



SIMULADOR DE AMOSTRAS CO-SUFICIENTES PARA MISTURAS ESCALONADAS DE NORMAIS

Wivian Cerqueira Ribeiro^{1*}, Prof. Dr. James Dean Oliveira dos Santos Junior^{2*}.

¹Universidade Federal do Amazonas, Departamento de Estatística, Av. Rodrigo Otávio Jordão Ramos, 6200, Coroadó I, 69080-900, Manaus AM, Brasil.

* wivian.ribeiro@ufam.edu.br; james@ufam.edu.br

Palavras-Chave: Amostras Co-suficientes, Misturas Escalonadas de Normais, Testes de Aderência, Simulador Metropolis-Hastings.

Introdução

Misturas de normais escalonadas são distribuições marginais obtidas a partir da integração do parâmetro de escala da distribuição normal, assumindo uma distribuição a priori [1]. Essas distribuições não possuem estatísticas suficientes com dimensão menor que a própria amostra, então a aplicação de amostras co-suficientes para a melhoria de estimadores e testes de aderência torna-se inviável.

O desenvolvimento de um simulador capaz de gerar amostras co-suficientes para misturas escalonadas de normais, com foco na distribuição Cauchy, permite a melhoria de testes de aderência e estimadores. A partir da geração dessas amostras, é possível aumentar a precisão de estimadores utilizando o Teorema de Rao-Blackwell e transformar testes conservadores em exatos.

O projeto propõe a utilização de uma estatística ancilar latente para gerar amostras co-suficientes, impactando positivamente o método de estimação nas famílias escalonadas de normais e aprimorando a precisão de estimativas e testes.

Material e Métodos

A metodologia foi baseada na representação aumentada (X, W) das misturas escalonadas, onde W é um vetor de variáveis latentes ancilares. Condicionado em W , a distribuição de X pertence à família exponencial, o que garante a existência de uma estatística suficiente latente $T(X, W)$. As amostras co-suficientes são então geradas a partir da distribuição condicional $X|T(x, W)$, que não depende de parâmetros [2].

O processo foi implementado por meio de um algoritmo de Metropolis-Hastings [3] para garantir que as amostras geradas converjam para o modelo original. As etapas foram:

1. **Geração da Proposta:** Um candidato $(y_1, \dots, y_n, w_1, \dots, w_n)$ é gerado resolvendo um sistema de equações que garante que as estatísticas suficientes $(\hat{\mu}, \hat{\sigma}^2)$ sejam preservadas.
2. **Correção de Metropolis-Hastings:** A proposta é aceita ou rejeitada com base em uma razão de aceitação para assegurar a convergência à distribuição correta.

Para o estudo de simulação, foram utilizadas $B = 6000$ iterações, com um período de aquecimento (*burn-in*) de 3000 amostras para reduzir o efeito do estado inicial.

Resultados e Discussão

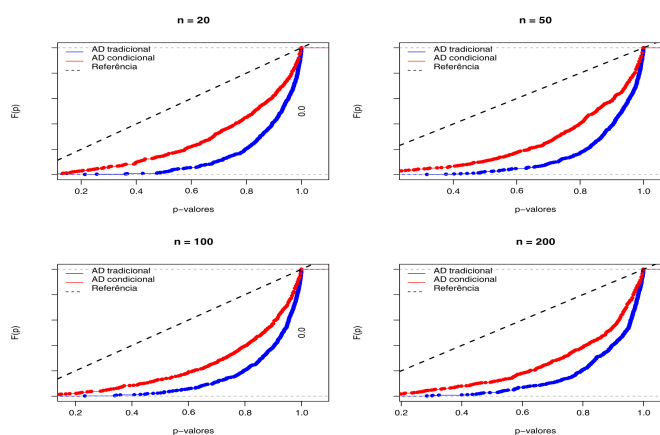


Figura 1. Funções de distribuição empíricas (ECDF) dos p-valores obtidos nos testes de Anderson-Darling tradicional e condicional, para diferentes tamanhos amostrais. A linha de referência corresponde à distribuição uniforme $(0,1)$.

O estudo, inicialmente proposto para a distribuição de Cauchy, foi adaptado para a distribuição t-Student com 5 graus de liberdade. A mudança foi necessária porque o estimador de máxima verossimilhança (EMV) para a Cauchy pode ser instável ou inexistente em amostras pequenas.

O desempenho do simulador foi avaliado comparando o teste de Anderson-Darling (AD) tradicional com sua versão condicional. Foram geradas 500 amostras de uma distribuição $t_5(0,1)$ para diferentes tamanhos amostrais $n \in \{20, 50, 100, 200\}$.

Os resultados (Figura 1) mostram que os p-valores do teste condicional (AD condicional) se aproximam mais da linha de referência que representa um teste exato. Embora ainda exista um leve conservadorismo, o método condicional corrige de forma significativa parte do viés do teste tradicional (AD tradicional), com uma melhora mais evidente em amostras menores, fornecendo resultados mais próximos do esperado teoricamente.

Conclusões

Embora a proposta inicial tivesse como foco a distribuição Cauchy, verificou-se que o estimador de máxima verossimilhança pode não existir para amostras pequenas, comprometendo a análise. Por esse motivo, optou-se por conduzir o estudo de simulação utilizando a distribuição t-Student com 5 graus de liberdade. Essa modificação permitiu a obtenção de estimativas estáveis e resultados consistentes, mantendo a viabilidade prática do simulador.

Os resultados obtidos mostram que o simulador de amostras co-suficientes melhora o desempenho do teste de Anderson-Darling em relação ao método tradicional. Embora o teste condicional ainda apresente certo conservadorismo, ele se aproxima mais da distribuição uniforme(0,1), resultando em conclusões mais próximas do valor esperado do teste. Essa melhoria representa uma vantagem relevante em comparação ao teste tradicional, principalmente em amostras pequenas.

Além disso, os resultados indicam que o simulador proposto pode ampliar a aplicação de testes de aderência em misturas escalonadas de normais e podem ser estendidos a outros testes além do Anderson-Darling.

Como limitações, destaca-se que o estudo restringiu-se à distribuição t-Student com $\nu = 5$ graus de liberdade e a um único teste de aderência. Como continuidade, recomenda-se explorar diferentes distribuições, valores de ν e avaliar o desempenho do simulador em outros cenários de teste.

Agradecimentos

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de iniciação científica concedida e ao Prof. Dr. James Santos pela orientação neste trabalho.

Referências

[1] Lee, Sharon X., and Geoffrey J. McLachlan. "Scale mixture distribution." *Wiley StatsRef: Statistics Reference Online* (2014): 1-16.

[2] O'Reilly, Federico, and Leticia Gracia-Medrano. "On the conditional distribution of goodness-of-fit tests." *Communications in Statistics Theory and Methods*, 35(3):541-549, 2006.

[3] Santos, James D., and Nelson L. Souza Filho. "A Metropolis algorithm to obtain co-sufficient samples with applications in conditional tests." *Communications in Statistics - Simulation and Computation*, 2018.