

Lista de Exercícios 5

GUSTAVO L. GILARDONI

Os exercícios numerados em verde são optativos.

1. Usando o método de Monte Carlo, obtenha uma aproximação para $\Gamma(2.5) = \int_0^\infty x^{1.5} e^{-x} dx$ com precisão inferior a 0.005 e confiança 95%.

2. Considere $n = 12$ ensaios de Bernoulli com probabilidade de sucesso p e $y = 9$ sucessos. Suponha que a distribuição a priori de p é especificada de forma que $\eta = \log[p/(1-p)]$ segue uma distribuição Normal com média $\mu = 0$ e variância $\sigma^2 = 1$. Obtenha aproximações para $E(p|y = 9)$, $DP(p|y = 9)$ e $\Pr(p > \frac{1}{2}|y = 9)$. Repita o exercício para os casos que $\sigma^2 = 4$ e 9 e compare com o caso $\sigma^2 = 1$.

3. (Amostragem por rejeição). Seja $p(\theta | D) \propto p(\theta) p(D | \theta)$ uma densidade *alvo* (da qual queremos amostrar) e $g(\theta)$ uma outra densidade *de proposta* tal que existe uma constante M tal que $p(\theta | D)/g(\theta) \leq M$. Considere o seguinte algoritmo:

1. Gere $\theta \sim g(\theta)$,
2. Gere $u \sim \text{Uniforme}(0,1)$,
3. Se $u < p(\theta | D)/[M g(\theta)]$, pare e aceite o valor de θ . Caso contrário, volte ao passo 1 (isto é, gere novamente $\theta \sim g(\theta)$, $u \sim \text{Uniforme}(0,1)$ etc.)

(i) Mostre que o valor de θ obtido no final do algoritmo tem distribuição $p(\theta | D)$. (ii) Mostre que para implementar esse algoritmo não é necessário conhecer a constante de normalização em $p(\theta | D)$ (isto é, é suficiente conhecer o núcleo dessa densidade, $p(\theta) p(D | \theta)$).

4. Use o exercício anterior para gerar uma amostra de tamanho $n = 100$ da distribuição cuja densidade é proporcional a $\sin(x)$ para $0 < x < \pi$.

5. Para $i = 1, \dots, n$ considere observações independentes $s_i | \lambda_i \sim \text{Poisson}(\lambda_i t_i)$, onde os tempos de observação t_i são fixos. Suponha que $\lambda_i | \beta \sim \text{iid Gama}(\alpha_0, \beta)$ e que $\beta \sim \text{Gama}(a, b)$.

(i) Calcule a distribuição condicional de $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) | D, \beta$. (ii) Considere a implementação de amostragem por importância amostral (a) no modelo reduzido $p(s_1, \dots, s_n; \beta) = p(s_1, \dots, s_n | \beta) p(\beta)$ obtido após integrar $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ com distribuição de proposta $p^*(\beta)$ e pesos de importância $w_i \propto p(\beta_i | D)/p^*(\beta_i)$ e (b) no modelo completo $p(s_1, \dots, s_n; \lambda_1, \dots, \lambda_n; \beta) = p(s_1, \dots, s_n | \lambda_1, \dots, \lambda_n) p(\lambda_1, \dots, \lambda_n | \beta) p(\beta)$ com distribuição de proposta $p^*(\lambda_1, \dots, \lambda_n; \beta) = p^*(\beta) p(\lambda_1, \lambda_n | D, \beta)$. Mostre que no caso (b) os pesos de importância são idênticos ao caso (a) e use esse resultado para aproximar $E(\lambda_i | D)$ e $\text{Var}(\lambda_i | D)$ por importância amostral para os dados sobre falhas em linhas de bombeio disponíveis nas notas de aula (ou em Gaver e O'Muircheartaigh, 1987, *Technometrics*, Vol. 29, pags. 1-15) usando $\alpha_0 = 0.166$, $a = 0.1$ e $b = 0.01$.

6. Obtenha as aproximações do Exercício 5 usando o amostrador de Gibbs.

7. Para os dados do Exemplo 6 das notas de aula, considere o modelo de regressão probit com uma única covariável e distribuição a priori não informativa $p(\beta) \propto 1$ e obtenha aproximações dos momentos a posteriori de β usando o amostrador de Gibbs.
8. Considere a distribuição a posteriori $(\theta_1, \theta_2, \theta_3) \sim \text{Dirichlet}(\alpha_1 = 468, \alpha_2 = 316, \alpha_3 = 110)$, discutida no contexto de uma pesquisa eleitoral no exercício 9 da Lista 2. Deseja-se obter as estimativas sob PQ para a diferença da intenção de voto entre o primeiro e o segundo candidato $(\theta_1 - \theta_2)$ e para a variância dessa diferença. (i) Obtenha as estimativas exatas usando as propriedades da distribuição de Dirichlet; (ii) Obtenha aproximações para essas estimativas usando (a) uma amostra iid de tamanho $M = 10000$ da distribuição a posteriori, (b) $M = 10000$ ciclos do amostrador de Gibbs e (c) o algoritmo de Metropolis–Hastings também com $M = 10000$. Comente os resultados obtidos.