Pythonで体験するベイズ推論
~PyMCによるMCMC入門~

第７章　ベイズA/Bテスト

石田研
M1 賀来智博
kaku@cb.cs.titech.ac.jp

目次

1. コンバージョンテスト（A/Bテストの復習）
2. 期待収益の解析
3. ベイズ推定によるt検定
4. 増加量の推定

目次

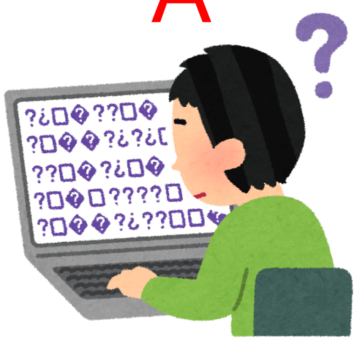
1. **コンバージョンテスト（A/Bテストの復習）**
   pymcmc1.ipyenv
2. 期待収益の解析
3. ベイズ推定によるt検定
4. 増加量の推定

A/Bテスト

２つの異なる処理の効果の差を検証する手法のこと

２つのウェブサイトのデザインをそれぞれA、Bと呼ぶ
ユーザーがサイトにやってきた時ランダムにどちらかを表示
それぞれのデザインについて**コンバージョン数**を記録する

A



visitors : 1300
conversion : 120

B

visitors : 1275
conversion : 125

コンバージョンをモデリング

- 事前分布
  コンバージョン確率：ベータ分布(α=1、β=1) ⇒ [0,1]の一様分布
- 観測値
  訪問者とコンバージョン数は二項分布に従うとする

共役関係

ベータ分布 ⟺ 二項分布

N回中X回の成功を観測したとする

事前分布		事後分布
$Beta(\alpha_0, \beta_0)$	$\Rightarrow$	$Beta(\alpha_0 + X, \beta_0 + N - X)$

AとBの事後確率の比較

サンプリング

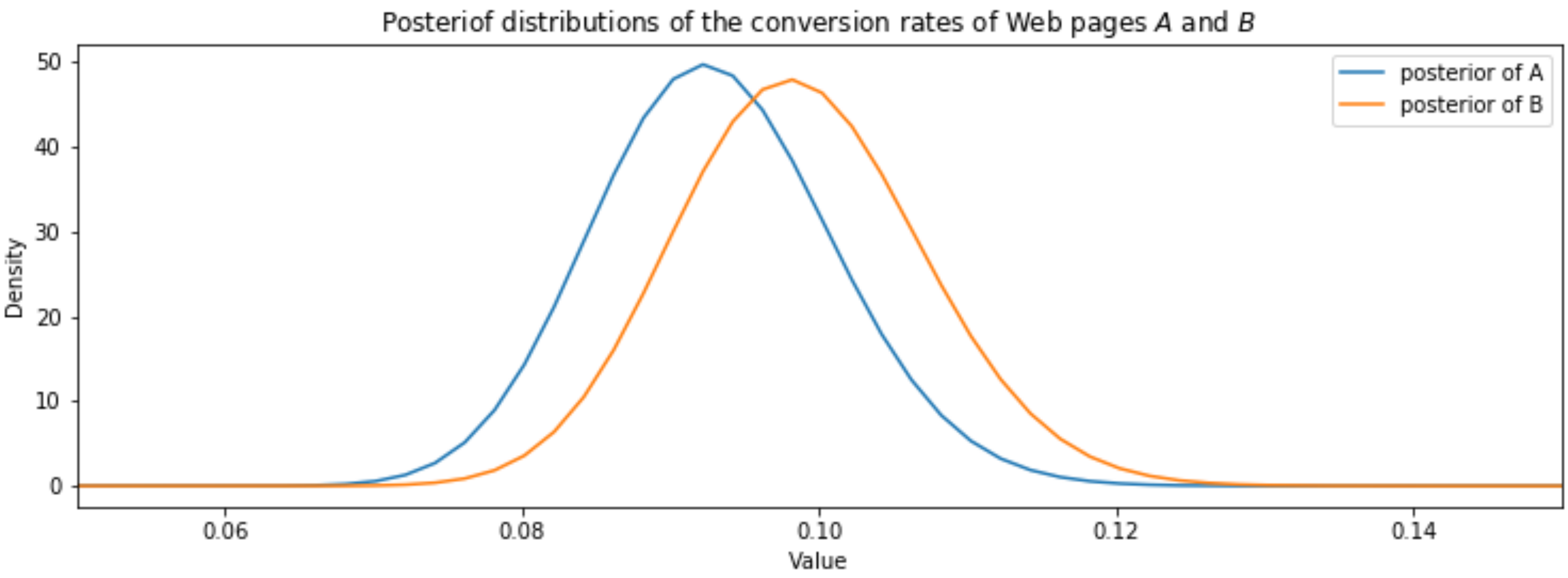
In [4]:
```
samples = 20000
samples_posterior_A = posterior_A.rvs(samples)
samples_posterior_B = posterior_B.rvs(samples)

print((samples_posterior_A > samples_posterior_B).mean())
```

