Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Sensibilidade da *priori* na analise bayesiana de dados provenientes da distribuição normal

Mariana Resende[1]
Carla Regina Guimarães Brighenti[2]

1 Introdução

O método das distribuições *a priori* conjugadas é o mais usado para ultrapassar as dificuldades de eliciação de distribuições *a priori* subjetivas em que se adota uma forma funcional conveniente. As distribuições *a priori* não informativas são utilizadas quando não existe informação *a priori* palpável, sendo essa informação de natureza objetiva ou subjetiva, ou seja, o chamado estado de "ignorância da *priori*". Ou quando o conhecimento *a priori* é pouco significativo em relação à informação amostral, ou o conhecimento é "vago" [2].

Ao se trabalhar com famílias conjugadas a distribuição *a posteriori* é fechada, isto é, possui a forma de um modelo probabilístico conhecido, então as amostras do parâmetro de interesse podem ser geradas de modo direto no software R via amostragem de Gibbs [1].

Assim, o objetivo deste trabalho foi estudar a influência de diferentes distribuições *a priori* na análise bayesiana de dados provenientes de distribuição normal via simulação.

2 Material e métodos

Para estimação de parâmetros no processo normal foram estabelecidos três casos:

(i) média desconhecida e variância conhecida,

(ii) média conhecida e variância desconhecida e

(iii) média e variância desconhecidas.

Os resultados iniciais utilizando *prioris* conjugadas informativas e vagas, foram gerados a partir de uma rotina desenvolvida no software R 3.0.2 pelo pacote RBugs [3].

Para obtenção das distribuições marginais *a posteriori* foram gerados 100.000 valores em um processo MCMC, considerando descarte de 1.000 valores iniciais e com um salto de 10 valores. A convergência das cadeias foi verificada usando os critérios de Heidelberger & Welch, Raftery e Lewis e, Geweke do pacote CODA (Convergence Diagnosis and Output Analysis) em linguagem S-Plus que estão implementados no software R. Foram gerados também os intervalos de credibilidade a 95% (ICr) para cada parâmetro do modelo [4].

[1] DEZOO - UFSJ e-mail: naninha_mr@yahoo.com.br
[2] DEZOO – UFSJ e-mail:carlabrighenti@ufsj.edu.br.

Foram geradas 30 amostras aleatórias de uma Normal, seguindo os valores paramétricos μ = -10, -1; 0; 1; e 10 e pelo menos dois valores de variância para cada parâmetro simulado.

(i) Média Desconhecida e variância conhecida

Foi estudado o modelo da *priori* conjugada Normal (μ, σ^2) para o parâmetro média que permite obter a distribuição a *posteriori* de modelo conhecido. Adotou-se como hiperparâmetros para *prioris* informativas os valores de μ = -1, -20, 0, -10, 1, 10 e 20 e para σ = 0, 2; 1; 2; 10; 0,25; 0,5; 1 ou 5. Para as *prioris* vagas adotou-se μ = 0 e σ =10^3 e 10^6. Utilizou-se também como *priori* a Uniforme (0,100).

(ii) Média Conhecida e variância desconhecida

Foi estudado também o modelo da *priori* conjugada para o parâmetro variância sendo Gama (α, β). Os hiperparâmetros adotados para *prioris* conjugadas informativas foram combinações dos valores 3 e 5 tanto para α quanto para β e para as *prioris* vagas α = 10^{-9}, 10^{-7}, 10^{-8}, 10^{-6} e β = 10^{-6}, 10^{-6}, 10^{-4}.

(iii) Média e variância desconhecidas

Considerou-se para os hiperparâmetros as *prioris* Normal (μ, σ) com μ = -1, -20, 0, -10, 1, 10 e 20 e σ = 0,2; 1; 2; 10; 0,25; 0,5; 1 e 5 e a *priori* Gama (α,β) com α = 3 e 5 e β = 3. Para as *prioris* vagas considerou-se μ = 0, σ = 10^3, α =10^{-9} e β = 10^{-6}.

3 Resultados e discussão

(i) Média Desconhecida e variância conhecida

Foram verificadas 8 situações para cada população simulada. Na Tabela 1 encontra-se parte dos resultados. Na população com média μ=-10, em apenas 1 das situações o intervalo de credibilidade continha o verdadeiro valor do parâmetro. Nesse caso o valor estimado para a média e a mediana foi exatamente igual ao valor simulado. Nas demais situações apesar do μ populacional não pertencer ao ICr as estimativas foram próximas do valor real simulado.

