MAT02034 - Métodos bayesianos para análise de dados

Introdução a computação bayesiana (continuação)

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2022

Introdução

Introdução

- Nos exemplos das aulas anteriores, fomos capazes de produzir amostras simuladas diretamente da distribuição *a posteriori*, uma vez que as distribuições possuíam formas funcionais familiares.
- Poderíamos obter estimativas de **Monte Carlo** da média *a posteriori* para qualquer função dos parâmetros de interesse.
- Mas em muitas situações, a distribuição *a posteriori* não tem uma forma familiar e precisamos usar um algoritmo alternativo para produzir uma amostra simulada.

Amostragem por rejeição

Amostragem por rejeição

- Um algoritmo geral para simular realizações (aleatórias) de uma dada distribuição de probabilidade é a **amostragem de rejeição**[1].
- Suponha que desejamos produzir uma amostra independente de uma densidade *a posteriori* $g(\theta|y)$ em que a constante de normalização **pode não ser conhecida**.
- O primeiro passo na amostragem de rejeição é encontrar outra densidade de probabilidade $p(\theta)$ tal que:
  - É fácil simular realizações de $p(\theta)$.
  - A densidade $p(\theta)$ assemelha-se à densidade *a posteriori* de interesse $g(\theta|y)$ em termos de localização e dispersão.
  - Para todo θ e uma constante c, $g(\theta|y) \leq cp(\theta)$.

[1] Ou **método da rejeição**, ou ainda, **método da aceitação/rejeição**.

Amostragem por rejeição

Suponha que sejamos capazes de encontrar uma densidade $p(\theta)$ com essas propriedades.

Em seguida, obtém-se as realizações de $g(\theta|y)$ usando o seguinte algoritmo de aceitação/rejeição:

1. Simule independentemente θ de $p(\theta)$ e uma variável aleatória uniforme U no intervalo unitário.
2. Se $U \leq \frac{g(\theta|y)}{cp(\theta)}$, então **aceite** θ como uma realização da densidade $g(\theta|y)$; caso contrário, **rejeite** θ.
3. Continue os passos 1 e 2 do algoritmo até que tenha coletado um número suficiente de θ "aceitos".

Amostragem por rejeição

- Por que esse método funciona? Um cálculo de probabilidade simples mostra que função distribuição acumulada (fda) da variável aleatória aceita, $\Pr\left(\theta_c \leq x | U \leq \frac{g(\theta|y)}{cp(\theta)}\right)$, é exatamente o fda de θ.

Amostragem por rejeição

- A amostragem por rejeição é um dos métodos mais úteis para simular realizações de uma variedade de distribuições.
- A principal tarefa no planejamento de um algoritmo de amostragem por rejeição é encontrar uma **densidade proposta** adequada $p(\theta)$ e uma constante c.
- Na etapa 2 do algoritmo, a probabilidade de aceitar um **candidato** é dada por $\frac{g(\theta|y)}{cp(\theta)}$.
  - Pode-se monitorar o algoritmo calculando a proporção de candidatos que são aceitos; um algoritmo de amostragem por rejeição eficiente tem uma alta taxa de aceitação.
- A escolha ótima para c é $\sup_{\theta}\{g(\theta|y)/p(\theta)\}$, mas mesmo essa escolha pode resultar em um número indesejavelmente grande de rejeições[2].

[2] Um algoritmo que rejeita muitos candidatos é pouco eficiente, e pode resultar em custo de tempo.

Amostragem por rejeição: exemplo

- Suponha que estamos estudando a distribuição do número de defeituosos X na produção diária de um produto.
- Considere o modelo $(X|Y,\theta) \sim \mathit{binomial}(Y,\theta)$, em que Y, a produção de um dia, é uma **variável aleatória** com uma distribuição de Poisson com média conhecida λ, e θ é a probabilidade de que qualquer produto seja defeituoso.
- A dificuldade, no entanto, é que Y não é observável, e a inferência deve ser feita apenas com base em X.
- A distribuição *a priori* é tal que $(\theta|Y=y) \sim \mathit{Beta}(\alpha,\gamma)$, com α e γ conhecidos independentes de Y.

