

副プロ「金融・保険+データ科学」合同説明会
 日時：2015年4月10日(金)18:00~20:00
 会場：法経講義棟1階 1番講義室



大学院等高度副プログラム

<http://www.sigmath.es.osaka-u.ac.jp/~Estat/subprogram.html>

狩野 裕

<http://www.sigmath.es.osaka-u.ac.jp/~kano/index.html>

「データ科学」

本プログラムは文部科学省平成24年度採択大学間連携共同教育推進事業の取組「データに基づく課題解決型人材育成に資する統計教育質保証」の活動の一環として実施されます。



今、統計学が熱い！



大学院等高度副プログラム 「データ科学」



- 選べる5つのコース
 - 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
- 修了要件
 - 10単位(5科目)
 - 主専攻とダブルカウント可能
 - 上限6単位
 - 詳細は後ほど

データ科学の目的

- データ科学の基本的な考え方と統計手法の数理的基礎を理解する
- 主専攻の研究分野に直結する統計手法を体系的に学ぶ
- 主専攻でない分野におけるデータ科学を知り学際的な視点を養う
- 最新の統計手法に関する情報を得る
- データ科学の教育における課題を発見し教育方法の改善に資する



各コースにおける構成科目



統計数理コース

授業科目名	単位数		開講学期	開講部局
	選択必修	選択		
データ科学特論I	2		1	基礎工
データ科学特論II	2		1	基礎工
統計的推測	2	2	2	基礎工
多変量解析	2	2	2	基礎工
時系列解析	2	2	2	基礎工
確率解析	2	1	1	基礎工
確率微分方程式	2	2	2	基礎工
行動統計科学特論II	2	1	1	人間科
統計・情報数学概論	2	1	1	理
変 数理計量ファイナンス特別講義I	2	1	1	基礎工

機械学習コース

授業科目名	単位数		開講学期	開講部局
	選択必修	選択		
データ科学特論I	2		1	基礎工
データ科学特論II	2		1	基礎工
データマイニング工学	2		2	工
統計解析	2		1	基礎工
リスク評価論		2	1	工
統計モデリング	2	1	1	基礎工
データ解析	2	2	2	基礎工
数理特論II	2	1	1	基礎工
数理計量ファイナンス特別講義I	2	1	1	基礎工 変



各コースにおける構成科目



人文社会統計学コース

授業科目名	単位数		開講 学期	開講 部局
	選択 必修	選択		
データ科学特論Ⅰ	2		1	基礎工
データ科学特論Ⅱ	2		1	基礎工
行動統計科学特講Ⅰ	2		2	人間科
経験社会学特講	2		2	人間科
行動統計科学特講Ⅱ		2	1	人間科
計量社会学特講		2	1	人間科
教育動態学特講		2	2	人間科
多変量解析		2	2	基礎工
標本調査		2	2	経済
変 数理計量ファイナンス特別講義Ⅰ		2	1	基礎工

保健医療統計学コース

授業科目名	単位数		開講 学期	開講 部局
	選択 必修	選択		
データ科学特論Ⅰ	2		1	基礎工
データ科学特論Ⅱ	2		1	基礎工
保健情報論	2		1	医学系
臨床統計疫学特論 A	2		1	医学系 変
臨床試験デザイン基礎		2	2	医学系
観察研究の統計的方法		2	1	医学系
臨床統計疫学特論 B			通年	医学系 変
リスク評価論		2	1	工
行動統計科学特講Ⅰ		2	2	人間科
行動統計科学特講Ⅱ		2	1	人間科
数理計量ファイナンス特別講義Ⅰ		2	1	基礎工 変



5

各コースにおける構成科目



経済経営統計学コース

授業科目名	単位数		開講 学期	開講 部局
	選択 必修	選択		
データ科学特論Ⅰ	2		1	基礎工
データ科学特論Ⅱ	2		1	基礎工
エコノトリックスⅠ	2		1	経済
行動統計科学特講Ⅰ	2		2	人間科
統計解析		2	1	経済
エコノトリックスⅡ		2	2	経済
マーケティング・サイエンス		2	2	経済
標本調査		2	2	経済
多変量解析		2	2	基礎工
データ解析		2	2	基礎工
変 数理特論Ⅲ（意思決定とデータ科学）		2	2	基礎工
変 数理計量ファイナンス特別講義Ⅰ		2	1	基礎工



