第10章ベイズ推定の数値計算法

10.3計算機による擬似乱数の発生法

07t4072f　岡崎　駿

目的

- 別の確率分布から得られる標本を用いて、所望の確率分布に従う標本を直接生成する手法を提案
- モンテカルロ積分を直接的に実行
- 提案手法
  - 逆関数サンプリング法
  - 棄却サンプリング法
  - マルコフ連鎖モンテカルロ法

逆関数サンプリング法(1)

- 1次元一様分布[0,1]に従う確率変数uから任意の1次元の確率密度関数 $p(\theta)$ に従う確率変数 θ を生成する方法

- 確率密度関数 $p(\theta)$ の累積分布関数の逆関数を用いる

逆関数サンプリング法(2)

- $p(\theta)$の累積分布関数

$$P(\theta) := \int_{-\infty}^{\theta} p(u)du$$

- $u = P(\theta)$ の逆関数

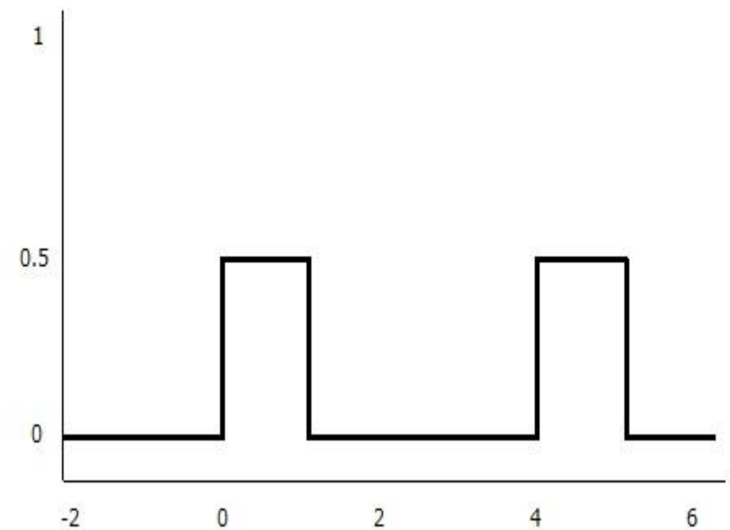
$$\theta = P^{-1}(u)$$

n個の[0,1]上の一様確率分布$\{u_i\}_{i=1}^{n}$ に対して、

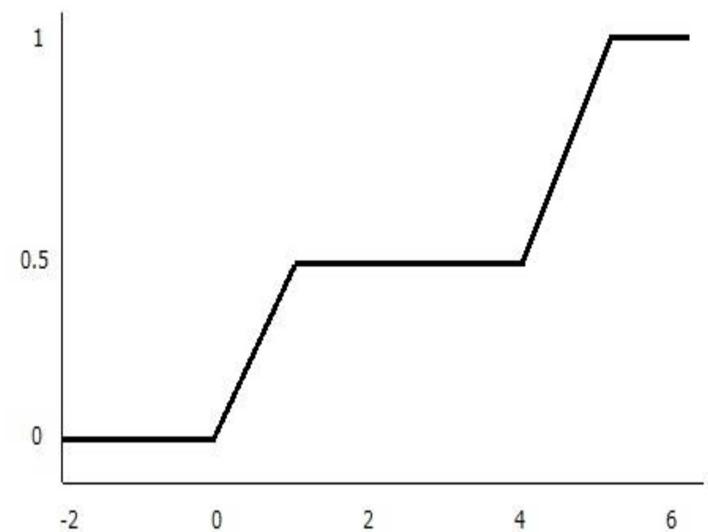
$\theta_i = P^{-1}(u_i)$を求めることで、

$p(\theta)$に従う確率変数が得られる

確率密度関数と累積分布関数

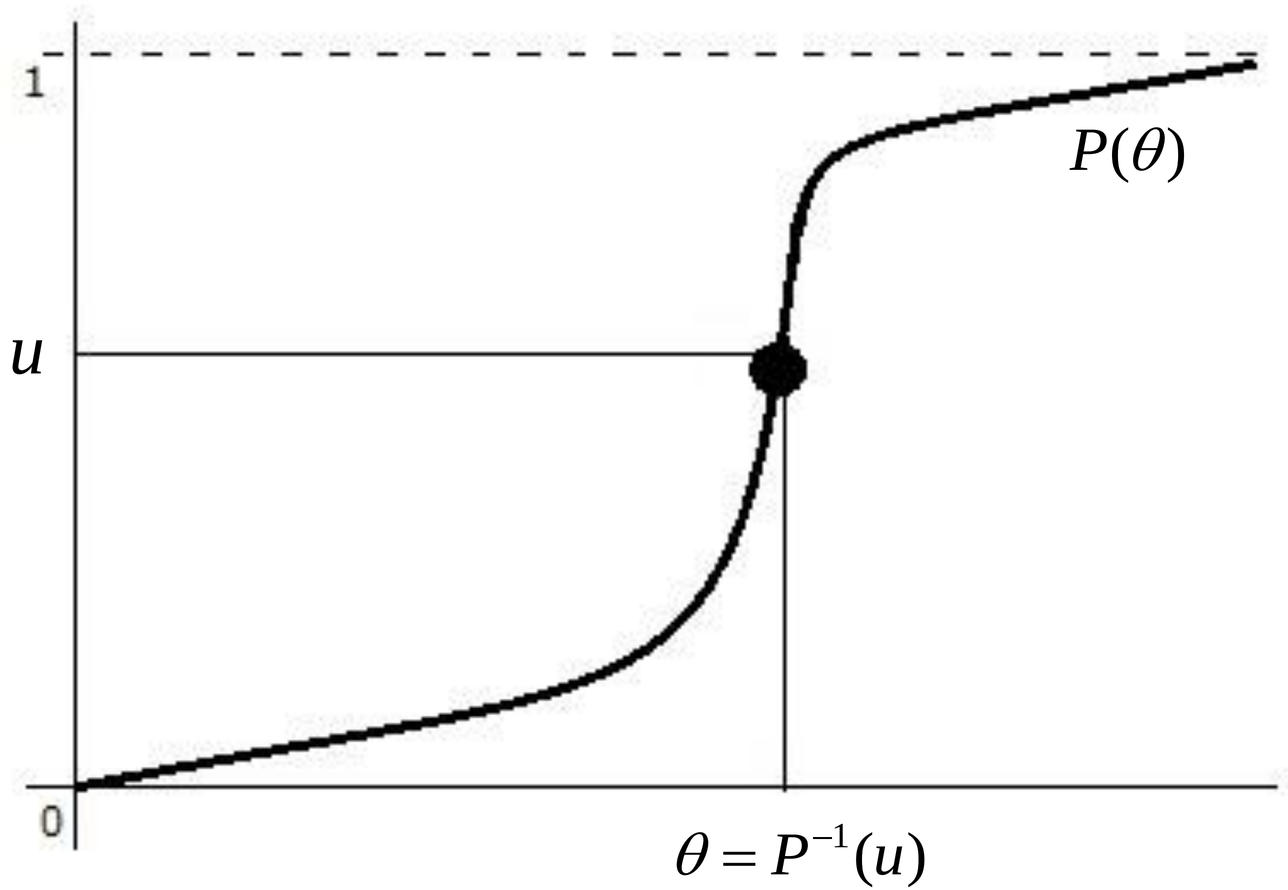


(a)確率密度関数 $p(\theta)$



(b)累積分布関数$P(\theta)$.

逆関数サンプリング法(3)



逆サンプリング法の正当性(1)

