



ESTATÍSTICA II

1 de agosto de 2024

Simulações do Estimador de Máxima Verossimilhança

Nome do Estudante: Kauã Dias Paula

1 Definindo as propriedades do estimador

Da função de probabilidade de massa

$$f(x; \theta) = \frac{1}{\theta} \mathbb{1}_{\{1, 2, \dots, \theta\}}(x),$$

onde $\theta = 1, 2, \dots$

Foi calculado o seguinte estimador de máxima verossimilhança do parâmetro θ :

$$\hat{\theta}_{EMV} = Y_n = \max\{x_1, x_2, x_3, \dots, x_n\}. \quad (1)$$

onde Y_n é máximo retirado da amostra.

Para encontrar a Função Massa de Probabilidade (FMP):

$$\mathbb{P}(Y_n = y) = \mathbb{P}(Y_n \leq y) - \mathbb{P}(Y_n < y),$$

para $\mathbb{P}(Y_n \leq y)$, sabe-se que por i.i.d.

$$\mathbb{P}(Y_n \leq y_1, Y_n \leq y_2, \dots, Y_n \leq y_n) = \mathbb{P}(Y_n \leq y_1) \times \mathbb{P}(Y_n \leq y_2) \times \dots \times \mathbb{P}(Y_n \leq y_n),$$

portanto a

$$\mathbb{P}(Y_n \leq y_n) = \left(\frac{y}{\theta}\right)^n.$$

Para o cálculo da $\mathbb{P}(Y_n < y)$, segue-se a mesma lógica, o resultado é $\left(\frac{y-1}{\theta}\right)^n$. Logo

$$\begin{aligned} \mathbb{P}(Y_n = y) &= \mathbb{P}(Y_n \leq y) - \mathbb{P}(Y_n < y) \\ &= \frac{y^n - (y-1)^n}{\theta^n}, \end{aligned} \quad (2)$$

para $\theta > 0$.

O primeiro momento é dado por:

$$\mathbb{E}(Y_n) = \theta - \frac{1}{\theta^n} \sum_{y=1}^{\theta-1} y^n. \quad (3)$$

Percebe-se que o estimador é viesado, pois

$$B(Y_n) = -\frac{1}{\theta^n} \sum_{y=1}^{\theta-1} y^n. \quad (4)$$

Nota-se que, o viés $B(Y_n) \leq 0$, portanto o estimador tende a subestimar o valor definitivo de θ .

Calcula-se também o $\mathbb{E}(Y_n^2)$:

$$\mathbb{E}(Y_n^2) = \theta^2 - \frac{1}{\theta^n} \sum_{y=1}^{\theta-1} y^n (2y + 1). \quad (5)$$

Nos revelando a $\text{Var}(Y_n)$:

$$\text{Var}(Y_n) = -\frac{1}{\theta^n} \sum_{y=1}^{\theta-1} y^n (2y+1) + \frac{2}{\theta^{n-1}} \sum_{y=1}^{\theta-1} y^n - \frac{1}{\theta^{2n}} \left(\sum_{y=1}^{\theta-1} y^n \right)^2. \quad (6)$$

Do qual nos fornece o Erro Quadrático Médio (EQM):

$$\text{EQM}(Y_n) = \text{Var}(Y_n) - [B(Y_n)]^2,$$

por fim

$$\text{EQM}(Y_n) = -\frac{1}{\theta^n} \sum_{y=1}^{\theta-1} y^n (2y+1) + \frac{2}{\theta^{n-1}} \sum_{y=1}^{\theta-1} y^n. \quad (7)$$

Para descobrir o tamanho da amostra n que satisfaça a propriedade $P(Y_n = \theta) = 1$, pela Equação (2), temos

$$\begin{aligned} P(Y_n = \theta) &= \frac{\theta^n - (\theta-1)^n}{\theta^n} \\ &= 1 - \left(\frac{\theta-1}{\theta} \right)^n, \end{aligned}$$

logo

$$\begin{aligned} 1 - \left(\frac{\theta-1}{\theta} \right)^n &= 1 \\ \left(\frac{\theta-1}{\theta} \right)^n &= 0, \end{aligned}$$

é fácil notar que isso ocorre quando $\theta = 1$, porém o termo $\left(\frac{\theta-1}{\theta} \right) < 1$, pelo fato do denominador ser sempre maior que o numerador. Então a expressão $\left(\frac{\theta-1}{\theta} \right)^n$ diminui quando n aumenta, portanto $\left(\frac{\theta-1}{\theta} \right)^n = 0$ quando $n \rightarrow \infty$. Entretanto como a $P(Y_n = \theta) = 1$ depende apenas do quão pequeno é o termo $\left(\frac{\theta-1}{\theta} \right)^n$, podemos criar um critério para estabelecer n de acordo com um erro tolerável para a probabilidade final, por exemplo:

