

統計学入門

～ランダムなデータと向き合う～

宮部賢志 明治大学数学科

2017年3月24日 中大連携講座

統計とは

- ❖ 一般的な統計のイメージ
- ❖ 科学における統計
- ❖ 偏りのあるサイコロを見分ける
- ❖ 仮説検定
- ❖ χ^2 分布

ベイズ統計

確率とゲーム

統計とは

一般的な統計のイメージ

- 日本の人口は1億2676万人（2017年3月1日時点）である。
- 年間の平均給与は420万円（2015年）である。
- 消費動向，食料26%，住居21%，光熱費8%，家具3%，服4%，医療4%，交通15%，教育3%，教養娯楽10%，雑費6%

すべて統計局より。

数字を使ってはいるけれど，数学と何か関係あるの？

科学における統計

データに意味を与える。

- 薬 A は薬 B よりもよく効くと言えるか？
- このサイコロは公平か？
- 国立大学より私立大学の方が就職しやすい，
は本当か？

ただの数字に意味を与えるのが統計学であり，高度な数学が使われる。

偏りのあるサイコロを見分ける

- 3が多く出過ぎている ⇒ 「多く」とはどのくらいか？
- 1から6まで万遍なく出ている ⇒ どの程度までの誤差を許してよいのか？

新しいことを主張する時には、「なんとなく」「私はこう思う」では科学にならない。相手が反論できないように説明する必要がある。

仮説検定

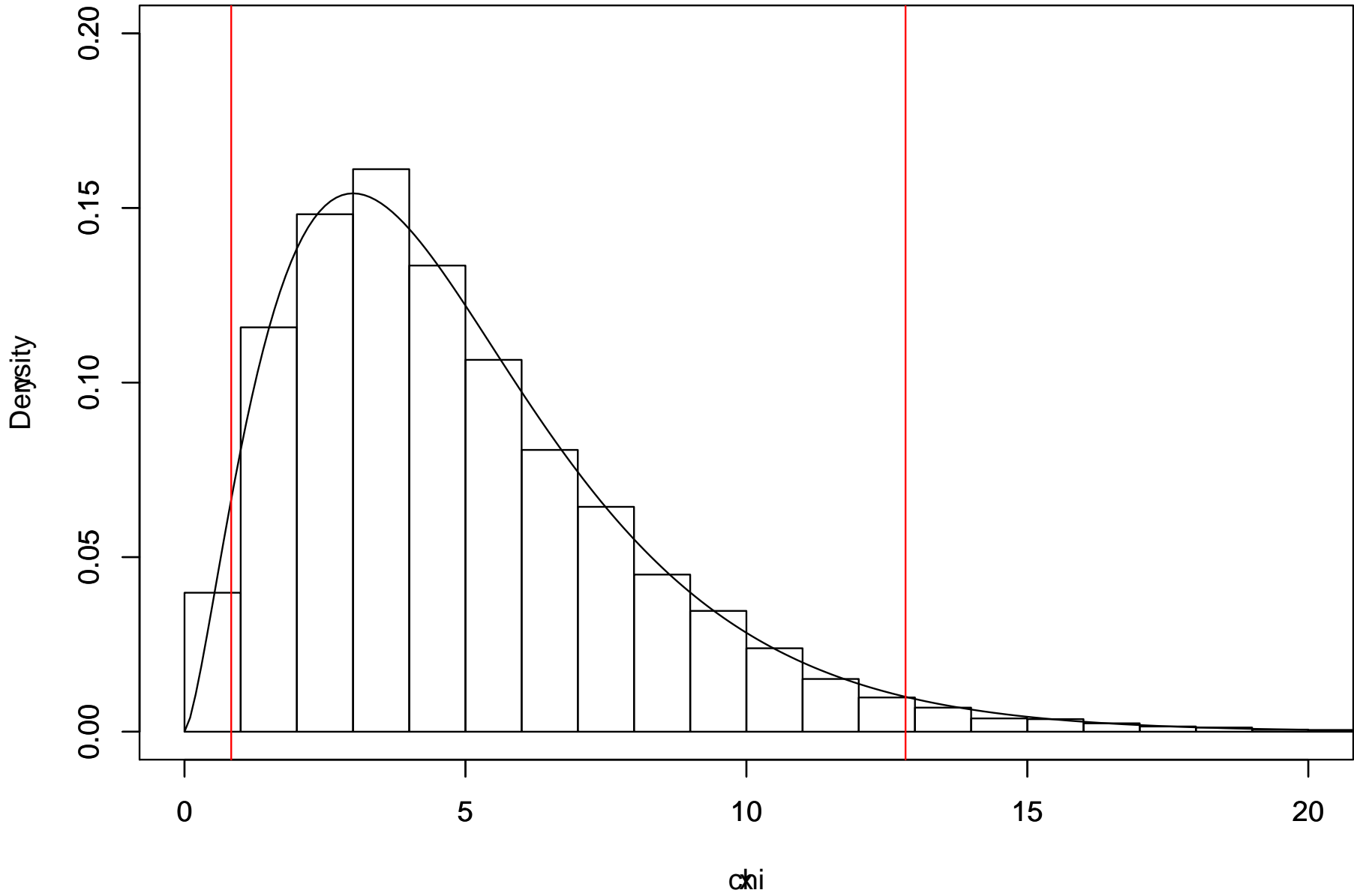
- 仮にサイコロが公平だったとしよう.
- 120回投げると,

$$\chi^2 = \sum_i^6 \frac{(x_i - 20)^2}{20}$$

の値は次のような分布をする.