0.30755

- サイトAがサイトBよりコンバージョン数が高い確率は約30%
- この本ではこの確率はそれほど有意ではないと言っている

コンバージョン率の事後分布



それほど変わらないように見える

目次

期待収益の解析

- 実際の企業でよくあるのは、自社サービスの登録数を増やすと共にユーザーが選択する契約プランも最適化するという目標
- もちろんなるべく高いプランを選ばせたい
- ユーザーに契約プランのページを２通り提示し、提示1回あたりの期待収益を決定することを考える
- 79ドル、49ドル、25ドルのプランを想定
  p_nはそれぞれのプランを選択する確率

$$E[R] = 79p_{79} + 49p_{49} + 25p_{25} + 0p_0$$
$$p_{79} + p_{49} + p_{25} + p_0 = 1$$

期待収益のモデリング

- 事前分布
  各プランの契約確率：ディリクレ分布⇒ベータ分布を一般化したもの
- 観測値
  ユーザーのプラン選択数は多項分布に従うとする
  ⇒二項分布を一般化したもの

共役関係

ディリクレ分布 ⇔ 多項分布

観測：$N_1, N_2, N_3 \cdots, N_m$

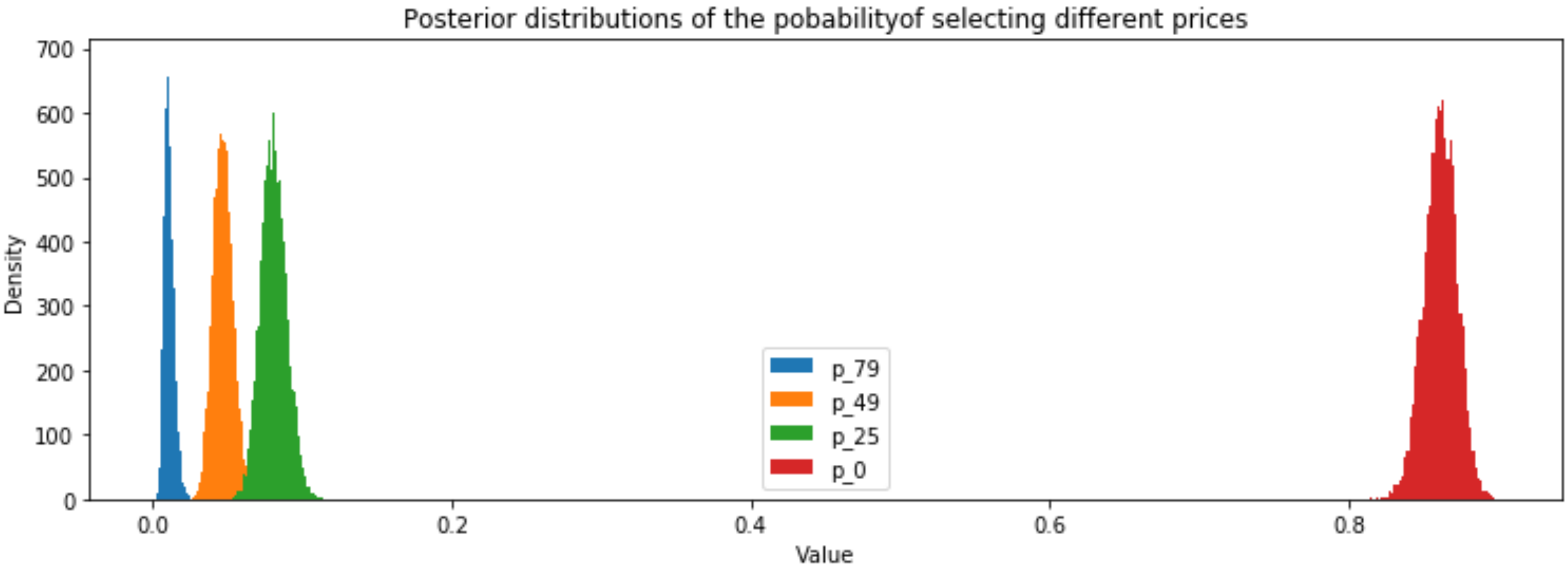
事前分布 ⇒ 事後分布

$$Dirichlet(1,1,1,\cdots,1) \Rightarrow Dirichlet(1+N_1, 1+N_2, 1+N_3, \cdots, 1+N_m)$$