Amostragem por rejeição: exemplo

- A análise bayesiana aqui não é um problema particularmente difícil porque a distribuição *a posteriori* $\theta|X = x$ pode ser obtida da seguinte forma.
- Primeiro, observe que $X|\theta \sim Poisson(\lambda\theta)$[3]. Em seguida, $\theta \sim Beta(\alpha, \gamma)$. Portanto,

$$g(\theta|X = x) \propto \exp(-\lambda\theta)\theta^{x+\alpha-1}(1-\theta)^{\gamma-1},\ 0 \leq \theta \leq 1.$$

- A única dificuldade é que esta não é uma **distribuição padrão** e, portanto, as quantidades *a posteriori* não podem ser obtidas de forma fechada.

[3] Lembre que $\Pr(X = x|\theta, \lambda) = \sum_{y=0}^{\infty} \Pr(X = x, Y = y|\theta, \lambda) = \sum_{y=0}^{\infty} \Pr(X = x|Y = y, \theta, \lambda) \times \Pr(Y = y|\theta, \lambda)$.

Amostragem por rejeição: exemplo

- Observando $g(\theta|X = x)$, uma boa escolha para a distribuição proposta $p(theta)$ deve ser a densidade de $Beta(x + \alpha, \gamma)$.
- Em nosso exemplo, suponha $X = 1$, $\alpha = 1$, $\gamma = 49$ e $\lambda = 100$.
- Note que $c = \sup_{\theta \in [0,1]}\{\exp(-100\theta)\} = 1$.
- Ainda, $\frac{g(\theta|X=x)}{cp(\theta)} = \exp(-100\theta)$.

Amostragem por rejeição: exemplo

```
metodo_rejeicao <- function(x, alpha, gamma, lambda, j = 0){
  repeat{
    j <- j + 1
# 1. Gerar de theta candidado de p e u de U(0,1)
  u <- runif(n = 1, 0, 1)
  theta_cand <- rbeta(
    n = 1, shape1 = x + alpha, shape2 = gamma)
# 2. Aceita theta candidato se u menor ou igual g/cp
    if (u <= exp(-lambda * theta_cand))
      return(c(theta_cand, j))
  }
}
```

Amostragem por rejeição: exemplo

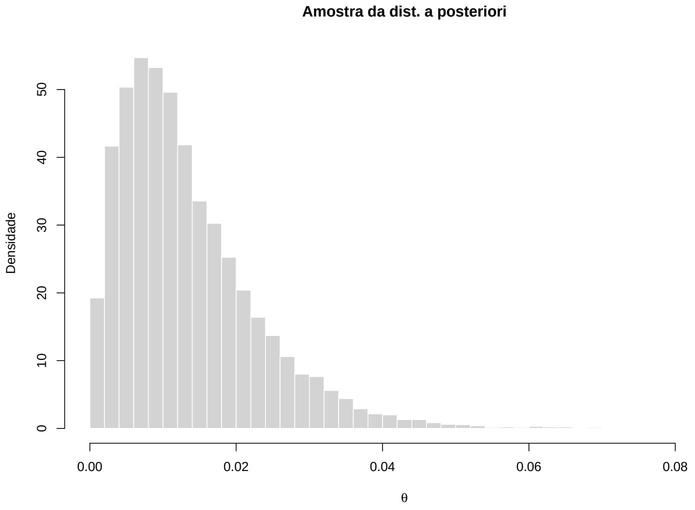
```
M_alvo <- 10000 # Número de amostras
theta <- c()
M_gerados <- c()

# Loop
for (i in 1:M_alvo){
  aux <- metodo_rejeicao(x = 1, alpha = 1,
                         gamma = 49, lambda = 100)
  theta[i] <- aux[1]
  M_gerados[i] <- aux[2]
}

taxa_aceita <- M_alvo/sum(M_gerados)
taxa_aceita
```

```
## [1] 0.1100461
```

Amostragem por rejeição: exemplo



Amostragem por rejeição: exemplo

- A partir desta amostra podemos obter estimar a média *a posteriori* de θ, intervalos de credibilidade, e probabilidades *a posteriori*.
- O método da rejeição pode ser utilizado para obter amostras de distribuições de qualquer dimensão (desde que p seja uma densidade no mesmo espaço que g).

