JINSE活動報告シンポジウム
「論より統計！統計教育の改善に向けた活動」
日時：2016 年2 月27 日(土)
会場：青山学院大学 本多記念国際会議場

これからの統計教育

大阪大学 基礎工学研究科

狩野　裕

Japanese Inter-university Network for Statistical Education

統計教育大学間連携ネットワーク

JINSE II

JINSE PART II

じんせいにあいがある

お題と，ひとこと回答

- 3年前と比較して統計に対する社会での認識は変わってきたか？
  - 統計や(ビッグ)データという用語は浸透した
  - 「データサイエンティスト」の求人が出てきた
    - ごく少数．キャリア形成は弱い
- 学生(または社会人)は社会で使う統計を理解しているか？
  - 大いに疑問
  - 講義コマ数が増加していない．カリキュラム改革も道半ば
  - 統計学を理解させるのは難しい！
- 対学生の統計教育が遅れると何が起こるのか？
  - リスク評価が改善しない．プアな意思決定方法が改善しない
  - 日本人の数理的素養を活かせないことは社会的損失
- 統計の分野で学生に絶対に教えておくべきことは何？
  - むちゃくちゃ多い

確率嫌いの日本人

- 河合薫氏（大阪ガス）
  - どうしようもない．諦めている
  - 日本の文化．大学教育だけでは不可能
- 狩野
  - 統計教育で何とかしたい
  - 大学教育だけでは不可能
- 例えば．．．
  - 数字で煙に巻くつもりか！
  - ある特定の宝くじ売り場に購買客が集中する
  - 頑張れば（原発）事故は起こらない
  - 根性を出せば必ず成功する
  - 世の中，数字ではないだよ！
  - 二者択一，確率？　結局どっちなんだ，白黒つけろ！
  - 麻雀，勝つと実力，負けると運

統計の分野で学生に絶対に教えておくべきことは何？
（参考 JINSE カリキュラム策定委員会WEB）

- データの見方（リテラシー）
  - 偶然の一致
  - 基礎率の誤り
  - 因果と相関，交絡変数
  - 確率的現象，回帰効果
  - 調査の設計，データの偏り
  - 他

- 統計基礎
  - 確率的現象の数理的理解
  - 統計的推測の基礎

- データによる課題解決
  - プロジェクトベース（PBL）
  - プレゼンテーション・コミュニケーション

統計の分野で学生に絶対に教えておくべきことは何？

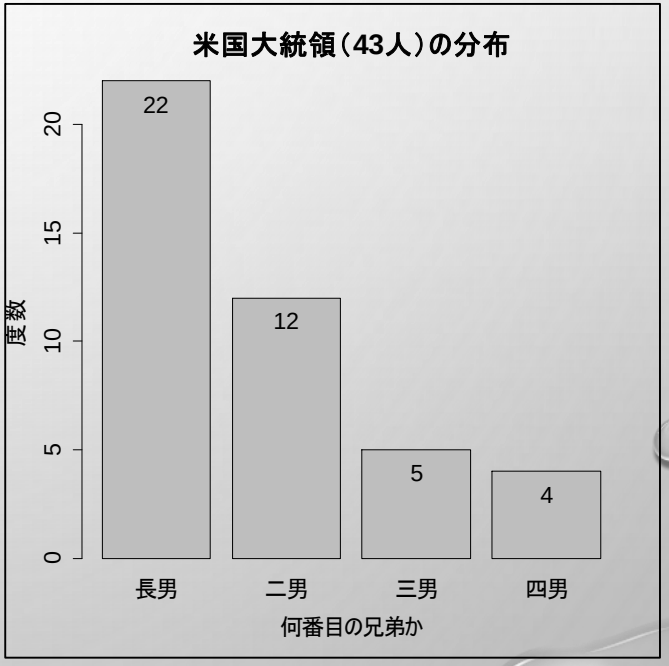
- データの見方（リテラシー）
  - 顧客満足度98.7%!
  - 「子」のつく名前の女の子は頭がいい
  - 修造がいれば気温が上がる
  - 偉大な文化人・芸術家・学者は長男・長女であることが多く，次男や次女は明らかに少ない

- 統計基礎
  - 確率変数と確率分布
  - 標本分布論
  - BLUE，検定の過誤，検出力
- データによる課題解決
  - 具体的な課題に接近することで以下を修得
    - PDCA，PPDAC
    - データ収集とデータハンドリングの基礎技術
    - 統計グラフ（visualization）
    - 統計解析と結論

小手試し
––データによる課題解決––

- 偉大な文化人･芸術家･学者は長男･長女であることが多く，次男や次女は明らかに少ない
  - 米国大統領の43人中22人は長男

- 追加課題
  - 明治維新の志士は？
  - ノーベル賞受賞者は？
  - M-1グランプリ優勝者は？



多様な教育が必要

- 全項目修得は不可能
  - リテ＋基礎＋課題
- 画一的教育は不適
  - 多様性が必要
  - 理工系
    - 統計＋数学＋情報技術のバランス
- 英語教育の必要性
  - 英語で統計コミュニケーション
  - 特に大学院