Vamos fixar o termo $\left(\frac{\theta-1}{\theta} \right)^n = 10^{-5}$, de acordo que $P(Y_n = \theta) = 1 - 0,00001 = 0,99999$. Logo

$$\begin{aligned} \left(\frac{\theta-1}{\theta} \right)^n &= 10^{-5} \\ n \log_{10} \left(\frac{\theta-1}{\theta} \right) &= -5 \log_{10} 10 \\ n &= -\frac{5}{\log_{10} \left(\frac{\theta-1}{\theta} \right)}. \end{aligned}$$

Percebe-se que o expoente -5 do qual indica quantas casas decimais nós queremos tolerar de "erro", se mantém intacto no numerador. Sendo assim, podemos chamar o expoente de d , do

qual indica quantas casas decimais iremos tolerar para nos aproximarmos de $P(Y_n = \theta) = 1$. A regra então se dá:

$$n = -\frac{d}{\log_{10}\left(\frac{\theta-1}{\theta}\right)}, \quad (8)$$

para $\theta \neq 1$.

2 O algoritmo utilizado para a simulação

1. Atribuindo n (os valores da amostra) a um vetor (50, 100, 200, 500, 1000);
2. Definindo REP o número de réplicas no valor de 1000;
3. Atribuindo θ a um vetor (5, 10, 100);
4. Definindo um *looping* que para um valor $j = 1, 2, 3$, θ_j percorra um valor de n_i , em que $i = 1, 2, 3, 4, 5$;
5. Dentro do *looping*:
 - a) Cria-se uma amostra com com função de distribuição uniforme discreta com réplicas de tamanho REP, amostras n_i e parâmetro θ_j ;
 - b) Calcula-se as estimativas utilizando a Equação (1);
 - c) Observa a distribuição do estimador;
 - d) Cria-se uma variável de tolerância $= \frac{1}{9e+307}$, para o resultado da expressão $\frac{1}{\theta^n}$;
 - e) Cria-se uma condição para que se a expressão for menor que a tolerância, $\mathbb{E}(Y_n) = \theta$, $\mathbb{E}(Y_n^2) = \theta^2$ e $B(Y_n) = 0$. Do contrário, calcula-se as propriedades utilizando suas respectivas equações, (3), (5) e (4);
 - f) Calcula-se o a raíz do Erro Quadrático Médio (*REQM*) teórico do estimador dado pela Equação (7);
 - g) Calcula-se um intervalo de confiança de 95% para θ da seguinte maneira:

$$IC(\hat{\theta}; 95\%) = \left[\hat{\theta} - \hat{\theta}_{0.975} \sqrt{\text{Var}(Y_n)}; \hat{\theta} + \hat{\theta}_{0.975} \sqrt{\text{Var}(Y_n)} \right], \quad (9)$$

onde $\hat{\theta}_{0.975}$ é o quantil 97.5% representante do estimador;

- h) Calcula-se a amplitude do intervalo, isto é, o limite superior - o limite inferior;
- i) Observa como o *REQM* de cada θ se comporta em relação ao número de amostras.

Com isso obtemos o viés, o *REQM* e o intervalo de confiança de 95% para cada tamanho de amostra e θ para o estimador.

Pela Equação (8) sabemos que n pode ser descrito em função de θ , logo para o vetor θ criado, criamos um novo vetor n com as funções $n = \left\lceil \frac{\theta}{2} \right\rceil$, $n = \theta$ e $n = 2\theta$, onde $\left\lceil \frac{\theta}{2} \right\rceil$ é a função teto de $\frac{\theta}{2}$, já que n há de ser escrito como $\{n \in \mathbb{Z} | n > 0\}$. Em resumo, a regra é $n(\theta) = \lceil 2^x \theta \rceil$, onde $\{x \in \mathbb{Z}\}$ e nesse caso $\{x \in [-1, 1]\}$.

Cria-se um novo *looping* para uma segunda simulação com as mesmas características do anterior, porém agora com o novo vetor n . Por fim, observa-se como a $P(Y_n = \theta)$ se aparenta quando n é descrito da forma previamente exposta.

3 Resultados

Os valores das Tabelas foram arredondados para 5 casas decimais.

Resultados da Primeira Simulação

| Número de Amostras | $\theta = 5$ | | | $\theta = 10$ | | | $\theta = 100$ | | |
|--------------------|--------------|---------|--------------|---------------|---------|--------------|----------------|---------|--------------|
| | Vício | REQM | REQM Teórico | Vício | REQM | REQM Teórico | Vício | REQM | REQM Teórico |
| 50 | -0,00001 | 0.03164 | 0.00378 | -0.00517 | 0.06328 | 0.07209 | -1.50229 | 2.38550 | 2.42375 |
| 100 | 0 | 0.00000 | 0.00001 | -0.00003 | 0.00000 | 0.00515 | -0.57211 | 1.01575 | 1.10024 |
| 200 | 0 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00003 | 0.00000 | 0.45187 | 0.00000 |
| 500 | 0 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.06328 | 0.00000 |
| 1000 | 0 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.03164 | 0.00000 |

* Onde *REQM* é a raiz quadrada do Erro Quadrático Médio

Tabela 1: Vício e REQM do Estimador $\hat{\theta}$.

