

Avaliação e Estimação dos Parâmetros da Distribuição Skellam

Antonino Alves Feitosa Neto¹

¹antonino_feitosa@yahoo.com.br

¹Universidade Federal do Rio Grande do Norte

15 de dezembro de 2023

Resumo

A distribuição de Skellam é aplicada em diferentes áreas como esportes, saúde, finanças, etc. Assim, é relevante a investigação do comportamento de seus estimadores pontuais em cenários de tamanho amostral pequeno, ou seja, em cenários que o desempenho deles pode ser afetado pela presença de viés. Desse modo, investigamos o desempenho dos estimadores obtidos pelo método de máxima verossimilhança e pelo método dos momentos, comparando-os com as suas respectivas versões corrigidas por Bootstrap. Eles são comparados em termos do viés e do erro quadrático médio obtidos por meio de simulações de Monte Carlo. Verificamos que o estimador pelo método de máxima verossimilhança apresenta o menor erro quadrático médio em todos os casos, apesar de ter um viés maior que os demais.

Palavras-chave: *Bootstrap*, Correção de viés, Máxima Verossimilhança, Método dos Momentos, Simulação de Monte Carlo

1 Introdução

A distribuição de Skellam foi proposta por Skellam [16] para modelagem da diferença entre duas contagens. Ela é derivada da diferença entre duas variáveis aleatórias (V.A.) independentes que seguem uma distribuição de Poisson e por esse motivo ela também é chamada de distribuição de diferenças de Poisson.

Ela foi aplicada em diferentes áreas como esportes, finanças, saúde, etc. Por exemplo: Catania et al. [8] utiliza a distribuição de Skellam na modelagem da variação de preços de ações; Jiang et al. [10] a aplica na identificação de padrões em expressão genética; já Karlis e Ntzoufras [7] a utilizam na previsão do resultado de partidas de futebol.

Além dessas aplicações, a plataforma arXiv [2] apresenta diferentes trabalhos em processo de publicação baseados na distribuição de Skellam. Destacamos o trabalho de Rave e Kauermann [13] que modela o fluxo de entrada e saída diários de uma unidade de tratamento intensiva durante a pandemia de Covid-19 na Alemanha. Também citamos o trabalho de Wang [18] que aplica a distribuição no desenvolvimento de um algoritmo de aprendizado de máquina para sistemas de recomendação. Assim, percebemos a importância dessa distribuição e sua contribuição em diferentes áreas.

O objetivo deste trabalho é avaliar o desempenho de estimadores pontuais da distribuição de Skellam em cenários de tamanho amostral pequeno. Nesse contexto, os estimadores usuais podem apresentar um desempenho insatisfatório devido à presença de tendências quanto utilizamos pequenas amostras. Logo, desejamos avaliar os estimadores pontuais pelo método de máxima verossimilhança (EMV) e pelo método dos momentos (EMM), comparando-os com as suas respectivas versões corrigidas por Bootstrap. Assim, ao final do trabalho, desejamos identificar em quais cenários (tamanhos amostrais) eles são mais adequados.

O restante do trabalho se divide nas seguintes seções: seção de Referencial Teórico, apresentando uma breve revisão da distribuição de Skellam e dos métodos utilizados; seguida pela seção de Metodologia, que apresenta como os métodos serão aplicados para alcançar os objetivos citados; a seção dos Resultados obtidos, incluindo a discussão de cada um; terminando pela seção de considerações finais.

2 Referencial Teórico

Irwin [6] apresenta a distribuição de probabilidade obtida da diferença de duas V.A. independentes e idênticas seguindo uma distribuição de Poisson. Skellam [16] generaliza essa distribuição para o caso de parâmetros diferentes. Assim, a distribuição de Skellam é definida na Definição 2.1 [7]:

Definition 2.1 (Distribuição de Skellam $PD(\theta_1, \theta_2)$). Considere o par de variáveis aleatórias (X, Y) tais que $X \sim Poisson(\theta_1)$ independente de $Y \sim Poisson(\theta_2)$. Então $Z = X - Y$ é dita ter uma distribuição de Skellam, denotada por $PD(\theta_1, \theta_2)$, e sua função massa de

probabilidade é dada por:

$$P(Z = z) = e^{-(\theta_1 + \theta_2)} \left(\frac{\theta_1}{\theta_2} \right)^{x/2} I_z(2\sqrt{\theta_1 \theta_2}) \quad , z = \dots, -1, 0, 1, \dots$$

onde $I_y(x) = \left(\frac{x}{2}\right)^y \frac{\sum_{k=0}^{\infty} \left(\frac{x^2}{4}\right)^k}{k!(y+k)!}$ é a função de Bessel modificada de primeiro tipo.