試しに計算してみる

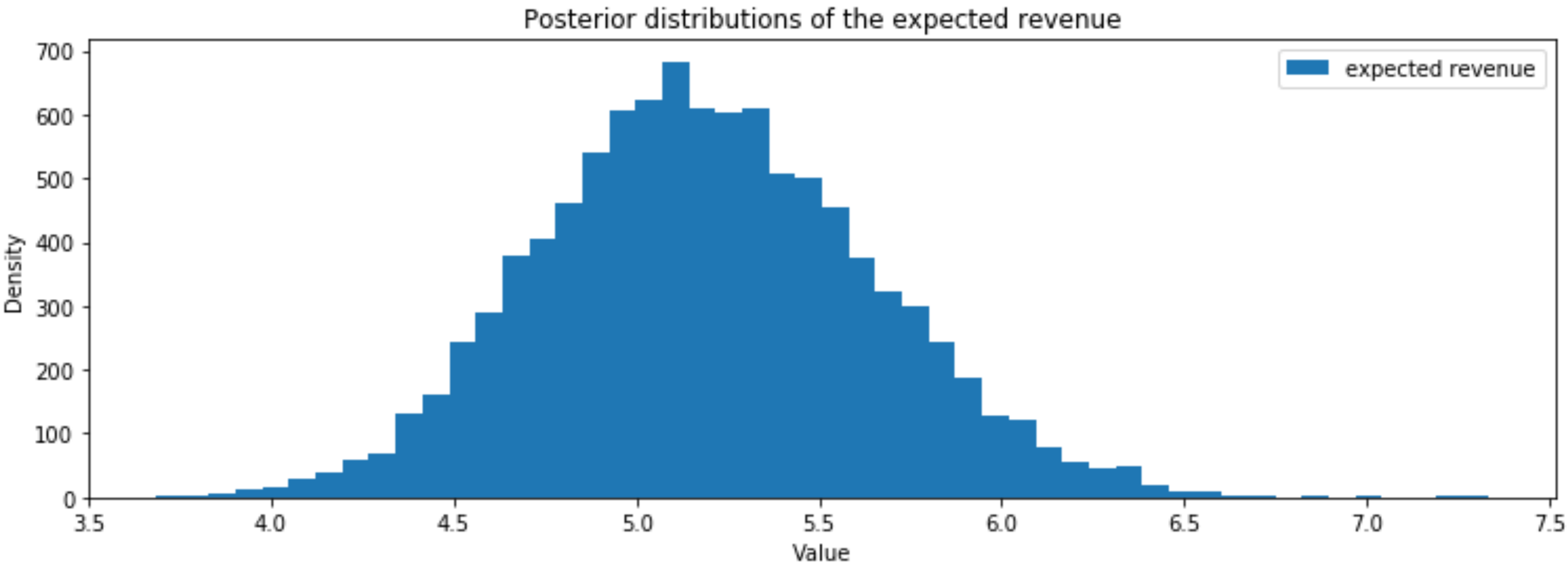
- $N_{79}: 10$、$N_{49}: 46$、$N_{25}: 80$、$N_0: 864$