| Número de Amostras | $\theta = 5$ | | $\theta = 10$ | | $\theta = 100$ | |
|--------------------|----------------|-----------|-----------------|-----------|------------------|-----------|
| | Intervalo | Amplitude | Intervalo | Amplitude | Intervalo | Amplitude |
| 50 | [4.981; 5.019] | 0.038 | [9.276; 10.714] | 1.438 | [-91.704; 288.7] | 380.404 |
| 100 | [5; 5] | 0 | [9.948; 10.052] | 0.104 | [5.448; 193.408] | 187.96 |
| 200 | [5; 5] | 0 | [10; 10] | 0 | [100; 100] | 0 |
| 500 | [5; 5] | 0 | [10; 10] | 0 | [100; 100] | 0 |
| 1000 | [5; 5] | 0 | [10; 10] | 0 | [100; 100] | 0 |

Tabela 2: Intervalo de Confiança de 95% para θ .

| θ | <i>REQM</i> |
|----------|-------------|
| 5 | 100% |
| 10 | 100% |
| 100 | 100% |

Tabela 3: Taxa de Queda de 50 a 1000 amostras do *REQM* do Estimador $\hat{\theta}$.

3. RESULTADOS

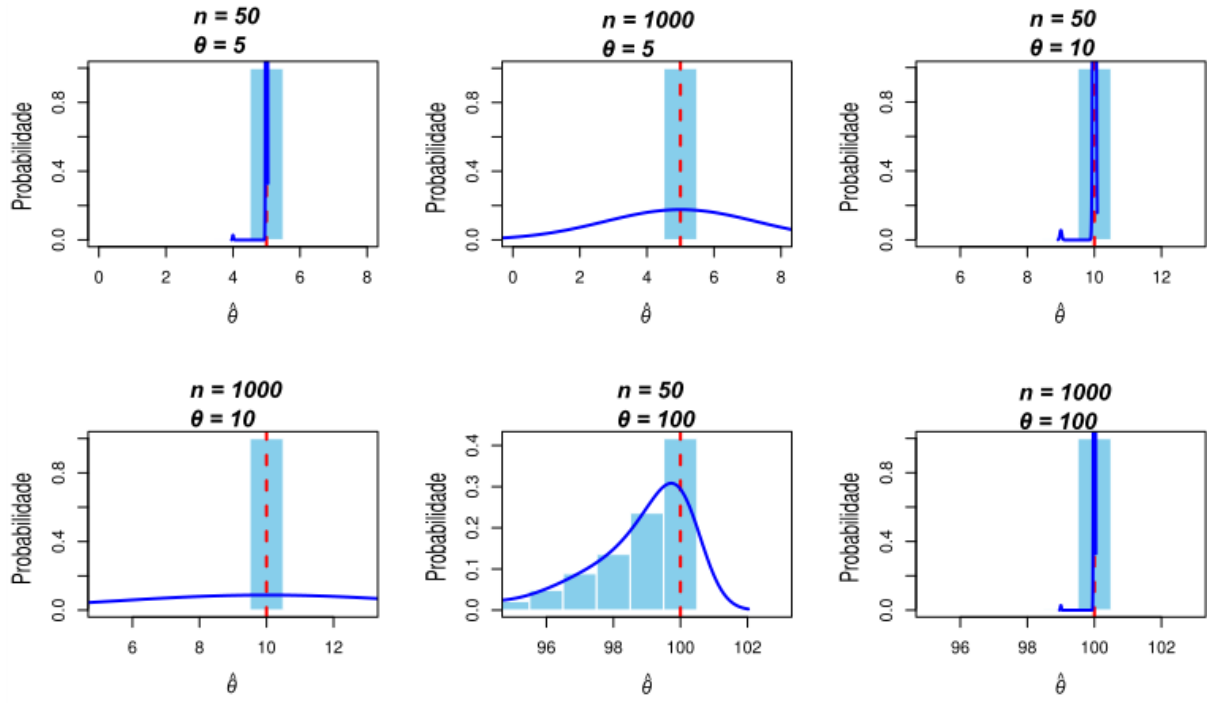


Figura 1: Gráficos da distribuição do estimador $\hat{\theta}_{EMV}$.

