

授業名	システム数理学基礎論	第11回
日時	12月06日（金）	2 限
担当	大和淳司	
連絡先	yamato@cc.kogakuin.ac.jp	
出席	提出物で代替	
資料配布	CoursePowerからのリンク（GoogleDrive）	
課題	CoursePowerで提出	12/11 20時
質問受付	CoursePower（質問登録）	
音声有無		
その他		

システム数学基礎論#11

前回課題から：ハッシュ表

- 素数を使う理由
通常の元データに偏りが生じていることが多い
 - アドレス = 偶数、8/16 の倍数 など
 - 学籍番号 十進数の各桁に意味をもたせた結果の偏り
 - スケジュール 曜日に依存する偏り
 - 金額 切りの良い数字、桁上り手前の数字 等
 - 偏り（特に周期的偏り）と一致した除数の使用は極度に集中し、ハッシュ衝突を引き起こす⇒素数が安全
- ランダム（一様乱数）な元データに対しては素数を用いる必要がない

- Python
 - モンテカルロ・シミュレーション（第16章）を実施
- ~~課題 1~~：
~~Crapsのゲームにおける良い戦略を考えて、機能するかどうかをシミュレーションする。例：勝ち逃げ戦略~~
- 課題 2：~~課題 1~~ とする
モンティ・ホール問題の正解はどちらかを、モンテカルロ・シミュレーションで判別せよ。（詳細別紙）Monty.py を参考に

- 決定論的世界
 - ニュートン力学
 - 計算
- 確率論的世界
 - 量子力学
 - サイコロ
- 情報不足なのか、本質的に非決定論なのか
 - 因果的非決定論 \Leftrightarrow 予言的非決定論
- 近似としての確率・統計

サイコロ dice

- コード 15.1

```
import random  
random.choice([1,2,3,4,5,6])
```
- n 個のサイコロの和

コインの表裏(Head or Tail)

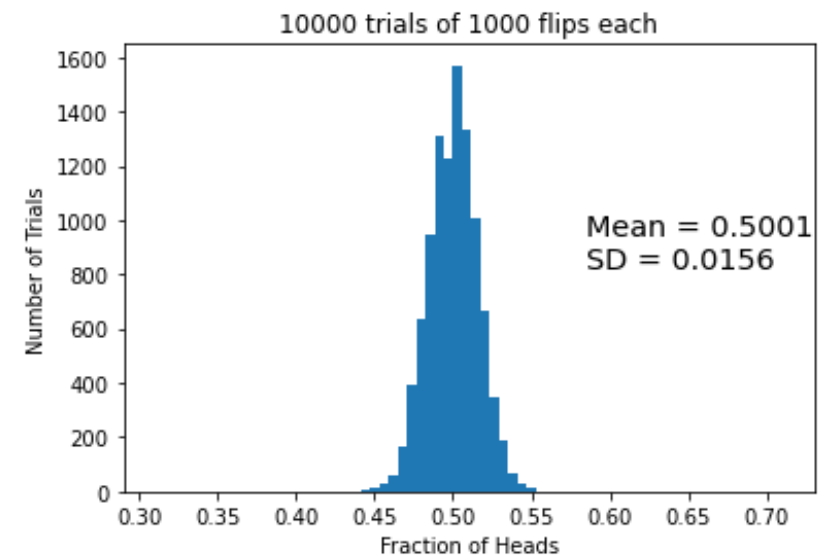
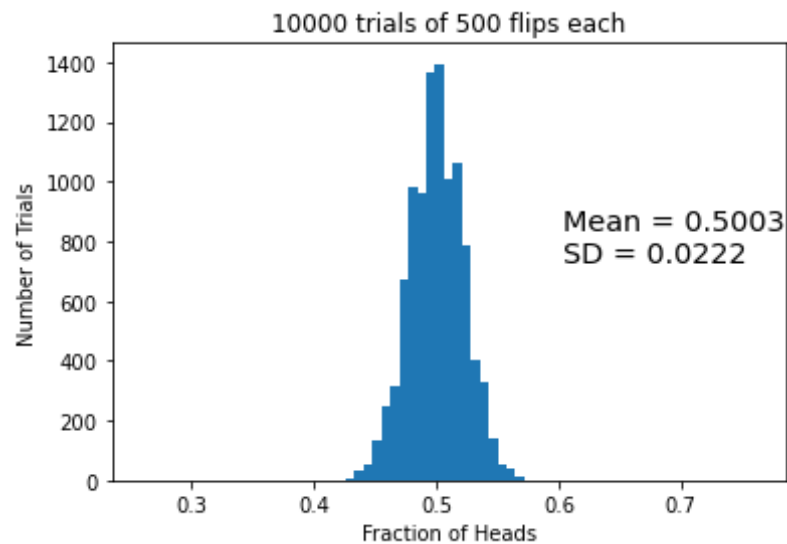
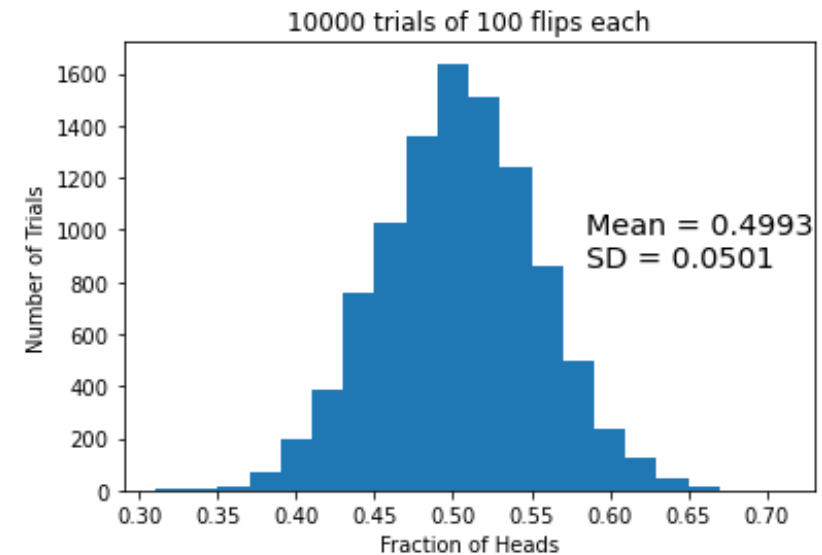
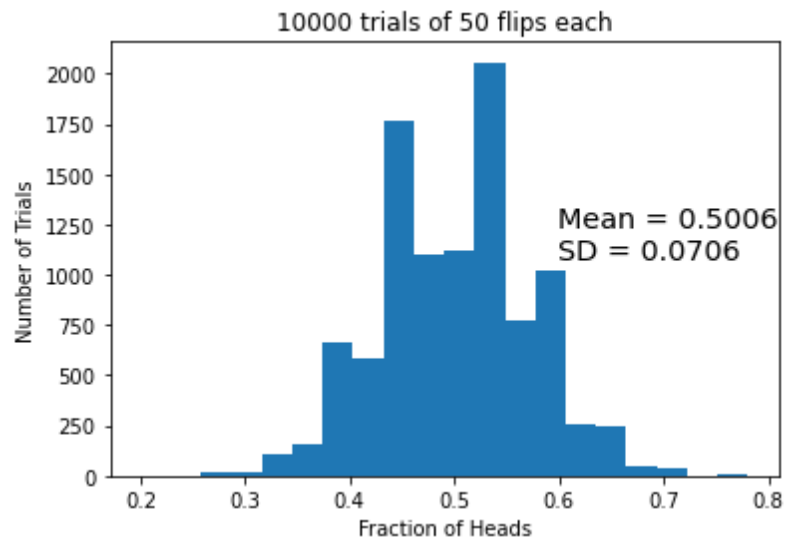
- `random.choice(['H', 'T'])`

