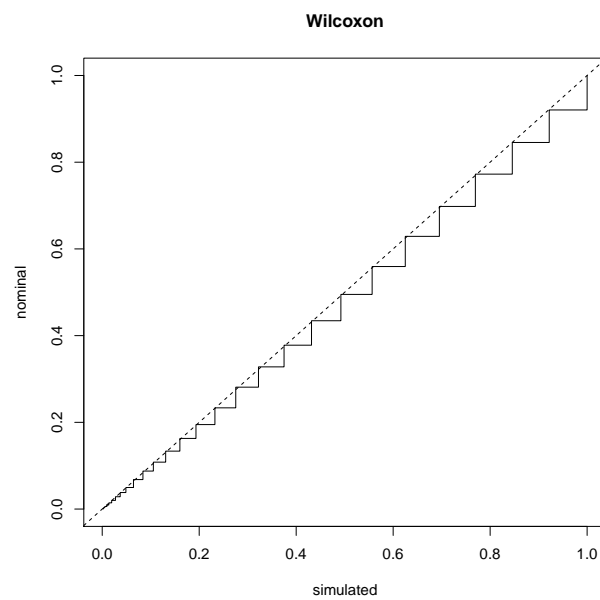


図 9 α エラーのプロット. 一様分布 $\text{Uniform}(0, 1)$. $n = 10$. Wilcoxon の符号付き順位検定.

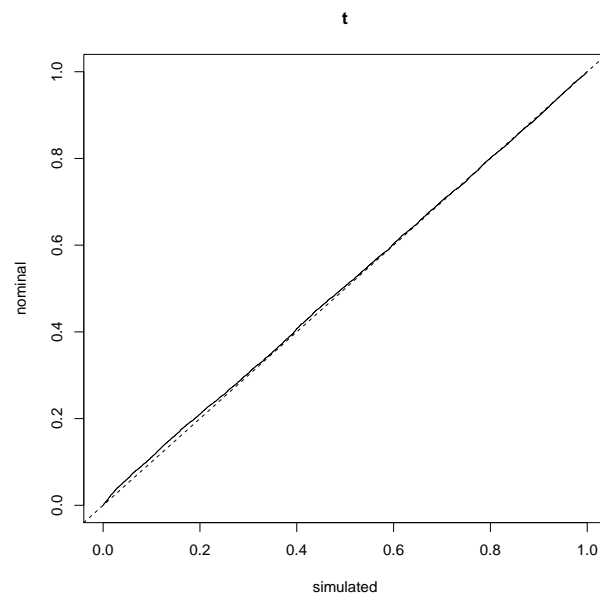


図 10 α エラーのプロット. ワイブル分布 Weibull(2, 1). $n = 5$. t 検定

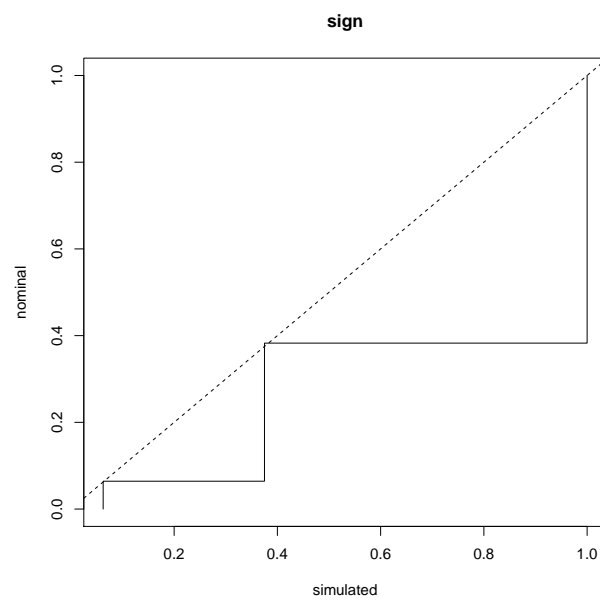


図 11 α エラーのプロット. ワイブル分布 Weibull(2, 1). $n = 5$. 符号検定.

