ベイズ推論概論

フワッと全体のイメージを持つベイズ

千葉

目次

- ベイズ推測とは

What's Bayes inference?

みんな大好き渡辺ベイズでは...

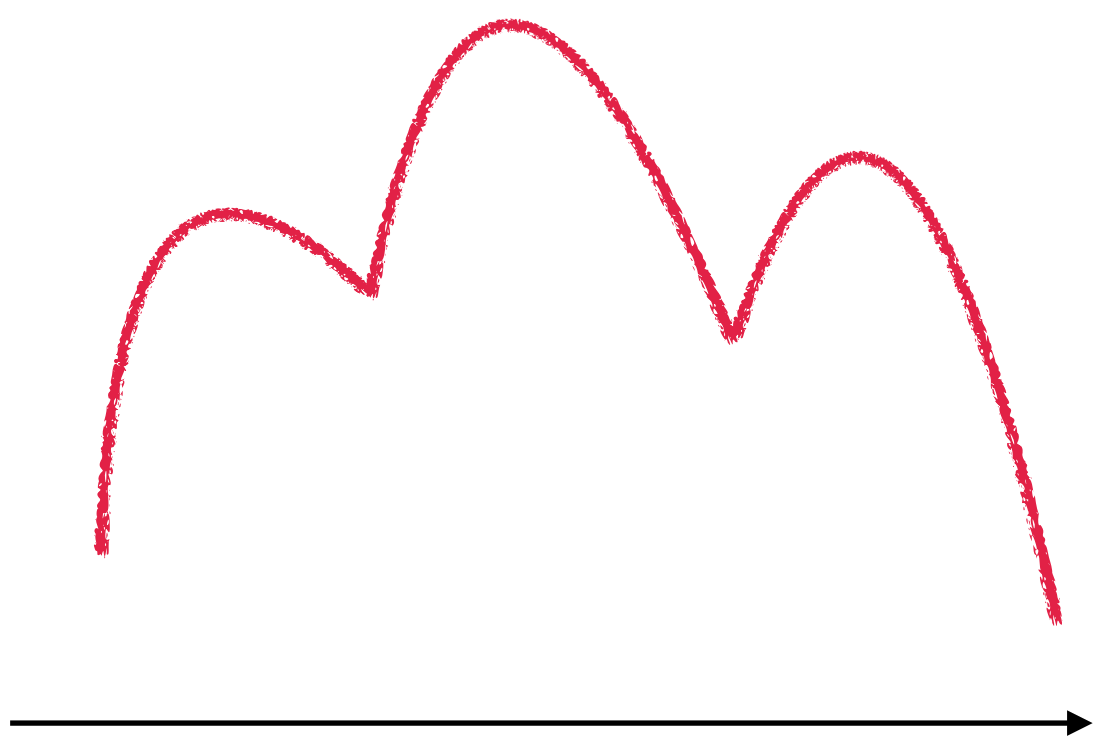
真の確率分布でq(x)はおおよそp*(x)くらいだろう

と推測することとベイズ推測を定義している。

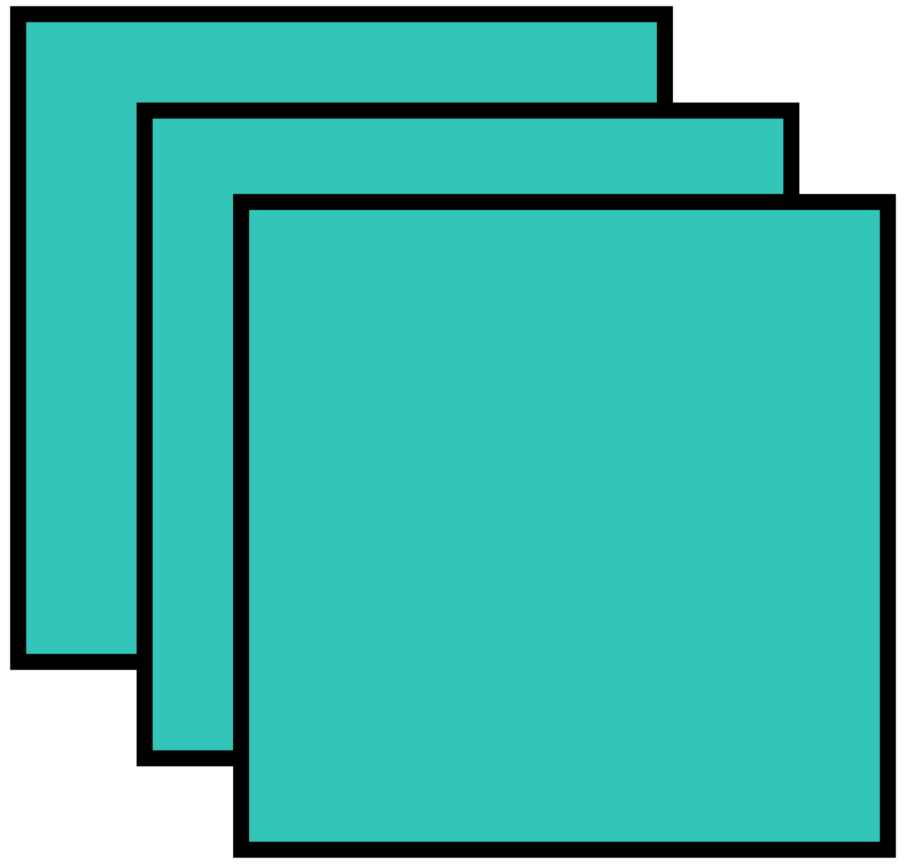
「ベイズ統計の理論と方法」渡辺澄夫 コロナ社(2012) P 5 参照

真の確率分布？$P^*(x)$?

ベイズ推測の考え方



生成



サンプル

真の確率分布q(x)

・データを生成しているルール

・確定しているが未知(観測はできない)

・一般にめちゃくちゃ複雑

推測

観測

<u>対象を生成してるルールは大体p*(x)だろうと推測する</u>

予測するルールのことを**予測分布**と言う

以後、　予測分布はp*(x)と書く

分析者

生成されたサンプル（データ）から
事象を支配するルールを推測する

定量的に扱う（真の確率分布・サンプル）

・n, N: 自然数

・N次元実ユークリッド空間$\mathbb{R}^N$

n 個の点の集合 $x_1, x_2, \ldots, x_n$ について考える

$$x^n = \underline{(x_1, x_2, \ldots, x_n)} \in \mathbb{R}^N \times \mathbb{R}^N \times \cdots \times \mathbb{R}^N$$