10回では？

10000回では？

- 大数の法則 law of large number
- 平均への回帰 regression to the mean
- 賭博者の誤謬 gambler's fallacy

- Fig. 15.9 実行例



モンテカルロ・シミュレーション

- 確率の計算を、
「実際にやってみる（計算機上で）」ことで置き換える
- サイコロで1が出るのは $1/6$
600回サイコロを振って100になるかどうか
- ゲームの勝率 先手後手どちらが有利？
 ルール変更の影響は？
- シミュレーションの一部
(流体計算などは決定論的or予言的非決定論)

通説では、のちに確率論とよばれるこの分野にパスカルが興味を持ったきっかけは、友人からある質問をされたことだといわれている。それは、2つのサイコロを24回振るうちに、両方とも6が出ることに賭けることで果たして儲かるのか、という質問であった。これは、17世紀半ばには難しい問題と考えられていた。パ

- コード 16.1
パスカルの計算は正しかったのか？
- もし五分以上のルールにしたい場合は、24回ではなく何回以上振るに設定すればよいか？
- `checkPascal()` のコードを修正して試してみよう

試行回数はどう影響するか？

- CrapsGame

- ルールが複雑
⇒場合分けでの確率計算が難しい
- モンテカルロシミュレーション
⇒場合分けの数をそれなりに網羅する試行回数が必要
計算複雑性の問題が表面化する

- crapsSim(handsPerGame, numGames)

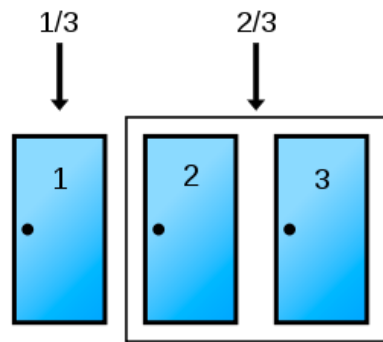
- GameのROI（収益率）を安定させるには？
- 勝つための戦略はありえるのか？
（なるべく負けない戦略）