$\theta = P^{-1}(u)$より任意の τ に対して

$$\mathrm{P}(\theta \le \tau) = \mathrm{P}(P^{-1}(u) \le \tau)$$

$\theta \le \theta'$ ならば$P(\theta) \le P(\theta')$なので

$$\mathrm{P}(P^{-1}(u) \le \tau) = \mathrm{P}(u \le P(\tau))$$

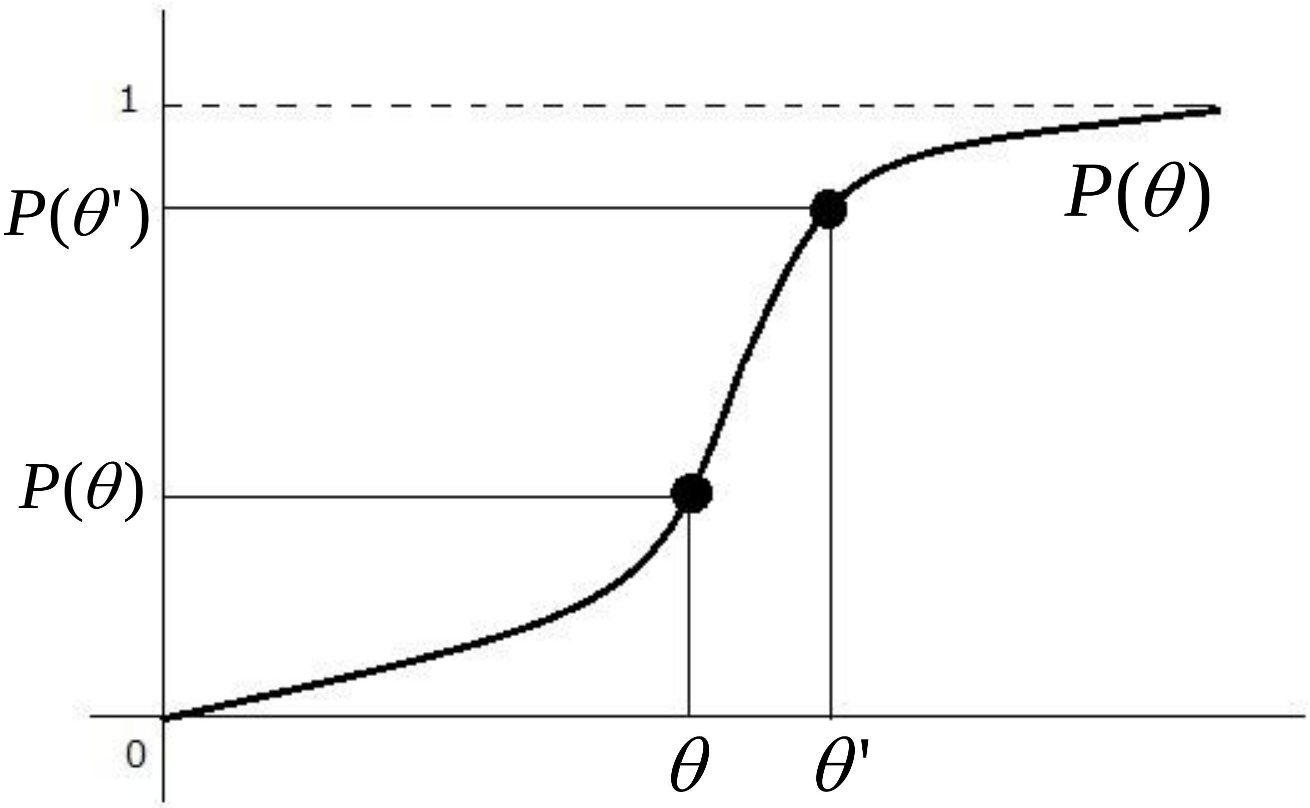
更に、u は[0,1]上の一様分布に従うので

$$\mathrm{P}(u \le P(\tau)) = \int_0^{P(\tau)} du = P(\tau)$$

が成り立ち、

$$\mathrm{P}(\theta \le \tau) = P(\tau)$$

逆サンプリング法の正当性(2)



逆サンプリング法の特徴

- 長所
  - 累積分布関数の逆関数に一様分布に従う確率変数を通すだけで所望の確率分布に従う確率変数が得られる
  - 要するに簡単な手法
- 短所
  - 1次元の確率分布にしか適応できない
  - 逆関数 $\theta = P^{-1}(u)$ を陽に計算する必要あり