6

構成科目の訂正



- 副プログラム紹介冊子に変更あり
- **数理計量ファイナンス特別講義Ⅰ**
 - 英語によるデータ科学
 - Antony Hayter 教授 (University of Denver)
 - 6月と7月集中, 夕方
 - 全コース対象
- **数理特論Ⅲ（意思決定とデータ科学）**
 - 第2学期 水1限
 - 河本薫氏 (招へい教授, 大阪ガス)
 - 経済経営統計学コース対象
 - 上限40名 (副プロ登録者優先)
- **保健医療統計学コースの構成科目にも変更あり**



7

コース修了要件



- 選択必修科目と選択科目
 - 合わせて5科目10単位以上修得
- 選択必修科目
 - 3科目6単位以上修得
- 副プログラム修了の規程
 - 所属専攻の修了単位を4単位(以上)上回る
- コア科目
 - データ科学特論Ⅰとデータ科学特論Ⅱ
 - 隔年開講, 夏季集中



8

副プロを履修する意義と価値

- 学際融合教育のための履修プログラム
 - 主専攻に加えて、学際的・俯瞰的な視点や複眼的視野を養うための履修プログラム
 - 体系的
 - 主専攻の専門性を活かすための関連分野を学ぶ
 - 学際的
 - 学生が所属する主専攻の教育課程以外の内容を学ぶ
 - <http://www.osaka-u.ac.jp/ja/education/fukusenkou>
- 副プロは大阪大学の特徴的なプログラム
 - 充実した51プログラム
 - 「大学院副専攻プログラム」 5プログラム
 - 「大学院等高度副プログラム」 46プログラム
 - 「データ科学」「金融・保険」は「**真の**」文理融合型プログラム
 - 数学・統計学＋応用
- 就職活動
 - 差別化, 特異性, 積極性, 幅広い知識

10



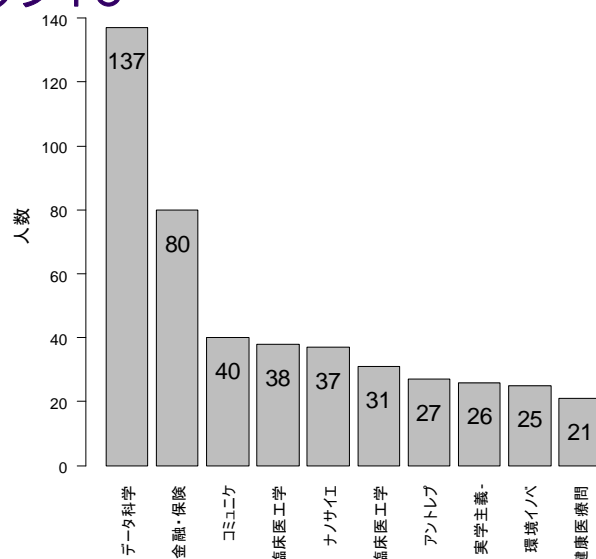
副プロ「データ科学」を履修する意義と価値

- データ科学
 - 研究・開発・実務のため重要な方法論
 - データリテラシーの獲得
 - 体系的に学習可能
 - 日本で唯一
 - 他分野のデータ解析を学ぶことができる