Observe que o suporte dessa distribuição abrange os conjunto dos valores inteiros. Os valores negativos representam situações em que a V.A. Y assume valores maiores que a X . Além disso, os parâmetros θ_1 e θ_2 assumem valores reais positivos, pois são parâmetros de uma distribuição de Poisson.

A partir da sua definição, como a diferença de duas V.A. independentes de Poisson, podemos obter a esperança e a variância de uma V.A. Z com distribuição $PD(\theta_1, \theta_2)$:

Theorem 1 (Esperança de $PD(\theta_1, \theta_2)$). Considere a V.A $Z \sim PD(\theta_1, \theta_2)$. O valor esperado $E[Z]$ e variância são definidos como $Var[Z]$:

$$E[Z] = E[X - Y] = E[X] - E[Y] = \theta_1 - \theta_2$$

$$Var[Z] = Var[X - Y] = Var[X] + (-1)^2 Var[Y] = \theta_1 + \theta_2,$$

em que X e Y são duas V.A. independentes de uma distribuições de Poisson com parâmetro θ_1 e θ_2 respectivamente.

Ademais, $PD(\theta_1, \theta_2)$ possui as seguintes propriedades:

$$P(Z = z|\theta_1, \theta_2) = P(Z = -z|\theta_2, \theta_1).$$

Sua função geradora de momentos é definida como:

$$M_z(t) = \exp\{-(\theta_1 + \theta_2) + \theta_1 e^t + \theta_2 e^{-t}\}.$$

E sua assimetria é definida como:

$$\beta_1 = \frac{\theta_1 - \theta_2}{(\theta_1 + \theta_2)^{\frac{3}{2}}}.$$

Desse modo, a distribuição apresenta uma assimetria positiva quando $\theta_1 > \theta_2$, negativa quando $\theta_1 < \theta_2$, e será simétrica quando $\theta_1 = \theta_2$. Para mais informações de como essas propriedades foram obtidas e também de propriedades adicionais, podem ser encontradas nos trabalhos de [7, 1].

Considerando os objetivos deste trabalho, também apresentaremos os estimadores obtidos pelo EMM e pelo EMV para os parâmetros θ_1 e θ_2 . Além disso, as próximas seções também apresentam um breve resumo das principais técnicas utilizadas para avaliar a qualidade desses estimadores de forma prática.

2.1 Estimadores de θ_1 e θ_2

Ao ajustar um modelo, devemos estimar os valores dos parâmetros de modo mais fidedigno possível aos dados observados. Em geral, podemos dividir os métodos de estimação em pontual e intervalar, ou seja, estimar um único ponto ou um intervalo com certo nível de confiança para o valor real do parâmetro.

A estimação pontual é mais simples e fácil de interpretar, sendo valiosa quando se deseja comunicar informações a um público não especializado. No entanto, ela não reflete a incerteza associada à estimação, além de ser sensível a valores aberrantes. Esses problemas podem ser tratados por uma estimação intervalar. Logo, cada uma tem o seu mérito e em muitos casos apresentamos as duas estimativas dos dois tipos para fornecer uma visão ampla da inferência realizada.

Assim, investigaremos a qualidade dos EMM e do EMV que são estimadores pontuais. Esses métodos podem ser encontrados Bolfarine [3] influenciando descrições de suas propriedades. Para a distribuição de Skellam, os EMM são definidos na Definição 2.2 [7].