Crapsで勝つ戦略はあるのか？

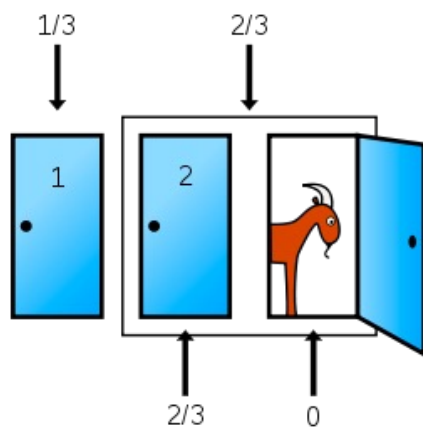
- 戦略の例：
 - 変動が大きい設定にして、勝ったらそこでやめる
 - 一回の掛け金を変動制とする。何度か負けたら次に勝つと信じて掛け金を上げる。
 - 他にも各人でアイデアを出してみる
(複数人で相談も可)

モンティ・ホール問題

「プレイヤーの前に閉まった3つのドアがあって、1つのドアの後ろには景品の新車が、2つのドアの後ろには、はずれを意味するヤギがいる。プレイヤーは新車のドアを当てると新車がもらえる。プレイヤーが1つのドアを選択した後、司会のモンティが残りのドアのうちヤギがいるドアを開けてヤギを見せる。



ここでプレイヤーは、最初に選んだドアを、残っている開けられていないドアに変更してもよいと言われる。**プレイヤーはドアを変更すべきだろうか？」**



ルールは簡単だが、正しいモデル化が難しい（確信がもてない）
→シミュレーションの方が楽（確実）

定式化の例

```
def NoChange():  
    CARis = random.choice(["A","B","C"]) #正解をランダムに選んでおく  
    PlayerChoice = random.choice(["A","B","C"]) #選択したのはこれ  
    if PlayerChoice == CARis:           #正解 = 選択 一致してたら  
        return "new_car"  
    else:  
        return "goat"
```

もちろん、上記と違う実装でも良い（むしろ推奨）

第2版の印刷ミス “(“不足

```
wins, losses, pushes = g.dpResults()
dpROIPerGame.append((wins - losses)/float(handsPerGame))

#統計値の概要を求めて表示する
meanROI = str(round(100*sum(pROIPerGame)/numGames), 4)) + '%'
sigma = str(round(100*stdDev(pROIPerGame), 4)) + '%'
print('Pass:', 'Mean ROI =', meanROI, 'Std. Dev. =', sigma)
meanROI = str(round((100*sum(dpROIPerGame)/numGames), 4)) + '%'
sigma = str(round(100*stdDev(dpROIPerGame), 4)) + '%'
print('Don\'t pass:', 'Mean ROI =', meanROI, 'Std Dev =', sigma)
```

コード 16.3 クラップスをシミュレート

修正された刷もあるかもしれません。各自確認してください。

第11回の授業と課題part2

- テキスト
17章～
- 課題2：お見合い問題（別紙）の解をモンテカルロ法で求めよ。
- 課題3：モンティ・ホール問題バリエーション（別紙）の実験と考察
上記2点ともに、レポートの形式を守ること。

提出先：CoursePower

提出期限：

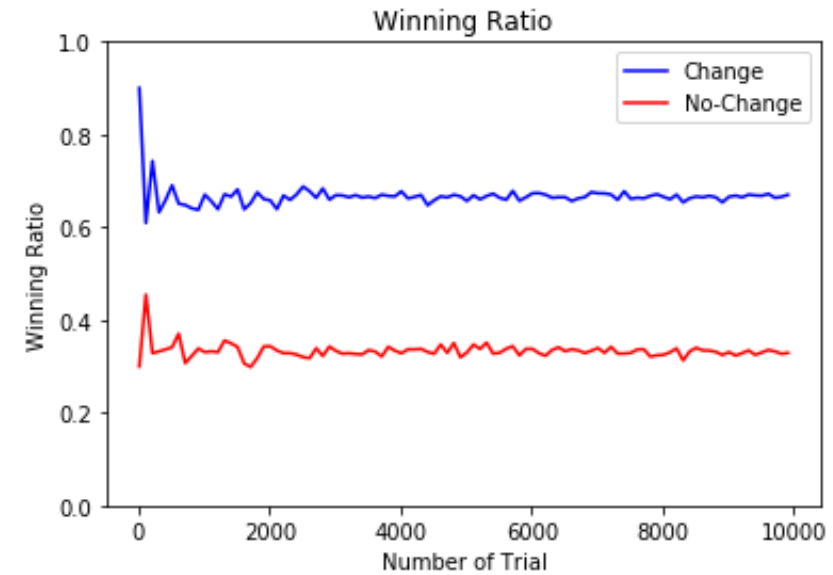
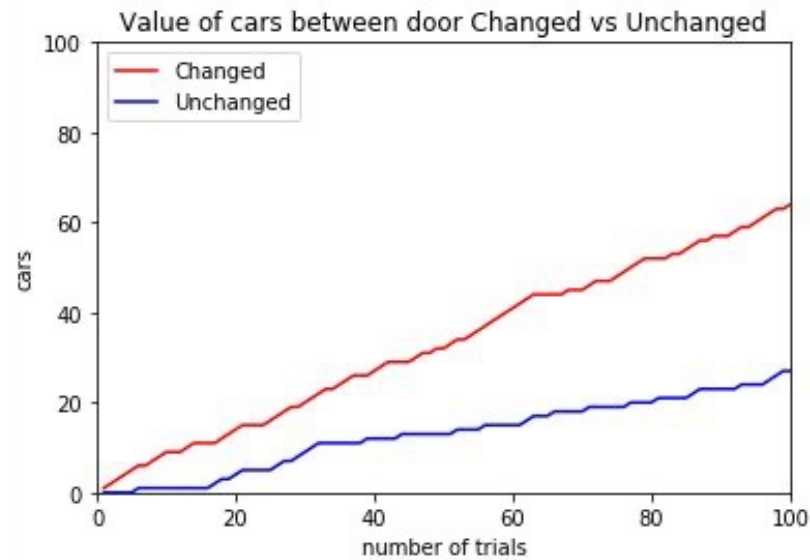
2024/12/11, 20:00 JST