- タイプⅠ
  - リテ：基礎：課題＝1：2：8
- タイプⅡ
  - リテ：基礎：課題＝1：8：2
- タイプⅢ
  - リテ：基礎：課題＝1：5：5

JINSEによる高度副プログラム「データ科学」@大阪大学

- 平成26年度に開設
  - 統計数理コース
  - 機械学習コース
  - 人文社会統計学コース
  - 保健医療統計学コース
  - 経済経営統計学コース
  - ビッグデータ＆データサイエンティストコース（H28年度より）
  - Statistics-in-English course (H28年度より)

続く

10分終了

JINSE活動報告シンポジウム
「論より統計！統計教育の改善に向けた活動」
日時：2016 年2 月27 日(土)
会場：青山学院大学 本多記念国際会議場

(続)これからの統計教育

「統計基礎」の実際

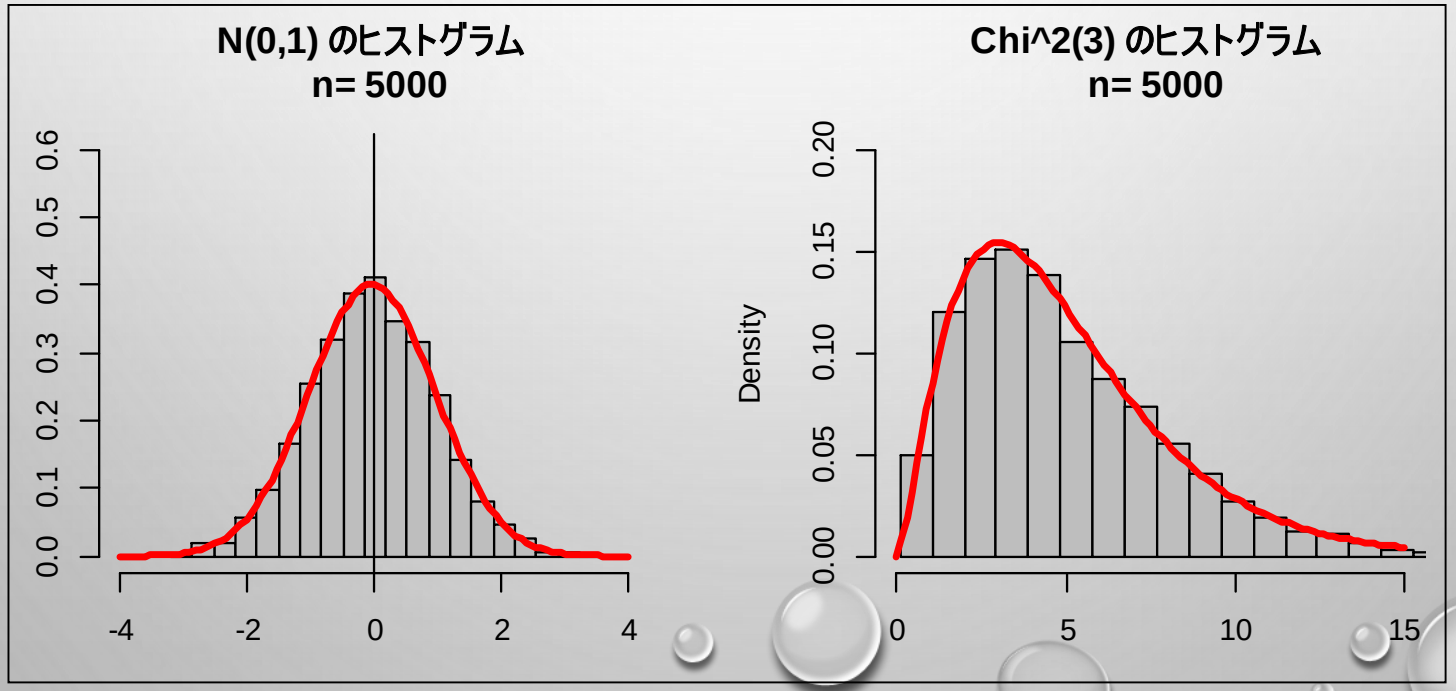
確率変数と確率分布

標本分布論

BLUE, 検定の過誤, 検出力

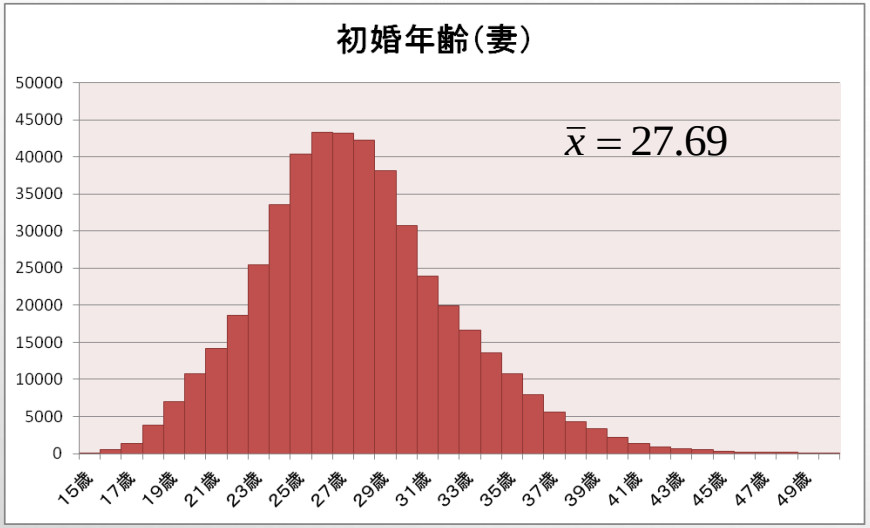
「標本分布論」事始め

- ランダムな現象(確率変数)の結果を正確に予測することはできないが，条件を整えて，多数回その現象を観測するとヒストグラム(分布)は安定する
  - ヒストグラムの極限を(確率)分布という

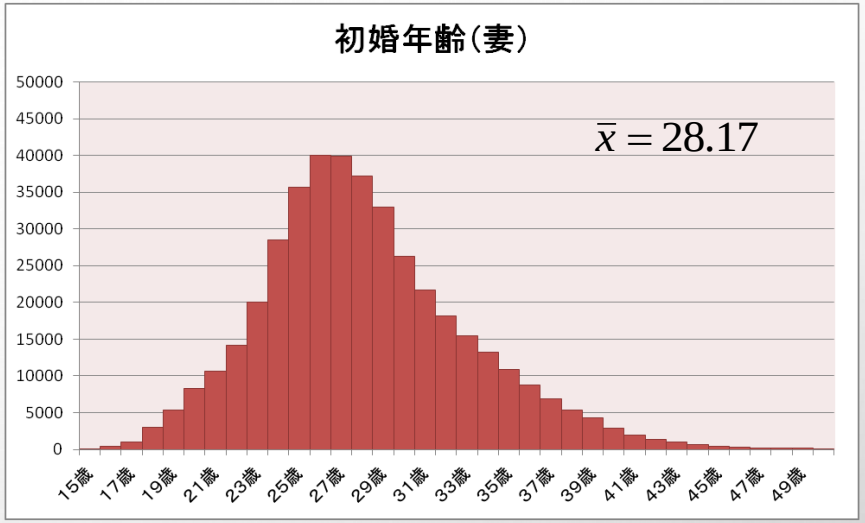


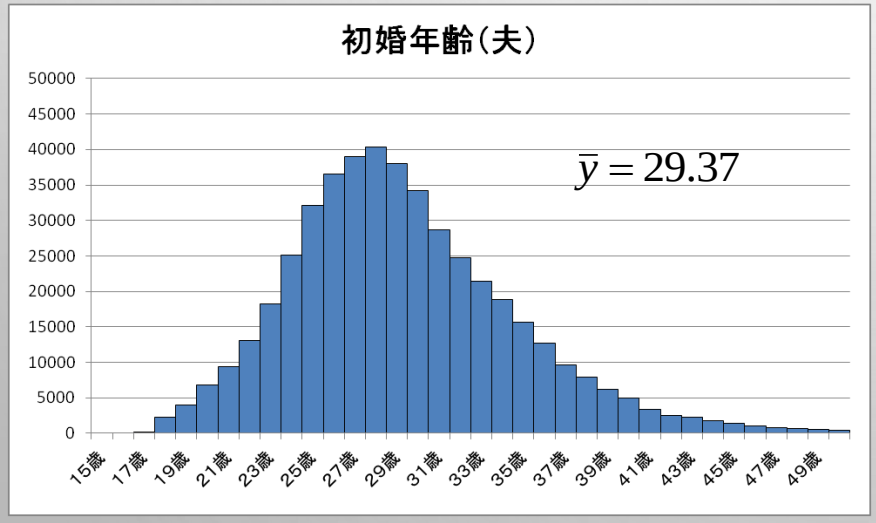
確率的現象の「安定性」
Nが大きいとき，確率的な現象の分布は安定する

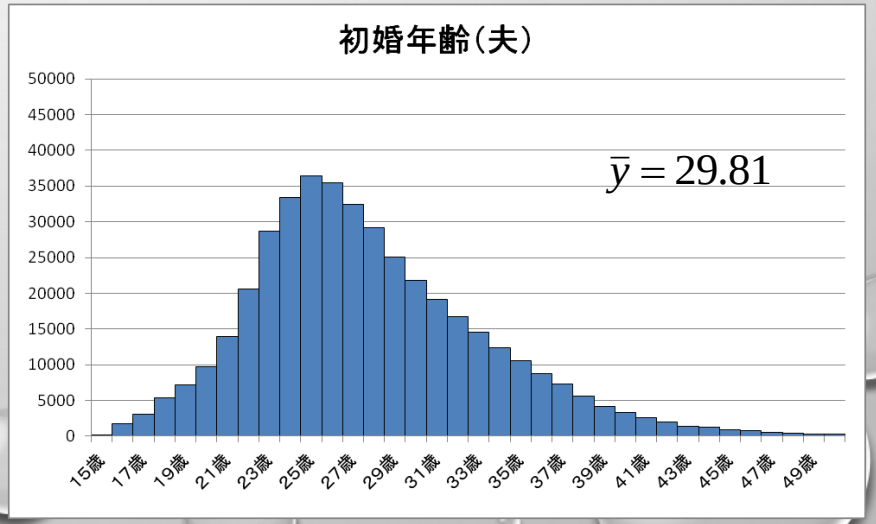
2008年度(n=466,609)



