

Relatório

Bruno Carvalho Silva Ribeiro

1 Requisitos da tarefa

1.1 Tarefa 1

- $n = 200$ vs $n = 1000$ Bayesiana
- Tabelas $n = 200$ e $n = 1000$
- Gráfico plotrix

1.2 Tarefa 2

- $n = 200$
- Bayesiano vs frequentista
- Tabelas bayes x Frequentista

1.3 Tarefa 3

- Desbalanceamento de $x[2]$ (covariável x_1) Bayesiana
- $n = 200$ prob = 0.5 e $n = 200$ prob = 0.1

1.4 Tarefa 4

- Usar a mesma matriz X e os mesmos valores reais
- $n = 200$
- link logit vs link probit
- tabelas logit vs probit
- plotrix logit vs probit

Table 1: Tabelas com resumo das configurações de cada cenário simulado

Índice na lista	Tamanho da amostra	Probabilidade de Sucesso	É link logit?
1	200	0.5	sim
2	1000	0.5	sim
3	200	0.1	sim
4	200	0.5	não

A tabela abaixo mostra a configuração de como foram gerados os dados para os cenários que usaremos nas tarefas.

Vale ressaltar que a probabilidade de sucesso dessa tabela se refere a covariável x_1 ou, no R, a segunda coluna da matriz de covariáveis ($x[,2]$), uma vez que ela foi gerada de uma distribuição bernoulli.

2 Panorama geral das estimações bayesianas

Table 2: Tabelas com as estimações dos 4 cenários de estudo

(a) Tamanho de amostra 200								(b) Tamanho de amostra 1000							
	true	mean	median	s.d.	HPD_inf	HPD_sup	Amplitude		true	mean	median	s.d.	HPD_inf	HPD_sup	Amplitude
beta0	1.5	1.5381	1.5278	0.2863	1.0181	2.1401	1.1220	beta0	1.5	1.2586	1.2570	0.1172	1.0298	1.4775	0.4477
beta1	0.5	0.2558	0.2556	0.3888	-0.5161	1.0021	1.5182	beta1	0.5	0.6575	0.6545	0.1797	0.3142	1.0093	0.6950
beta2	-0.5	-1.3068	-1.3161	0.3517	-1.9568	-0.5844	1.3724	beta2	-0.5	-0.6021	-0.6005	0.1444	-0.8812	-0.3167	0.5645
beta3	1.0	0.9274	0.9236	0.3460	0.3043	1.6981	1.3937	beta3	1.0	1.0313	1.0224	0.1563	0.7312	1.3336	0.6025
beta4	-1.0	-0.8458	-0.8532	0.3658	-1.5743	-0.1553	1.4191	beta4	-1.0	-0.8299	-0.8288	0.1470	-1.1115	-0.5444	0.5671

(c) Tamanho de amostra 200 prob = 0.1								(d) Tamanho de amostra 200 probit							
	true	mean	median	s.d.	HPD_inf	HPD_sup	Amplitude		true	mean	median	s.d.	HPD_inf	HPD_sup	Amplitude
beta0	1.5	1.3245	1.3245	0.1966	0.9450	1.7008	0.7558	beta0	1.5	1.5794	1.5745	0.2230	1.1280	1.9812	0.8533
beta1	0.5	0.0924	0.0616	0.7552	-1.3376	1.5234	2.8610	beta1	0.5	0.1007	0.0982	0.2810	-0.4395	0.6419	1.0814
beta2	-0.5	-1.2774	-1.2790	0.3329	-1.9522	-0.6552	1.2970	beta2	-0.5	-0.6428	-0.6405	0.2559	-1.1062	-0.0688	1.0374
beta3	1.0	0.8504	0.8468	0.3150	0.2270	1.4480	1.2210	beta3	1.0	0.8421	0.8383	0.2560	0.3477	1.3584	1.0107
beta4	-1.0	-0.7784	-0.7728	0.3327	-1.4055	-0.1468	1.2587	beta4	-1.0	-0.7492	-0.7369	0.2849	-1.3043	-0.1995	1.1048

3 Comparação: Bayesiana vs frequentista

Na comparação que se segue, faremos usando $n = 200$. O intuito é verificar o quão próxima a estimação pelo método bayesiano se encontra do clássico frequentista