- もし赤線の外側であれば, 仮定が間違っていると考えるほうが自然.
- もし赤線の内側であれば, 仮定に不自然さは見当たらない.
- 0.8312116 ~ 12.8325

Histogram of chi



統計とは

ベイズ統計

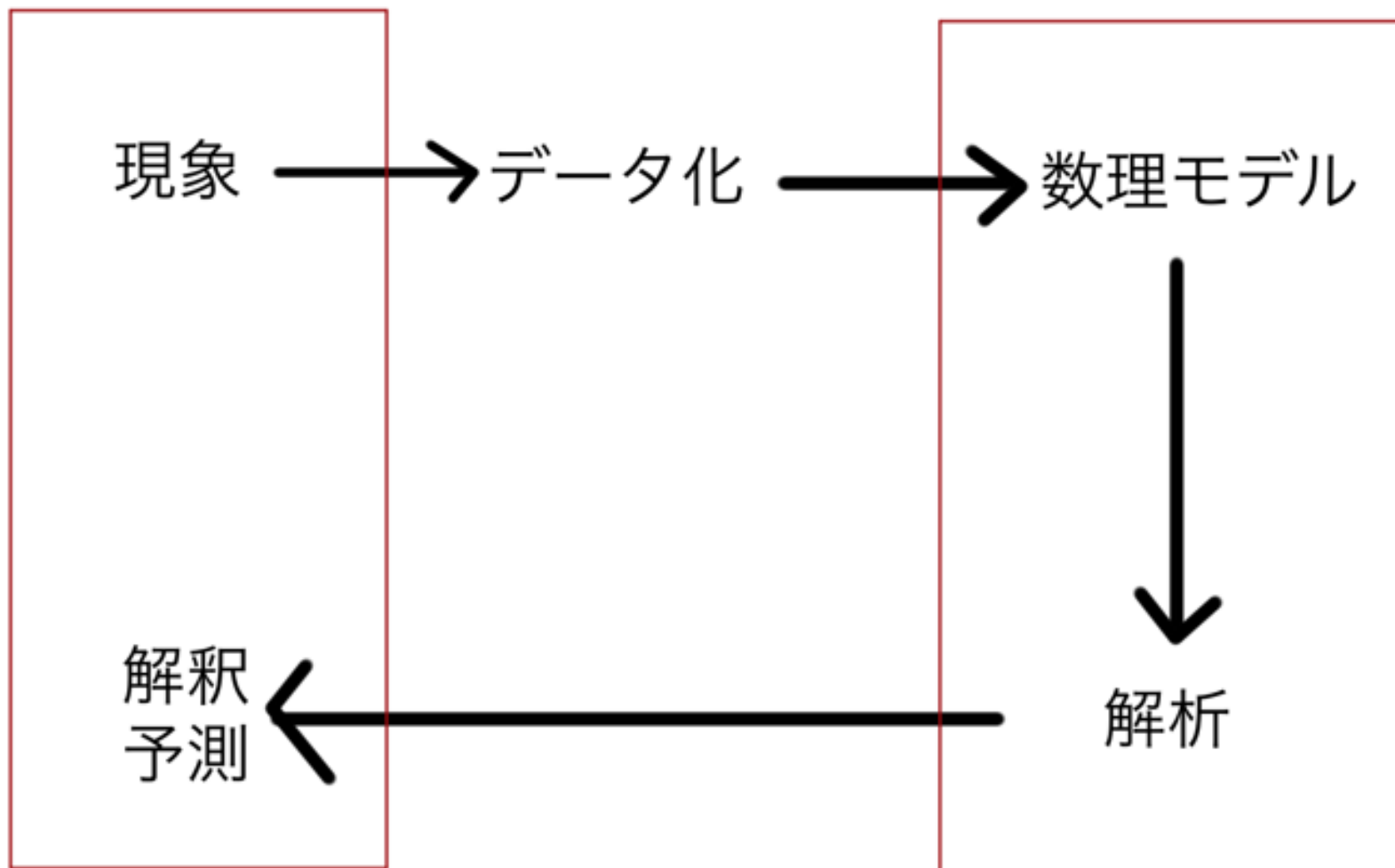
- ❖ 数理モデル
- ❖ 統計モデル
- ❖ 01 列の予測
- ❖ モデル化

確率とゲーム

ベイズ統計

現実世界

数学の世界



統計モデル

- 死亡数を予測する ⇒ 生命保険, 年金
- 株価を予測する ⇒ 数理ファイナンス
- 天気予報
- 生物の個体数を予測する
- OCR(光学的文字認識)
- 写真から花の種類を予測する
- 音楽を種類ごとに分類する
- 車の自動運転
- amazon でのお勧め商品

01列の予測

- 次の値が0か1かを当てるゲームをする
- ただし、皆さんはカジノ側で、当てられないようにランダムに0と1を書く
- ランダム=儲けられない

モデル化

- 000 ~ 111 までの 8 種類の頻度を調べる
- 個数の偏り = 個人の癖
- 前の 2 桁を見て次の桁をベイズの公式で予測する
- Kelly の公式

$$b = 2 \times p - 1$$

により, 掛け金を定める

統計とは

ベイズ統計

確率とゲーム

- ❖ 賭けゲーム
- ❖ ゲームと確率
- ❖ 終わり

確率とゲーム

賭けゲーム

- 今度は2人で対戦する。
- 11回のコインで6回以上表を出す必要がある。
- 予測する人はできるだけ儲けられるように，コインを決める人はできるだけ資金を抑えられるように動く。
- 最終的な資金はいくらになるか？

ゲームと確率

定理 1. このゲームで要求される事象の確率が p ならば、資金を $1/p$ 倍するような戦略が存在する。また、これが最適である。

情報を上手く使えますか？

問題 2. このゲームにおいて、表が6回以上出るという情報には、いくら価値があるだろうか？
このゲームの参加費はいくらが適当だろうか？

終わり

終わり