2011年度(n=418,223)

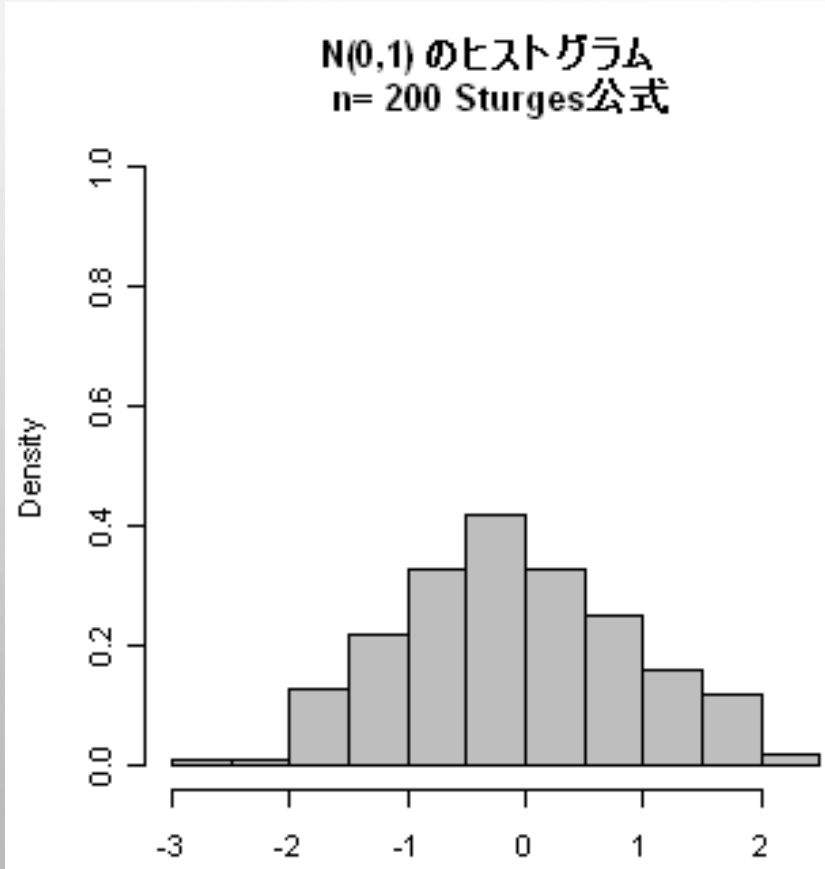


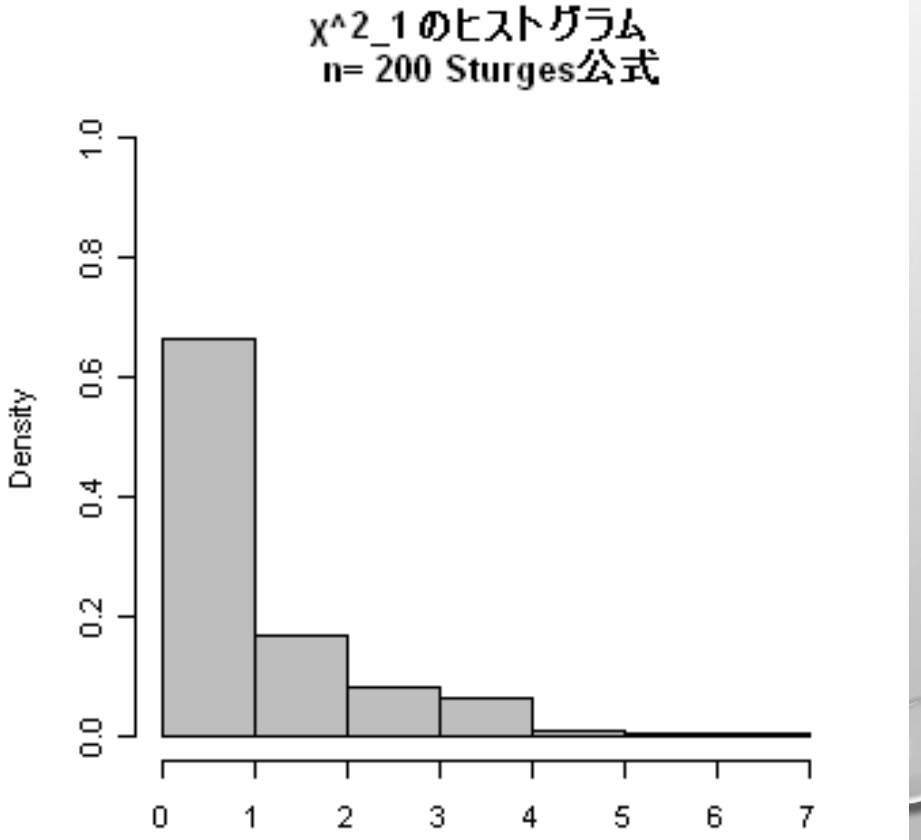




カイ二乗分布はどんな分布？

- 確率分布（密度関数の形）を予想させる



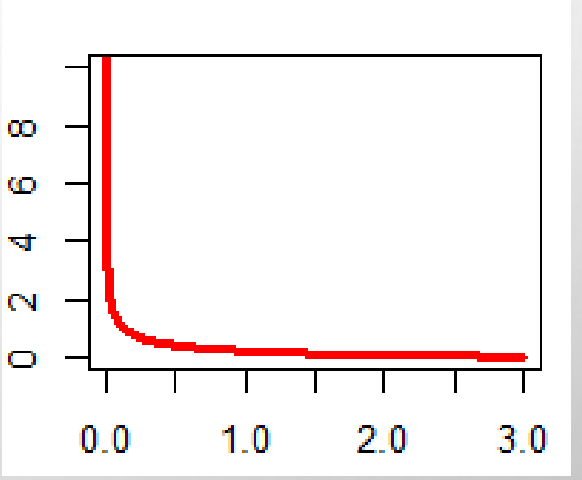


密度関数の形を予想
--学生に聞いてみる--

	H19	H20	H21	H24
0	a few	11	1	2
有限値	many	many	15	many
+∞	some	12	4	3

$$Y = X^2 \text{ with } X \sim N(0,1)$$

$$\begin{aligned} h(y) &= \frac{d}{dy}P(Y \le y) = \frac{d}{dy}P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \frac{d}{dy}\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{d}{dy} 2 \int_{0}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-\frac{y}{2}} \qquad (y > 0) \end{aligned}$$



ゆえに

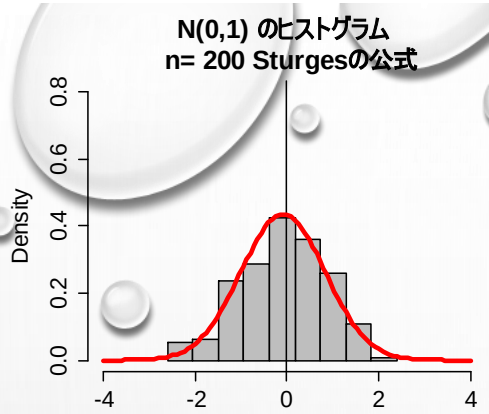
$$h(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, & \text{if } y > 0 \\ 0, & \text{if } y \le 0 \end{cases}$$

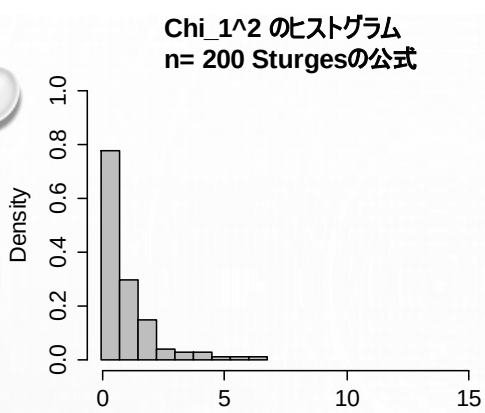
意外に出来ない（高校の数学）

$$\frac{d}{dy}\int_{-\infty}^{g(y)} f(x)dx = f(g(y))g'(y)$$

学生への課題

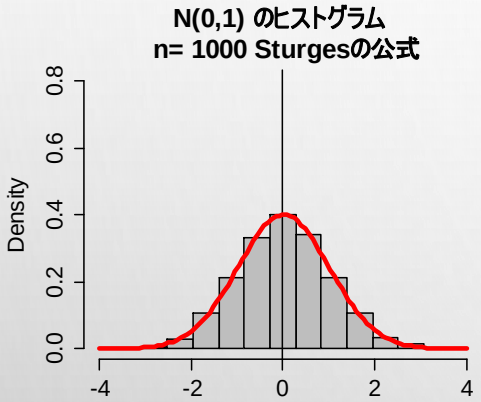
1. シミュレーションによってカイ二乗分布 X^2 を確認せよ
   - 標本サイズを色々と変えてみよ
   - e.g., n=200, 1000, 50000
2. 自由度1のカイ二乗分布の確率密度関数を描け
   - dchisq(x,1)