各契約プランが選択される確率の事後分布



試しに計算してみる

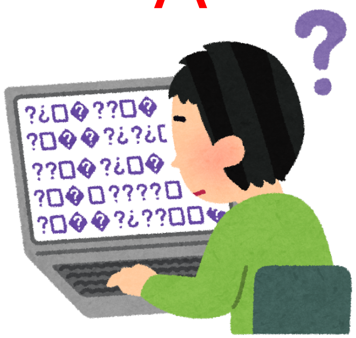
- 各契約プランが選択される事後確率から計算された期待収益の分布



大体４～６ドルくらい

A/Bテストに拡張

A

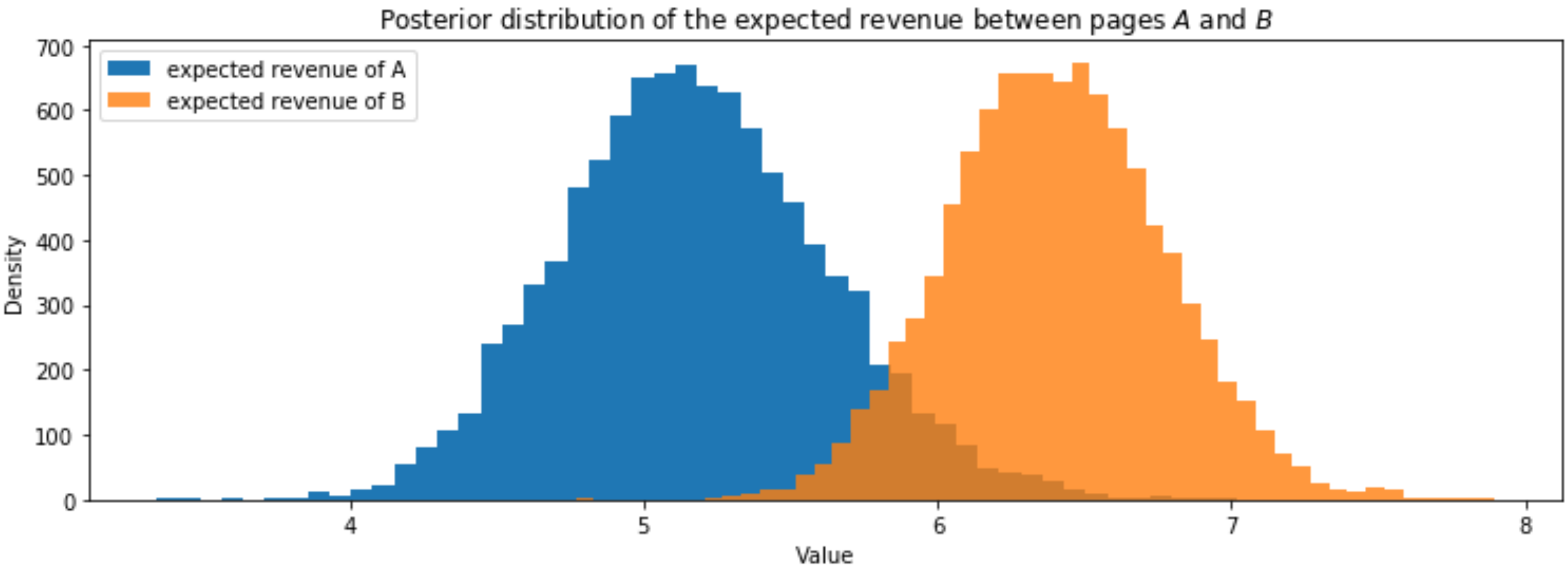


N : 1000
N_{79} : 10
N_{49} : 46
N_{25} : 80
N_0 : 874

B

N : 2000
N_{79} : 45
N_{49} : 84
N_{25} : 200
N_0 : 1671

AとBの期待収益の分布を比較



- サイトAの収益はサイトBの収益に比べて１ドル少ないように見える
- ページビュー毎に１ドル差が出るのは収益に大きな影響がある

Bの収益がAよりも大きい確率は？

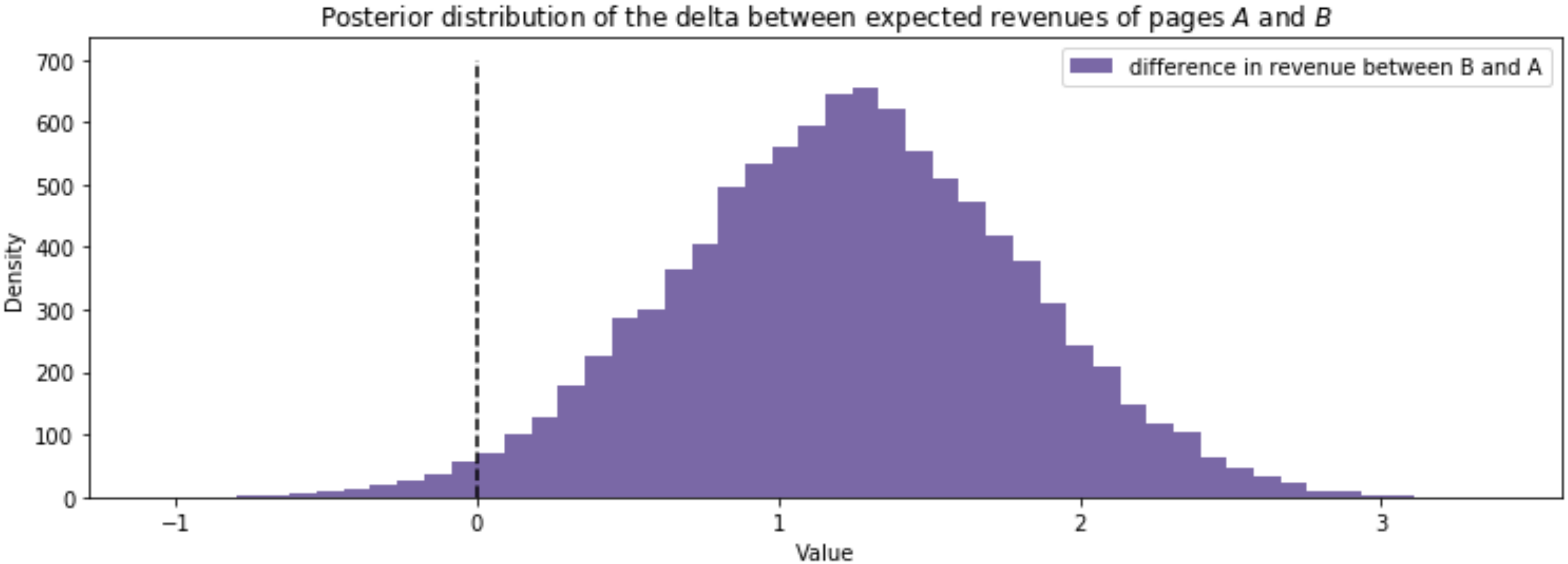
```
[8]: p = (posterior_expected_revenue_b > \
          posterior_expected_revenue_a).mean()

     print("Probability that page B has "
           "a higher rebenue than page A: %.3f" % p)
```

Probability that page B has a higher rebenue than page A: 0.983

- 98%の確率でサイトBの収益はサイトAより大きい
  有意に大きそう

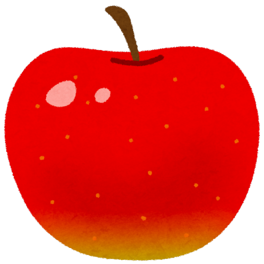
サイト間の収益の差の事後分布



- 50%の確率で１ドル以上儲かりそう
- もしBを採用するという結論が間違っていたとしてもそこまで損は大きくなさそう
  ⇒分布の伸び具合

目次

t検定



- 母平均に対する検定
  ex)平均値が 0 と異なるかどうか
  　リンゴAとリンゴBの大きさの平均は異なるか
- 制約
  対象データが**正規分布**に従っている必要がある
  正規分布に従っていればt値がt分布に従うので、p値の計算ができる
  2群間の平均の検定を行なう場合は**分散も等しい**必要がある