Definition 2.2 (EMM). Seja Z_1, Z_2, \dots, Z_N uma amostra aleatória da V.A. Z tal que $Z \sim PD(\theta_1, \theta_2)$. Então, os EMM de θ_1, θ_2

$$\tilde{\theta}_1 = (S_z^2 + \bar{Z})/2$$

$$\tilde{\theta}_2 = (S_z^2 - \bar{Z})/2,$$

em que \bar{Z} é a média e S_z^2 a variância amostral de Z_1, Z_2, \dots, Z_N , respectivamente. Os estimadores não são definidos caso $S_z^2 - |\bar{Z}| < 0$.

Apesar da obtenção dos EMM serem analiticamente simples, não é uma tarefa trivial definir os EMV, pois envolve a resolução de um sistema de equações não lineares. Desse modo, podemos recorrer a métodos numéricos para obtenção dos EMV.

Uma vez que as estimativas foram obtidas, precisamos decidir qual delas é mais fidedigna ao valor real do parâmetro, ou seja, decidir entre as estimativas de EMM ou EMV. Podemos verificar na literatura que a medida de erro quadrático médio (EQM) pode ser utilizada como métrica de comparação da qualidade de estimadores. O erro quadrático médio de um

estimador $\hat{\theta}$ de um parâmetro θ pode ser definido em termos do viés $Viés(\hat{\theta})$ e variância $Var(\hat{\theta})$ desse estimador como na Equação 2.1 [3]:

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + Viés(\hat{\theta}).$$

Porém, calcular o viés e a variância de um estimador de modo analítico pode ser uma tarefa complexa, pois pode resultar na resolução de equações não lineares. Novamente, podemos recorrer a métodos numéricos para a determinação desses valores, em particular, podemos recorrer ao método estatístico de Monte Carlo para determinação da esperança e da variância de um estimador, e por consequência, a determinação do erro quadrático médio.

2.2 Método de Monte Carlo

Esta seção descreve brevemente como podemos utilizar o método de Monte Carlo para obter a esperança e a variância de um estimador. Uma descrição detalhada desse método, incluindo teoria e exemplos, pode ser encontrada em [14].

Desse modo, considere o problema de estimar o valor esperado de uma V.A. X qualquer e suponha que possam ser geradas amostras aleatórias com a mesma distribuição de probabilidade de X . Chamamos a geração de um valor de simulação e caso sejam geradas r simulações, isto é, X_1, X_2, \dots, X_r , podemos calcular a média amostral \bar{X} . Logo, como consequência da lei dos grandes números [15], \bar{X} converge para $E[X]$ quando $r \rightarrow \infty$.

Portanto, se desejamos estimar a esperança de um estimador $E[\hat{\theta}]$, precisamos de um método de geração de variáveis aleatórias com a mesma distribuição de probabilidade de $\hat{\theta}$ e de um número de réplicas r grande de amostras. E assim podemos obter uma estimativa numérica de $E[\hat{\theta}]$ e também de $Var[\hat{\theta}]$, ou seja, temos como calcular o EQM de um estimador $\hat{\theta}$.

2.3 Método de Bootstrap

Os métodos de Bootstrap foram introduzidos por Efrom [4]. Correspondem a uma classe de métodos de Monte Carlo para aproximação de uma distribuição de probabilidade por meio de uma função empírica obtida de uma amostra finita.

O termo *Bootstrap* é oriundo da expressão da língua inglesa “*to pull oneself up by one’s bootstrap*” que passa a ideia de alcançar sucesso por esforço próprio, iniciando de circunstâncias muito difíceis e sem ajuda, como emergir de um afogamento puxando pela alça do próprio sapato [17]. Na estatística, pode ser interpretado como a capacidade de obter propriedades de uma população a partir de poucas observações.

Desse modo, considere que desejamos estimar um parâmetro de interesse θ , que temos uma amostra de observações x_1, x_2, \dots, x_n e que dispomos de um estimador para θ de forma $\hat{\theta} = f(x_1, x_2, \dots, x_n)$. Assim, a estimativa $\hat{\theta}^*$ para θ é obtida pela média amostral $\bar{\hat{\theta}}^*$ dos estimadores de B amostras bootstrap. Uma amostra bootstrap é uma amostra com reposição obtida de x_1, x_2, \dots, x_n (de mesmo tamanho que a observada), assumindo que cada x_i possui a mesma probabilidade de ocorrência $1/n$, usada para obter uma estimativa $\hat{\theta}^{(b)}$. A distribuição de $\bar{\hat{\theta}}^*$ converge para a distribuição de θ quando B tende ao infinito [5].