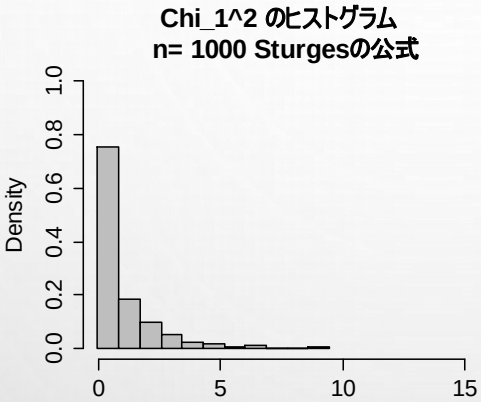




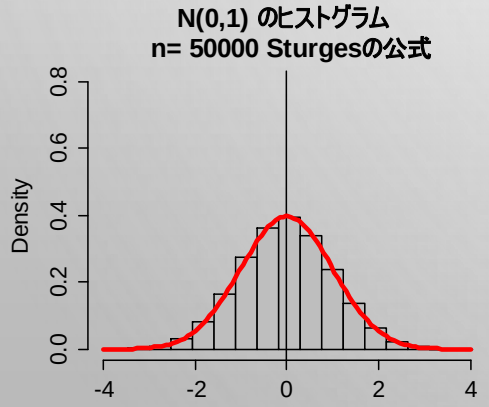
学生のクレーム

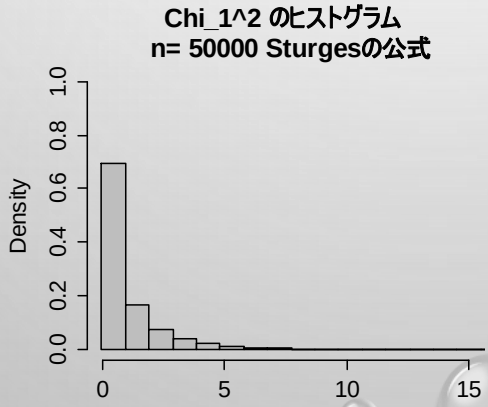
- nを大きくしてヒストグラムを描いてみましたが，天から降りてきません！





どやっ！





- Sturgesの公式

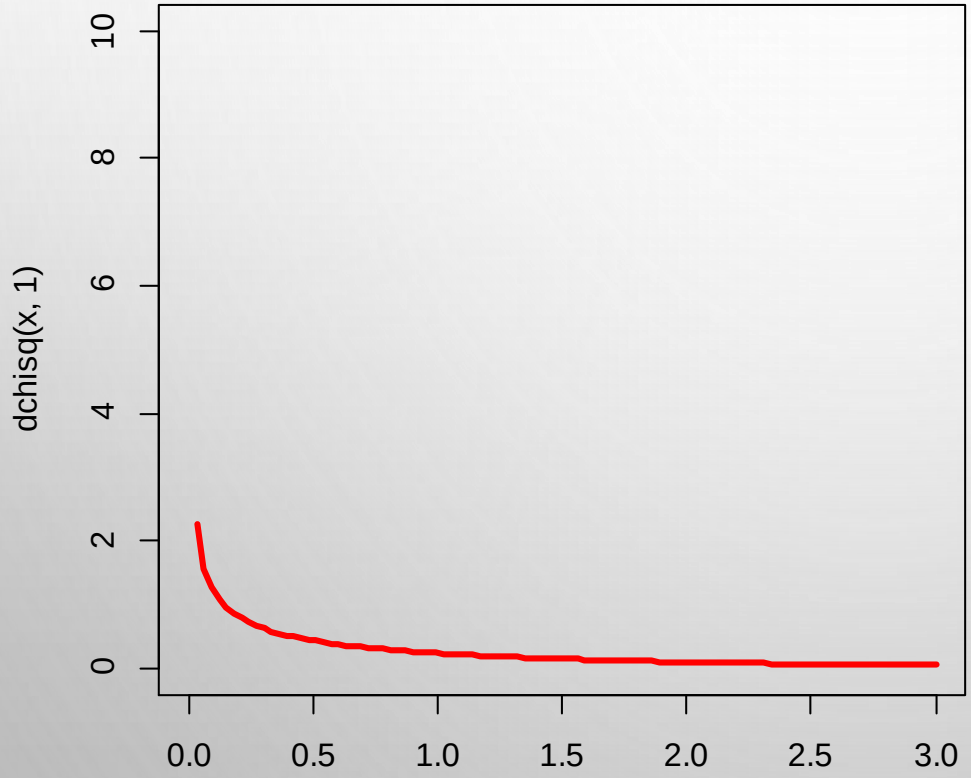
$$k_n = 1 + \log_2 n$$

$$h_n = (X_{max} - X_{min})/k_n$$

確率密度関数のグラフ

- 天から降りてきません！

どやっ！




```
curve(
    dchisq(x,1),
    from=0,to=3,
    col="red", lwd=3,
    ylim=c(0,10)
    )
```

