数理統計学2022

記号法などについては次のメモを参照:

Memo.md (https://github.com/genkuroki/Statistics/blob/master/2022/Memo.md)

すべてをまとめたPDF:

<u>Statistics2022.pdf (https://github.com/genkuroki/Statistics/blob/master/2022/Statistics2022.pdf)</u> [download (https://github.com/genkuroki/Statistics/raw/master/2022/Statistics2022.pdf)]

これはおまけで作っているPDFファイルなので更新は遅れる予定、以下の個別のPDFファイルの方が常に新しいバージョンになっている。

目標

- 確率変数の扱い方
- 大数の法則と中心極限定理
- よく使われる確率分布達に関する知識
- 尤度と推定と要約統計
- 検定と信頼区間を表裏一体のものとして理解すること
- P値函数について理解すること
- 簡単な場合について具体的な数値の計算を行えるようになること(重要!)

資料

<u>Jupyter (https://jupyter.org/)</u> notebook をオンラインで読むなら <u>nbviewer (https://nbviewer.org/)</u> 経由の閲覧がおすすめです。オフラインでも読めるようにするためには pdf ファイルをダウンロードしてください.

<u>Julia言語 (https://julialang.org/)</u>を使って数値の計算やグラフの作画を行っています。必要最小限の使い方については次の資料を見て下さい:

• 07-2 Julia言語を使った統計学の勉強の仕方 (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/07-2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.jpynb)

WolframAlpha (https://www.wolframalpha.com/) も併用しています.

明らかな誤りをまだ大量に含んでいるものと思われます。 ごめんなさい。 適切に訂正しながら読んでください。 随時訂正更新しています。

01 Bernoulli試行と関連確率分布

- Jupyter notebook
 - (https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.ipynb)
 [nbviewer]
- (https://nbviewer.org/qithub/qenkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.ip
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.pdf)
 - (https://github.com/genkuroki/Statistics/raw/master/2022/01%20Bernoulli%20trial%20and%20related%20distributions.pdf)]
- hand-written note
 - (https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% Idownload
 - (https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E
- hand-written note
 - (https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% Idownload
 - (https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E

定義された確率分布

- カテゴリカル分布: $Categorical(p_1, \ldots, p_n)$
- ・ 成功確率 p の Bernoulli分布: Bernoulli(p)
- ・ 試行回数 n, 成功確率 p の Bernoulli試行の分布: $Bernoulli(p)^n$
- 試行回数 n, 成功確率 p の **二項分布**: Binomial (n,p) (n 回のBernoulli試行での成功回数の分布)
- 成功回数 k, 成功確率 p の **負の二項分布**: NegativeBinomial $(n,p) = \operatorname{NegBin}(n,p)$ (Bernoulli試行をちょうど k 回成功するまで続けたときの失敗の回数の分布)

- 成功確率 p の **幾何分布**: Geometric (p) = NegativeBinomial(1, p) (Bernoulli試行を成功するまで続けたときの失敗の回数の分布)
- 0から1のあいだの **一様分布**: Uniform(0,1) (コンピュータでの一様乱数函数 rand() のモデル化)
- 試行回数 n の 一様乱数生成の繰り返しの分布: $Uniform(0,1)^n$
- 平均 μ , 分散 σ^2 (標準偏差 σ)の **正規分布**: Normal(μ , σ)
- 標準正規分布: Normal() = Normal(0,1)
- 分布 D の アフィン変換: aD+b
- 分布 D の スケール変換: aD (a > 0)
- 分布 D の 平行移動: D+b
- 分布 D の 逆数: 1/D

02 Gauss積分, ガンマ函数, ベータ函数

Jupyter notebook

(https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20fu

(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Be

pdf

(https://github.com/genkuroki/Statistics/blob/master/2022/02%20Gaussian%20integrals%2C%20Gamma%20and%20Beta%20fulldownload

· hand-written note

(https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% [download]

(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E

· hand-written note

(https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% [download]