なお、過去レポートの再提出を受け付けますが再提出は提出先は別です。気をつけてください。

調べる、人に聞くなどしてとにかく実施することが大事です。

ChatGPT等の利用も規程(Google Driveで資料とともに配布済)に従えば可です

Mondy課題について



直感的な説明

- 最初の選択がハズレだったら
→ 変えれば当たる
- 最初の選択が当たりだったら
→ 変えれば外れる

2/3の確率だった

1/3の確率だった

直感的な説明 2

- ドアがたくさんある場合で考える
- ドアが100枚
- 正解 1、不正解99枚
- 司会者は、選択されたドアと 1 枚 を除く 98枚の
ドアを開ける（すべて不正解）
- ここで、変更するのとししないのはどちらが有利？

複雑化したバリエーション

- 男『ここに A B C D 4 枚のカードがあります。』
男『4 枚のうち 1 枚が当たりです。』
男『私はどれが当たりか知っています。』
男『さあ、好きなの 1 枚選んで。』
女『じゃあ A』
男『では、貴方の選ばなかった B C D のうち D はハズレであることを教えよう。』
（D をめくる。確かにハズレだった。）
男『もう一度残った A B C の 3 枚から選び直していいよ。変えてみる？』
女『（モンティホールの応用だから変えたほうが若干得そうね）じゃあ B。』
男『選ばなかった A C のうち C もハズレなことを教えよう。』
（C をめくる。確かにハズレだった。）
男『ラストチャンス。A B どっち？』
女『...(やっぱり A に戻したくなってきたw) 』
- 女は A に変えるべきだろうか？

複雑化したバリエーション

- 男『ここに A B C D 4 枚のカードがあります。』
男『4 枚のうち 1 枚が当たりです。』
男『私はどれが当たりか知っています。』
男『さあ、好きなの 1 枚選んで。』
女『じゃあ A』
男『では、貴方の選ばなかった B C D のうち D はハズレであることを教えよう。』
(D をめくる。確かにハズレだった。)
男『もう一度残った A B C の 3 枚から選び直していいよ。変えてみる?』
女『(モンティホールの応用だから変えたほうが若干得そうね) じゃあ B。』
男『選ばなかった A C のうち C もハズレなことを教えよう。』
(C をめくる。確かにハズレだった。)
男『ラストチャンス。A B どっち?』
女『...(やっぱり A に戻したくなってきたw) 』
- 女は A に変えるべきだろうか？

Q1. これは正しいか？

Q2. これは正しいか？

要検討点

「変更するのが有利」と「最初の A に戻す」は矛盾してないか？

- 確率の問題として考えてみよう
 - シミュレーションで確認してみよう
(課題3とします)
- 参考用にMonty.pyを配布します。

標本抽出

- 17章で使う `bm_results2012.txt` を配布します。
- 2万件程度なので、標本する必要は（今どきは）ない
- サンプルングと全数での統計の比較を試みる

sampleTimes(times, 40)