Observe que não fazemos suposições sobre a distribuição de probabilidade da amostra de observações, ou seja, é um método não paramétrico. No entanto, caso a distribuição seja conhecida, as amostras bootstrap podem ser substituídas por amostras aleatórias e independentes da distribuição, ou seja, temos um método de Monte Carlo. Nesse último caso, classificamos o método Bootstrap como paramétrico.

Apesar de $\hat{\theta}^*$ ser uma estimativa de θ , geralmente ele não aproxima bem a locação da distribuição e, portanto, não deve ser usado como estimador para θ . Dentre as várias aplicações, podemos usar o método de Bootstrap para avaliação e correção do viés de um estimador [5].

Assim, considere o problema de remover o viés de um estimador $\hat{\theta}$, ou seja, desejamos obter $\hat{\theta}^c = \hat{\theta} - \text{Viés}(\hat{\theta})$ tal que $\theta = \hat{\theta}^c$. Porém, não temos como calcular $\text{Viés}(\hat{\theta}) = E[\hat{\theta}] - \theta$, pois θ é desconhecido. No entanto, podemos utilizar a estimativa de Bootstrap para $\hat{\theta}$ como mostrado na Equação 2.3[5].

$$\hat{\theta}^c = \hat{\theta} - \text{Viés}(\hat{\theta}^*) = \hat{\theta} - (E[\hat{\theta}^*] - \hat{\theta}) = 2\hat{\theta} - \bar{\hat{\theta}}^*.$$

Ademais, também efetuaremos a correção de viés por Bootstrap, de modo paramétrico, dos estimadores de EMM e EMV, ou seja, serão comparados quatro métodos.

3 Metodologia

Consideraremos quatro cenários para a avaliação dos estimadores em função dos valores reais dos parâmetros, no seguinte conjunto $PD(\theta_1 = 10, \theta_2 = 10)$, $PD(\theta_1 = 10, \theta_2 = 100)$, $PD(\theta_1 = 100, \theta_2 = 10)$ e $PD(\theta_1 = 100, \theta_2 = 100)$. Escolhemos esses valores baseados em aplicações de esportes, por exemplo, cada equipe marca cerca de 1 a 2 pontos em jogos de futebol, já para o handebol são cerca de 20 pontos, enquanto no basquete são cerca de 100 pontos. Os resultados das partidas desses esportes podem ser modelados por distribuições de Skellam e por isso escolhemos os valores como combinações de 10 e 100.

Para cada par de valores dos parâmetros, o EQM de cada estimador será aproximado pelo método de Monte Carlo gerando 2000 amostras com os valores reais dos parâmetros. Isto é,

obtermos o EQM dos estimadores de EMM ($\hat{\theta}_1$), EMV e de suas versões com viés corrigido por Bootstrap paramétrico, usando 500 amostras bootstrap. Essas quantidades foram escolhidas considerando o tempo e recursos computacionais para a realização do trabalho.

Além disso, cada cenário será executado com tamanhos amostrais n de 25, 50, 75 e 100. Esses valores foram escolhidos por representarem amostras pequenas e que provavelmente apresentarão um alto viés. Por fim, os resultados serão comparados em termos da magnitude do EQM, variância e viés de cada estimador.

A implementação foi efetuada no R [12] com o auxílio do pacote Skellam [9] para geração de valores aleatórios e função de probabilidade da distribuição. O software está disponível no repositório Skellam-Distribution [11].

4 Resultados

A Tabela 1 apresenta as métricas dos estimadores para a distribuição de probabilidades $PD(\theta_1 = 10, \theta_2 = 10)$. Observa-se que o EMV $\hat{\theta}$ apresenta o menor EQM para todos os tamanhos amostrais, porém, com o maior valor de viés. Apesar de sua versão corrigida apresentar um menor viés em todos os tamanhos amostrais, ele gera um aumento da variância e um maior EQM.

