Pythonで体験するベイズ推論
4.1-4.3.1

秋山研　修士1年　黄毅聰

大数の法則

$Z_1, Z_2, \cdots, Z_N$を、ある確率分布からサンプリングした N 個の独立したサンプルとする。大数の法則によれば、期待値 $E[Z]$ が無限大に発散しないかぎり、以下の式が成り立つ。

$$\frac{1}{N}\sum_{i=1}^{N} Z_i \to E[Z], N \to \infty$$

つまり、同じ分布から得られた確率変数の集合の平均は、その分布の期待値に収束する！

直感的に理解するには

二つの値c_1とc_2だけをとる確率変数 Z を考えよう。Z の中のあるサンプルを Z_i と表す。 Z の平均を取ってみよう。

$$\frac{1}{N}\sum_{i=1}^{N} Z_i$$

Z_iはc_1とc_2のどちらかの値しかとらないので、この総和を二つに分けることができる。

$$\frac{1}{N}\sum_{i=1}^{N} Z_i = \frac{1}{N}\left(\sum_{Z_i=c_1} c_1 + \sum_{Z_i=c_2} c_2\right)$$

$$= c_1 \sum_{Z_i=c_1} \frac{1}{N} + c_2 \sum_{Z_i=c_2} \frac{1}{N}$$

$$= c_1 \times (c_1\text{の近似的な頻度}) + c_2 \times (c_2\text{の近似的な頻度})$$

$$\approx c_1 \times P(Z = c_1) + c_2 \times P(Z = c_2)$$

$$= E[Z]$$

平均するサンプルを多く集めれば集めるほど、この近似の精度は良くなる

ポアソン分布に従う確率変数の収束

ポアソン分布に従う確率変数の三つのデータ列に対して、それらが大数の法則に従う様子を示してみよう。

条件：

- パラメータ = 4.5
- sample_size = 10,000

nを1からsample_sizeまで変えて、最初からnサンプルの平均をプロットする

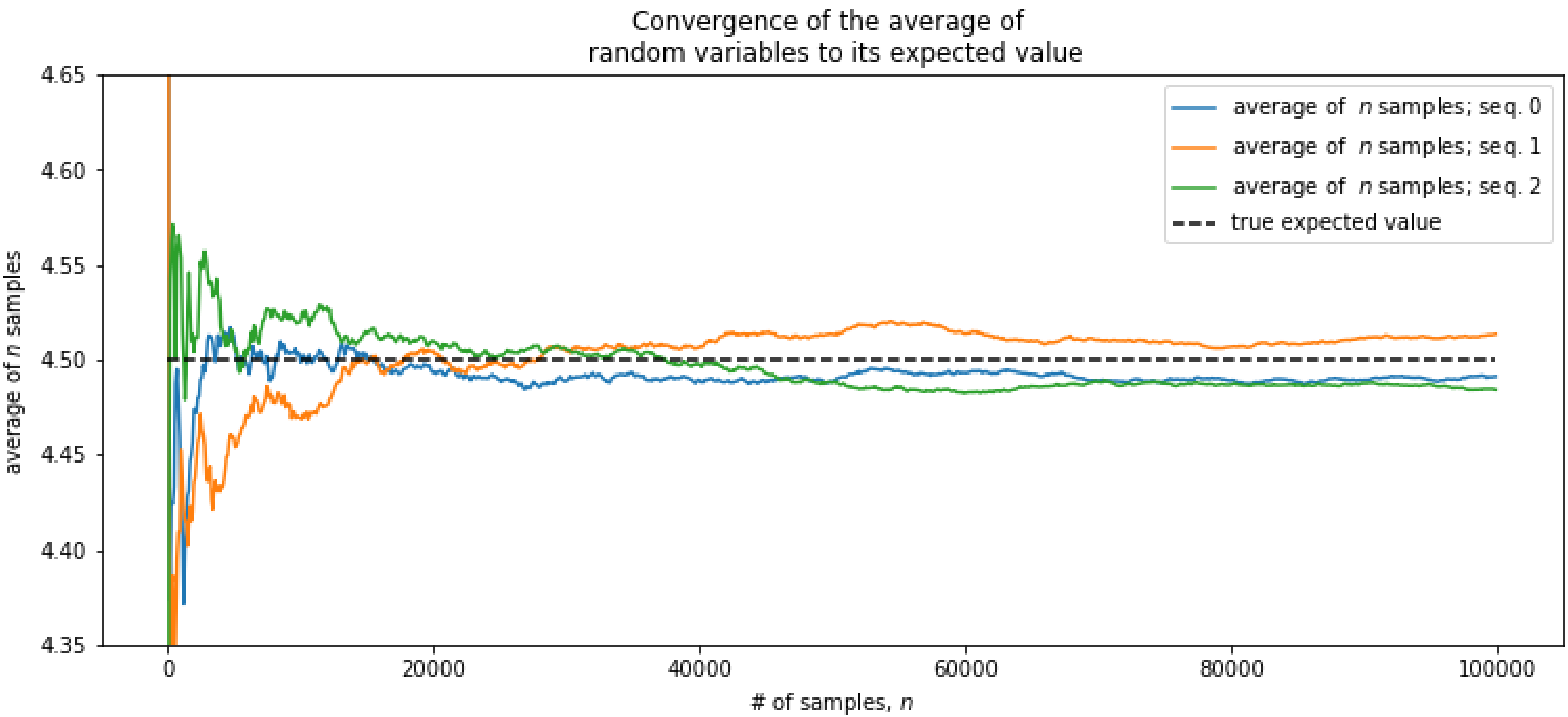
ポアソン分布

定数 λ>0 に対して、自然数を値にとる確率変数 X が次を満たすとき、確率変数 X はパラメータのポアソン分布に従う。