- sampleSizeの影響を確認しよう
- sampleSize=40 図17.2と比較
40から変更する。大きく、小さく

ヒストグラムの形は？平均、分散は？
複数回実行したときの変化に着目 安定性？

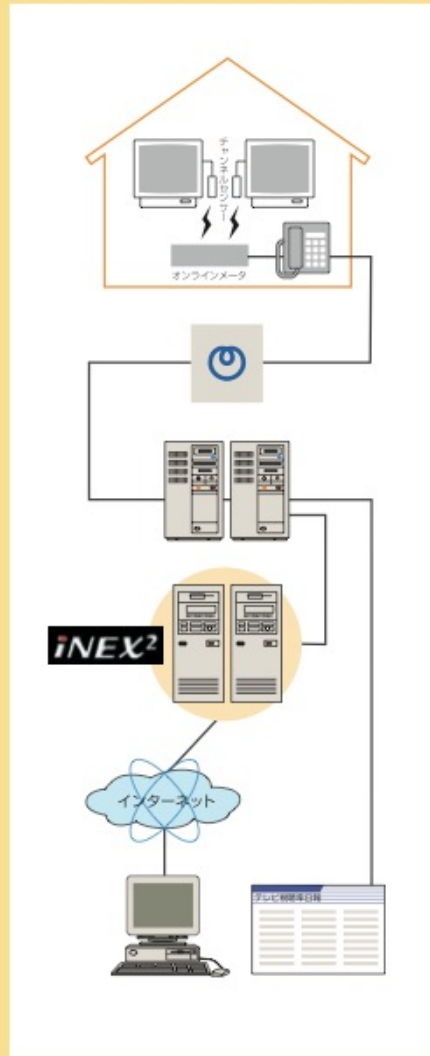
- P.271, コード17.5

`testSamples(1000, 40)`

母集団の平均を知りたい。40サンプルで平均値を求めた。
これはどの程度安定しているのか？
何度もやって分布を調べよう。(1000回！)

- SD1からは少数サンプルでも安定
- SD100からは少数サンプルでは？
- `testSamples(10,40)`では？
 - 100, 40 では？
 - 10, 400 では？
- 試行回数 vs. サンプルサイズ のトレードオフ
- 中心極限定理を確認しよう

オンラインメータシステムによる世帯視聴率調査



STEP-1. 調査対象世帯内での視聴状況測定

家庭内の最大3台までのテレビにそれぞれ接続されたチャンネルセンサーから、オンラインメータに無配線でデータが転送され、1日の視聴状況が記録されます。

STEP-2. データ送信

記録されたデータは、データ通信回線を利用して毎日、早朝に自動ダイヤルによりコンピュータセンターに送信されます。

STEP-3. 集計

世帯視聴率の最小単位は1分。それら毎分視聴率をもとに番組視聴率や時間区分視聴率を集計します。

STEP-4. 視聴率データ提供サービス

ビデオリサーチのWebによるデータサービスシステムであるiNEX²により、契約企業やテレビ局、広告会社などに視聴率データを提供しています。

STEP-5. 印刷

世帯視聴率データは、翌営業日に視聴率日報として印刷し、発行しています。

機械式世帯視聴率調査地区

■PMシステム調査地区(3地区)

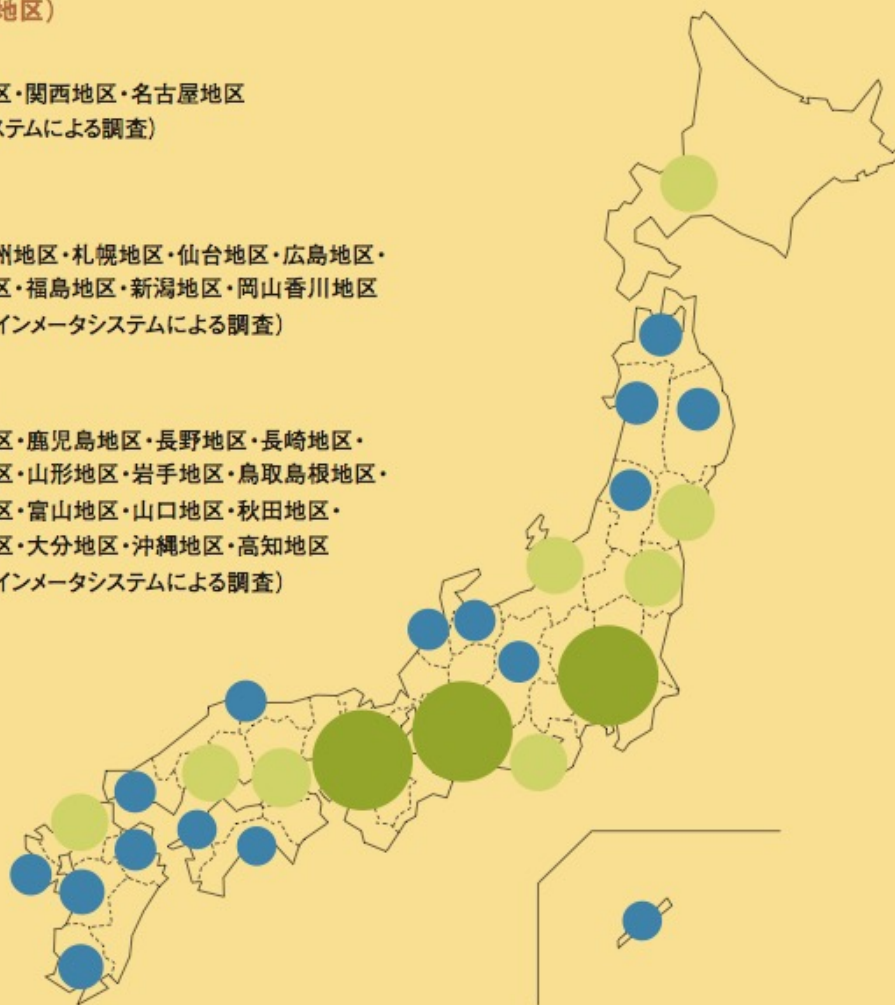
● 600世帯 関東地区・関西地区・名古屋地区
(PMシステムによる調査)