Nota-se que as diferenças de magnitudes entre as métricas dos $\tilde{\theta}$ e dos $\tilde{\theta}^*$ são praticamente desprezíveis, menores que 10 vezes em relação à magnitude do EQM. Isso sugere que os EMM para a distribuição de Skellam não são viesados nesse cenário. Além disso, com o aumento do tamanho amostral, verificamos uma diminuição do EQM em todos os estimadores.

Desse modo, conclui-se que devemos utilizar o EMV $\hat{\theta}$ para a estimação de parâmetros quando a distribuição real é $PD(\theta_1 = 10, \theta_2 = 10)$ e que não há vantagens em aplicar correção de viés nesse caso. Os resultados da Tabela 2 são similares aos da Tabela 1, com os dados escalados por um fator de 100. Logo, obtêm-se as mesmas conclusões para a distribuição real $PD(\theta_1 = 100, \theta_2 = 100)$.

Porém, para valores distintos dos parâmetros, em que os resultados são apresentados nas Tabelas 3 e 4, nota-se um comportamento diferente. O EMV $\hat{\theta}$ ainda apresenta o menor EQM e também verifica-se uma diminuição do EQM com o aumento do tamanho amostral. No entanto, os estimadores corrigidos $\hat{\theta}^*$ e $\tilde{\theta}^*$ apresentam maior viés e EQM que suas respectivas versões $\hat{\theta}$ e $\tilde{\theta}$ em quase todos os casos. Isso é mais evidente para o parâmetro assumindo o menor valor, $\theta_1 = 10$ em 3 e $\theta_2 = 10$ em 4.

Desse modo, conclui-se que a correção de viés apresenta pouca ou quase nenhuma melhora em relação aos estimadores não corrigidos e que o EMV $\hat{\theta}$ apresenta os menores valores de EQM em todos os tamanhos amostrais.

n	Est.	Estimativas de $\theta_1 = 10$				Estimativas de $\theta_2 = 10$			
		Média	Viés	Var	EQM	Média	Viés	Var	EQM
25	$\hat{\theta}$	9.69	-0.31	8.19	8.29	9.69	-0.31	8.34	8.44
	$\hat{\theta}^*$	10.05	0.05	8.88	8.88	10.05	0.05	9.04	9.04
	$\tilde{\theta}$	10.07	0.07	8.86	8.87	10.07	0.07	9.02	9.02
	$\tilde{\theta}^*$	10.07	0.07	8.85	8.85	10.06	0.06	9.00	9.00
50	$\hat{\theta}$	9.79	-0.21	4.35	4.39	9.78	-0.22	4.26	4.31
	$\hat{\theta}^*$	9.98	-0.02	4.52	4.53	9.96	-0.04	4.44	4.44
	$\tilde{\theta}$	9.98	-0.02	4.52	4.53	9.96	-0.04	4.44	4.44
	$\tilde{\theta}^*$	9.98	-0.02	4.55	4.55	9.96	-0.04	4.45	4.45
75	$\hat{\theta}$	9.86	-0.14	2.84	2.86	9.86	-0.14	2.87	2.89
	$\hat{\theta}^*$	9.98	-0.02	2.90	2.90	9.98	-0.02	2.93	2.94
	$\tilde{\theta}$	9.98	-0.02	2.92	2.92	9.98	-0.02	2.96	2.96
	$\tilde{\theta}^*$	9.98	-0.02	2.93	2.93	9.98	-0.02	2.96	2.96
100	$\hat{\theta}$	9.92	-0.08	2.08	2.09	9.91	-0.09	2.10	2.11
	$\hat{\theta}^*$	10.02	0.02	2.13	2.13	10.00	0.00	2.15	2.15
	$\tilde{\theta}$	10.02	0.02	2.13	2.13	10.00	-0.00	2.15	2.15
	$\tilde{\theta}^*$	10.01	0.01	2.14	2.14	10.00	-0.00	2.15	2.15