11



H26年度副プロ受講申請者数トップ10



12



H26年度「データ科学」申請者数内訳

研究科別申込人数

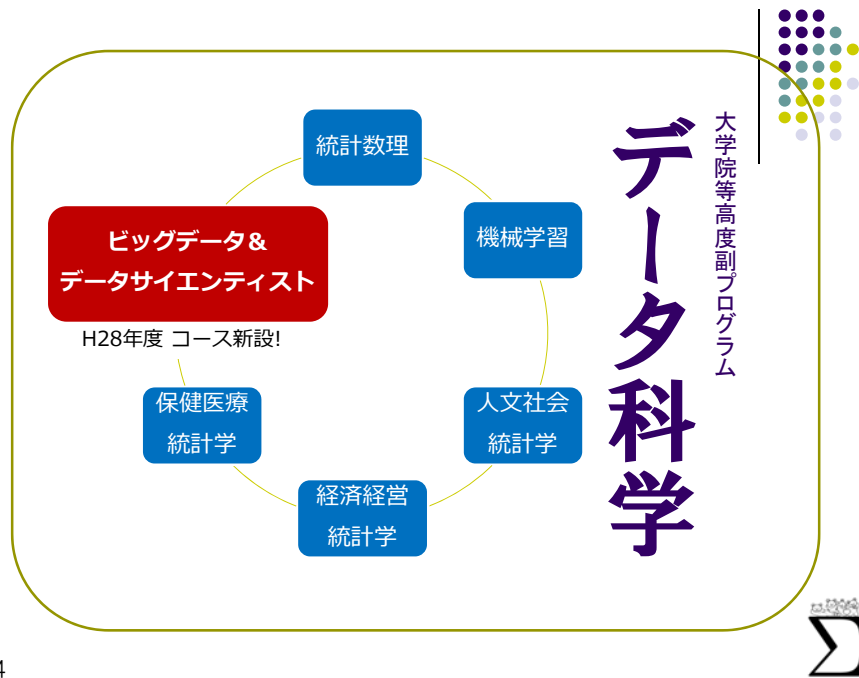
人間科学研究科	14人
経済学研究科	4人
理学研究科	5人
医学系研究科	11人
工学研究科	1人
基礎工学研究科	40人
国際公共政策研究科	2人
生命機能研究科	3人
情報科学研究科	2人
計	82人

コース別申込人数

統計数理コース	50人
機械学習コース	26人
人文社会統計学コース	17人
保健医療統計学コース	22人
経済経営統計学コース	22人
計	137人

13





ビッグデータ&データサイエンティストコース

- 平成28年度よりビッグデータ&データサイエンティストコースを新設します。それに先立ち同コースの**コア科目を平成27年度に先行開講**します。
- 平成27年度取得の単位は平成28年度に新コース開講の際には修了単位として認定されます。
- 数理特論IIIについての受講者は演習等の関係から最大40人と制限されています。受講希望者が上限を超えた場合は基本的には抽選ですが、本副プログラムの登録者は優先して受講できます。
- コースWebページ <http://osku.jp/w082>



Two Big Professors



鬼塚 真 教授(H26/7着任)

- 情報科学研究科
ビッグデータ工学講座
- 元NTTソフトウェアイノベーションセンター主幹研究員
- 専門
 - 知識発見のための分散データ処理, ソーシャルネットワークに関する研究開発
- 講義科目(第1学期)
 - ビッグデータ工学
 - ビッグデータ解析

河本 薫 招聘教授(H27/4着任)

- 大阪ガス(株) ビジネスアナリシスセンター所長
- 工学博士, 経済学博士
- データサイエンティスト・オブ・ザ・イヤー賞. 日経BP社, 2013
- 著作
 - 会社を変える分析の力
 - 真実を見抜く分析力: ビジネスエリートは知っているデータ活用の基礎知識 2014
- 講義科目
 - 数理特論III(意思決定とデータ科学)
 - 上限40名(副プロ登録者優先)
 - 第2学期: 月5限 → 水1限

構成科目(案)

時間割コード	授業科目名	単位数		開講学期	開講部局	備考
		選択必修	選択			
290724	データ科学特論Ⅰ	2		1	基礎工学研究科	奇数年度開講
290729	データ科学特論Ⅱ	2		1	基礎工学研究科	偶数年度開講
新規	ビッグデータ工学	2		1	情報科学研究科	奇数年度開講
新規	ビッグデータ解析	2		1	情報科学研究科	偶数年度開講
新規	意思決定とデータ科学(数理特論Ⅲ)	2		2	新センター	
280834	データマイニング工学		2	2	工学研究科	
290617	Data Science and Case StudiesⅠ(数理計量ファイナンス特別講義Ⅰ)		2	1	新センター	集中(6月,7月)