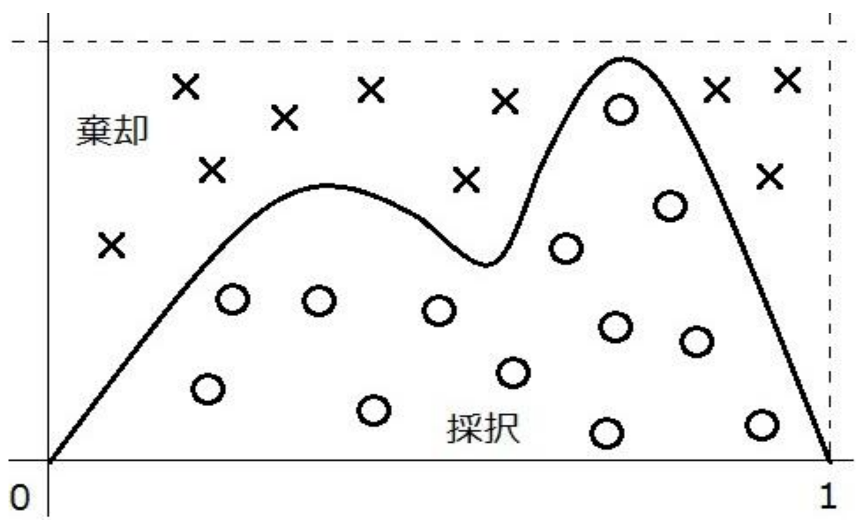
棄却サンプリング法(1)

- 適当な確率密度関数$p'(\theta)$に従う標本を用いて、 $p(\theta)$ に従う標本を生成する方法
- 多次元の確率分布にも適応可能
- $\max_{\theta}\left[\frac{p(\theta)}{p'(\theta)}\right] < \kappa$ を満たす有限のκ が存在し、その値が既知である必要がある

棄却サンプリング法(2)

$i=1,\ldots,n$に対して、以下を繰り返す

1. 代理分布$p'(\theta)$から提案点θ'を発生させる.
2. $[0,\kappa]$上の一様分布から指標vを発生させる.
3. もし$v > p(\theta')/p'(\theta')$ならば提案点$\theta'$を棄却し、1に戻る.
4. 提案点θ'を採択し、$\theta_i \leftarrow \theta'$とする.
5. カウンタを増やす：$i \leftarrow i+1$



棄却サンプリング法の特徴

- 長所
  - ユーザがあらかじめ計算するのは κ だけで良い
  - 代理分布 $p'(\theta)$ から標本を発生させるだけで簡単に所望の確率分布 $p(\theta)$ に従う確率変数が得られる

- 短所
  - $p(\theta)$ の形によっては提案点 θ' が採択される確率が低い（ $p(\theta)$ に依存しやすい）
  - n個の標本を採択するまでに時間がかかる

マルコフ連鎖モンテカルロ法

- 制限のない汎用的な手法
- マルコフ連鎖モンテカルロ法の1種である「メトロポリス・ヘイスティングス法」の紹介

独立に提案点を発生（棄却サンプリング法）

マルコフ連鎖を用いて提案点を逐次的に発生

(メトロポリス・ヘイスティングス法)

※代理分布が現在の標本 θ_i に依存

メトロポリス・ヘイスティングス法(1)

代理分布：$p'(\theta|\theta')$

生成された提案点θ'に対して[0,1]上の一様分布に従う指標υを用いて提案点θ'が適切かどうかを判定する.

もし、

$$\upsilon \leq \frac{p(\theta')p'(\theta_i|\theta')}{p(\theta_i)p'(\theta'|\theta')}$$

ならば提案点θ'を採択し、

$$\theta_{i+1} = \theta'$$

とする一方、

$$\upsilon > \frac{p(\theta')p'(\theta_i|\theta')}{p(\theta_i)p'(\theta'|\theta')}$$

ならば提案点θ'を棄却し、現在の値を保持する.

$$\theta_{i+1} = \theta_i$$

メトロポリス・ヘイスティングス法(2)

- 注意点
  - 初期値 θ_0 はユーザが適当に決定する。
  - メトロポリス・ヘイスティングス法で得られる標本 $\{\theta_i\}_{i=1}^n$ は $n \to \infty$ の極限での平衡分布が確率密度 $p(\theta)$ を持つ。
  - ユーザの定める初期値に依存することを回避するために、標本 $\{\theta_i\}_{i=1}^n$ のうち最初の適当な割合を捨てる方法もある。