Tabela 1: Métricas dos estimadores para $\theta_1 = 10$ e $\theta_2 = 10$.

n	Est.	Estimativas de $\theta_1 = 100$				Estimativas de $\theta_2 = 100$			
		Média	Viés	Var	EQM	Média	Viés	Var	EQM
25	$\hat{\theta}$	96.02	-3.98	804.51	820.35	95.98	-4.02	806.14	822.27
	$\hat{\theta}^*$	99.81	-0.19	870.50	870.54	99.77	-0.23	872.19	872.24
	$\tilde{\theta}$	99.99	-0.01	872.92	872.92	99.96	-0.04	874.58	874.59
	$\tilde{\theta}^*$	100.01	0.01	878.74	878.74	99.97	-0.03	880.13	880.13
50	$\hat{\theta}$	98.14	-1.86	397.35	400.82	98.11	-1.89	397.59	401.17
	$\hat{\theta}^*$	100.06	0.06	413.97	413.98	100.04	0.04	413.98	413.98
	$\tilde{\theta}$	100.13	0.13	413.60	413.62	100.10	0.10	413.85	413.86
	$\tilde{\theta}^*$	100.14	0.14	414.67	414.69	100.11	0.11	414.87	414.88
75	$\hat{\theta}$	98.47	-1.53	253.60	255.93	98.50	-1.50	254.04	256.28
	$\hat{\theta}^*$	99.78	-0.22	260.74	260.79	99.81	-0.19	261.23	261.27
	$\tilde{\theta}$	99.79	-0.21	260.46	260.50	99.82	-0.18	260.90	260.94
	$\tilde{\theta}^*$	99.79	-0.21	260.74	260.79	99.82	-0.18	261.29	261.33
100	$\hat{\theta}$	98.61	-1.39	197.15	199.09	98.54	-1.46	197.16	199.29
	$\hat{\theta}^*$	99.56	-0.44	201.36	201.55	99.50	-0.50	201.26	201.51
	$\tilde{\theta}$	99.59	-0.41	201.06	201.23	99.53	-0.47	201.07	201.29
	$\tilde{\theta}^*$	99.61	-0.39	201.64	201.79	99.54	-0.46	201.58	201.79

Tabela 2: Métricas dos estimadores para $\theta_1 = 100$ e $\theta_2 = 100$.

n	Estimador	Estimativas de $\theta_1 = 10$				Estimativas de $\theta_2 = 100$			
		Média	Viés	Var	EQM	Média	Viés	Var	EQM
25	$\hat{\theta}$	10.24	0.24	150.68	150.74	100.18	0.18	153.41	153.44
	$\hat{\theta}^*$	8.90	-1.10	216.30	217.51	98.84	-1.16	218.42	219.76
	$\tilde{\theta}$	11.96	1.96	178.46	182.30	99.66	-0.34	257.08	257.19
	$\tilde{\theta}^*$	9.28	-0.72	224.19	224.71	99.66	-0.34	258.18	258.30
50	$\hat{\theta}$	9.71	-0.29	78.53	78.62	99.74	-0.26	80.64	80.70
	$\hat{\theta}^*$	9.00	-1.00	107.70	108.70	99.03	-0.97	109.56	110.51
	$\tilde{\theta}$	10.64	0.64	86.24	86.65	99.79	-0.21	113.24	113.29
	$\tilde{\theta}^*$	9.18	-0.82	108.84	109.52	99.79	-0.21	113.79	113.83
75	$\hat{\theta}$	9.72	-0.28	64.64	64.71	99.75	-0.25	66.26	66.33
	$\hat{\theta}^*$	9.29	-0.71	83.93	84.44	99.32	-0.68	85.57	86.03
	$\tilde{\theta}$	10.37	0.37	69.57	69.71	99.90	-0.10	84.34	84.35
	$\tilde{\theta}^*$	9.37	-0.63	84.87	85.26	99.89	-0.11	84.41	84.42
100	$\hat{\theta}$	9.61	-0.39	51.82	51.97	99.61	-0.39	52.28	52.43
	$\hat{\theta}^*$	9.31	-0.69	64.62	65.09	99.31	-0.69	64.98	65.45
	$\tilde{\theta}$	10.16	0.16	55.24	55.26	99.86	-0.14	63.07	63.09
	$\tilde{\theta}^*$	9.47	-0.53	66.21	66.49	99.88	-0.12	63.34	63.35