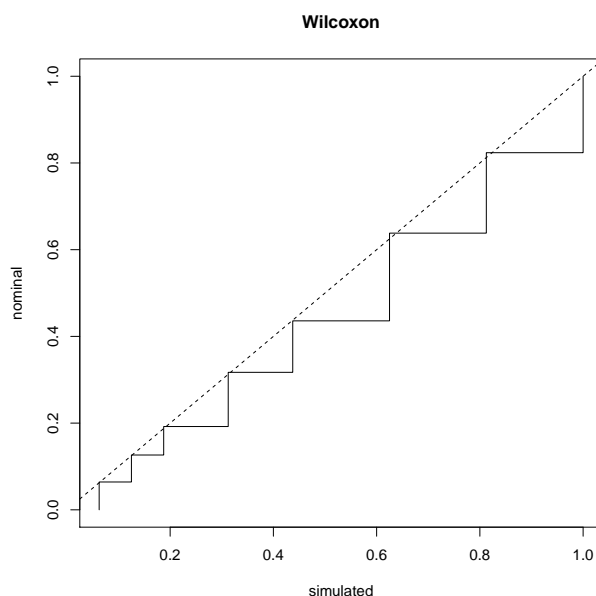


図 12 α エラーのプロット. ワイブル分布 Weibull(2, 1). $n = 5$. Wilcoxon の符号付き順位検定.

次に Weibull(0.5, 1) からサンプリングして $n = 5$ のときを図 13-15 に示す.

t 検定がガタガタになった. Wilcoxon の符号付き順位検定も崩れはじめた.

サンプルサイズを増やして $n = 10$ のときを図??-??に示す.

Wilcoxon の符号付き順位検定はサンプルサイズを増やすとむしろはっきり棄却するようになる.

もっとサンプルサイズを増やして $n = 100$ のときを図??-??に示す.

やはり Wilcoxon の符号付き順位検定はサンプルサイズを増やすとむしろはっきり棄却するようになる.

これは Wilcoxon の符号付き順位検定が分布が左右対称であるという仮定を棄却していると解釈できる.

■ここまでのシミュレーションのまとめ. 平均値や中央値に対して検定をやりたいとして, 分布の裾が明らかに重いときは, いまのところは符号検定をやるしかない.

Wilcoxon の符号付き順位検定は, 分布が左右対称でないときは適さない.

分布が左右対称に近いときは t 検定で概ね間に合う.

多くの人がノンパラメトリックの検定をやりたいがるのは分布が左右対称でないときだと思う.