Table 3: Tabelas com as estimações dos 4 cenários de estudo

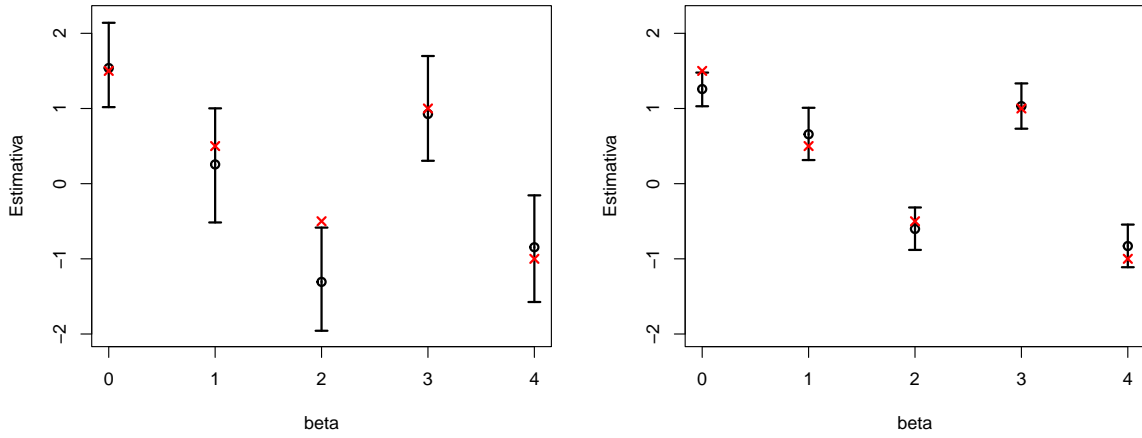
(a) Método Bayesiano

	true	mean	median	s.d.	HPD_inf	HPD_sup	Amplitude
beta0	1.5	1.5381	1.5278	0.2863	1.0181	2.1401	1.1220
beta1	0.5	0.2558	0.2556	0.3888	-0.5161	1.0021	1.5182
beta2	-0.5	-1.3068	-1.3161	0.3517	-1.9568	-0.5844	1.3724
beta3	1.0	0.9274	0.9236	0.3460	0.3043	1.6981	1.3937
beta4	-1.0	-0.8458	-0.8532	0.3658	-1.5743	-0.1553	1.4191

(b) Método Bayesiano

	Estimate	Std. Error	z value	Pr(> z)	IC_inf	IC_sup	Amplitude
beta0	1.4890	0.2835	5.2524	0.0000	0.9334038	2.0446903	1.111287
beta1	0.2585	0.3944	0.6554	0.5122	-0.5145081	1.0315173	1.546025
beta2	-1.2798	0.3487	-3.6697	0.0002	-1.9633173	-0.5962622	1.367055
beta3	0.9005	0.3472	2.5936	0.0095	0.2199989	1.5810668	1.361068
beta4	-0.8154	0.3558	-2.2915	0.0219	-1.5127978	-0.1179688	1.394829

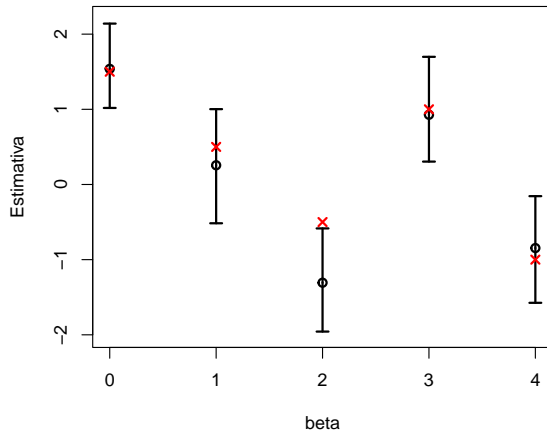
4 Gráficos



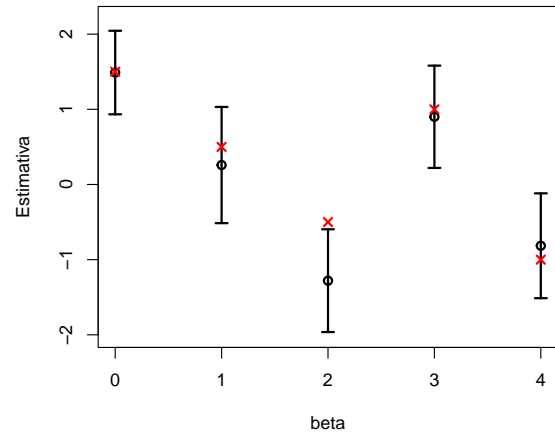
(a) Bayes $n = 200$

(b) Bayes $n = 1000$

Figure 1: Estudo do efeito do tamanho amostral na qualidade da estimação bayesiana