Tabela 3: Métricas do estimadores para $\theta_1 = 10$ e $\theta_2 = 100$.

n	Estimador	Estimativas de $\theta_1 = 100$				Estimativas de $\theta_2 = 10$			
		Média	Viés	Var	EQM	Média	Viés	Var	EQM
25	$\hat{\theta}$	100.61	0.61	147.29	147.66	10.62	0.62	143.63	144.01
	$\hat{\theta}^*$	99.38	-0.62	211.99	212.37	9.39	-0.61	208.90	209.28
	$\tilde{\theta}$	100.45	0.45	244.03	244.23	12.46	2.46	171.45	177.51
	$\tilde{\theta}^*$	100.45	0.45	244.86	245.06	9.87	-0.13	216.92	216.94
50	$\hat{\theta}$	99.62	-0.38	89.31	89.45	9.60	-0.40	86.70	86.86
	$\hat{\theta}^*$	98.85	-1.15	120.01	121.33	8.82	-1.18	117.49	118.87
	$\tilde{\theta}$	99.56	-0.44	126.80	126.99	10.58	0.58	94.94	95.28
	$\tilde{\theta}^*$	99.58	-0.42	127.24	127.41	9.08	-0.92	118.52	119.36
75	$\hat{\theta}$	99.82	-0.18	66.13	66.17	9.84	-0.16	65.33	65.36
	$\hat{\theta}^*$	99.38	-0.62	85.18	85.57	9.40	-0.60	84.54	84.90
	$\tilde{\theta}$	99.97	-0.03	84.38	84.38	10.51	0.51	69.75	70.01
	$\tilde{\theta}^*$	99.98	-0.02	84.68	84.68	9.54	-0.46	85.42	85.63
100	$\hat{\theta}$	99.71	-0.29	52.30	52.39	9.72	-0.28	51.27	51.35
	$\hat{\theta}^*$	99.45	-0.55	65.15	65.45	9.46	-0.54	64.16	64.45
	$\tilde{\theta}$	99.97	-0.03	62.25	62.25	10.25	0.25	54.42	54.49
	$\tilde{\theta}^*$	99.98	-0.02	62.19	62.20	9.57	-0.43	64.77	64.96

Tabela 4: Métricas do estimadores para $\theta_1 = 100$ e $\theta_2 = 10$.

5 Considerações Finais

Este trabalho avaliou o desempenho dos estimadores pontuais pelo método de máxima verossimilhança e pelo método dos momentos para a distribuição de Skellam em cenários de tamanho amostral pequeno. Além disso, os estimadores foram comparados com suas versões corrigidas por Bootstrap para os tamanhos amostrais de 25, 50, 75 e 100, com combinações dos valores de 10 e 100 para os parâmetros das distribuições reais.

Observou-se que o estimador de máxima verossimilhança apresentou o menor erro quadrático médio em todos os cenários e que não houve melhorias no EQM ao corrigi-lo por Bootstrap. Além disso, verificou-se um aumento do viés dos estimadores corrigidos em relação às suas respectivas versões não corrigidas ao empregar valores diferentes dos parâmetros das distribuições reais, ou seja, a correção não foi efetiva nesses cenários.

Assim, recomendamos o uso do estimador de máxima verossimilhança para tamanhos amostrais pequenos sem a necessidade de correção de viés. No entanto, seria interessante verificar esses resultados para amostras ainda menores, dependendo da necessidade da aplicação.

Referências

- [1] A. Alzaid e Maha Omair. “On The Poisson Difference Distribution Inference and Applications”. Em: *Bulletin of the Malaysian Mathematical Sciences Society. Second Series* 1 (jan. de 2010).
- [2] *arXiv*. <https://arxiv.org/>. Disponível em: <https://arxiv.org>. Acesso em: 28 nov. 2023.
- [3] Heleno Bolfarine e Monica Carneiro Sandoval. *Introdução a inferência estatística*. SBM, 2010.
- [4] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. Em: *The Annals of Statistics* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552. URL: <https://doi.org/10.1214/aos/1176344552>.
- [5] Bradley Efron e Robert J. Tibshirani. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57. Boca Raton, Florida, USA: Chapman & Hall/CRC, 1993.