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

ポアソン分布の期待値はパラメータ λ に等しい

確率変数列の平均が期待値に収束する様子



サンプルの個数が大きいほど、平均のばらつきが大きい　⇒　最終的に4.5に近づく

じゃ、期待値にどれぐらい収束するの？

考え方：ある n に対して、このようなサンプル列を何千回も生成して、真の期待値から平均的にどれぐらい離れているかを計算してプロットすればよさそう…!

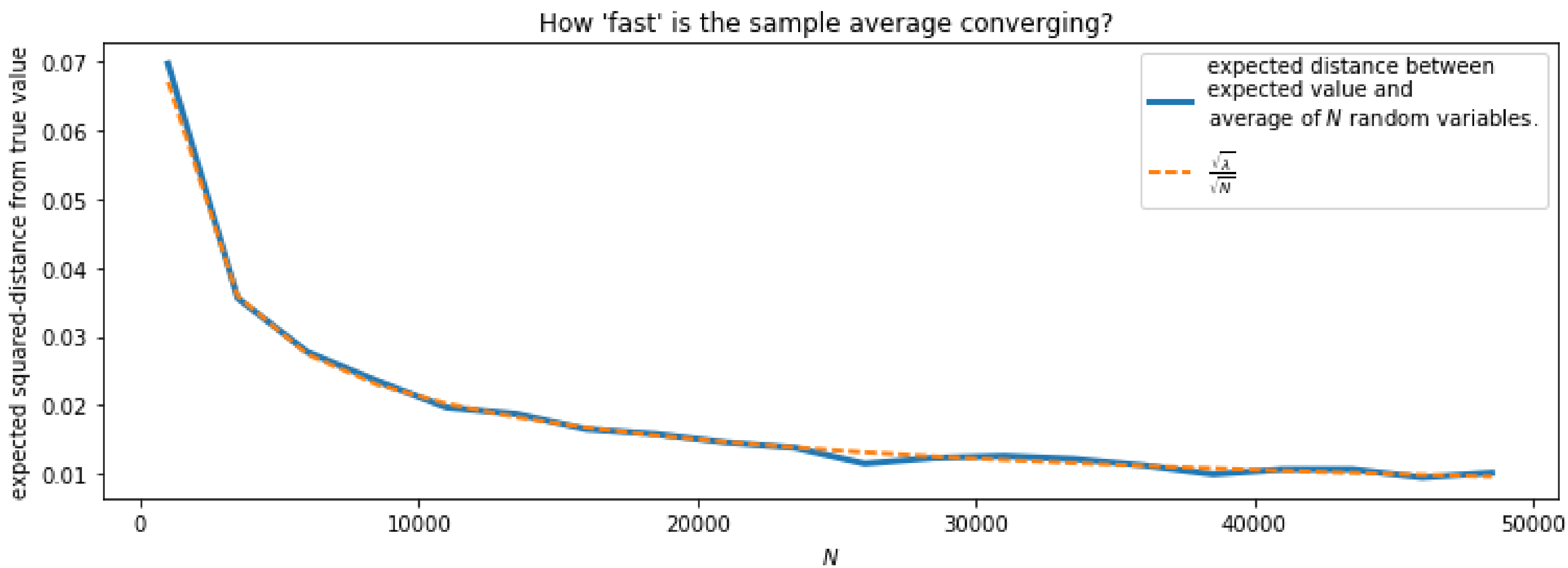
つまり、

$$D(n) = \sqrt{E[(\underbrace{\frac{1}{n}\sum_{i=1}^{n} z_i - 4.5)^2}_{Y_{n,k}}]} = \sqrt{E[Y_n]}$$

これは期待値なので、大数の法則を使って近似することができる

$$D(n) \approx \sqrt{\frac{1}{N}\sum_{k=1}^{N} Y_{n,k}}$$

サンプル平均はどのぐらい速く収束する？



Nが大きくなると、サンプ平均と真の期待値との距離の期待値は小さくなるが、「収束率」が減っている！

Z のような確率変数のサンプル列が与えられたら、大数の法則の *E[Z]* への収束率は以下で与えられる

$$\frac{\sqrt{Var(Z)}}{\sqrt{N}}$$

期待値と確率

期待値と確率の推定の関係は次の指示関数で定義しよう

$$1_A(x) = \begin{cases} 1(if\ x \in A) \\ 0(else) \end{cases}$$

大数の法則により、たくさんのサンプル χ_i があれば、イベント A の確率 $P(A)$ を推定できる

$$\frac{1}{N}\sum_{i=1}^{N} 1_A(X_i) \rightarrow E[1_A(X_i)] = P(A)$$

これを利用して、例えば、Z ~ Exp(0.5) が10よりも大きい確率を推定したいとしよう。指数分布 Exp(0.5) からたくさんの値をサンプリングして、以下のようにすることができる

$$P(Z > 10) = \sum_{i=1}^{N} 1_{z>10}\,(Z_i)$$

じゃ、ベイズ統計と何の関係?