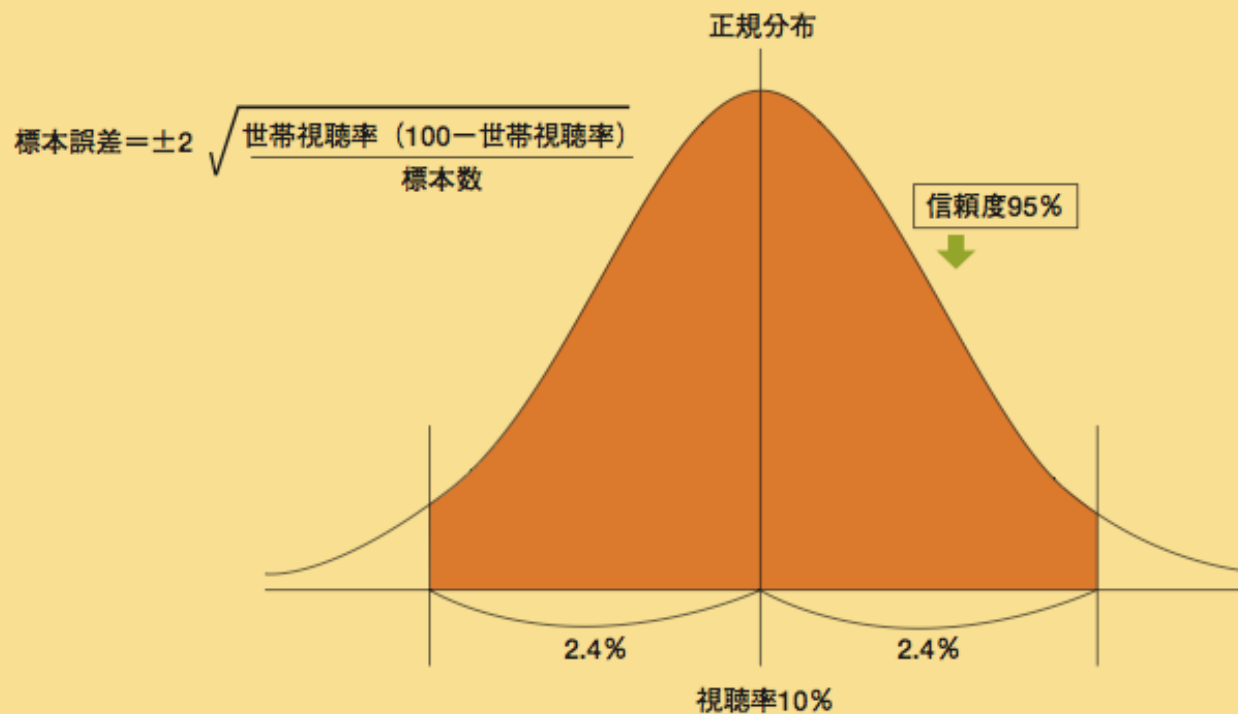
■52週調査地区(8地区)

● 200世帯 北部九州地区・札幌地区・仙台地区・広島地区・
静岡地区・福島地区・新潟地区・岡山香川地区
(オンラインメータシステムによる調査)

■24週調査地区(16地区)

● 200世帯 熊本地区・鹿児島地区・長野地区・長崎地区・
金沢地区・山形地区・岩手地区・鳥取島根地区・
愛媛地区・富山地区・山口地区・秋田地区・
青森地区・大分地区・沖縄地区・高知地区
(オンラインメータシステムによる調査)





世帯視聴率	標本数 600	標本数 200
5%・95%	±1.8%	±3.1%
10%・90%	2.4	4.2
20%・80%	3.3	5.7
30%・70%	3.7	6.5
40%・60%	4.0	6.9
50%	4.1	7.1

中心極限定理

- 同じ母集団から抽出された標本サイズが十分に大きいと、標本の平均（標本平均）はおおよそ正規分布に従う。
- この正規分布の平均は母集団の平均に非常に近い。
- 標本平均の分散（定義については 15.3 節参照）は、母集団の分散を標本サイズで割ったものに非常に近い。

- コード 17.6
[0,5]の実数ができるサイコロ 一様分布
平均と分散は？
- このサイコロを100個振る
100個の平均値 はどのように分布するか？
平均と分散は？
- 平均と分散を比較する