$$t\text{値} = \frac{\text{期待値の差の大きさ}}{\sqrt{\text{分散}/\text{サンプルサイズ}}}$$

t値が大きい

- 期待値の差が大きい
- 分散が小さい（期待値を比較に使える）
- サンプルサイズ大（期待値が信用できる）

⇒ 有意差あり

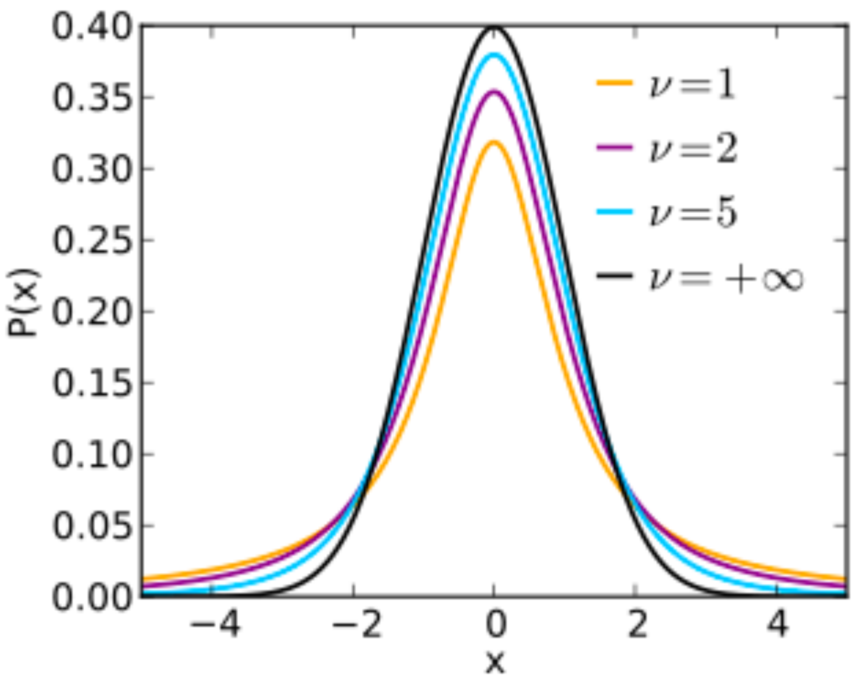
t分布

- 確率密度関数

$$P(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

Γはガンマ分布

νは自由度　$\nu \to \infty$なら正規分布



- 平均：0
- 分散：

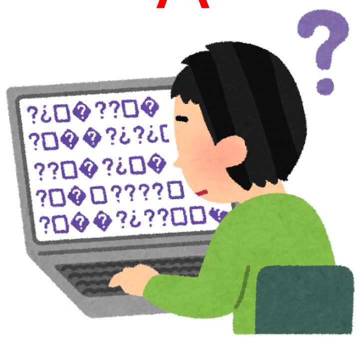
$\nu > 2$の場合、$\dfrac{\nu}{\nu-2}$

$1 < \nu \leq 2$の場合、∞ (無限大)

サンプルデータ

- 各サイト250サンプルを用意
- 各ユーザーの滞在時間は正規分布に従う

A



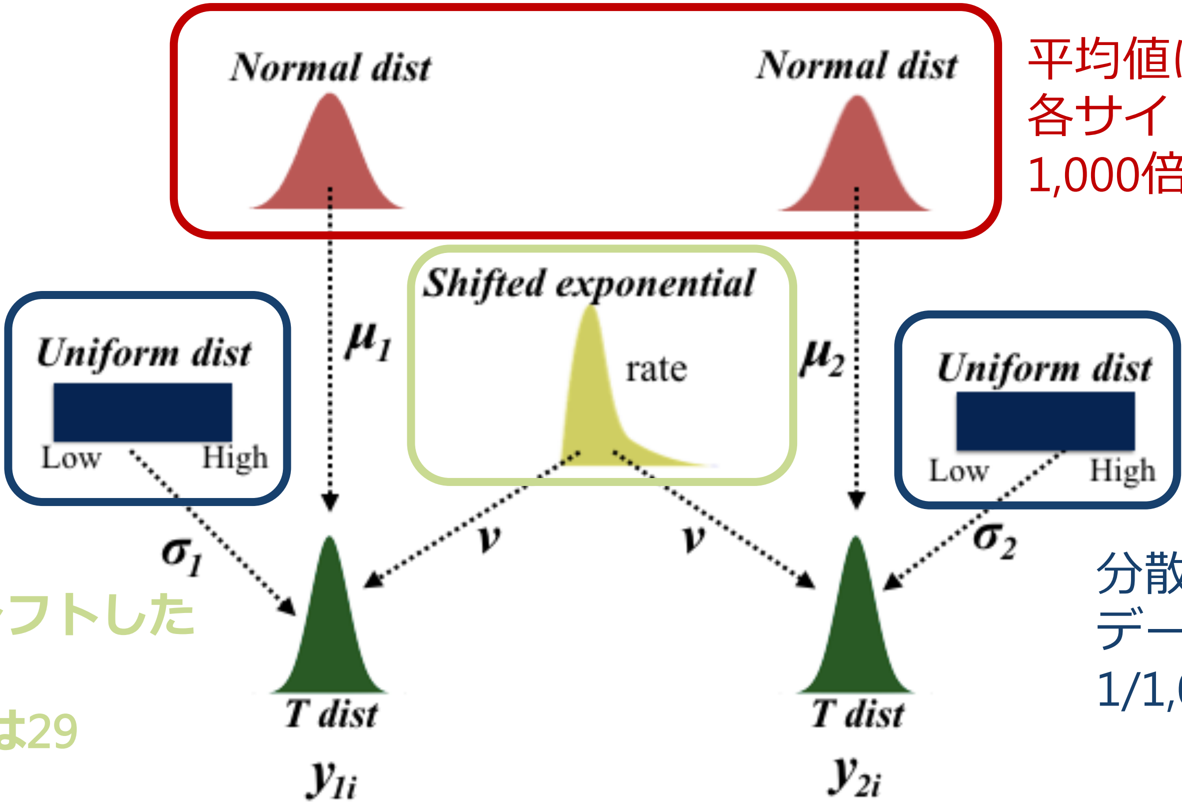
平均：30
分散：4

B

平均：26
分散：7

ベイズt検定(BESTモデル)