(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E

定義された確率分布

- 正規分布 (再): $Normal(\mu, \sigma) = \mu + \sigma Normal(0, 1)$
- 形状 α , スケール θ の **ガンマ分布**: $\operatorname{Gamma}(\alpha, \theta) = \theta \operatorname{Gamma}(\alpha, 1)$ (平均的なイベント発生間隔が θ のときにちょうど α 回イベントが生じるまでにかかる時間の分布)
- 平均 θ の **指数分布**: Exponential(θ) = Gamma($1, \theta$) (平均的なイベント発生間隔が θ のときにイベントが発生するまでにかかる時間の分布)
- 中央値 a, スケール b の Laplace分布: Laplace(a,b)
- ・ 形状 α , スケール θ の **逆ガンマ分布**: Inverse $Gamma(\alpha, \theta) = 1/Gamma(\alpha, 1/\theta)$
- 自由度 ν の \mathbf{x}^2 分布(カイ二乗分布): $\mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2,2)$
- 形状 α , スケール θ の Weibull分布: Weibull(α , θ)
- パラメータ α,β の ベータ分布: $\mathrm{Beta}(\alpha,\beta)$ ($\alpha+\beta-1$ 個の一様乱数の中で α 番目に小さな(= β 番目に大きな)値の分布)
- ・ 自由度 ν の **t分布**: $TDist(\nu)$ (平均 0 の正規分布で分散の逆数が $Chisq(\nu)/\nu$ に従う場合, $BetaPrime(1/2,\nu/2)$ と密接に関係)
- ・ パラメータ α,β の ベータプライム分布: $\operatorname{BetaPrime}(\alpha,\beta)$ (確率 t に関するベータ分布 $\operatorname{Beta}(\alpha,\beta)$ をオッズ u=t/(1-t)=1/(1-t)-1 に関する分布に変換したもの)
- 自由度 ν_1, ν_2 の **F分布**: $\mathrm{FDist}(\nu_1, \nu_2) = (\nu_2/\nu_1) \, \mathrm{BetaPrime}(\nu_1/2, \nu_2/2)$
- Dirichlet分布: Dirichlet $(\alpha_1,\ldots,\alpha_n)$ (n=2) の場合がベータ分布と一致)

03 確率分布達の解釈

Jupyter notebook

(https://github.com/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.ipynb)
[nbviewer

 $\underline{(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/03\%20Interpretation\%20of\%20probability\%20distributions.ipyroperation with the properties of the propert$

• pdf (https://github.com/genkuroki/Statistics/blob/master/2022/03%20Interpretation%20of%20probability%20distributions.pdf) Idownload

(https://github.com/genkuroki/Statistics/raw/master/2022/03%20Interpretation%20of%20probability%20distributions.pdf)]

• hand-written note

(https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% Idownload

(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%I

定義された確率分布

• 対数正規分布: $LogNormal(\mu, \sigma) = exp(Normal(\mu, \sigma))$

- 平均 λ の Poisson分布(ポアソン分布): Poisson(λ)
- **負の二項分布**: NegativeBinomial $(\alpha, 1/(1+\theta))$ (Poisson分布の平均パラメータ λ がガンマ分布に従う場合)
- a から b のあいだの 一様分布: Uniform(a,b)
- ベータニ項分布: BetaBinomial (n,α,β) (二項分布の成功確率パラメータがベータ分布に従う場合)
- 超幾何分布: Hypergeometric (s, f, n) (非復元抽出の分布)
- ベータ負の二項分布: NegativeBetaBinomial (n,α,β) (負の二項分布の成功確率パラメータがベータ分布に従う場合)
- 試行回数 n の Pólyaの壺の確率分布: Pólya (n,α,β) (Bernoulli試行の分布で成功確率パラメータがベータ分布に従う場合)

04 標本分布について

- <u>Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.ipynb)</u> [<u>nbviewer (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.ipynb)</u>]
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.pdf) [download (https://github.com/genkuroki/Statistics/raw/master/2022/04%20Distribution%20of%20samples.pdf)]
- · hand-written note

(https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% Idownload

(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%I

演習用サンプル

- Anscombeの例(アンスコムの例)
 - csv file (https://github.com/genkuroki/Statistics/blob/master/2022/data/anscombe.csv), raw version (https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/anscombe.csv)
 - <u>transposed version (https://github.com/genkuroki/Statistics/blob/master/2022/data/anscombe_transposed.csv)</u>,
 <u>transposed raw version</u>

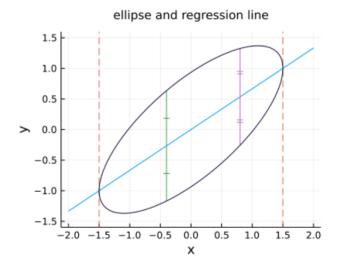
(https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/anscombe_transposed.csv)