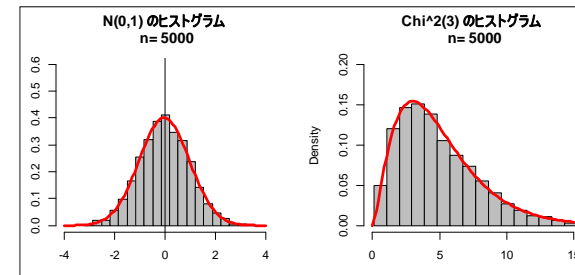
- 分布論の基礎
 - 初級者にどう教える?
 - 統計教育
- シンプソンのパラドックス
- 統計的予測

「データ科学」三題

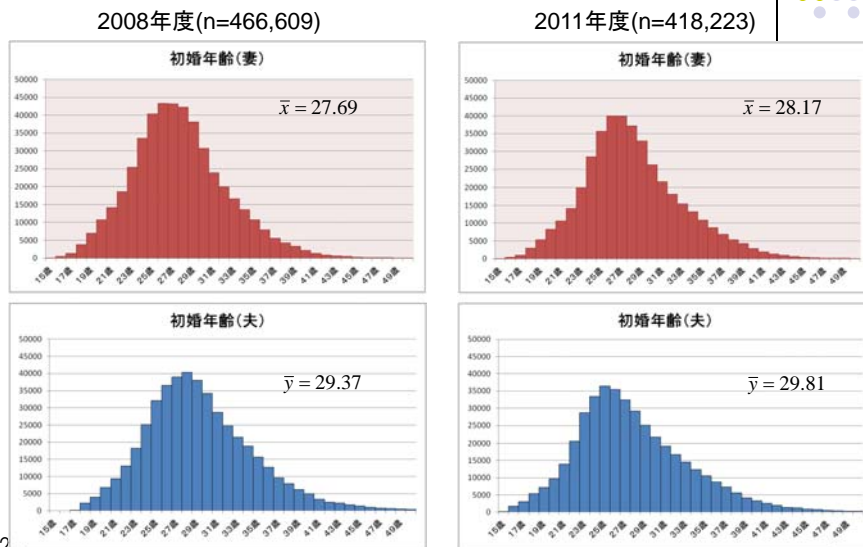


分布論の基礎

- 近年、統計学で標本分布論はあまり教えない
 - その昔は主要テーマ
 - 概念と基本的なスキルは身に付けておきたい
 - PCが手軽に使えるようになりシミュレーションが容易になった
- 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

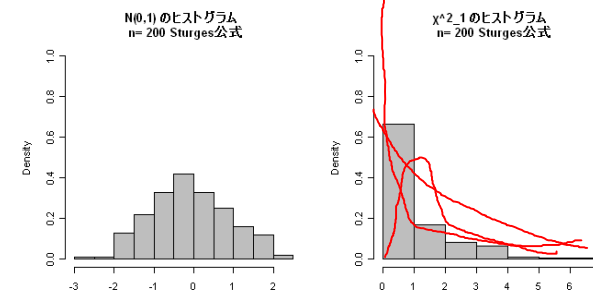


確率的現象の「安定性」 nが大きいとき、確率的な現象の分布は安定する



カイ2乗分布(df=1)

- $x=0$ における密度の値を予想



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



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

$$\begin{aligned} h(y) &= \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} P(\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \frac{d}{dy} \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{d}{dy} 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-\frac{y}{2}} \quad (y > 0) \end{aligned}$$