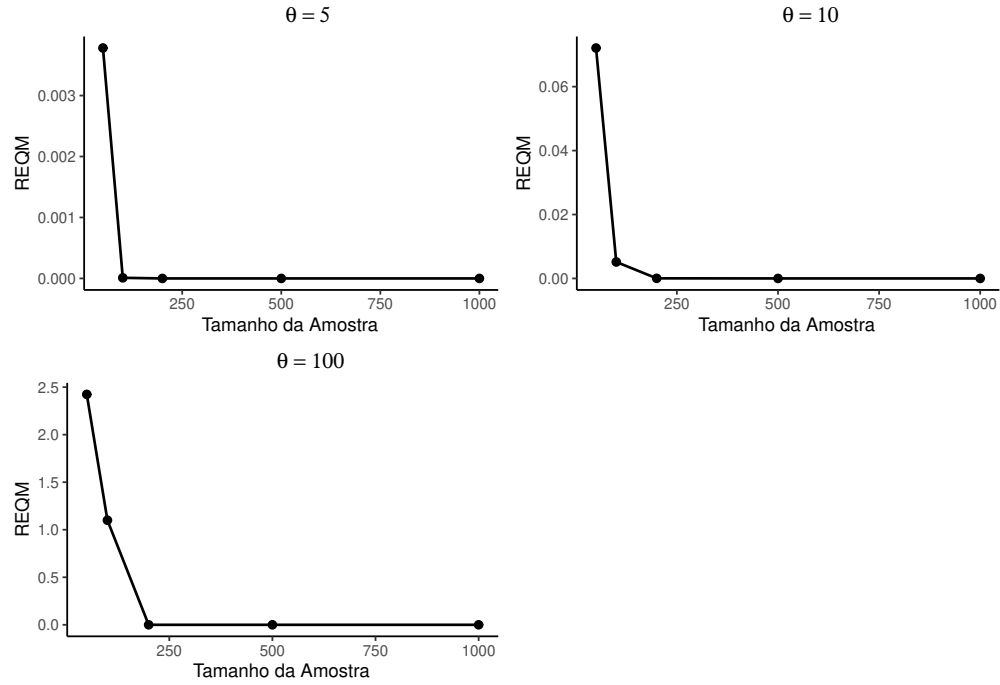


Figura 2: Gráficos do REQM de cada θ em função do número de amostras.

Resultados da Segunda Simulação

| Número de Amostras | $\theta = 5$ | | | $\theta = 10$ | | | $\theta = 100$ | | |
|------------------------|--------------|---------|--------------|---------------|---------|--------------|----------------|---------|--------------|
| | Vício | REQM | REQM Teórico | Vício | REQM | REQM Teórico | Vício | REQM | REQM Teórico |
| $n = \frac{\theta}{2}$ | -0.8 | 1.29111 | 1.23935 | -1.20825 | 1.85632 | 1.84083 | -1.50229 | 2.28848 | 2.42375 |
| $n = \theta$ | -0.416 | 0.78642 | 0.78384 | -0.49143 | 1.00684 | 0.92972 | -0.57211 | 1.09312 | 1.10024 |
| $n = 2\theta$ | -0.11353 | 0.37165 | 0.35502 | -0.13394 | 0.3926 | 0.40052 | 0 | 0.41971 | 0 |

* Onde $REQM$ é a raiz quadrada do Erro Quadrático Médio

Tabela 4: Vício e REQM do Estimador $\hat{\theta}$ Para a Segunda Simulação.

| Número de Amostras | $\theta = 5$ | | $\theta = 10$ | | $\theta = 100$ | |
|------------------------|-----------------|-----------|-----------------|-----------|------------------|-----------|
| | Intervalo | Amplitude | Intervalo | Amplitude | Intervalo | Amplitude |
| $n = \frac{\theta}{2}$ | [-0.533; 8.933] | 9.466 | [-5.096; 22.68] | 27.776 | [-91.704; 288.7] | 380.404 |
| $n = \theta$ | [1.262; 7.906] | 6.644 | [1.616; 17.401] | 15.785 | [5.448; 193.408] | 187.96 |
| $n = 2\theta$ | [3.205; 6.568] | 3.363 | [6.091; 13.641] | 7.55 | [100; 100] | 0 |

Tabela 5: Intervalo de Confiança de 95% para θ Para a Segunda Simulação.

| θ | REQM |
|----------|---------|
| 5 | 71.354% |
| 10 | 78.242% |
| 100 | 100% |

Tabela 6: Taxa de Queda de $n = \frac{\theta}{2}$ a $n = 2\theta$ amostras do REQM do Estimador $\hat{\theta}$.