定義された確率分布

- 分布達 D_1, \ldots, D_n の 積: $D_1 \times \cdots \times D_n$
- 分布 D の 累乗: Dⁿ
- 分布 D のサイズ n の 標本分布: Dⁿ

作成された動画

楕円と回帰直線の関係 (https://github.com/genkuroki/Statistics/blob/master/2022/images/ellipse_and_regressionline.gif)



05 大数の法則と中心極限定理

- <u>Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/05%20Central%20limit%20theorem.ipynb)</u> [nbviewer_(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/05%20Central%20limit%20theorem.ipynb)]
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/05%20Central%20limit%20theorem.pdf) [download (https://github.com/genkuroki/Statistics/raw/master/2022/05%20Central%20limit%20theorem.pdf)]
- hand-written note
 (https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%
 [download

(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%I

大数の法則の数値例 ($(X_1+X_2+\cdots+X_n)/n$ の分布)

Bernoulli試行, 正規分布の標本分布, ガンマ分布 Gamma(2, 3) の標本分布

大数の法則が成立しない場合の数値例

Cachy分布の標本分布、Pólyaの壺=事前分布がベータ分布のBernoulli試行

ランダムウォークの数値例 $X_1+X_2+\cdots+X_n$ の分布

Bernoulli試行, 正規分布の標本分布, ガンマ分布 Gamma}(2, 3) の標本分布, Cachy分布の標本分布, Pólyaの壺 = 事前分布がベータ 分布のBernoulli試行

中心極限定理の数値例

対称な分布 二項分布, 一様分布の標本分布, ベータ分布 Beta(0.5, 0.5), Bernoulli試行 p = 0.5, Laplace分布の標本分布

非対称な分布 ガンマ分布 Gamma(3, 1) の標本分布, Bernoulli試行 (p = 0.25) の標本分布, Poisson分布 (期待値1) の標本分布, 指数分布 (期待値1) の標本分布, χ^2 分布 (自由度1) の標本分布, 混合正規分布の標本分布, 対数正規分布の標本分布

正規分布で近似される確率分布の数値例

デルタ法の応用 自由度が大きなx2分布の逆数,自由度が大きなx2分布の対数,2つのパラメータが両方とも大きなベータ分布

適当なパラメータ領域において正規分布で近似される分布 二項分布, 負の二項分布, Poisson分布, ガンマ分布, χ²分布, 逆ガンマ分布, ベータ分布, t分布, ベータプライム分布, F分布, 超幾何分布, ベータ二項分布, ベータ負の二項分布

06 条件付き確率分布, 尤度, 推定, 記述統計

Jupyter notebook

(https://github.com/genkuroki/Statistics/blob/master/2022/06%20Conditional%20distribution%2C%20likelihood%2C%20estimatic Inbyiewer

 $\underline{(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/06\%20Conditional\%20distribution\%2C\%20likelihood\%2C\%20li$

pdf

(https://github.com/genkuroki/Statistics/blob/master/2022/06%20Conditional%20distribution%2C%20likelihood%2C%20estimatic Idownload

(https://github.com/genkuroki/Statistics/raw/master/2022/06%20Conditional%20distribution%2C%20likelihood%2C%20estimatio

• hand-written note

(https://github.com/genkuroki/Statistics/blob/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88% [download]

(https://github.com/genkuroki/Statistics/raw/master/2022/handwritten/%E6%95%B0%E7%90%86%E7%B5%B1%E8%A8%88%E

演習用サンプル

- <u>Datasaurusの例 (データサウルスの例) (http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html)</u>
 - csv file (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_data.csv), raw version (https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/Datasaurus_data.csv)
 - x座標のみ (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus X.txt), y座標のみ (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus Y.txt)
 - x座標のみコンマ付き (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus X with commas.txt), y
 座標のみコンマ付き (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus Y with commas.txt)

この例に関する詳しい解説が次の場所にある:

 Justin Matejka, George Fitzmaurice. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. Honourable Mention, ACM SIGCHI Conference on Human Factors in Computing Systems, 2017 [link (https://www.autodesk.com/research/publications/same-stats-different-graphs)]

関連の情報が datasaurus same stats を検索 (https://www.google.com/search?g=datasaurus+same+stats) すれば得られる.