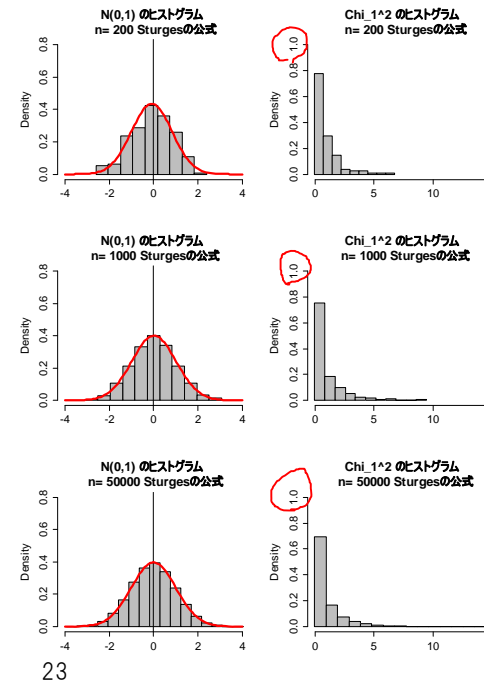
ゆえに

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

意外に出来ない (高校の数学)

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

22



学生の質問

- 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!



- Sturgesの公式

$$k_n = 1 + \log_2 n$$

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



極値統計学

$$X \sim N(0, 1)$$

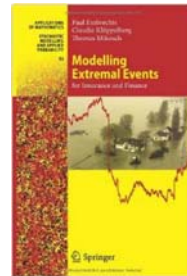
$$X_{max} = \max\{X_1, \dots, X_n\} = O_p(\sqrt{\log n})$$

$$h_n = O_p\left(\frac{1}{\sqrt{\log n}}\right) \xrightarrow{P} 0$$

$$X \sim \chi_1^2$$

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

$$h_n = O_p(1)$$



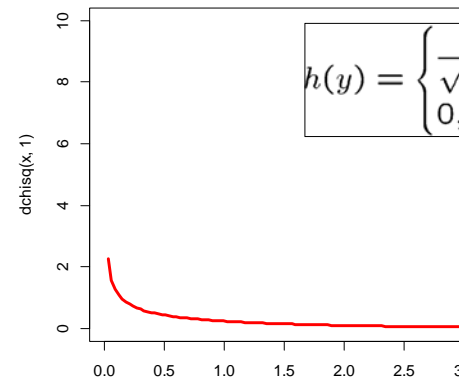
- Embrechts et al. (2011). Modelling Extremal Events for Insurance and Finance. Springer pp.548.

24



再度学生からの質問！ Rによるpdfのグラフ

- (再び)天から降りてきません!

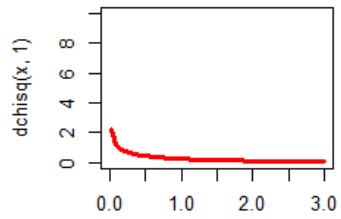


$$h(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, & \text{if } y > 0 \\ 0, & \text{if } y \leq 0 \end{cases}$$

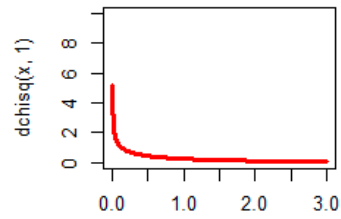
```
curve(
  dchisq(x,1),
  from=0,to=3,
  col="red", lwd=3,
  ylim=c(0,10)
)
```

25



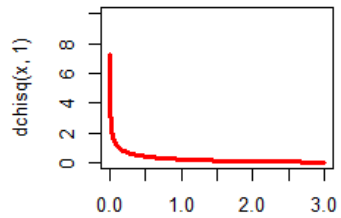


デフォルト

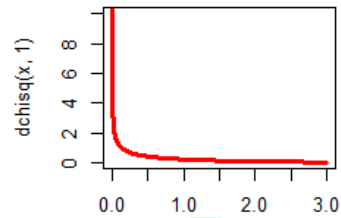


n=500

`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
 - 統計学の概念と手法
 - 数学
 - 情報技術
 - 現象の理解
- Data Science
 - Intelligence
 - Integration
 - Interdisciplinary
 - Interaction



THANK YOU FOR YOUR
ATTENTION