- サイトAとサイトBのユーザー滞在時間がt分布に従うと仮定
- 各パラメーターの事前分布



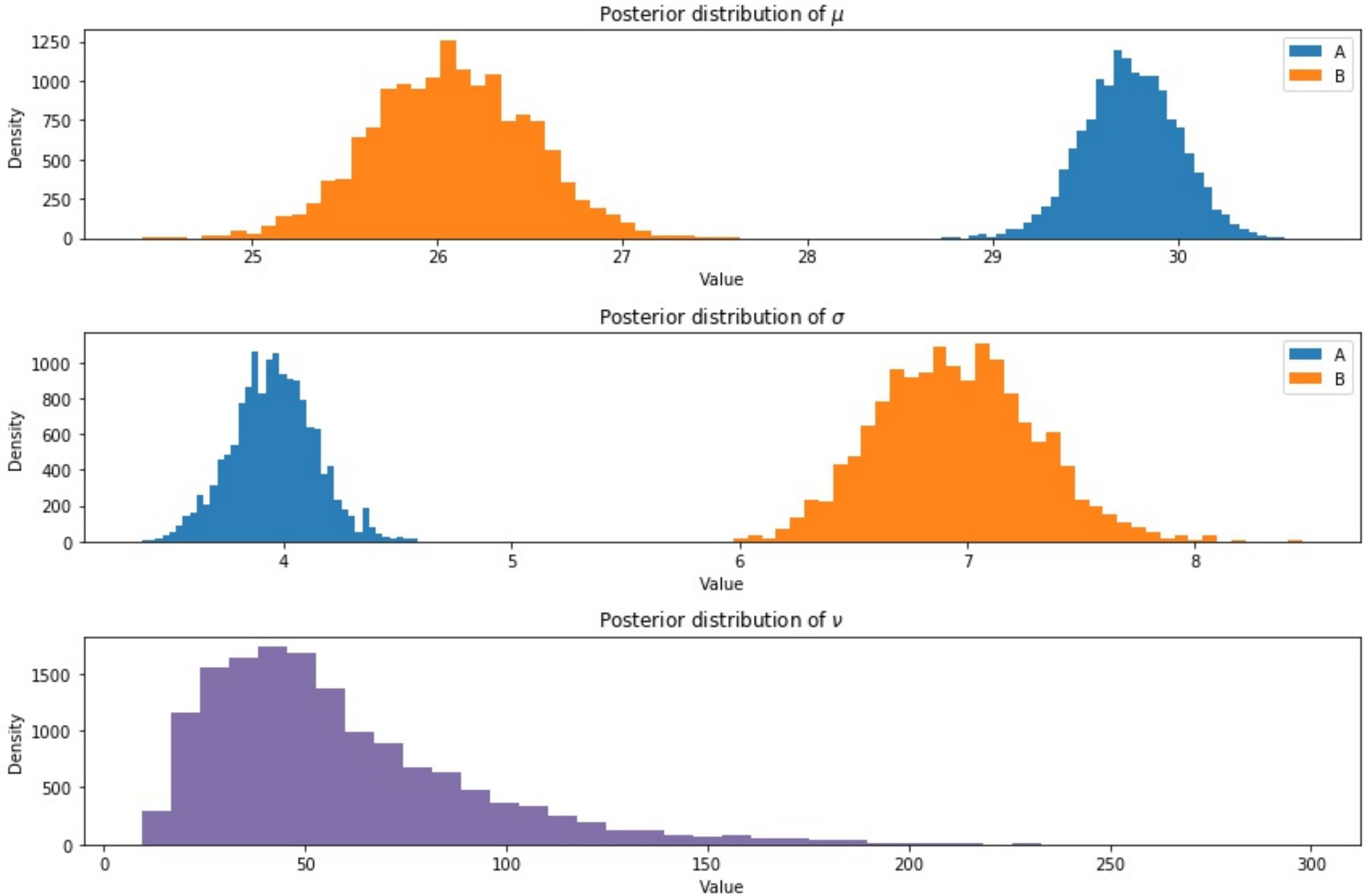
平均値は、
各サイトの平均値、
1,000倍の分散の正規分布

自由度は
右に1だけシフトした
指数分布
パラメータは29

分散は一様分布
データの標準偏差の
1/1,000から1,000倍

※分布の形はあくまでイメージです

各パラメータの事後分布



普通のt検定

一般のt検定

```
In [16]: t, p = ttest_rel(durations_a, durations_b)
```

```
In [17]: print( "p値 = %(p)s" %locals() )
```

```
p値 = 2.31597943405e-11
```

- p値が0.05以下なのでAとBの平均は有意に差がある

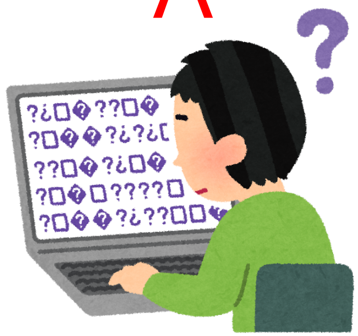
目次

増加量の推定

- A/Bテストを見た上司が
  「サイトAの方がいいのはわかったけど
  　　実際にどれくらいもうかるの？」
  と聞いてくるかもしれない
- A/Bテストは2値問題を解くための手法
  増加量を求める時は注意が必要