平均の標準誤差

$$SE = \frac{\sigma}{\sqrt{n}}$$

n サンプルサイズ

σ 母集団の標準偏差

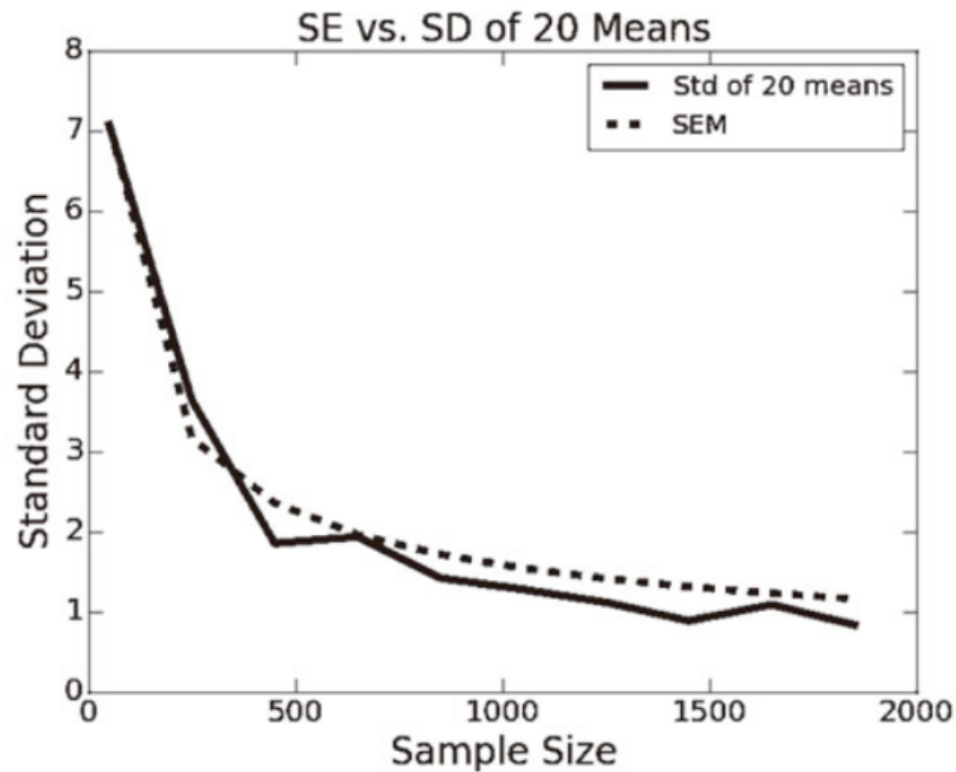


図 17.6 平均の標準誤差

平均の標準誤差

$$SE = \frac{\sigma}{\sqrt{n}}$$

n サンプルサイズ

σ 母集団の標準偏差

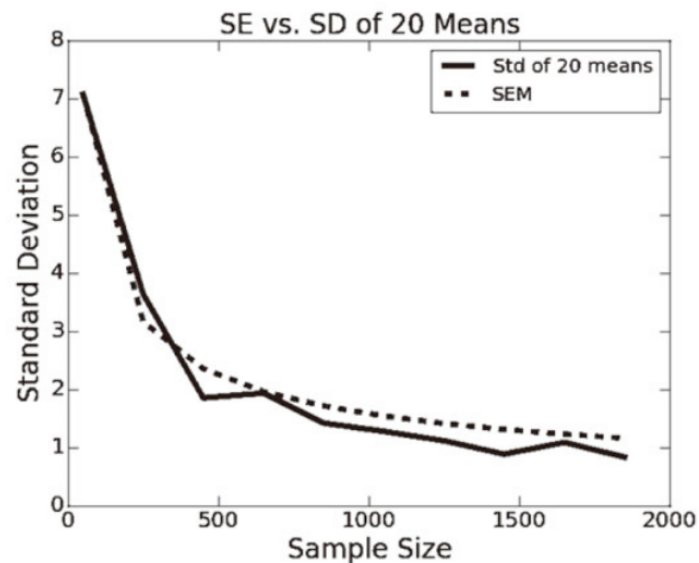


図 17.6 平均の標準誤差

平均の標準誤差

$$SE = \frac{\sigma}{\sqrt{n}}$$

n サンプルサイズ

σ 母集団の標準偏差

普通はわからない

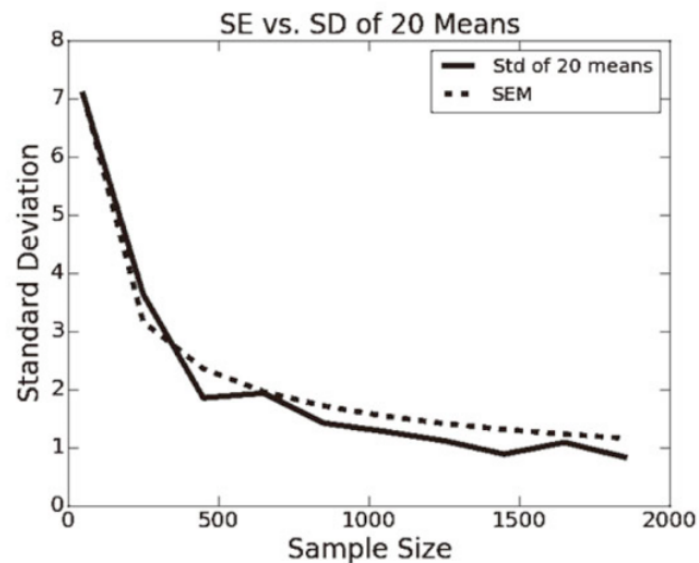


図 17.6 平均の標準誤差

平均の標準誤差

$$SE = \frac{\sigma}{\sqrt{n}}$$

n サンプルサイズ

σ 母集団の標準偏差

普通はわからない

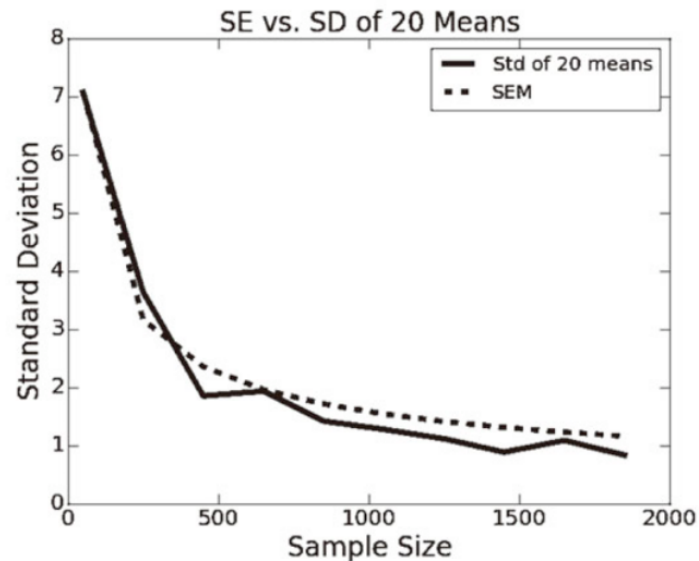


図 17.6 平均の標準誤差

ある程度以上のサイズの標本標準偏差で代用

どのくらいなら可？ コード17.9

第11回の授業と課題後半

- テキスト
17章～
- 課題3：お見合い問題（別紙）の解をモンテカルロ法で求めよ。
- 課題4：モンティ・ホール問題バリエーションの実験と考察
上記2点ともに、レポートの形式を守ること。

提出先：CoursePower

期限：12/11(水) 20:00

調べる、人に聞くなどしてとにかく実施することが大事です。

注意点 変動させるパラメータ

- コード 17.3
 - `sampleSize = 40`
- コード 17.5
 - `testSamples(1000, 40)`
- コード 17.6
 - `plotMeans(100, 100000, 11,)`

- 確認したいことは何なのか？ どの値が変わったらどうなる？