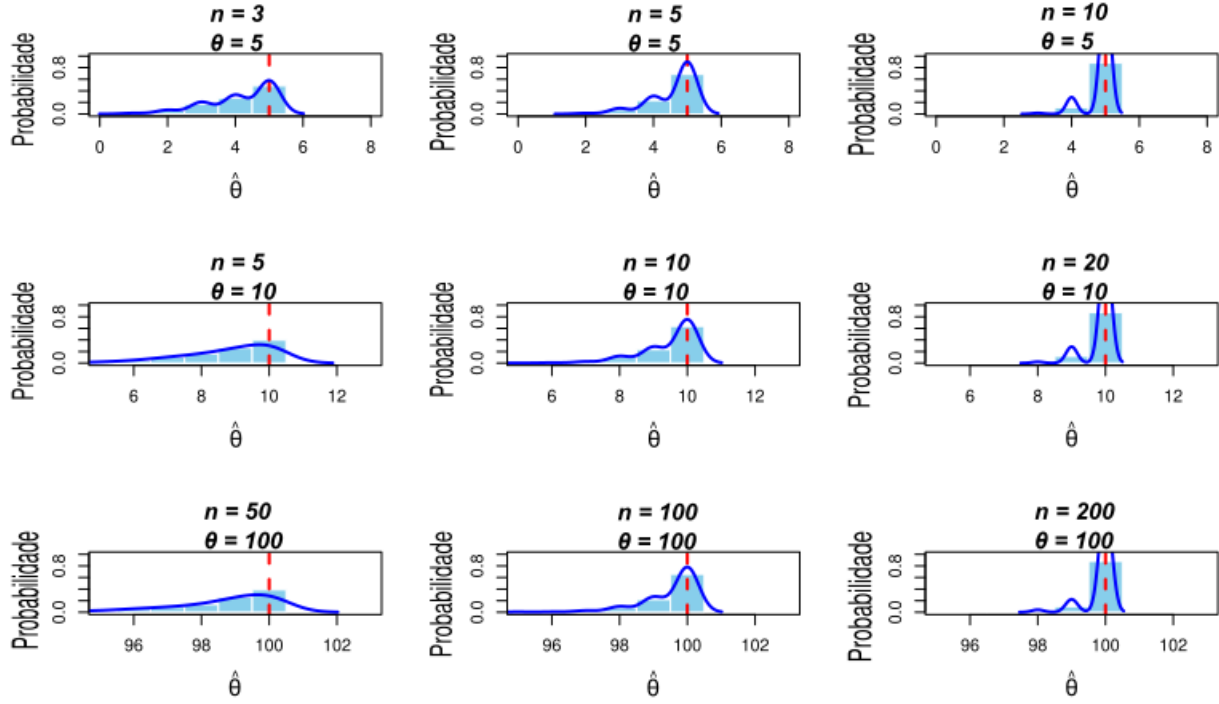


Figura 3: Gráficos da distribuição do estimador $\hat{\theta}_{EMV}$ para n em função de θ .

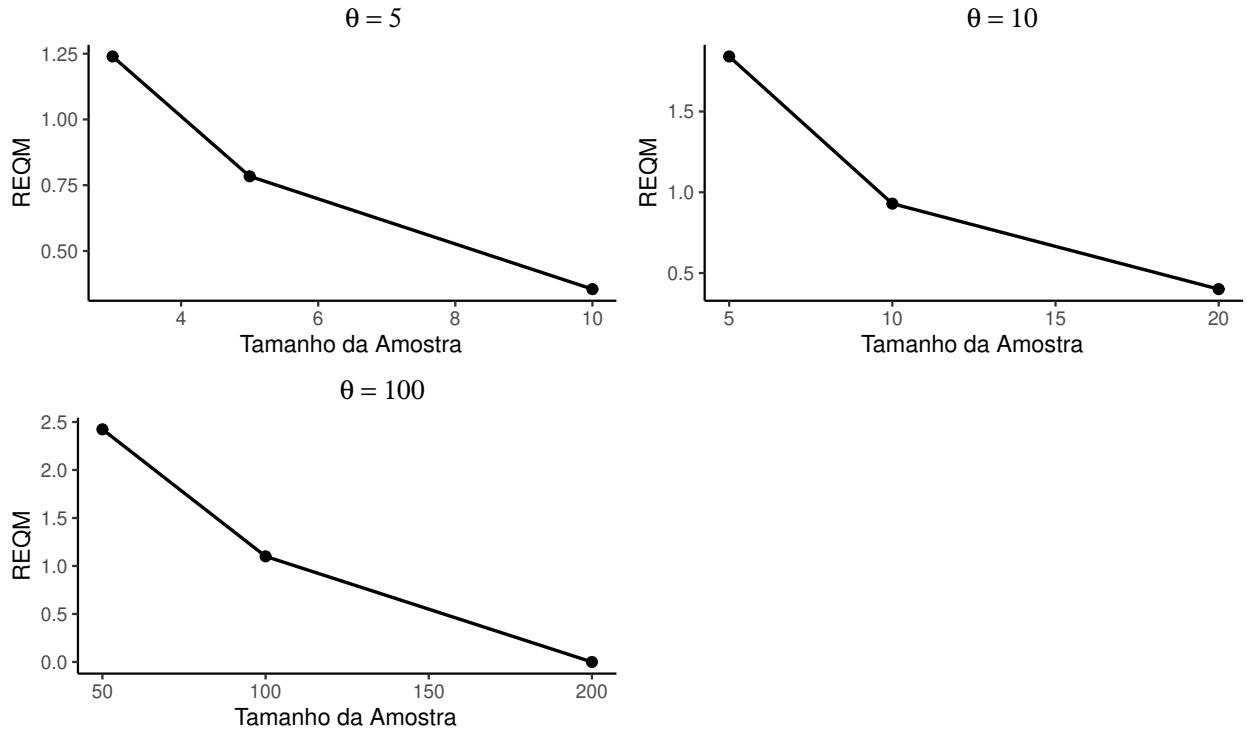


Figura 4: Gráficos da distribuição do estimador $\hat{\theta}_{EMV}$ para n em função de θ .