⇩

サンプル $x_1, x_2, \ldots, x_n$ が、

真の確率分布q(x)に独立に従う

確率変数の実現値であるとする

i.i.dの仮定

x^nを$(\mathbb{R}^N)^n$上の分布

$$q(x^n) = \prod_{i=1}^{n} q(x_i) = q(x_1)q(x_2)\cdots q(x_n)$$

を持つ確率変数

$$X^n = (X_1, X_2, \ldots, X_n)$$

の実現値として考える。

$\boxed{\text{サンプルを確率変数の実現値}}$

定量的に扱う（予測分布）

β : 逆温度と呼ばれるハイパーパラメータ

p*(x)を構成する

・**確率モデル**： $p(x|w)$

パラメータ$w \in W \subset \mathbb{R}^d$が与えられたときの

$x \in \mathbb{R}^N$の上の条件付き確率分布

・**事前分布**：$\phi(w)$

$w \in W$上の確率分布

・**パラメータの事後分布**

$$p(w|X^n) = \frac{1}{Z_n(\beta)}\phi(w)\prod_{i=1}^{n} p(X_i)|w)^{\beta}$$

・**分配関数（規格化定数）**

$$Z_n(\beta) = \int_W \phi(w)\prod_{i=1}^{n} p(X_i|w)^{\beta} dw$$

統計力学同様、この分配関数を使って色々なことがわかる。

予測分布p*(x)はこのパラメータの事後分布を使って定義する

定量的に扱う（予測分布）

分析者

<u>対象を生成してるルールは大体p*(x)だろうと推測する</u>

大体こんな感じ〜

期待値〜

平均値〜

$$p^*(x) = \mathbb{E}_w[p(x|w)] = \int p(x|w)p(w|X^n)dw$$

予測分布 := 事後分布によって確率モデルを平均したもの

では、「真の分布q(x)は大体p*(x)くらいだろう。」と予測したとする

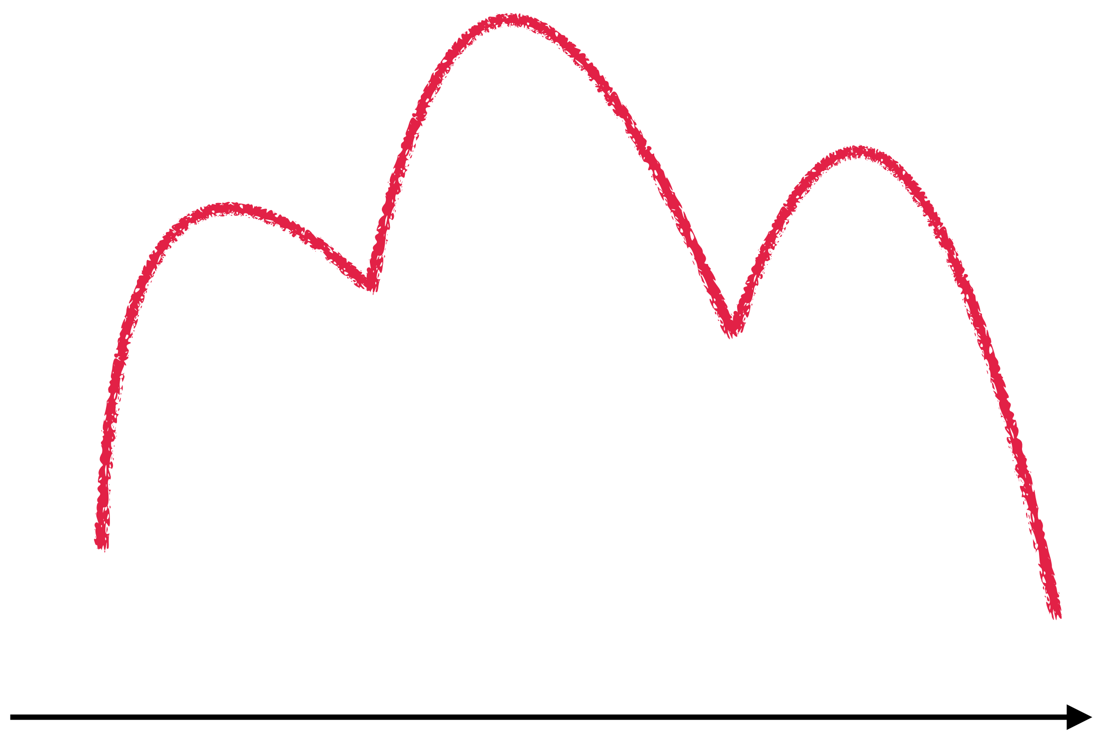
次に行うのは予測の評価だろう。

しかし、真の分布、データを支配しているルールを知りたいけど知りようがない。

むしろ、知れるなら統計的に予測なんてしなくていい。

どうやって、評価したらいいのか、なんの道具を使えばいいのか。。

真の確率分布q(x)



大体p*(x)これやな

この違いを測りたい

汎化損失と呼ぶ

これは直接測れない。。

真の確率分布含んでるし。。

この違いを推定するものがある。

それが**WAIC**

とかAIC

・WAIC

$$W_n = T_n + \frac{\beta V_n}{n}$$

・T_n : 経験損失

$$T_n = -\frac{1}{n}\sum_{i=1}^{n} log\mathbb{E}_w[p(X_i|w)]$$

・V_n : 汎関数分散

$$V_n = \sum_{i=1}^{n} \{\mathbb{E}_w[(logp(X_i|w))^2] - \mathbb{E}_w[logp(X_i|w)]^2\}$$