サンプル

A

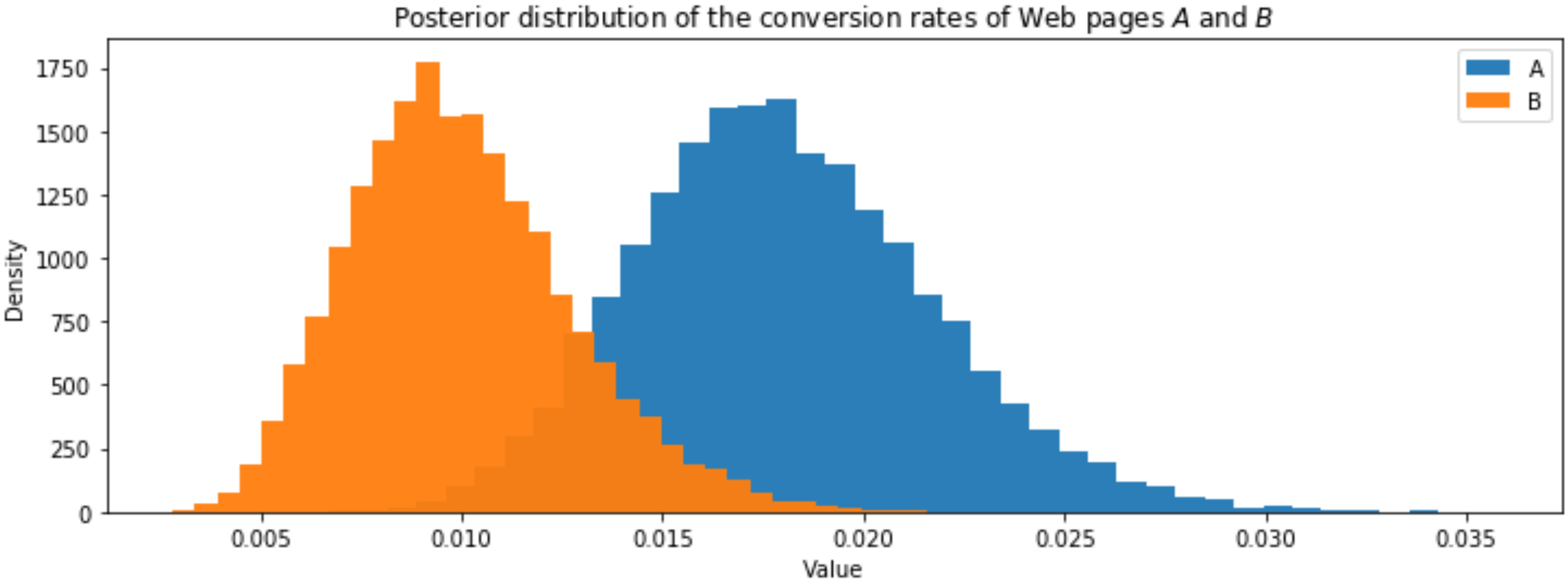


visitors : 1275
conversion : 22

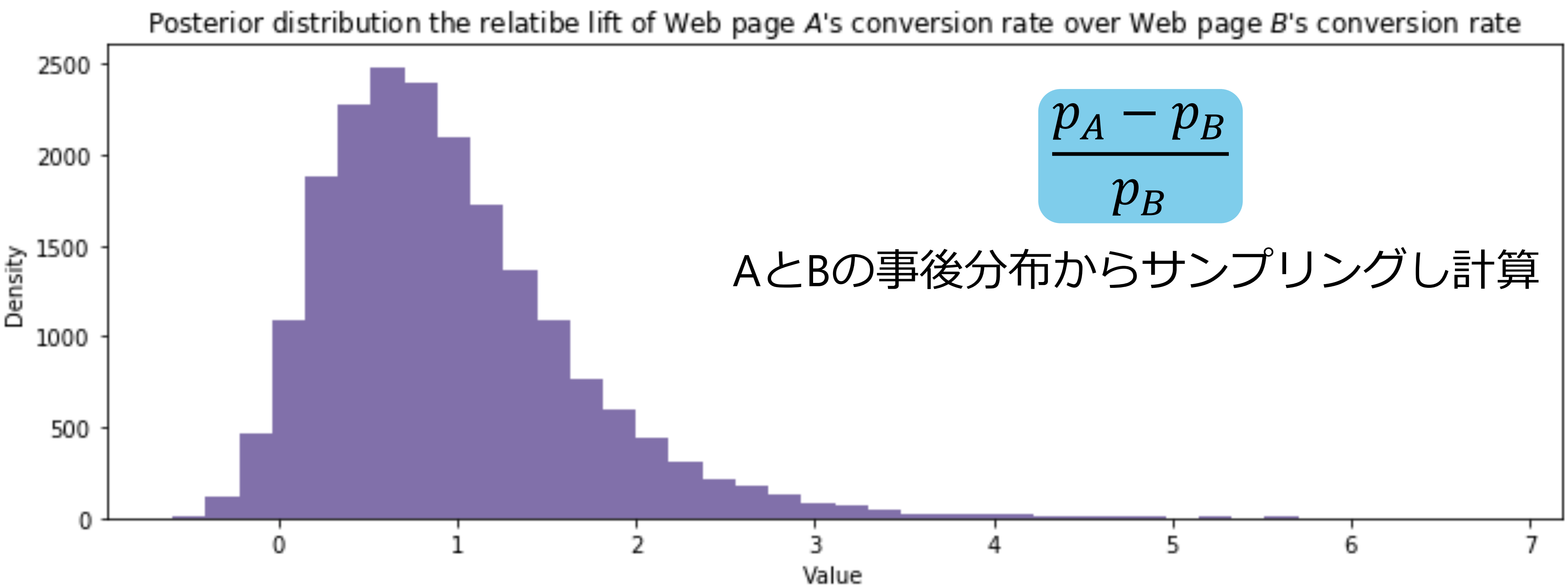
B

visitors : 1300
conversion : 12

コンバージョン率の事後分布



相対増加量の事後分布



$$\frac{p_A - p_B}{p_B}$$

AとBの事後分布からサンプリングし計算

```
print((posterior_rel_increase > 0.2).mean())
print((posterior_rel_increase > 0.5).mean())
```

```
0.8916
0.72045
```

89%の確率で相対増加量が20%以上ある

72%の確率で相対増加量が50%以上ある

悪い例

- サイトAのサイトBに対する相対的な増加量を単純に点推定

$$\frac{\hat{p}_A - \hat{p}_B}{\hat{p}_B}$$

$\hat{p}_A, \hat{p}_B$：サイトAとBの事後分布の平均

$\hat{p}_A$と$\hat{p}_B$の値が0に近いととんでもなく大きい値が出てしまう

```
In [32]: print("bad increase rate = %.3f" %