Amostragem por importância

Amostragem por importância

- Voltemos ao problema básico de calcular uma integral na inferência bayesiana.
- Em muitas situações, a constante de normalização da densidade *a posteriori* $g(\theta|y)$ será desconhecida, então a média posterior da função $h(\theta)$ será dada pela razão de integrais

$$\mathrm{E}\left[h(\theta)|y\right] = \frac{\int h(\theta)g(\theta|y)d\theta}{\int g(\theta|y)d\theta}.$$

Amostragem por importância

- Se pudéssemos simular uma amostra $\{\theta^j\}$ diretamente da densidade *a posteriori* $g(\theta|y)$, poderíamos aproximar esse valor esperado por uma estimativa de Monte Carlo.
- No caso em que não podemos gerar uma amostra diretamente de $g(\theta|y)$, suponha que podemos construir uma densidade de probabilidade $p(\theta)$ que podemos simular e que se aproxime da densidade *a posteriori* $g(\theta|y)$.
- Reescrevemos a média *a posteriori* como

$$\mathrm{E}\left[h(\theta)|y\right] = \frac{\int h(\theta)\frac{g(\theta|y)}{p(\theta)}p(\theta)d\theta}{\int \frac{g(\theta|y)}{p(\theta)}p(\theta)d\theta} = \frac{\int h(\theta)w(\theta)p(\theta)d\theta}{\int w(\theta)p(\theta)d\theta},$$

em que $w(\theta) = g(\theta|y)/p(\theta)$ é a **função peso**.

Amostragem por importância

- Se $\theta^1, \ldots, \theta^M$ são uma amostra simulada da densidade de aproximação $p(\theta)$, então a **estimativa de amostragem por importância** da média *a posteriori* é

$$\overline{h}_{AI} = \frac{\sum_{j=1}^{M} h(\theta^j) w(\theta^j)}{\sum_{j=1}^{M} w(\theta^j)}.$$

- Esta é chamada de **estimativa de amostragem por importância** porque estamos amostrando valores de θ que são importantes no cálculo das integrais no numerador e no denominador.

Amostragem por importância

- Como na amostragem por rejeição, a principal questão ao planejar uma boa estimativa de amostragem por importância é encontrar uma densidade de amostragem adequada $p(\theta)$.
- Essa densidade deve ser de uma forma funcional familiar para que as realizações simuladas estejam disponíveis.
- A densidade deve imitar a densidade *a posteriori* $g(\theta|y)$ e ter caudas relativamente planas (*flat*) para que a função peso $w(\theta)$ seja limitada por cima.
- Pode-se monitorar a escolha de $p(\theta)$ inspecionando os valores dos pesos simulados $w(\theta^j)$.
    - Se não houver pesos muito grandes, é provável que a função de peso seja limitada e o amostrador de importância esteja fornecendo uma estimativa adequada.

Amostragem por importância

Retornando ao exemplo:

```
amostragem_importancia <- function(x, alpha, gamma, lambda, M = 1000){

  # 1. Gerar de theta de p
  theta <- rbeta(n = M, shape1 = x + alpha, shape2 = gamma)
  # 2. Calcula pesos de importância
  w.theta <- exp(-lambda * theta)
  # 3. Estimativa da média a posteriori
  media_post <- weighted.mean(x = theta, w = w.theta)
  return(media_post)
}

amostragem_importancia(x = 1, alpha = 1,
                       gamma = 49, lambda = 100, M = 10000)
```

```
## [1] 0.01352942
```

Sampling Importance Resampling

Sampling Importance Resampling

- Na amostragem por rejeição, simulamos realizações a partir de uma proposta de densidade $p(\theta)$ e aceitamos um subconjunto desses valores para serem distribuídos de acordo com a densidade *a posteriori* de interesse $g(\theta|y)$.
- Existe um método alternativo de obtenção de uma amostra simulada a partir da densidade *a posteriori* $g(\theta|y)$ motivada pelo algoritmo de amostragem por importância.