W_nの期待値が

理論的に汎化損失の期待値と

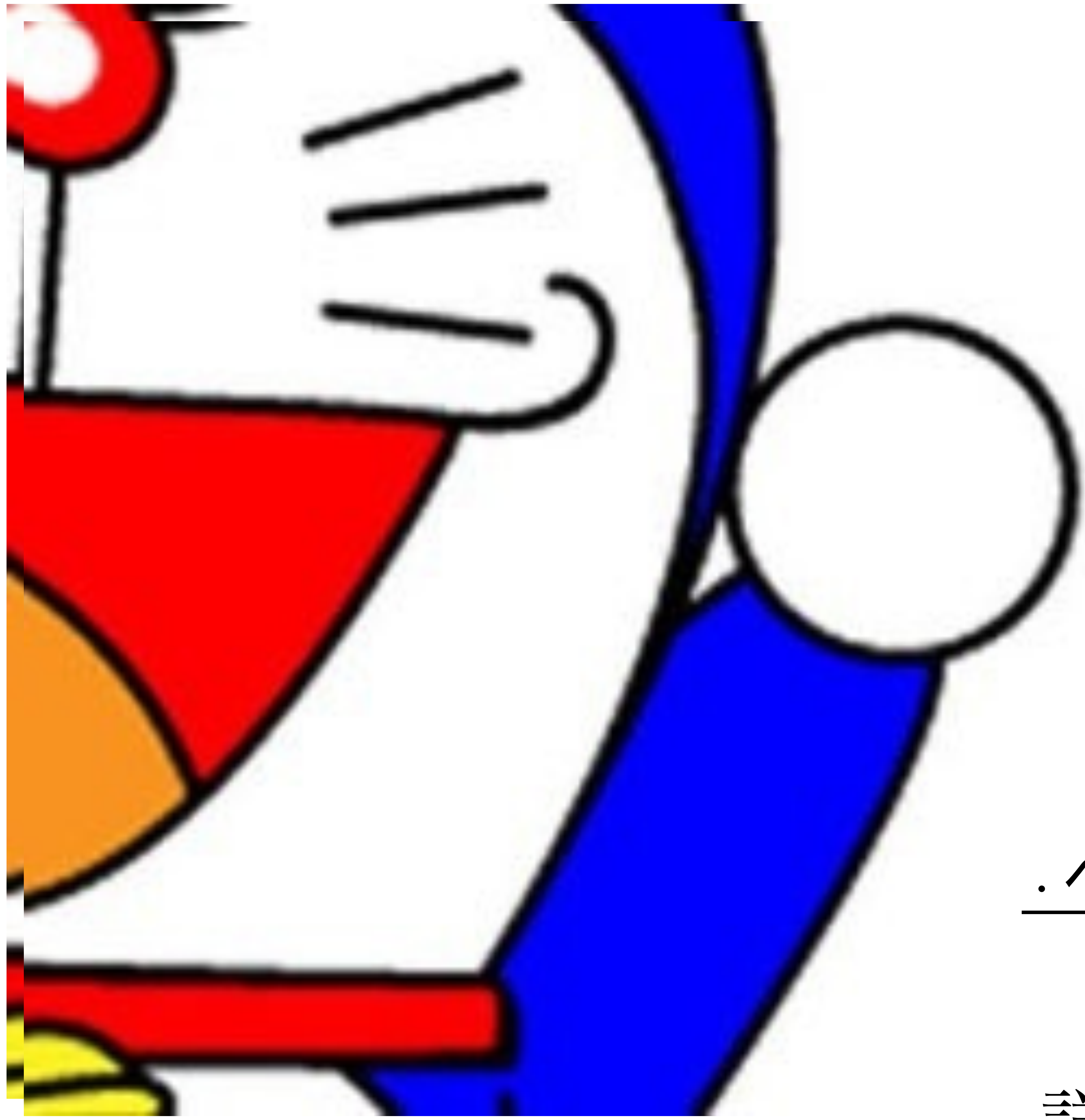
漸近的に一致する

↓

小さければ、

汎化損失の期待値も小さい

いい感じの予測