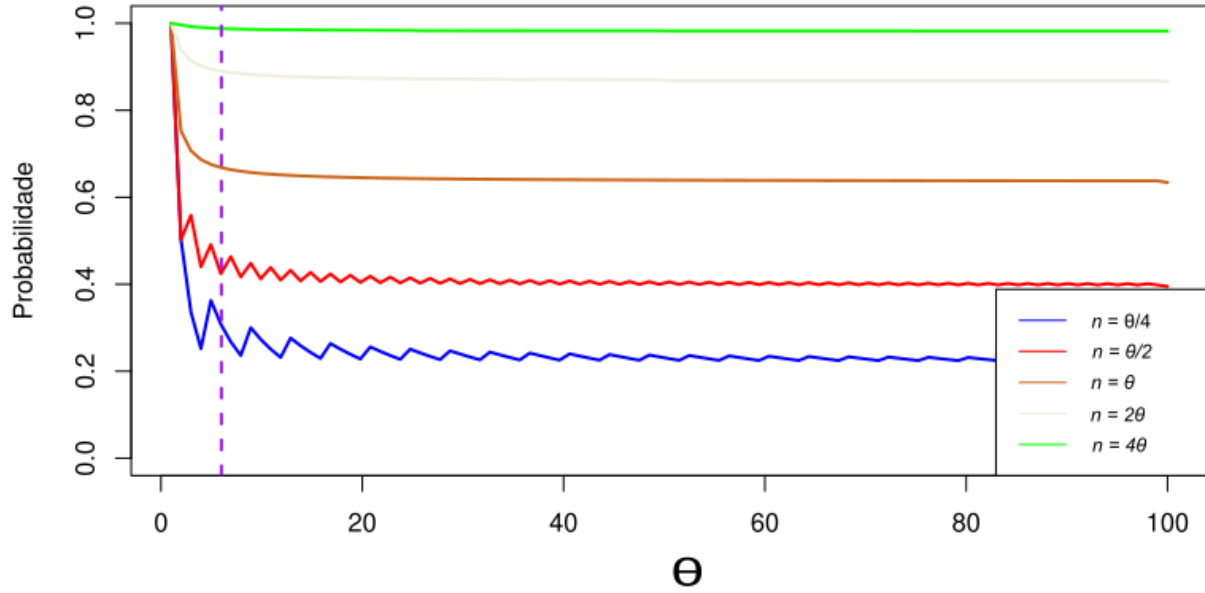


Figura 5: Gráfico de $P(Y_n = \theta)$ quando $n = \lceil 2^x \theta \rceil$ para $x \in [-4, 4]$.

4 Discussão

Antes de comentar sobre os resultados, vale ressaltar que, para a simulação correr de forma com que não haja erros de números absurdos, foi necessário criar uma condição de tolerância para a expressão $\frac{1}{\theta^n}$, da qual aparece nas Equações (3) e (5), isso se deu pelo fato de que a linguagem R, em específico, computa valores até 10^{308} , ou seja, caso um número seja $> 10^{308} = \infty$, caso seja $< 10^{308} = -\infty$, agora caso este valor esteja no denominador, $< \frac{1}{10^{308}} = 0$, dito isso, em alguns momentos na simulação ocorre que a expressão da soma das equações (3) e (5) apontam o valor de ∞ enquanto que o termo $\frac{1}{\theta^n} = 0$, ao final da conta, o computador tenta fazer o produto entre $0 \times \infty$ e aponta valores *NaN*. Como $\theta^n > \sum_{y=1}^{\theta-1} y^n$ e $\sum_{y=1}^{\theta-1} y^n(2y+1)$, para contornar o cenário descrito, caso o valor de $\frac{1}{\theta^n}$ seja menor que o valor tolerado escolhido, o produto das somas e da expressão $\frac{1}{\theta^n}$ será igual a zero, sendo assim, $\mathbb{E}(Y_n) = \theta$ e $\mathbb{E}(Y_n^2) = \theta^2$, de acordo com as Equações (3) e (5).

Discussão Acerca da Primeira Simulação

Da Tabela 1 é possível observar que o vício de cada θ quando diferente de zero, é sempre um valor negativo, o que concorda com o viés calculado na Equação (4) e também que o viés para amostras de tamanho 200 ou maior, é zero. $\theta = 100$ chama atenção, pois os vícios para tamanho da amostra iguais a 50 e 100 são os maiores valores computados. Para o *REQM*, nota-se que o valor também diminui de acordo com o Número de Amostra, é possível analisar que o valor Teórico do *REQM* é bastante próximo dos valores simulados, nos dando

certa garantia de que não houve erro nos cálculos e nem na simulação. Uma vez garantido a segurança entre os valores teóricos e os valores simulados, percebe-se que para cada valor θ os valores das propriedades do estimador aumentam, em $\theta = 100$ os valores do *REQM* aumentam de forma preocupante, além disso, a Figura 2 demonstra que a relação do *REQM* e o Tamanho da Amostra para cada θ se comporta de maneira similar, porém para $\theta = 100$ o *REQM* demora mais para chegar a zero, isso se da ao fato de que $\theta = 100$ é o único valor de θ que é maior que os primeiros valores da amostra. Bem como a Tabela 3, que certifica a afirmação de que o *REQM* tem uma taxa de queda ao longo do Tamanho da Amostra com cada valor de θ próxima, nota-se também que para todos os θ de 50 a 1000 amostras o *REQM* vai para zero.