Tabela 1 – Valores estimados *a posteriori* da média $\boldsymbol{\mu}$ quando a variância é conhecida

Parâmetro		**Hiperparâmetros**		**Média**	**Mediana**	**Desvio Padrão**	**Intervalo de Credibilidade**	**Erro de Monte Carlo**
μ	**σ**	**μ**	**σ**					
-10	2,5	-1	1	-10,00	-10,00	0,1147	[-10,23; -9,776]	0,00122
-1	1	-10	1	-1,280	-1,279	0,179	[-1,634; 0,926]	0,00191
1	0,25	0	0,5	1,011	1,012	0,3535	[0,3137; 1,707]	0,00376
			0,25	0,9414	0,9425	0,3591	[0,233; 1,649]	0,003829
	1		0,5	0,9303	0,9309	0,181	[0,5734; 1,287]	0,00193
		10	1	1,338	1,339	0,1796	[0,984; 1,692]	0,00191
10	2,5	20	2	10,15	10,15	0,1139	[9,924; 10,37]	0,00121

Apenas para a *priori* N(-20,10) houve um distanciamento maior do valor paramétrico simulado indicando que a *priori* utilizada teve grande influência. Para as amostras provenientes de população com médias simuladas $\mu = 10$ ou $\mu = -1$ também em apenas 1 situação o verdadeiro valor pertencia ao ICr a 95%.

Com o intuito de constatar a influência da informação a *priori* no resultado da distribuição a *posteriori*, também foram utilizadas *prioris* vagas. Entre as 32 situações simuladas, em nenhuma o verdadeiro valor do parâmetro pertencia ao ICr gerado e as estimativas ficaram distantes do valor real simulado. Isto evidencia que mesmo quando se utiliza *prioris* vagas elas terão algum tipo de influência nos valores estimados. Como foi utilizada média igual a zero as estimativas *a posteriori* ficaram em torno desse valor.

Já quando utilizadas *prioris* Uniforme (a, b), em apenas 1 das situações o verdadeiro valor não pertencia ao ICr, mesmo assim sua estimativa estava próxima do valor simulado. Isso mostra que no caso de *prioris* vagas para o processo Normal com média desconhecida e variância conhecida é mais indicado adotar como *a priori* Uniforme (Tabela 2).

Tabela 2 – Valores estimados *a posteriori* da média **μ** quando a variância é conhecida para *prioris* não vagas Uniforme (a,b).

Parâmetro		**Hiperparâmetros**		**Média**	**Mediana**	**Desvio Padrão**	**Intervalo de Credibilidade**	**Erro de Monte Carlo**
μ	**σ**	**μ**	**σ**					
1	0,25	0	100	0,9544	0,9585	0,3437	[0,332; 1,641]	0,01095
	1			0,8291	0,8268	0,1863	[0,468; 1,18]	0,006057
10	2,5	0	100	10,15	10,15	0,1178	[9,922; 10,37]	0,003831
	10			11,92	11,92	0,0589	[11,81; 12,03]	0,001915

(ii) Média Conhecida e variância desconhecida

Na Tabela 3 estão os resultados em que o verdadeiro valor estava dentro do intervalo de credibilidade gerado, sendo que, para o parâmetro populacional $\mu = -10$ em apenas 1 dos casos o verdadeiro valor do parâmetro pertencia ao ICr. Esse resultado implica que as *prioris* utilizadas tiveram grande influência nos resultados obtidos. Para média $\mu = 10$ em apenas 1 caso também o valor do parâmetro pertencia ao ICr. Para as médias iguais a $\mu = -1$ e $\mu = 1$ em todas as 8 situações simuladas o verdadeiro valor pertencia ao ICr, mostrando que nesses casos a *priori* teve influência positiva.

Considerando *prioris* vagas (Tabela 4) e média populacional $\mu = -10$ ou $\mu = 10$ em nenhuma das situações simuladas o verdadeiro valor do parâmetro pertencia ao intervalo de credibilidade. Quando o parâmetro populacional era $\mu = -1$ ou $\mu = 1$ em 50% das situações o verdadeiro valor paramétrico simulado pertencia ao ICr, sendo estes quando utilizado desvio padrão igual a 1.

Tabela 3 – Valores estimados *a posteriori* da variância σ^2 quando a média é conhecida.

Parâmetro		**Hiperparâmetros**		**Média**	**Mediana**	**Desvio Padrão**	**Intervalo de Credibilidade**	**Erro de Monte Carlo**
μ	**σ**	**μ**	**σ**					
-10	2,5	5	3	3,601	3,474	0,865	[2,304; 5,634]	0,0076
-1	0,25	3	3	0,2326	0,2238	0,057	[0,1462; 0,367]	0,00056
	1			0,9468	0,59499	0,2322	[0,5949; 1,494]	0,00229
	0,25		5	0,332	0,3194	0,08141	[0,2086; 0,524]	0,00080
	1			0,9545	0,9184	0,2341	[0,5998; 1,507]	0,00232
	0,25	5	3	0,1917	0,1849	0,0460	[0,1227; 0,300]	0,00040
	1			0,7471	0,7207	0,1794	[0,478; 1,169]	0,00157
	0,25		5	0,2839	0,2739	0,0682	[0,1817; 0,444]	0,00059
	1			1,098	1,059	0,2636	[0,7023; 1,717]	0,00231
1	0,25	3	3	0,2514	0,2419	0,06165	[0,158; 0,3968]	0,00061
	1			0,8677	0,8349	0,2128	[0,5453; 1,37]	0,00210
	0,25		5	0,3778	0,3635	0,09265	[0,237; 0,5963]	0,00091
	1			0,8987	0,8647	0,2204	[0,564; 1,418]	0,00218
	0,25	5	3	0,1991	0,1921	0,04782	[0,1274; 0,311]	0,00041
	1			0,8196	0,7906	0,1968	[0,5244; 1,282]	0,00172
	0,25		5	0,3049	0,2942	0,07324	[0,195; 0,4771]	0,00064
	1			0,7516	0,725	0,1805	[0,4809; 1,176]	0,00185
10	2,5	3	3	3,768	2,368	0,9241	[2,368; 5,948]	0,00915