極値統計学

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(0,1)$$

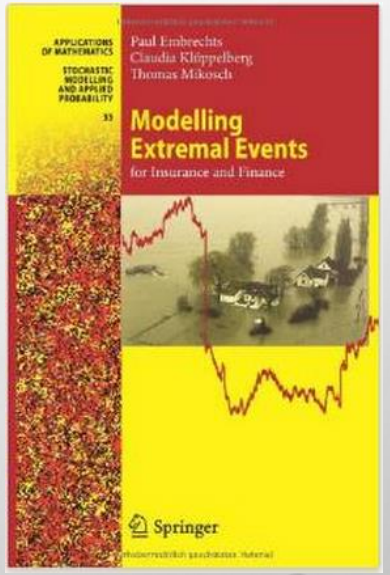
$$X_{max} = \max\{X_1, \ldots, X_n\} = O_p(\sqrt{\log n})$$

$$h_n = O_p\left(\frac{1}{\sqrt{\log n}}\right)$$

$$Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \chi_1^2$$

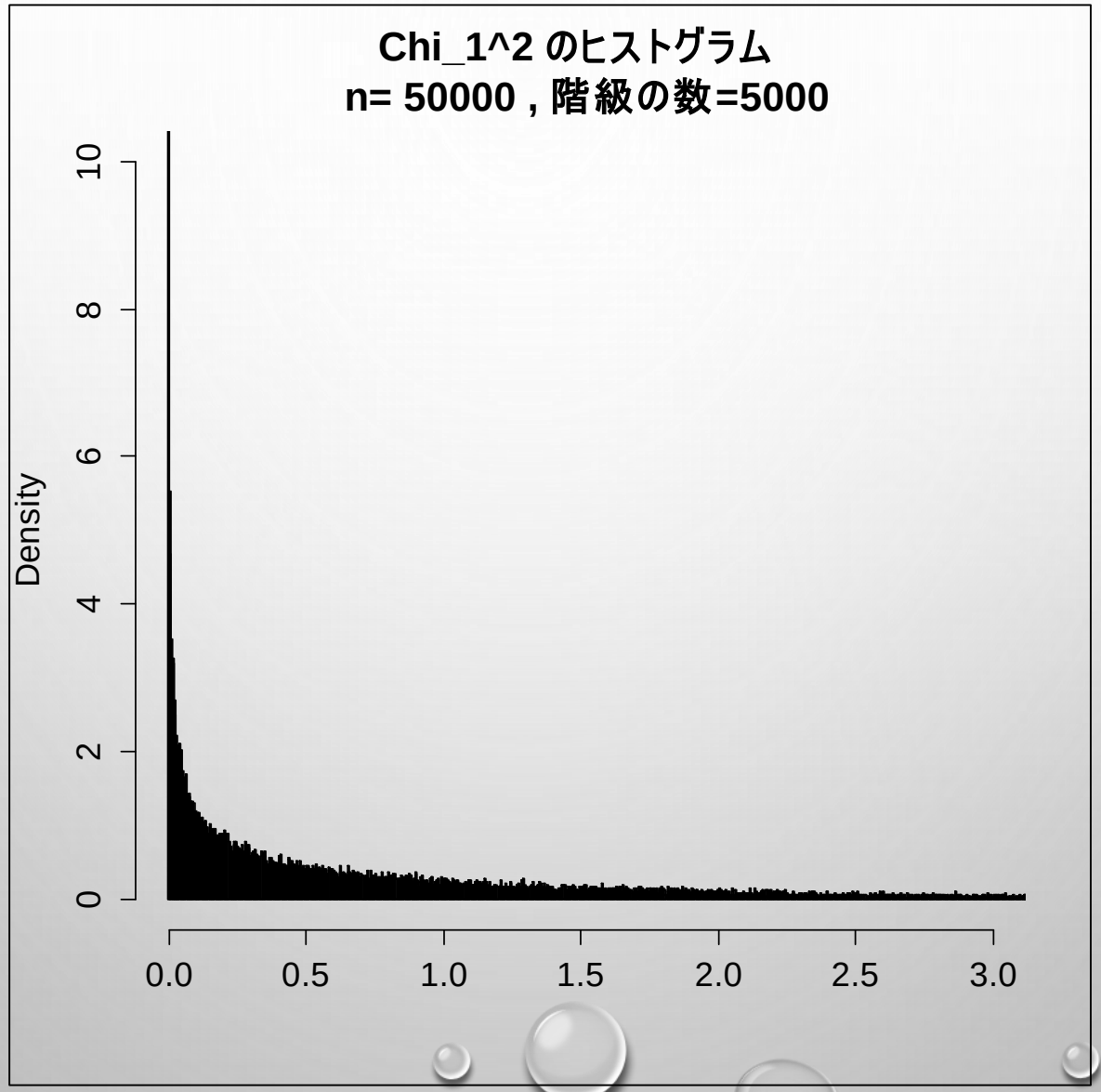
$$Y_{max} = \max\{Y_1, \ldots, Y_n\} = O_p(\log n)$$