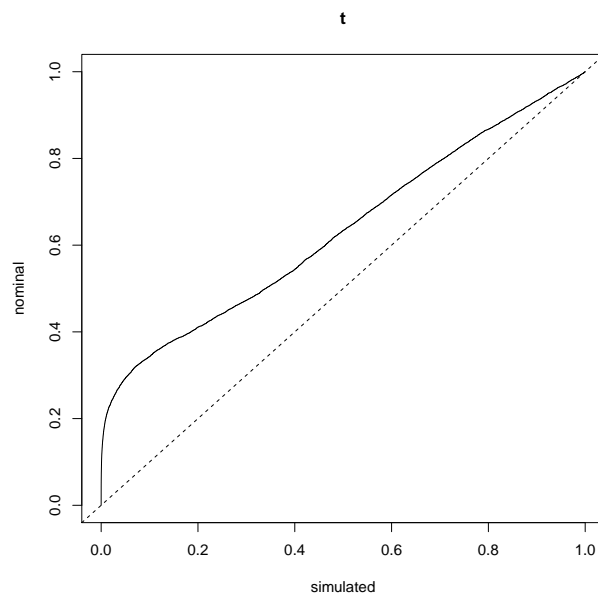


図 13 α エラーのプロット. ワイブル分布 $\text{Weibull}(0.5, 1)$. $n = 5$. t 検定

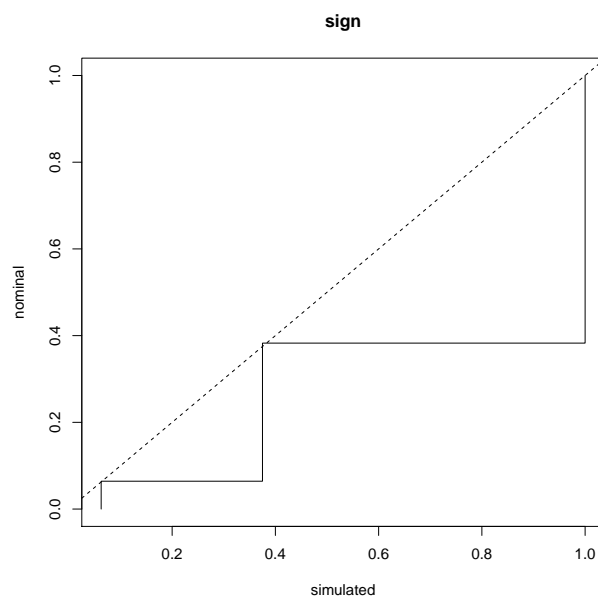


図 14 α エラーのプロット. ワイブル分布 $\text{Weibull}(0.5, 1)$. $n = 5$. 符号検定.

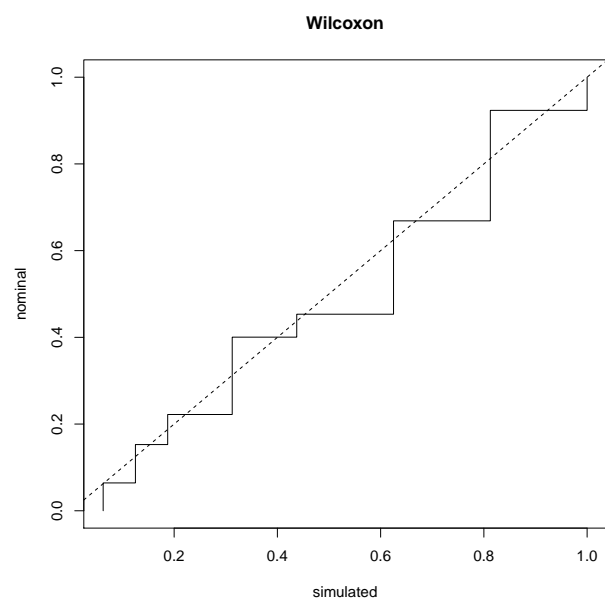


図 15 α エラーのプロット. ワイブル分布 $\text{Weibull}(0.5, 1)$. $n = 5$. Wilcoxon の符号付き順位検定.