Tabela 4 – Valores estimados da distribuição *a posteriori* da variância σ^2 quando a média é conhecida para *prioris* conjugadas não informativas.

Parâmetro		**Hiperparâmetros**		**Média**	**Mediana**	**Desvio Padrão**	**Intervalo de Credibilidade**	**Erro de Monte Carlo**
μ	**σ**	**μ**	**σ**					
-1	1	10^{-9}	10^{-6}	0,971	0,9316	0,2657	[0,5787; 1,608]	0,0022581
		10^{-7}	10^{-5}	0,9024	0,8658	0,247	[0,5378; 1,495]	0,002398
		10^{-8}	10^{-5}	0,986	0,946	0,2698	[0,587; 1,633]	0,00262
		10^{-6}	10^{-4}	1,055	1,012	0,2886	[0,6286; 1,747]	0,002803
1		10^{-9}	10^{-6}	0,6405	0,6145	0,1753	[0,381; 1,061]	0,001702
		10^{-7}	10^{-5}	1,035	0,9928	0,2832	[0,616; 1,714]	0,00275
		10^{-8}	10^{-5}	0,7678	0,7367	0,2101	[0,4576; 1,272]	0,002041
		10^{-6}	10^{-4}	1,269	1,217	0,3473	[0,7562; 2,102]	0,003372

(iii) Média e variância Desconhecidas

Quando μ = - 10 em 3 situações das 16 simuladas considerando o parâmetro μ, o verdadeiro valor dos parâmetros pertencia ao ICr. Quando considerado μ =10 em nenhuma das situações o valor paramétrico real pertencia ao ICr. Para μ = -1 em 10 das 16 situações os valores reais simulados tanto para μ quanto para σ estavam dentro do intervalo de credibilidade. Já para μ = 1, em 12 das simulações os valores reais de μ e σ pertenciam ao ICr.

Para as *prioris* vagas provenientes da Normal para o parâmetro média, em todas as situações simuladas em nenhuma o valor do parâmetro pertencia ao ICr e as estimativas ficaram distantes do valor simulado. Para o parâmetro σ, em duas situações o verdadeiro valor pertencia ao ICr, sendo esses quando $\mu = -1$ ou $\mu = 1$ e $\sigma = 1$. Nos demais casos as estimativas não se aproximaram do valor real. Para *prioris* de uma Uniforme não houve convergência.

Com relação ao erro de Monte Carlo, para os casos em que houve convergência, este foi muito pequeno, confirmando que o tamanho da cadeia amostral utilizada foi suficiente para obtenção de estimativas a *posteriori* das distribuições marginais, ou seja, a convergência foi alcançada pela amostragem de Gibbs. Quando a convergência foi atingida pelo critério de Geweke, o valor de $|k|$ foi menor que 1,96 e para o critério de Heidelberger e Welch, a palavra *"passed"* foi apresentada indicando a estacionariedade e convergência [3].

4 Conclusões

A *priori* selecionada teve grande influência na estimativa do parâmetro sendo que os melhores resultados foram obtidos no caso de média desconhecida e variância conhecida, quando utilizou-se o valor paramétrico positivo e *prioris* normal informativa ou a uniforme.

Para estimação da variância populacional desconhecida quando a média é conhecida a distribuição Gama, adotada como *priori*, foi mais adequada quando as médias populacionais foram iguais a -1 e 1. Não ocorrendo um padrão de quais hiperparâmetros foram melhores.

No caso de média e variância desconhecida quando adotadas *prioris* vagas as estimativas não foram adequadas. Quando adotadas *prioris* conjugadas informativas as estimativas foram melhores quando $\mu = -1$ ou $\mu = 1$.

5 Agradecimentos

A FAPEMIG pelo apoio financeiro.

6 Bibliografia

[1]BERGER, J. O. **Statistical Decision Therory and Bayesian Analysis**. Second Edition. New York: Springer-Verlag, 1985.

[2]PAULINO,C.D.; TURKMAN, M.A.A.; MURTEIRA, B. **Estatística Bayesiana**. Fundação Caloustre Gulbenkian. Lisboa. 2003.

[3]R DEVELOPMENT CORE TEAM. R: **A language and environment for statistical computing. R Foundation for Statistical Computing.** Vienna, Austria. 2009.

[4]ROSSI, R. M. **Introdução aos métodos Bayesianos na análise de dados zootécnicos com uso do WinBUGS e R.** 2010.