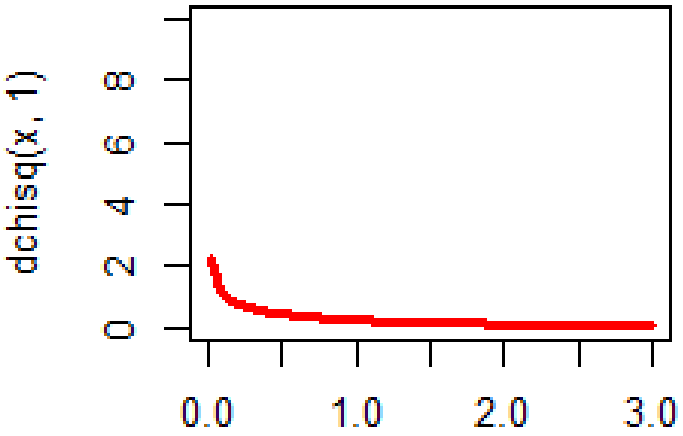
$$h_n = O_p(1)$$



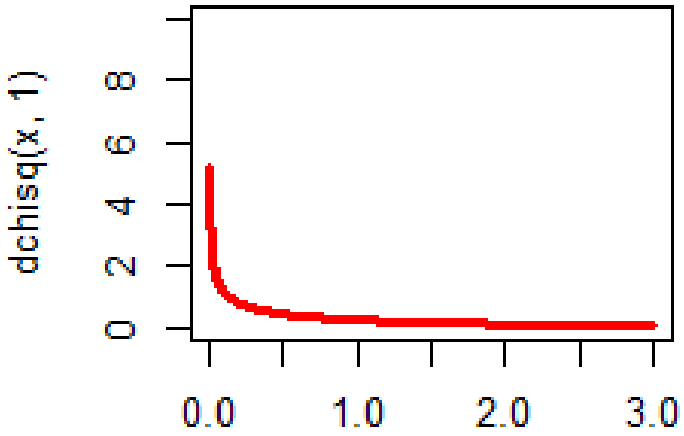
- Embrechts et al. (2011). Modelling Extremal Events for Insurance and Finance. Springer pp.548.

級の幅を小さくする


Chi_1^2 のヒストグラム
n= 50000 , 階級の数=5000
Density
0
2
4
6
8
10
0.0
0.5
1.0
1.5
2.0
2.5
3.0

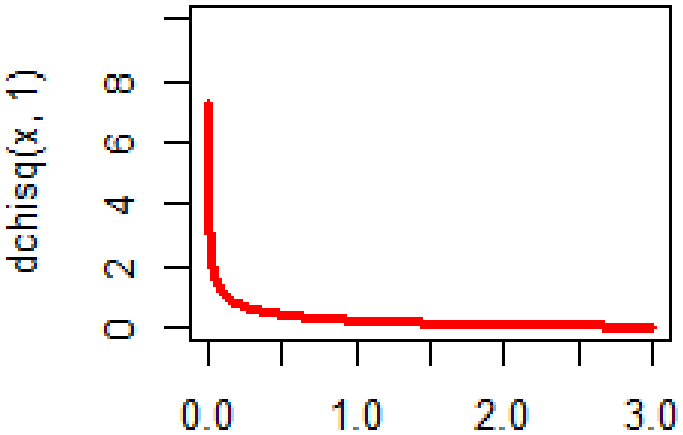


デフォルト

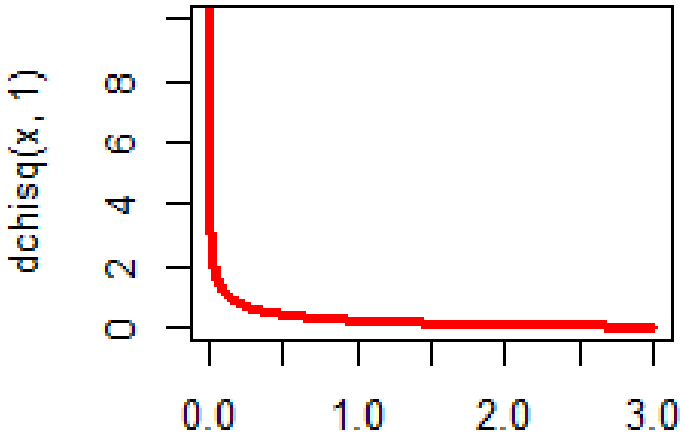


n=500

```
curve(dchisq(x,1),from=0,to=3,col="red",
      n=500, ylim=c(0,10),lwd=3,xlab="")
```



n=1000



n=5000

最後に

- データ科学の4要素
  1. 統計学の概念と手法
  2. 現象の理解
  3. 数学
     - 最大値の分布
     - 収束のオーダー
     - 関数の連続の一様性
  4. 情報技術
     - グラフ描画の基礎

- Four “I’s” by the Data Science Institute at Imperial College
  - Intelligence
  - Integration
  - Interdisciplinary
  - Interaction

まとめ2013

おとなの学問，統計学

もっと見よう数字とデータ

て”ーたとお友達になろう

なんと最強統計学

ぉしまい

本日のまとめ2015

とても大事な基礎統計

リテラシーとしての統計学

プロジェクトベースの課題解決

ルールを守って統計解析

スルーできない情報技術

リろんと現実つなぐ統計

いちおく総活躍「統計学」

日本統計学会会長
JINSE運営委員

内閣総理大臣

おわり

二人の子ども問題

封筒問題