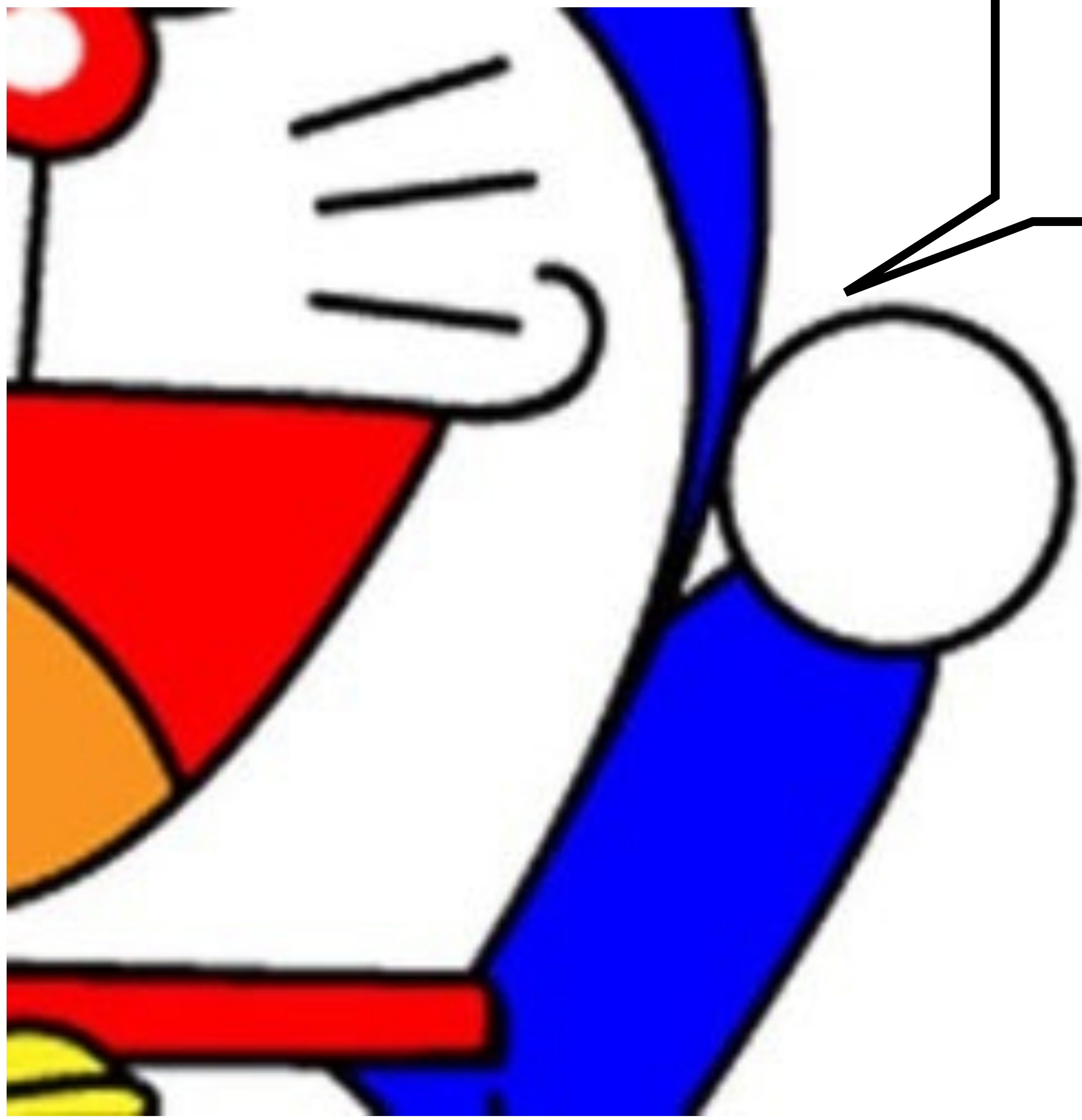


・ベイズ統計の理論と方法

詳しくはこの本を読むか著者のHPへ

書籍名クリックでAmazonへHPクリックでHPへ飛びます

douyatte?