定義された確率分布

- 2変量正規分布: $\operatorname{MvNormal}(\mu,\Sigma)$ (多変量正規分布 でも同じ記号を使う.)
- 2×2の分割表の分布
 - 4つのPoisson分布の積
 - 四項分布
 - 2つの二項分布の積
 - Fisherの非心超幾何分布

07 例

07-1 例:ベータ函数と二項分布の関係とその応用

- <u>Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/07-1%20Relationship%20between%20beta%20and%20binomial%20distributions.ipynb) [nbviewer (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/07-</u>
 - 1%20Relationship%20between%20beta%20and%20binomial%20distributions.ipynb)]
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/07-
 - 1%20Relationship%20between%20beta%20and%20binomial%20distributions.pdf) [download
 - (https://github.com/genkuroki/Statistics/raw/master/2022/07-
 - 1%20Relationship%20between%20beta%20and%20binomial%20distributions.pdf)]

07-2 Julia言語を使った統計学の勉強の仕方

- Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/07-
 - 2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.ipynb) [nbviewer
 - (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/07-
 - 2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.ipynb)]
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/07-
 - 2%20How%20to%20use%20Julia%20language%20for%20learning%20statistics.pdf) [download
 - (https://github.com/genkuroki/Statistics/raw/master/2022/07-
 - 1%20Relationship%20between%20beta%20and%20binomial%20distributions.pdf)]

08 検定と信頼区間(1) 一般論

- Jupyter notebook
 - (https://github.com/genkuroki/Statistics/blob/master/2022/08%20Hypothesis%20testing%20and%20confidence%20interval%20-%20General%20theory.ipynb) Inbviewer
- (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/08%20Hypothesis%20testing%20and%20confidence%20inter%20General%20theory.ipynb)]
- ndf
 - $\underline{(https://github.com/genkuroki/Statistics/blob/master/2022/08\%20Hypothesis\%20testing\%20and\%20confidence\%20interval\%20-\\ \underline{\%20General\%20theory.pdf}) \ \underline{[download]}$
 - (https://github.com/genkuroki/Statistics/raw/master/2022/08%20Hypothesis%20testing%20and%20confidence%20interval%20-%20General%20theory.pdf)]

P値函数 の概念が最も重要である.

この一連のノートでは **信頼区間は常に対応するP値函数を使って定義される** ことになる。

関連講義動画

 京都大学大学院医学研究科 聴講コース 臨床研究者のための生物統計学「仮説検定とP値の誤解」 佐藤 俊哉 医学研究科教授 https://youtu.be/vz9cZnB1d1c (https://youtu.be/vz9cZnB1d1c)

この講義動画は「P値に関するASA声明 (https://www.biometrics.gr.jp/news/all/ASA.pdf)」の考え方の非常に良い解説になっている.

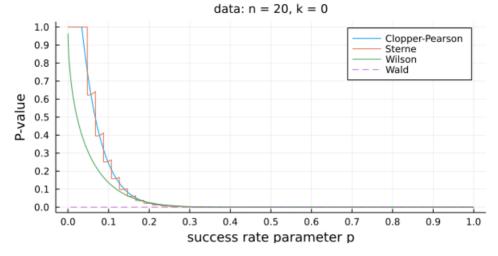
09 検定と信頼区間(2) 比率

- Jupyter notebook
 - (https://github.com/genkuroki/Statistics/blob/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Proportion.ipynb) [nbviewer]
 - $\underline{(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/09\%20Hypothesis\%20testing\%20and\%20confidence\%20inter\%20Proportion.ipynb)]}$
- pdf
 - (https://github.com/genkuroki/Statistics/blob/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Proportion.pdf) [download
 - (https://github.com/genkuroki/Statistics/raw/master/2022/09%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Proportion.pdf)]

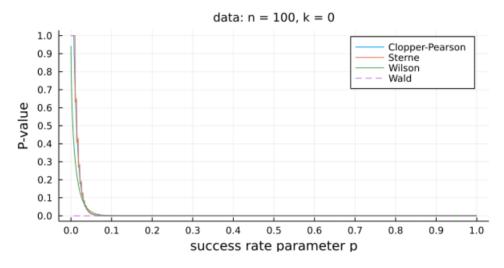
定義された信頼区間

- 二項検定に付随する4種の信頼区間
 - Clopper-Pearsonの信頼区間
 - Sterneの信頼区間
 - Wilsonの信頼区間
 - Waldの信頼区間