Sampling Importance Resampling

- Como antes, simulamos M realizações de θ a partir da densidade proposta $p(\theta)$ denotada por $\theta^1, \ldots, \theta^M$ e calculamos os pesos $\{w(\theta^j) = g(\theta^j|y)/p(\theta^j)\}$.
- Agora, convertemos os pesos em probabilidades usando a fórmula

$$p^j = \frac{w(\theta^j)}{\sum_{j=1}^{M} w(\theta^j)}.$$

Sampling Importance Resampling

- Suponha que tomemos uma nova amostra $\theta^{*1}, \ldots, \theta^{*M}$ da distribuição discreta sobre $\theta^1, \ldots, \theta^M$ com respectivas probabilidades $p^1, \ldots, p^m$.
- Então os $\{\theta^{*j}\}$ serão aproximadamente distribuídos de acordo com a distribuição *a posteriori* $g(\theta|y)$.
- Este método, chamado de ***sampling importance resampling*** **(SIR)**[4], é um procedimento de ***bootstrap*** **ponderado** em que **amostramos com reposição** da amostra $\{\theta^j\}$ com probabilidades de amostragem desiguais.

[4] Alguns autores traduziram o termo para o português como **reamostragem ponderada**.

Sampling Importance Resampling

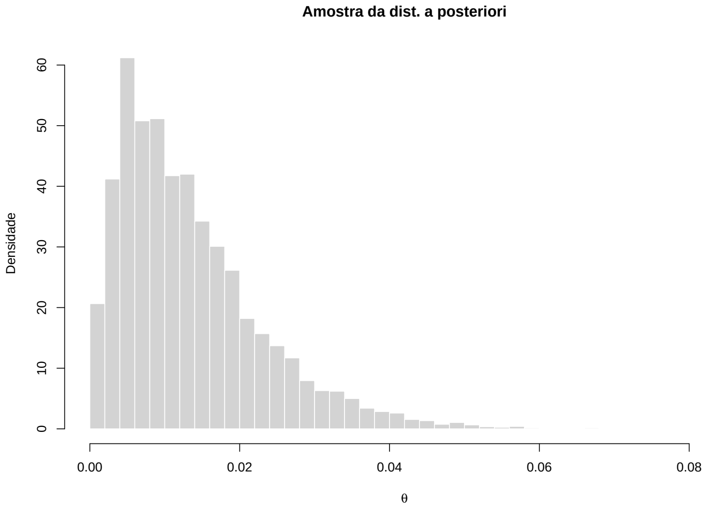
Retornando ao exemplo:

```
amostragem_sir <- function(x, alpha, gamma, lambda, M = 1000){

  # 1. Gerar de theta de p(theta)
  theta <- rbeta(n = M, shape1 = x + alpha, shape2 = gamma)
  # 2. Calcula pesos de importância
  w.theta <- exp(-lambda * theta)
  # 3. Converte pesos em probabilidades
  p <- w.theta/sum(w.theta)
  # 4. Reamostra theta de acordo com p
  ind <- sample(x = 1:M, size = M, replace = TRUE, prob = p)
  theta_s <- theta[ind]
  return(theta_s)
}

theta_estrela <- amostragem_sir(x = 1, alpha = 1,
                                gamma = 49, lambda = 100, M = 10000)
```

Sampling Importance Resampling



Para casa

- Rodar os códigos dos exemplos de aula.
  - Trazer as dúvidas para o Fórum Geral do Moodle e para a próxima aula.
- Exercício (Moodle).

Próxima aula

- Métodos de Monte Carlo via Cadeias de Markov.

Por hoje é só!

Bons estudos!