              ((samples_posterior_A.mean() - samples_posterior_B.mean()) / samples_posterior_B.mean()))

bad increase rate = 0.807
```

AのBに対する相対増加率は80%

値の不確実性についての情報が消滅

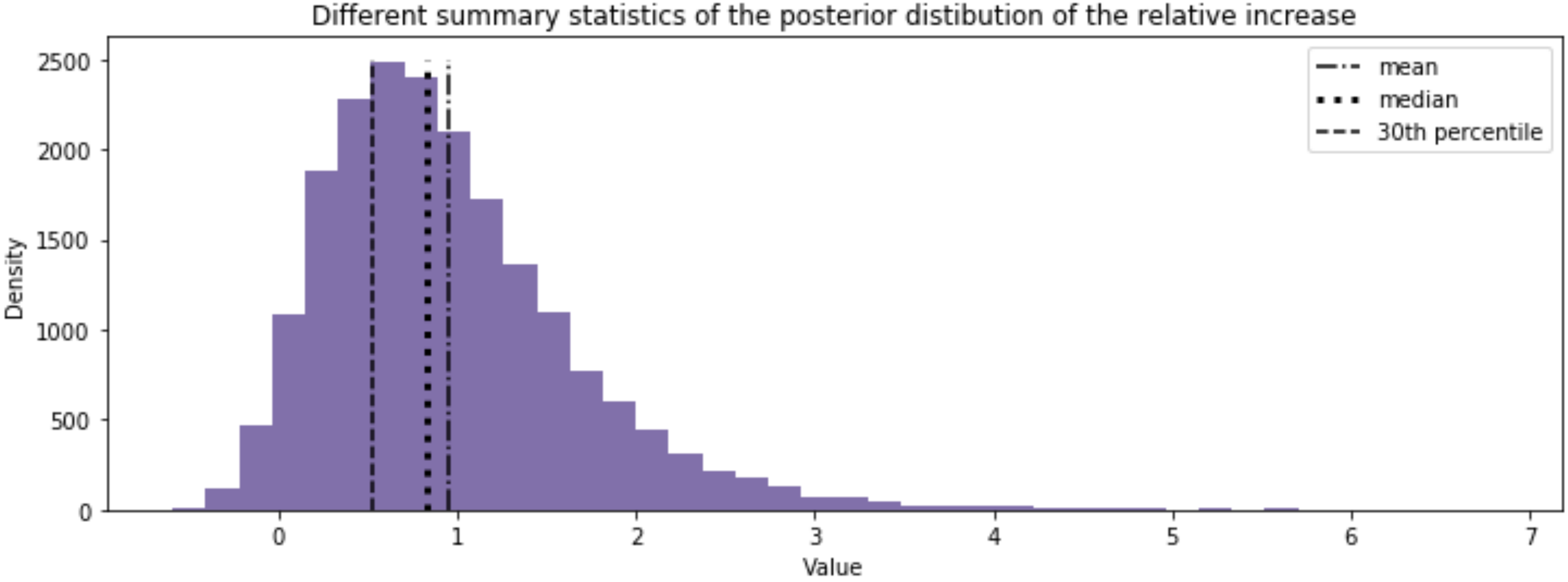
それでも点推定がしたい時は

1. 相対増加量の事後分布の平均を返す ×
   →事後分布のグラフがロングテールのため平均はよろしくない

2. 相対増加量の事後分布の中央値を返す △
   →歪んだ分布に対しても、平均よりは影響を受けにくい

3. 相対増加量の50%以下のパーセンタイルを返す ○
   →損失関数と同じ効果をもつ
   過小評価より過大評価にペナルティを与えることになるため
   推定値がより慎重な値になる
   →実験でより多くのデータを得るにつれて、
   相対増加量の事後分布の幅が狭くなるので、
   どのパーセンタイルを選んでも同じ値に収束する

要約統計量のプロット



Different summary statistics of the posterior distibution of the relative increase
mean
median
30th percentile
Density
Value
2500
2000
1500
1000
500
0
0
1
2
3
4
5
6
7