作成された動画



n=100の二項分布モデルのP値函数 (https://github.com/genkuroki/Statistics/blob/master/2022/images/pvaluefunction100.gif)



10 検定と信頼区間(3) 平均

Jupyter notebook

(https://github.com/genkuroki/Statistics/blob/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Mean.ipynb) [nbviewer

(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/10%20Hypothesis%20testing%20and%20confidence%20inter%20Mean.ipynb)]

pdf

(https://github.com/genkuroki/Statistics/blob/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Mean.pdf) [download

(https://github.com/genkuroki/Statistics/raw/master/2022/10%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Mean.pdf)]

定義された信頼区間

平均の信頼区間

- 標準正規分布を使って計算される信頼区間
- t分布を使って計算される信頼区間

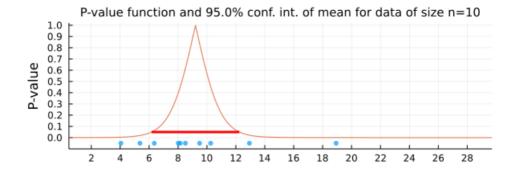
後者のみを使用する.

定義された確率分布

• 自由度vのx分布: $Chi(\nu) = \sqrt{Chisq(\nu)}$

作成された動画

n→∞でのP値函数と信頼区間の挙動 (https://github.com/genkuroki/Statistics/blob/master/2022/images/confint_of_mean.gif)



11 検定と信頼区間(4) 比率の違い

· Jupyter notebook

(https://github.com/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Two%20proportions.ipynb) [nbviewer

(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20inter%20Two%20proportions.ipynb)]

pdf

(https://github.com/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Two%20proportions.pdf) [download

(https://github.com/genkuroki/Statistics/raw/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Two%20proportions.pdf)]

定義された信頼区間

オッズ比の信頼区間

- Wald型の信頼区間
- Pearsonのx²検定に付随する信頼区間
- Fisher検定に付随する信頼区間
 - Sterne型
 - Clopper-Pearson型

リスク比の信頼区間

- Wald型の信頼区間
- Pearsonのx²検定に付随する信頼区間

比率の差の信頼区間

• Wald型の信頼区間

定義された確率分布

• Fisherの非心超幾何分布: FisherNoncentralHypergeometric (m, n, r, ω)

関連講義動画

- <u>聴講コース 臨床研究者のための生物統計学 (https://ocw.kyoto-u.ac.jp/course/328/)</u>
 - 佐藤俊哉, ランダム化ができないとき, 2018/10/25, 長さ 1:02:31) [https://youtu.be/tUkyjZXU1vc (https://youtu.be/tUkyjZXU1vc)]
 - 佐藤俊哉, 交絡とその調整, 2018/12/20, 長さ 1:00:52) [https://youtu.be/ybdkQFEdCPM (https://youtu.be/ybdkQFEdCPM)]
 - 佐藤俊哉, 回帰モデルと傾向スコア, 2019/02/21, 長さ 1:04:44) [https://youtu.be/cOHN444kBlo (https://youtu.be/cOHN444kBlo)]

「検定と信頼区間(4) 比率の違い」のノートブックの内容をJulia言語で直接実行するためのヒント

「検定と信頼区間(4) 比率の違い」のノートブック

(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/11%20Hypothesis%20testing%20and%20confidence%20interval%%20Two%20proportions.jpynb) の中の

```
a, b, c, d = 49, 965, 26, 854
Oshow a, b, c, d
\chi^2 = (a+b+c+d)*(a*d-b*c)^2/((a+b)*(c+d)*(a+c)*(b+d))
@show χ²
Qshow pvalue = ccdf(Chisq(1), \chi^2)
\alpha = 0.05
\mathbb{Q}show \alpha
confint = find_zeros(0.5, 3.0) do \omega
     A, B, C = 1-\omega, a+d+\omega*(b+c), a*d-\omega*b*c
     \delta = 2C/(B + \sqrt{(B^2 - 4A*C)})
     \chi^2 = \delta^2 \times (1/(a-\delta) + 1/(b+\delta) + 1/(c+\delta) + 1/(d-\delta))
     \operatorname{ccdf}(\operatorname{Chisq}(1), \chi^2) - \alpha
end
@show confint;
(a, b, c, d) = (49, 965, 26, 854)
\chi^2 = 4.368235947209686
pvalue = ccdf(Chisq(1), \chi^2) = 0.03661487827607113
\alpha = 0.05
confint = [1.031778110800563, 2.6957054358850328]
```