統計モデリング〜

pythonでベイズ推論を行うには大きく二つのパッケージがある。

・pystan

stanファイルに確率モデルを記述してベイズ推論を行う

ドキュメント：https://pystan.readthedocs.io/en/latest/

・PyMC3

pythonのみで完結する。速度はまあ、、

ドキュメント：https://docs.pymc.io/

どちらもMCMCを行なってベイズ推論を行う。

定量的に扱う（予測分布）

分析者

<u>対象を生成してるルールは大体p*(x)だろうと推測する</u>

大体こんな感じ〜

期待値〜

平均値〜

$$p*(x) = \mathbb{E}_w[p(x|w)]$$
$$= \int p(x|w)p(w|X^n)dw$$

再掲

予測分布 := 事後分布によって確率モデルを平均したもの

定量的に扱う（予測分布）

$$
\begin{aligned}
p^*(x) &= \mathbb{E}_w[p(x|w)] \\
&= \int p(x|w)p(w|X^n)dw \\
&= \int p(x|w)\boxed{\frac{\phi(w)\prod_{i=1}^{n} p(X_i)|w)^\beta}{\int_W \phi(w)\prod_{i=1}^{n} p(X_i|w)^\beta dw}}dw
\end{aligned}
$$

学習したパラメータの分だけ多重積分になっている