第5章で紹介する点推定は、ベイズ推論では期待値の計算に帰着↓

もし事後分布から直接サンプリングできるなら、単にその平均を求めれば、期待値計算ができる

精度を上げるには、サンプリングをたくさんすればよい→実際case by case

大数の法則が成り立つのはサンプルサイズ N が大きいときだけである

サンプルサイズが小さいというわざわい

例：集約されたデータの扱い

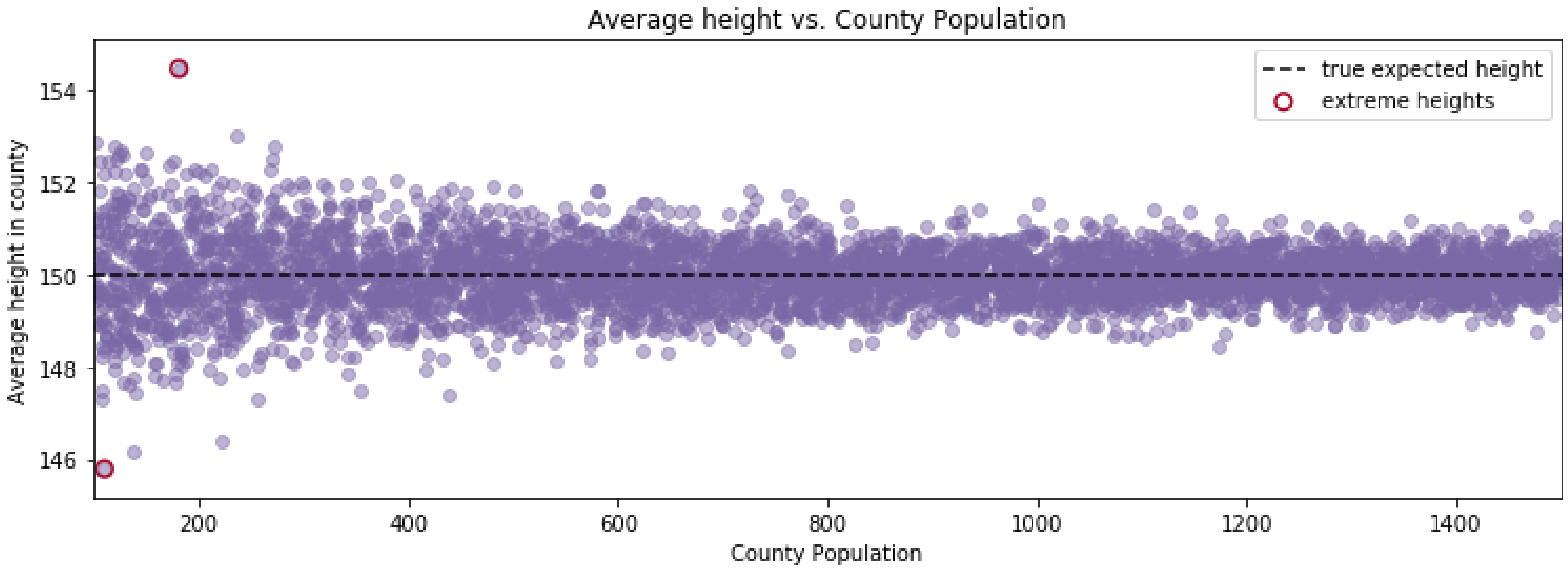
ある5,000の地域があり、各地域の人口は100から1,500の間で一様分布すると仮定する場合、地域ごとの平均身長を測定しよう

仮定より、身長はどの地域でも同じ、つまり各個人の身長の分布は住んでる地域によらず同じ

$$\text{身長} \sim \text{Normal}(150,15)$$

大数の法則を適用し、地域レベルで身長データを集約すると、データとして得られるのは各地域の平均だけ

人口に対する平均身長のプロット



人口が多いか少ないかを考えなかったら、過ちを犯す!

丸のついたデータが極値になった間違った推論の理由
- サンプルサイズ N が大数の法則に適用するには小さすぎ
- 地域の人口に対して、人口は一様分布ではない

まとめ

- 大数の法則：同じ分布から得られた確率変数の集合の平均は、その分布の期待値に収束
- ベイズ統計との関係：点推定は、ベイズ推論では期待値の計算に帰着
- サイプルサイズが小さいとき、大数の法則は適用できないかも