- [6] J. O. Irwin. “The Frequency Distribution of the Difference between Two Independent Variates Following the Same Poisson Distribution”. Em: *Journal of the Royal Statistical Society* 100.3 (dez. de 1937), pp. 415–416. ISSN: 0952-8385. DOI: 10.1111/j.2397-2335.1937.tb04518.x. eprint: https://academic.oup.com/jrsssa/article-pdf/100/3/415/49708933/jrsssa_100_3_415.pdf. URL: <https://doi.org/10.1111/j.2397-2335.1937.tb04518.x>.
- [7] Dimitris Karlis e Ioannis Ntzoufras. “Bayesian modelling of football outcomes: using the Skellam’s distribution for the goal difference”. Em: *IMA Journal of Management Mathematics* 20.2 (set. de 2008), pp. 133–145. ISSN: 1471-678X. DOI: 10.1093/imaman/dpn026. eprint: <https://academic.oup.com/imaman/article-pdf/20/2/133/2118878/dpn026.pdf>. URL: <https://doi.org/10.1093/imaman/dpn026>.
- [8] Roberto Di Mari Leopoldo Catania e Paolo Santucci de Magistris. “Dynamic Discrete Mixtures for High-Frequency Prices”. Em: *Journal of Business & Economic Statistics* 40.2 (2022), pp. 559–577. DOI: 10.1080/07350015.2020.1840994. eprint: <https://doi.org/10.1080/07350015.2020.1840994>. URL: <https://doi.org/10.1080/07350015.2020.1840994>.
- [9] Jerry W Lewis, Patrick E Brown e Michail Tsagris. *Densities and Sampling for the Skellam Distribution*. R package version 0.2.0. 2016. URL: <http://r-forge.r-project.org/projects/healthqueues>.
- [10] Ke Mao Libo Jiang e Rongling Wu. “A skellam model to identify differential patterns of gene expression induced by environmental signals”. Em: *BMC Genomics* 15.772 (1 2014). DOI: 10.1186/1471-2164-15-772.
- [11] Antonino A. Feitosa Neto. *Skellam-Distribution*. Dez. de 2023. URL: <https://github.com/antonino-feitosa/Skellam-Distribution>.
- [12] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: <https://www.R-project.org/>.
- [13] Martje Rave e Göran Kauermann. *The Skellam Distribution revisited - Estimating the unobserved incoming and outgoing ICU COVID-19 patients on a regional level in Germany*. 2023. arXiv: 2305.15301 [stat.AP].
- [14] Christian Robert e George Casella. “Monte Carlo Statistical Method”. Em: *Technometrics* 42 (nov. de 2000). DOI: 10.2307/1270959.
- [15] S. Ross. *Probabilidade: Um Curso Moderno com Aplicações*. Bookman, 2009. ISBN: 9788577806881. URL: <https://books.google.com.br/books?id=ZKXtwsmyPFcC>.

- [16] J. G. Skellam. “The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations”. Em: *Journal of the Royal Statistical Society* 109.3 (1946), pp. 296–296. DOI: <https://doi.org/10.1111/j.2397-2335.1946.tb04670.x>. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2397-2335.1946.tb04670.x>. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2397-2335.1946.tb04670.x>.
- [17] *to pull oneself up by one’s bootstraps*. <https://www.collinsdictionary.com/dictionary/english/to-pull-yourself-up-by-ones-bootstraps>. Disponível em: <https://www.collinsdictionary.com/dictionary/english/to-pull-yourself-up-by-ones-bootstraps>. Acesso em: 28 nov. 2023.
- [18] Hao Wang. *Skellam Rank: Fair Learning to Rank Algorithm Based on Poisson Process and Skellam Distribution for Recommender Systems*. 2023. arXiv: 2306.06607 [cs.IR].