この部分が複雑で計算が難しいので、**MCMC**を使う

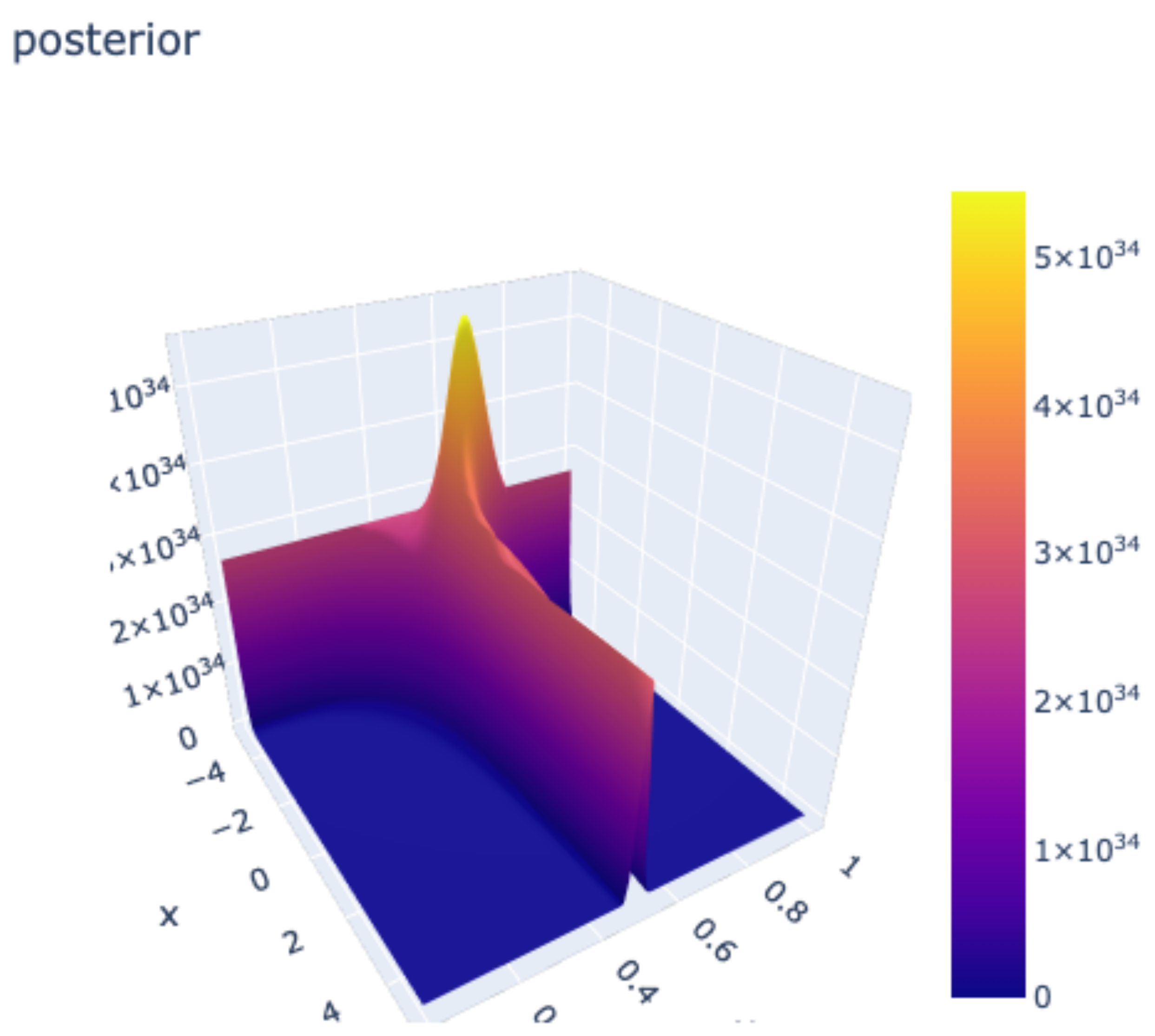
MCMCとは

これには条件があるが本スライドでは割愛

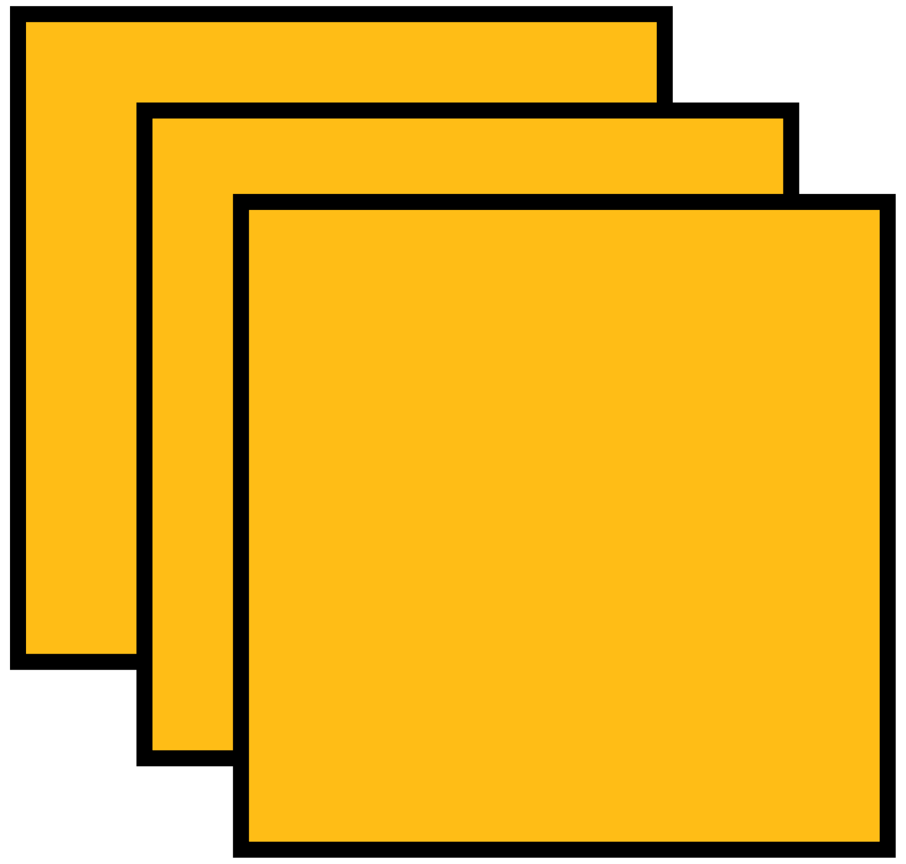
乱数を使って発生させたサンプルが

パラメータの事後分布からのサンプルとみなせるアルゴリズム

パラメータの事後分布は複雑！

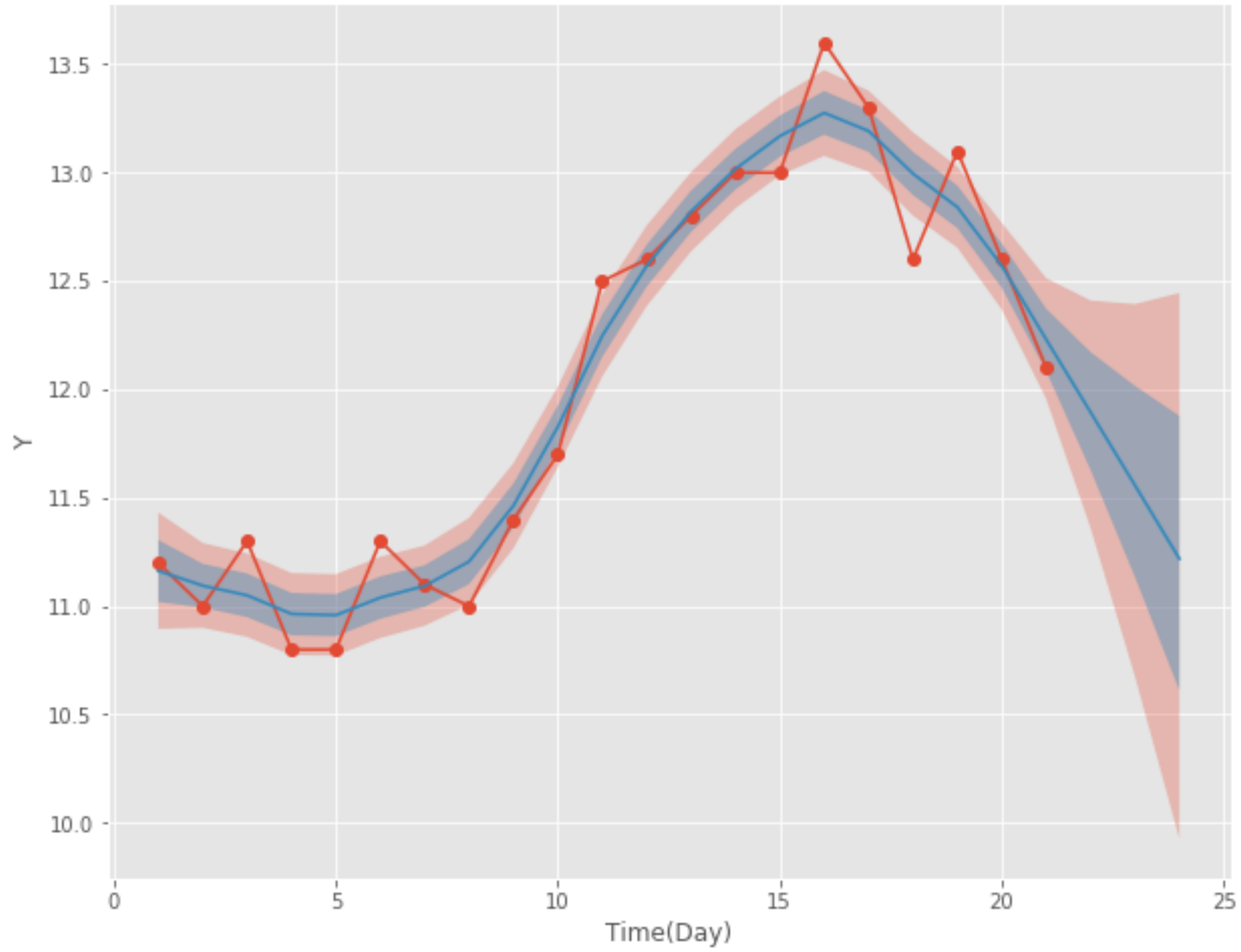


MCMCサンプル



生成プログラムはこちら

MCMCサンプルを得れば、得たサンプルと確率モデルを使って予測分布が作れる



こういうの