の部分を直接 Julia (download (https://julialang.org/downloads/))で実行するためには以下のようにすればよい.

```
MSYS2 MINGW64 Shell
                                                                                       ×
                                                                                Documentation: https://docs.julialang.org
                            Type "?" for help, "]?" for Pkg help.
                            Version 1.7.3 (2022-05-06)
                            Official https://julialang.org/ release
julia> using Distributions, Roots
julia> a, b, c, d = 49, 965, 26, 854
(49, 965, 26, 854)
julia> \chi^2 = (a+b+c+d)*(a*d-b*c)^2/((a+b)*(c+d)*(a+c)*(b+d))
4.368235947209686
julia> pvalue = ccdf(Chisq(1), \chi^2)
0.03661487827607113
julia> \alpha = 0.05
0.05
julia> confint = find_zeros(0.5, 3.0) do ω
            A, B, C = 1-\omega, a+d+\omega*(b+c), a*d-\omega*b*c
            \delta = 2C/(B + \sqrt{(B^2 - 4A*C)})
            \chi^2 = \delta^2 * (1/(a-\delta) + 1/(b+\delta) + 1/(c+\delta) + 1/(d-\delta)
            ccdf(Chisq(1), \chi^2) - \alpha
       end;
julia> @show confint;
confint = [1.031778110800563, 2.6957054358850328]
help?> χ²
 \chi^2" can be typed by \phi^2
search: χ²
```

注意するべきポイントは using Distributions, Roots を忘れないことである.

しかし、どのくらい前もって using しておけばよいのか分かり難いので、上のノートブックの上の方にある using 達をすべて 前持って実行しておくという手もある. using には少し時間が取られるので、そのようにしたJuliaの窓をそのまま閉じずに放置 しておき、超高級電卓として利用し続けると便利だと思われる.

 χ^2 などの入力の仕方は、プロンプト julia> で ? と入力して、ヘルプモードのプロンプト help?> を表示させて、 χ^2 を貼り付けてエンターキーを押せばよい、その結果

" χ^2 " can be typed by \chi<tab>\^2<tab>

と表示されている。 すなわち、 \chi と入力してタブキーを押し、 さらに \^2 と入力してタブキーを押せば χ^2 と入力できる.

12 検定と信頼区間(5) 平均の差

Jupyter notebook

(https://github.com/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Two%20means.jpynb) [nbviewer

(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20inter%20Two%20means.ipynb)]

• pdf

(https://github.com/genkuroki/Statistics/blob/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Two%20means.pdf) Idownload

(https://github.com/genkuroki/Statistics/raw/master/2022/12%20Hypothesis%20testing%20and%20confidence%20interval%20-%20Two%20means.pdf)]

定義された信頼区間

2群の平均の差の信頼区間

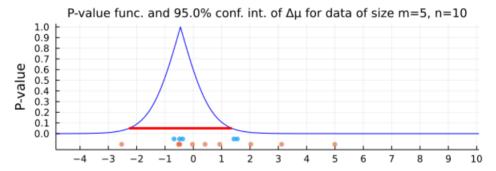
• Welchのt検定に付随する信頼区間

Studentのt検定は使用しない.

作成された動画

データサイズを大きくしたときのP値函数と信頼区間の挙動

(https://github.com/genkuroki/Statistics/blob/master/2022/images/confint_of_diffmeans.gif)



13 誤用を避けるための注意

- <u>Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/13%20How%20to%20avoid%20misuse.ipynb)</u> [nbviewer_(https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/13%20How%20to%20avoid%20misuse.ipynb)]
- pdf () [download ()]

P値に関する誤用を避けるために最初に読むべき基本文献は

• 統計的有意性とP値に関するASA声明 (https://www.biometrics.gr.jp/news/all/ASA.pdf)

である. ただし, 以下の部分の「データ」は「統計モデル内でランダムに生成された仮想的なデータ」の意味であると解釈する必要がある. (「現実で得た観察データ」であると解釈してはいけない.)