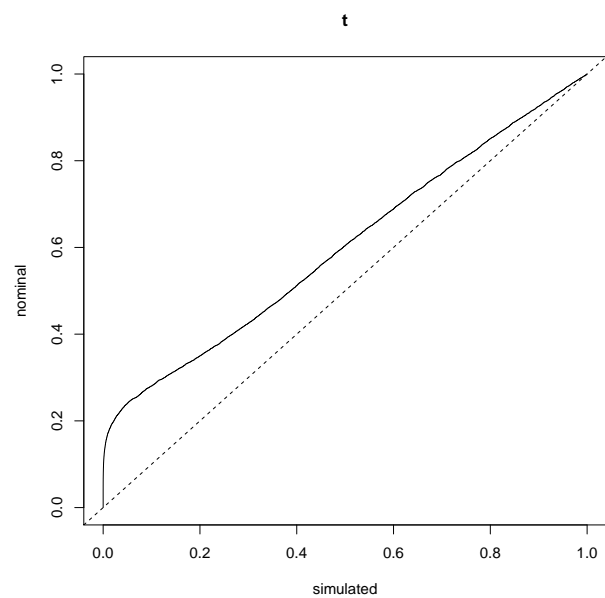


図 16 α エラーのプロット. ワイブル分布 $\text{Weibull}(0.5, 1)$. $n = 10$. t 検定

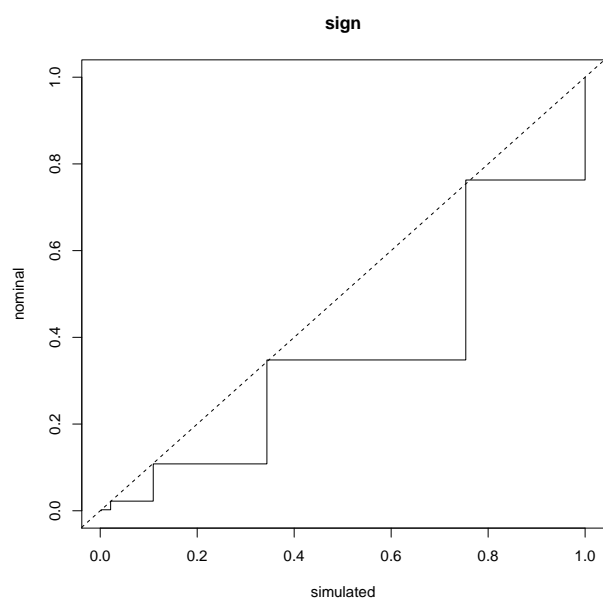


図 17 α エラーのプロット. ワイブル分布 Weibull(0.5, 1). $n = 10$. 符号検定.

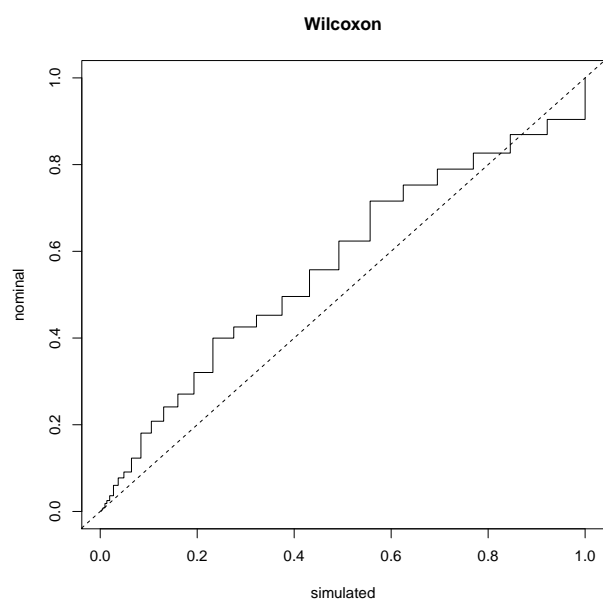


図 18 α エラーのプロット. ワイブル分布 Weibull(0.5, 1). $n = 10$. Wilcoxon の符号付き順位検定.

