Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 8, n. 1, 2021.

Trabalho apresentado no XL CNMAC, Evento Virtual - Co-organizado pela Universidade do Mato Grosso do Sul (UFMS).

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Estatística para dados direcionais

Saul de A. Souza[1]
UFPE, Recife, PE
Abraão D. C. Nascimento[2]
DEST/UFPE, Recife, PE
Getúlio Jose Amorim do Amaral[3]
DEST/UFPE, Recife, PE

A análise estatística de dados direcionais é amplamente utilizada em diversas áreas da ciência, tais como: Climatologia, Astrofísica e Oceanografia [5]. Na prática, existem muitos sistemas de coordenadas que podem ser utilizadas para representar pontos esféricos, além de métodos para obter suas projeções sobre o plano. Um deles é baseado em sistemas de coordenadas polares.

Imagine que queiramos analisar um conjuntos de direções ou vetores unitários na hiperesfera unitária $\Omega_q := \{x \in \mathbb{R}^q : x^\top x = 1\}$. A princípio a distribuição von Mises-Fisher (vMF) pode ser um excelente candidato e é amplamente usada para modelar dados que se concentram nos polos. Ela possui dois parâmetros responsáveis por regular a locação, μ, e a concentração, $\kappa \in [0, \infty)$.

A densidade da vMF tem a forma (para $\mu, x \in \Omega_q$)

$$f_{\text{vMF}}(x; \mu, \kappa) \;=\; a_q^{-1}(\kappa) \exp(\kappa\, [\mu^\top x]), \tag{1}$$

em que $a_q^{-1}(\kappa) = \kappa^{q/(2-1)} / \left[(2\pi)^{q/2} I_{(q/2)-1}(\kappa)\right]$ é a constante de normalização e $I_q(\cdot)$ representa a função de Bessel modificada de primeiro tipo e ordem q [2]. Neste caso dizemos que $x \sim \text{vMF}_q(\mu, \kappa)$. Quando $q = 3$, configurando um cenário tridimensional, as observações são plotadas na esfera unitária. Contudo, quando $q = 2$ os pontos são identidicados sobre a circuferência de raio unitário.

Conforme discutido em [2], se $x \sim \text{vMF}_q(\mu, \kappa)$ então o produto interno estocástico $T := T(\mu, x) := \langle \mu, x \rangle = \mu^\top x$ tem densidade (para $t \in [-1, 1]$)

$$f_T(t) \;=\; a_q^{*-1}(\kappa)\, \exp(\kappa t)\, (1 - t^2)^{\frac{q-3}{2}}, \tag{2}$$

em que $a_q^{*-1}(\kappa) = w_{q-1}\, a_q^{-1}(\kappa)$, $w_q = 2(\pi)^{q/2}/\Gamma(q/2)$ e $\Gamma(\cdot)$ é a função gamma. Desde que X seja rotacionalmente simétrico sobre μ, a densidade da projeção $\mu^\top X$ é dada por (2) [3]. Essa densidade será a chave para determinar a distribuição da medida de distância proposta neste trabalho.

Observe que T descreve a associação entre possíveis resultados da vMF e seu parâmetro de localização, indicando quão concentrado é o resultado. É conhecido por [4] que a concentração é um dos fenômenos mais importantes na teoria de dados direcionais. Embora muitos pesquisadores tenham dedicado seus estudos a concentração, não conhecemos nenhuma estatística ou mecanismo para checar alta ou baixa concentração, lidando com este fenômeno diretamente. Neste artigo, entendemos que a proposta de distribuição de uma medida de distância em termos de T pode ser uma boa maneira de estudar a concentração.

[1] saul$_a$*souza*@*hotemail.com*
[2] abraao@de.ufpe.br
[3] gjaa@de.ufpe.br

A partir da Equação (2), é possível calcular a função de distribuição acumulada (fda) da distância estocástica

$$D := D(x, \mu) := 1 - T^2. \tag{3}$$

Na prática, cada resultado gerado a partir deste modelo pode ser entendido como uma medida da distância $D(\cdot, \cdot)$ entre uma observação esférica distribuída de von Mises-Fisher e seu parâmetro de localização. Dessa forma, ao invés de realizar uma análise multivariada dos dados em uma esfera unitária tridimensional, configurando um cenário mais complexo, derivamos uma medida de distância para capturar a concentração de pontos esféricos.

Nossa abordagem trata da análise descritiva de uma medida de distância em $(0, 1)$. Para esse fim, fornecemos uma nova distribuição derivada de uma medida de distância entre pontos vMF distribuídos na esfera unitária. A partir dessa nova lei, também desenvolvemos uma nova abordagem para identificar fenômenos de alta concentração em dados direcionais. Nós mostramos que se os dados esféricos seguem a lei de von Mises-Fisher, então sua concentração pode ser modelada por nossa distribuição. Algumas de suas propriedades são derivadas e discutidas: função geradora de momento, curtose, assimetria e matriz de informação de Fisher. São forneceidos procedimentos inferenciais baseados em probabilidade, ambos: estimação pontual e teste de hipótese. Os estudos de simulação mostram que as estimativas de máxima verossimilhança desempenham assintoticamente bem, mesmo em amostras pequenas. Observamos que o teste da razão de verossimilhanças supera frequentemente os testes de Wald e score para a distribuição TD.

Fizemos uma aplicação para dados paleomagnéticos e ilustramos como nosso modelo é empregado para analisar a concentração de dados esféricos. Para tanto, utilizamos análises gráficas e testes de hipóteses. Os resultados mostram que a medida de distância aplicada aos cossenos direcionais é capaz de fornecer evidências a respeito da dispersão de pontos esféricos.

Agradecimentos

Os autores agradecem à CAPES, CNPq e FACEPE pelo financiamento da pesquisa.

Referências

[1] Fisher, N., Lewis, T. e Embleton, B. *Statistical Analysis of Spherical Data, 1a. edição*. Cambridge University Press, 1993.

[2] Ko, D. Robust estimation of the concentration parameter of the Von Mises-Fisher distribution, *Annals of Statistics*, 20:917–928, 1992. DOI: 10.1214/aos/1176348663.

[3] Ley, C. e Verdebout, T. *Modern Directional Statistics, 1a. edição*. Chapman and Hall, 2017.

[4] Mardia, K. e Jupp, P. *Directional Statistics, 1a. edição*. Wiley, 1999.