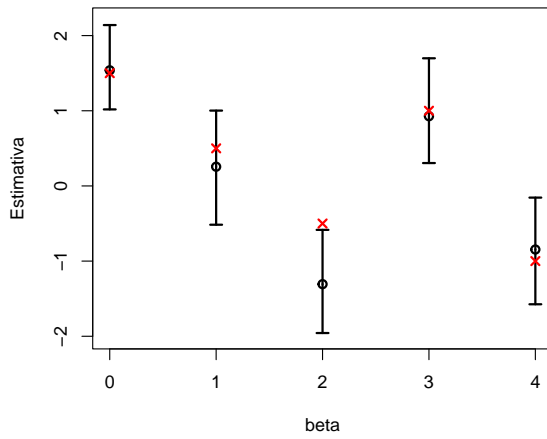


(a) Método Bayesiano

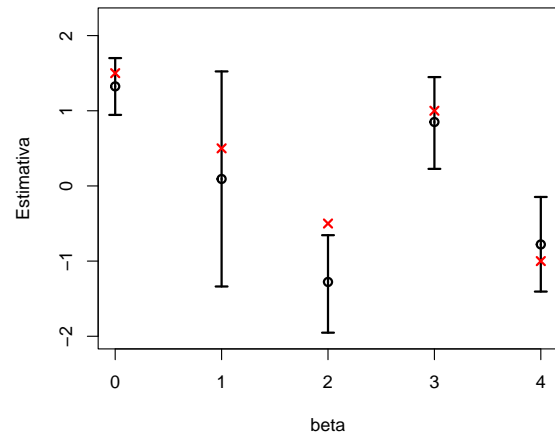


(b) Método Frequentista

Figure 2: Comparação bayesiano vs frequentista

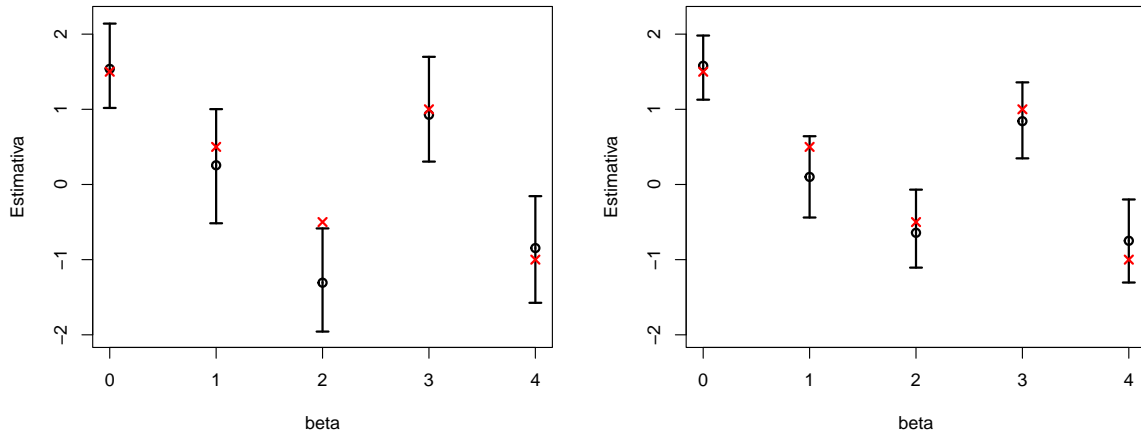


(a) $n = 200$ e $\Pr(X_2=x) = 0.5$



(b) $n = 200$ e $\Pr(X_2=x) = 0.1$

Figure 3: Estudo do desbalanceamento da covariável x1 pelo método bayesiano



(a) $n = 200$ e logit

(b) $n = 200$ e probit

Figure 4: Comparação logit vs probit pelo método bayesiano

5 Observações

Eu achei muito estranho na Table 2 o fato de que para β_0 q quando usamos $n = 200$ a estimação foi de 1.5381 enquanto que, ao aumentarmos o tamanho da amostra para $n = 1000$, a estimação, apenas para esse parâmetro, piorou passando a valer 1.2586. Inicialmente, pensei que poderia ser um problema no código stan porém, conforme é possível ver nas observações 1 e 2 abaixo, quando eu coloco um tamanho amostral muito grande, de $n = 10000$, as cadeias seguem a lei de, ao aumentar o tamanho da amostra, a qualidade da estimação melhora. O que poderia estar causando esse erro?

- Observação 1
- Observação 2

A função pmap funciona da seguinte maneira: ela combina os parâmetros elemento a elemento. Desse modo, na primeira posição, ela irá pegar o elemento na posição 1 das listas `n`, `prob_X2`, `logit` e efetuar a computação. Desse modo, tem-se `gera_bayes(n = 200, prob_X2 = 0.5, logit = TRUE)`, `gera_bayes(n = 1000, prob_X2 = 0.5, logit = TRUE)` e assim sucessivamente.