2. P 値とは?

おおざっぱにいうと、P 値とは特定の統計モデルのもとで、データの統計的要約(たとえば、2 グループ比較での標本平均の差)が観察された値と等しいか、それよりも極端な値をとる確率である。

一連の講義では,通常の統計学入門の教科書とは違って,「統計モデル」という用語を多用した. このP値の定義の具体例については他のノートで詳しく説明した.

14 回帰

- <u>Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/14%20Regression.ipynb) [nbviewer (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/14%20Regression.ipynb)]</u>
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/14%20Regression.pdf) [download (https://github.com/genkuroki/Statistics/raw/master/2022/14%20Regression.pdf)]

定義された信頼推定

線形回帰の場合:

- 回帰函数の値の信頼区間
- 回帰函数+残差の予測区間

線形回帰の場合:

- $t_*=eta_0+eta_1x_*$ の信頼区間
- ・ $p_* = \operatorname{logistic}(\beta_0 + \beta_1 x_*)$ の信頼区間
- $x_* = 0$ の場合から β_0 の信頼区間が得られる.
- β₁ の信頼区間

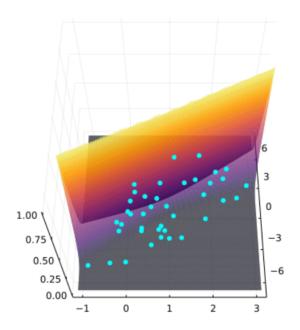
定義された確率分布

- 多変量正規分布: $MvNormal(\mu, \Sigma)$
- ロジスティックモデル: LogisticModel (x, β)

作成された動画

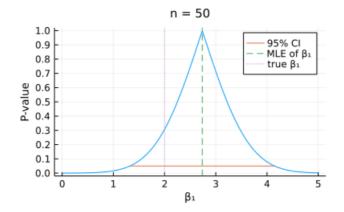
線形回帰の信頼区間に対応するP値函数

(https://github.com/genkuroki/Statistics/blob/master/2022/images/anim_linreg_confint_pvalfunc_3d.gif) (x軸は-2から2まで)



ロジスティック回帰のβιの信頼区間に対応するP値函数

(https://github.com/genkuroki/Statistics/blob/master/2022/images/anim_logisticreg_beta1_confint_pvalfunc.gif)



15 まとめ

- <u>Jupyter notebook (https://github.com/genkuroki/Statistics/blob/master/2022/15%20Summary.ipynb) [nbviewer (https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/15%20Summary.ipynb)]</u>
- pdf (https://github.com/genkuroki/Statistics/blob/master/2022/15%20Summary.pdf) [download (https://github.com/genkuroki/Statistics/raw/master/2022/15%20Summary.pdf)]
- 1. 二項分布モデルでのClopper-Pearsonの信頼区間
- 2. P値函数について
- 3. Welchの t 検定について
- 4. 数学的な補足: 大数の法則と中心極限定理について

参考資料

以上で解説できなかったことに関するノート

- Kullback-Leibler情報量とSanovの定理 (2016-2018) [pdf (https://genkuroki.github.io/documents/20160616KullbackLeibler.pdf)]
- ガンマ分布の中心極限定理とStirlingの公式 (2016-2018) [pdf (https://genkuroki.github.io/documents/20160501StirlingFormula.pdf)]
- 一般化されたLaplaceの方法 (2016) [pdf (https://genkuroki.github.io/documents/20161014GeneralizedLaplace.pdf)]
- 最尤法とカイ二乗検定の基礎 (2017-2020) [pdf (https://genkuroki.github.io/documents/IntroMLE.pdf)]
- 確率論入門 (2017) [pdf (https://genkuroki.github.io/documents/IntroProbability.pdf)]
- ベイズ統計入門 (手書きのノート, 2019) [pdf (https://genkuroki.github.io/documents/2019-09-03 BayesianStatistics.pdf)]
- ベイズ統計の枠組みと解釈について (2019) [Jupyter notebook (https://nbviewer.org/github/genkuroki/Statistics.ipynb)]
- Kullback-Leibler情報量と記述統計 (2019-2020) [Jupyter notebook (https://nbviewer.org/github/genkuroki/Statistics/blob/master/KL%20information%20and%20descriptive%20statistics.ipynb)]

←