 - それを精度良く確認できるようにするために増やすべき値がどれなのか？
- この違いについて注意してパラメータを変化させ、結果についてまとめること

注意

- 17章のコードで使用している関数、
 - stdDev
 - variance

については、P.229, コード15.5にある定義を使用すること。

期末試験について

- 最終回1月10日（金）授業内試験実施
- 持込可能なものについての注意
 - A4の紙1枚のみ持込可能
 - 本人の手書きのみ認める。印刷・コピーなど不可
 - 両面使用可
 - 表面右上に学籍番号と氏名を記入すること
 - 試験後回収する

課題再提出について

- これまでの一部のレポートの再提出を受け付けます。
- 期限：2024/12/27, 20:00 JST
- CoursePowerの再提出先の、正確な場所に提出すること。
- IS数字(回数)_学籍番号_氏名.pdf とする
例：IS13_J323432_情報太郎.pdf
- PDFファイルのみ（docx, pptxなど禁止）
- .py .ipynb などのコード提出
 - ただし、レポート内容はPDFファイルで完結、完備であること。zip禁止
- 理由を問わず期限以降は受け付けません
 - 理由1：これまでに十分に期間があったこと
 - 理由2：採点の時間がないこと

今回の課題 まとめると

- 課題 1, モンティ・ホール問題
 - 課題 2, お見合い問題
 - 課題 3, モンティ・ホール問題改
-
- 1つのPDFファイルで提出してください
 - 期限は12/11(水), 20:00 JST

この後

- 12/13, 20はD社データサイエンス部門の方の特別講義、演習、発表会となります
- 事前課題pandasの基本はしっかりやっておくこと
- Python と Computationの初歩はおおよそ学べたと思いますので、ネット上の様々なリソースを使ってこの先の学習を各自で進められます
- 教科書の残りの一部（機械学習系）は、3年生の授業（パターン認識演習）で扱いますので、教科書をすぐに処分などしないこと