As informações da Figura 1 nos mostra a distribuição de $\hat{\theta}_{EMV} = 5, 10$ e 100 para amostras de tamanho $n = 50, 1000$. Repare a peculiaridade do estimador em relação aos seus parâmetros, $\hat{\theta}_{EMV}$ não segue uma distribuição normal quando n aumenta, na verdade para n suficientemente maior que θ , a $P(Y_n = \theta) = 1$ pela Equação (8), logo o estimador converge para uma distribuição degenerada. Em específico, nota-se que para $\theta = 100$ e $n = 50$, existe uma maior distribuição dos dados. Da Tabela 2, a amplitude do intervalo de cada θ diminui de acordo com o Número de Amostra, comprovando a acurácia do estimador, $\theta = 100$ chama atenção novamente, pois apresenta intervalos muito grandes para $n = 50$ e 100, e inclusive, para $n = 50$ o valor de $\theta = 100$ é estatisticamente zero, porém como debatido anteriormente, é coerente que para $\theta \geq n$ algumas propriedades tenham valores altos. Para $n > 100$, todos os intervalos têm amplitudes iguais a zero, para $\theta = 5$ isso ocorre a partir de $n \geq 100$, de novo, o resultado é compreensível, dado que para n suficientemente maior que θ a $P(Y_n = \theta) = 1$.

Discussão Acerca da Segunda Simulação

Na Tabela 4, nota-se que os valores do *REQM* teóricos e simulados também são muito próximos, nos garantindo mais uma vez a ausência de erros nos cálculos e simulações. Também da Tabela 4, para $n = \frac{\theta}{2}$ o vício e *REQM* do estimador $\hat{\theta}_{EMV}$ são exorbitantes, para $n = \theta$ as propriedades diminuem, porém, seguem altas, já para $n = 2\theta$, os valores diminuem ainda mais, chegando em resultados aceitáveis para o *REQM* e bons para vício, para $\theta = 100$ esses valores chegam a zero. A Figura 4 mostra que o *REQM* tem uma taxa de queda semelhante para cada θ e $n(\theta)$, porém, apenas para $\theta = 100$ o *REQM* vai para zero, oque é comprovado pela Tabela 6, onde para $\theta = 5$ e $\theta = 10$ o *REQM* tem uma taxa de queda aproximada, mas para $\theta = 100$ a taxa de queda é de 100%.

A Figura 3 nos mostra que para todos os θ , se n é descrito como $n(\theta)$ então as distribuições são aproximadamente iguais, no caso específico de quando θ é ímpar como é o caso dos gráficos de $\theta = 5$, a distribuição se difere um pouco quando $n = \frac{\theta}{2}$ pelo fato de que, obviamente a metade de um número ímpar não pode ser inteiro, oque desafia a lógica para descrever n , logo $n(\theta) = \lceil 2^x \theta \rceil$ para um θ ímpar não é exatamente a metade de θ como quando θ se revela como par. A Tabela 5 mostra que para $n = \frac{\theta}{2}$ o valor de θ é estatisticamente zero com uma amplitude absurda com ênfase para quando $\theta = 100$, quando $n = \theta$ a amplitude diminui porém ainda sim é de certa forma ineficaz, ainda com destaque para $\theta = 100$ que continua com uma amplitude extremamente elevada, quando $n = 2\theta$ os intervalos se tornam mais eficazes, para $\theta = 100$ a amplitude vai para zero.

A figura 5 nos traz visualmente como a $P(Y_n = \theta)$ se comporta com $n(\theta)$ nas condições previamente descritas, nota-se que para $\theta = 1$ a $P(Y_n = \theta) = 1$ para qualquer n como já

4. DISCUSSÃO

explicado, também é possível observar que para cada $n(\theta)$ as probabilidades seguem modelos parecidos e que aparentam se estabilizar para quando $\theta > 5$. Quando $n = 2\theta$ e $\theta > 5$, por exemplo, a $P(Y_n = \theta) \approx 0,86$. A linha que representa $n = \frac{\theta}{2}$ difere um pouco das outras pelo mesmo motivo antes explicado, como para calcular n nessas condições fora usado a função teto, a $P(Y_n = \theta)$ para um θ ímpar será sempre maior que o próximo θ par, logo causando uma pequena oscilação na linha que não se observa nas outras, que se apresentam sempre suaves, para $n = 4\theta$ a $P(Y_n = \theta) \approx 1$ para qualquer θ de um a cem.

$\hat{\theta}_{MM}$ VS $\hat{\theta}_{EMV}$

Os resultados das Tabelas (7), (8) e (9) foram resgatados de um estudo anterior sobre o estimador $\hat{\theta}$ pelo método de momentos para a função de probabilidade de massa $f(x; \theta) = \frac{1}{\theta} \mathbb{1}_{\{1,2,\dots,\theta\}}(x)$, onde $\theta = 1, 2, \dots$ (Para mais detalhes acesse <https://drive.google.com/file/d/1cx6kLNwMN6gHiRflTVqyI54KXMPt9ZYc/view?usp=sharing>).

