pythonで体験するベイズ推論

秋山研 M1 山澤まりな

ベイズの定理（復習）

$$P(A \mid X) = \frac{P(X \mid A)\, P(A)}{P(X)}$$

Where: データ: X
事象: A

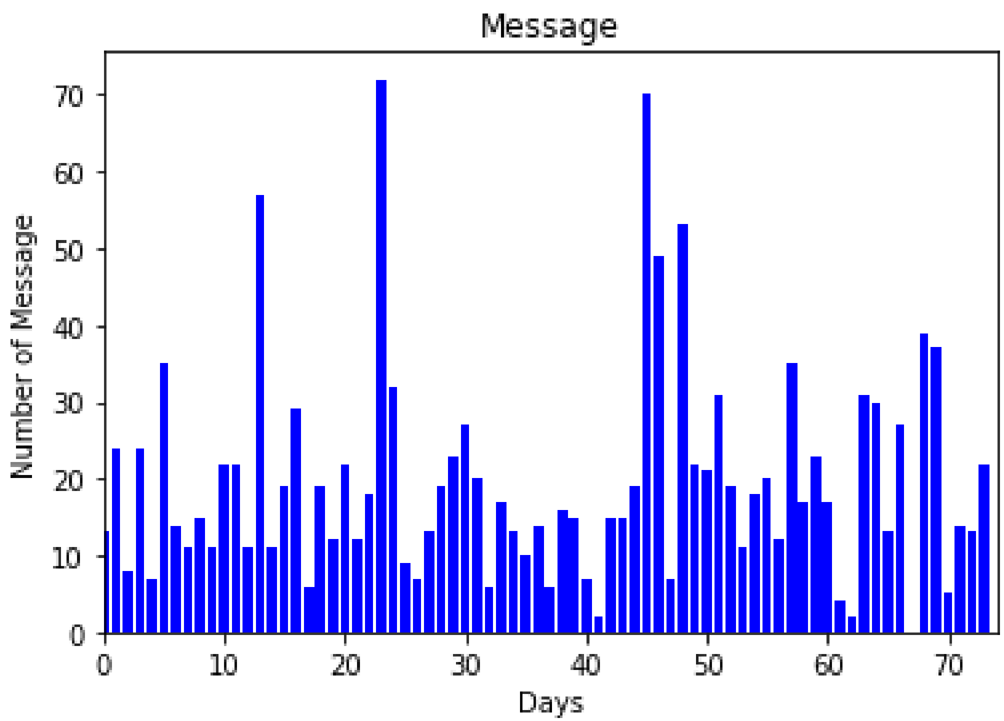
$P(A)$: 事象Aが成立する確率
$P(X)$: データXが取れる確率
$P(A \mid X)$: データXが与えられた時事象Aが起こる確率
$P(X \mid A)$: 事象Aが成立する時データXが取れる確率

$P(X)$は事象Aによらないため

$$\underbrace{P(A \mid X)}_{\text{事後確率}} \propto P(X \mid A)\, \underbrace{P(A)}_{\text{事前確率}}$$

例題（先週の続き）

あるユーザーの毎日のメールの受信数



* https://git.io/vXTVC

仮説：ある日を境にメールの受信数が増えているのではないか？

モデリング（事前分布の設定）

Step 1. モデリングする分布の設定

メールの受信数は自然数なので，ポアソン分布が妥当

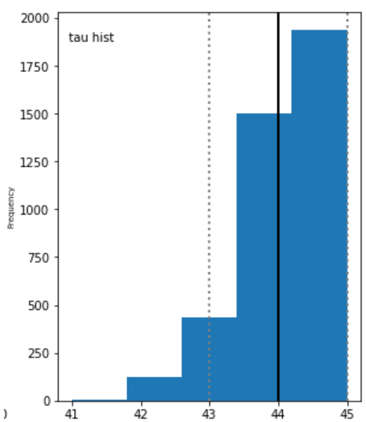
Step 2. 事前分布の設定

- 受信メール数が増えた日 τ
- τ 日までのポアソン分布の母数 λ_1
- τ 日以降のポアソン分布の母数 λ_2

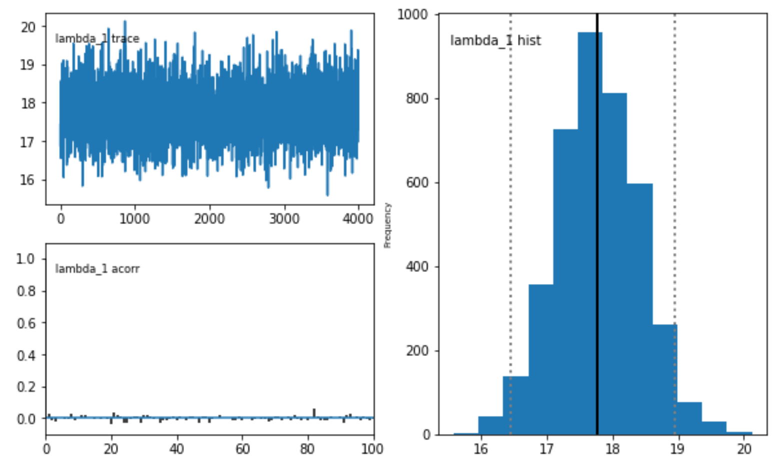
先週やったこと

$\lambda_1, \lambda_2, \tau$ の事後分布を推定し，メール受信数増加日を推定した

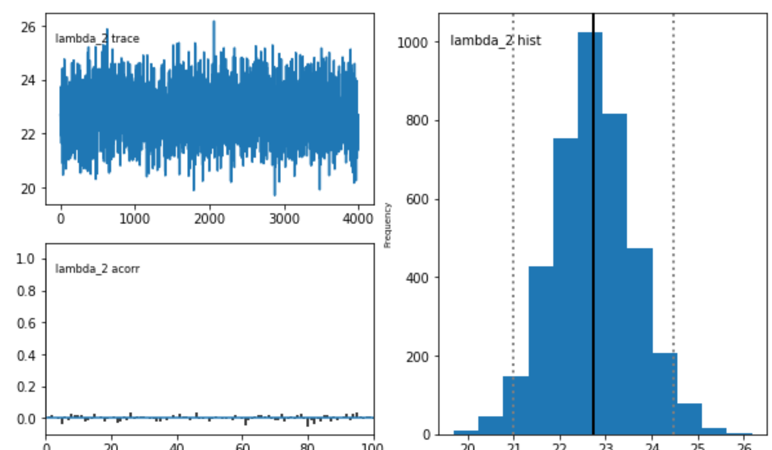
τ



λ_1



λ_2



今週やること その１

問題：2つのλは本当に統計的に異なるのか

先週の推定の結果，λ_1, λ_2の事後分布が異なることが分かったが，

統計的に異なることを示したい

解決法：$\mathrm{P}(\lambda_1 < \lambda_2 \mid data)$を計算する

すなわち，λ_1がλ_2よりも小さい確率を考えるとよい

今週やること その２

問題：そもそもメール受信数の変化点は1点と仮定していいのか

解決法：変化点の個数についても事前分布を作り，

モデル推定を行えばよい

- 今回はλを3つ設定し，変化点を2点に拡張するのみにとどめる

今週やること その３

演習：メッセージ受信増加率の期待値を求める

（やることその１のコードの下に解答あり）