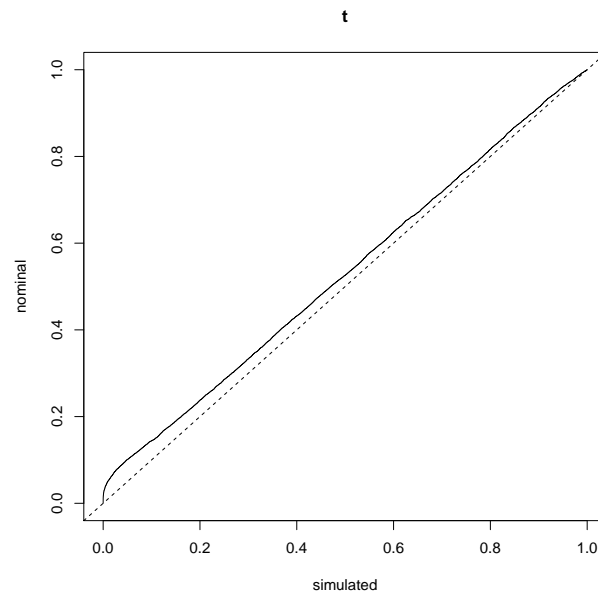


図 19 α エラーのプロット. ワイブル分布 Weibull(0.5, 1). $n = 100$. t 検定

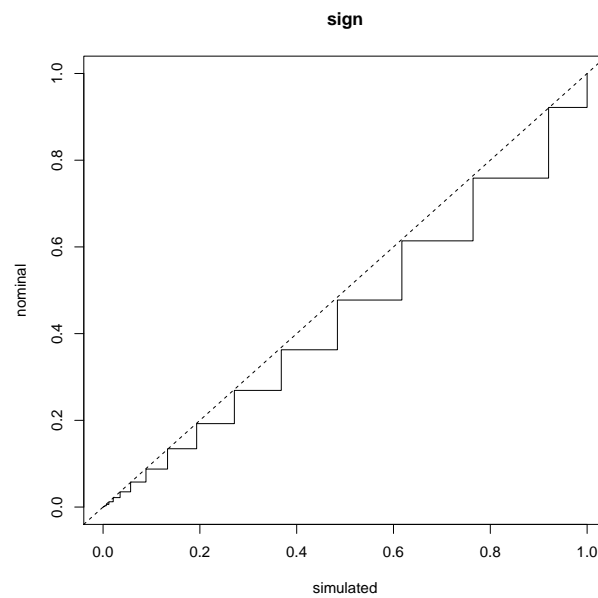


図 20 α エラーのプロット. ワイブル分布 Weibull(0.5, 1). $n = 100$. 符号検定.

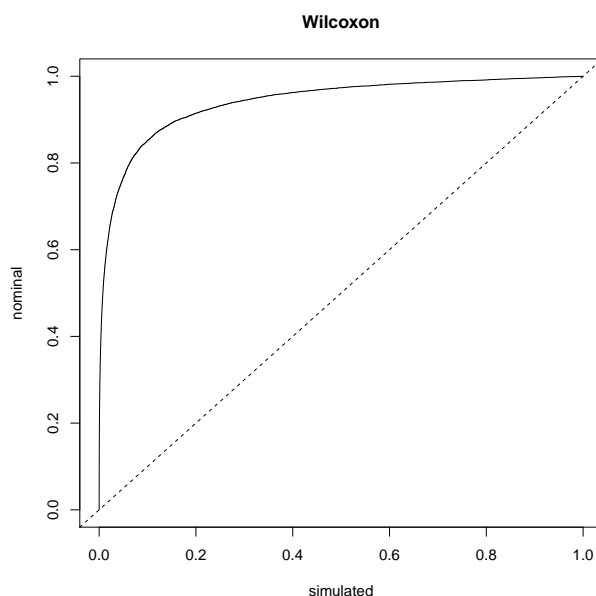


図 21 α エラーのプロット. ワイブル分布 $\text{Weibull}(0.5, 1)$. $n = 100$. Wilcoxon の符号付き順位検定.

しかし, そういう場合に Wilcoxon の符号付き順位検定が教えてくれるのは, 分布が左右対称でないという事実であり, 中央値について判断するのには適さない.

Wilcoxon の符号付き順位検定は正直, 使い所が難しいと思う.

そこで, Wilcoxon の符号付き順位検定を擁護するために, もうひとつシミュレーションを追加しよう.

本節のシミュレーションに用いたコードは, R フォルダにある `SignTestsim.R` である.

7 2 群の比較

ここでは Wilcoxon の順位和検定 (Wilcoxon rank sum test), あるいは マン・ホイットニー (Mann-Whitney) の U 検定と呼ばれる方法について述べる. 分布 F の密度関数が 0 を中心に左右対称であることを仮定する (意識されにくい仮定).

データ X_1, \dots, X_m が独立に $F(x - \mu)$ に従い, データ Y_1, \dots, Y_n が独立に $F(x - \nu)$ に従うとする. つまり分布の形状は同じで, ロケーション (中央値) だけがずれている状況を仮定する (意識されにくい仮定).

多くの場合, 興味があるのは, $\Delta = \nu - \mu > 0$ であるかどうかである.

R_i を X_i と Y_i を合併したときの, Y_i の順位 (小さい順) とする. フォーマルに書くと

$$R_i = \#\{j; X_j \leq Y_i\} + \#\{j; Y_j \leq Y_i\} \quad (6)$$

である. 次の統計量 U を考える.

$$U = \sum_{i=1}^n R_i. \quad (7)$$

U が十分大きい時は, Y_i の順位が後ろのほうに集まっていることなので, Δ は大きそうである. このことを使った検定を考える.

帰無仮説を $H_0 : \Delta = 0$ ($\nu = \mu$) と置くと, (つまり順位が一様にランダムであれば) $k = U - n(n+1)/2$ の確率分布は次の漸化式から計算できる.

$$p_{m,n}(k) = \frac{n}{m+n} p_{m,n-1}(k-m) + \frac{m}{m+n} p_{m-1,n}(k) \quad (8)$$

7.1 シミュレーション