

## Lista de exercícios 2 - métodos MCMC e convergência

Exercícios dos slides e comentados em sala de aula, também fazem parte dessa lista de exercícios.

**Exercício 1:** Utilize o método GS para simular amostras da distribuição Normal multivariada com  $\mu = (-2, 0, 2)^T$  e  $\Sigma = \begin{pmatrix} 1 & 0.7 & -0.2 \\ 0.7 & 1 & 0.5 \\ -0.2 & 0.5 & 1 \end{pmatrix}$ . Verifique a convergência, a autocorrelação, o valor ideal de *burn in* e saltos.

**Exercício 2:** Utilize o método GS para simular amostras da distribuição Multinomial( $n = 100, \mathbf{p} = (0.15, 0.30, 0.25, 0.30)$ ). Verifique a convergência, a autocorrelação, o valor ideal de *burn in* e saltos.

**Exercício 3:** Conduza a estimação Bayesiana do modelo de regressão logística, vista em aula, usando o algoritmo MH com propostas independentes. Você conseguiu resultados melhores ou piores do que os vistos em aula? Justifique. Verifique a convergência e a eficiência do método e compare com o algoritmo visto em sala.

**Exercício 4:** A tabela abaixo mostra a frequência observada  $y_i$  do grupo de sangue observado para uma amostra de  $n = 435$  indivíduos. Aqui os índices  $i = 1, 2, 3, 4$  indexam, respectivamente, os grupos de sangue O, A, B, AB. As probabilidades  $p_i$  são determinadas pelas leis genéticas, com  $p, q$  e  $r$  sendo as probabilidades dos genes do tipo A, B e O, respectivamente, com  $p + q + r = 1$ .

Tabela 1: Dados do exercício 4.			
$i$	Grupo de sangue	Frequência $y_i$	Probabilidade $p_i$
1	O	176	$r^2$
2	A	182	$p^2 + 2pr$
3	B	60	$q^2 + 2qr$
4	AB	17	$2pq$

1. Encontre a função de verossimilhança  $p(\mathbf{y}|\boldsymbol{\theta})$  para  $\boldsymbol{\theta} = (p, q)$  e usando  $p + q + r = 1$ ;
2. O grupo de sangue observado depende do genótipo que não pode ser diretamente observado. Siga a relação entre genótipo de grupo de sangue abaixo:

$k$	Grupo de sangue	Genótipo	Probabilidade $p_i$
1	O	OO	$r^2$
2	A	AA	$p^2$
3	A	AO	$2pr$
4	B	BB	$q^2$
5	B	BO	$2qr$
6	AB	AB	$2pq$

Seja  $z_j = 1, 2, 3, 4, 5, 6$  o genótipo não observável do indivíduo  $j$  e seja  $\mathbf{z} = (z_1, \dots, z_n)$ .

Escreva a função de verossimilhança aumentada desses dados,  $p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})$ .

3. Proponha um algoritmo GS para simular da distribuição a posteriori de  $(\boldsymbol{\theta}, \mathbf{z})$  e conduzir a inferência.