| Número de Amostras | $\theta = 5$ | | | $\theta = 10$ | | | $\theta = 100$ | | |
|--------------------|--------------|---------|--------------|---------------|---------|--------------|----------------|---------|--------------|
| | Vício | REQM | REQM Teórico | Vício | REQM | REQM Teórico | Vício | REQM | REQM Teórico |
| 50 | 0.00396 | 0.42405 | 0.40000 | -0.02820 | 0.79419 | 0.81240 | 0.14212 | 8.02637 | 8.16456 |
| 100 | -0.00572 | 0.28917 | 0.28284 | -0.00202 | 0.60130 | 0.57446 | -0.08526 | 5.73433 | 5.77321 |
| 200 | 0.00120 | 0.19777 | 0.20000 | 0.01046 | 0.42170 | 0.40620 | 0.02897 | 4.07050 | 4.08228 |
| 500 | 0.00085 | 0.12245 | 0.12649 | 0.00567 | 0.26107 | 0.25690 | 0.03497 | 2.50347 | 2.58186 |
| 1000 | 0.00012 | 0.08801 | 0.08944 | 0.00114 | 0.18293 | 0.18166 | -0.03406 | 1.84123 | 1.82565 |

* Onde REQM é a raiz quadrada do Erro Quadrático Médio

Tabela 7: Vício e REQM do Estimador $\hat{\theta}_{MM}$.

| Número de Amostras | $\theta = 5$ | | $\theta = 10$ | | $\theta = 100$ | |
|--------------------|--------------------|-----------|---------------------|-----------|-----------------------|-----------|
| | Intervalo | Amplitude | Intervalo | Amplitude | Intervalo | Amplitude |
| 50 | [4.17283; 5.83509] | 1.66226 | [8.41518; 11.52842] | 3.11324 | [84.41043; 115.87381] | 31.46338 |
| 100 | [4.42751; 5.56105] | 1.13354 | [8.81942; 11.17654] | 2.35712 | [88.67546; 111.15402] | 22.47856 |
| 200 | [4.61358; 5.38882] | 0.77524 | [9.18393; 10.83699] | 1.65306 | [92.05079; 108.00715] | 15.95636 |
| 500 | [4.76086; 5.24084] | 0.47998 | [9.49398; 10.51737] | 1.02339 | [95.12817; 104.94176] | 9.81359 |
| 1000 | [4.82763; 5.17261] | 0.34498 | [9.64259; 10.35969] | 0.7171 | [96.35713; 103.57476] | 7.21763 |

Tabela 8: Intervalo de Confiança de 95% para $\hat{\theta}_{MM}$.

| θ | REQM |
|----------|---------|
| 5 | 79.246% |
| 10 | 76.966% |
| 100 | 77.060% |

Tabela 9: Taxa de queda de 50 a 1000 Amostras do REQM do Estimador $\hat{\theta}_{MM}$.

A priori, foi calculado que

$$\hat{\theta}_{MM} \xrightarrow{p} 2\bar{X}_n - 1,$$

enquanto que para $\hat{\theta}_{EMV}$ segue a Equação (1), é fácil ver que para $\hat{\theta}_{MM}$, o cálculo exige mais tempo pelo fato de haver mais passos até o resultado da estimativa.

A distribuição de $\hat{\theta}_{MM}$ converge para uma Normal, uma vez que o estimador está fortemente ligado a \bar{X} , em oposição, para $\hat{\theta}_{EMV}$ o estimador converge para uma distribuição degenerada. $\hat{\theta}_{MM}$ é um estimador não-viesado, em contrapartida o viés de $\hat{\theta}_{EMV}$ é dado pela Equação (4) e como já discutido, tende a subestimar o valor θ .

Visto que

$$EQM(\hat{\theta}_{MM}) = \frac{\theta^2 - 1}{3n},$$

torna-se indiscutível o fato de que há maior facilidade em calcular o $EQM(\hat{\theta}_{MM})$ que $EQM(\hat{\theta}_{EMV})$ devido a Equação (7).

Os resultados em comparação das Tabelas (1) e (7), mesmo que $\hat{\theta}_{EMV}$ seja viesado, mostram-se absurdamente favoráveis, uma vez que o vício se torna insignificante ao passo que n aumenta, o mesmo ocorre para o $REQM$. Para as Tabelas (2) e (8), o fenômeno descrito se difere apenas para quando $\theta = 100$ e $n \leq 100$. Pelas Tabelas (3) e (9), a taxa de queda do estimador $EQM(\hat{\theta}_{EMV})$ também se apresenta maior para todos os θ .

5 Conclusões

O estudo analisou o comportamento do estimador $\hat{\theta}_{EMV}$ usando simulações na linguagem R, destacando a importância de definir uma condição de tolerância para evitar erros numéricos. Os resultados mostraram que o vício do estimador é negativo para valores de θ diferentes de zero e se aproxima de zero para amostras maiores (200+). O $REQM$ diminui com o aumento do tamanho da amostra, mas é elevado para $\theta = 100$ em amostras menores (50 e 100).

A distribuição do estimador converge para uma distribuição degenerada com $P(Y_n = \theta) = 1$ para grandes valores de n em relação a θ . Os intervalos de confiança diminuem com amostras maiores, mas são maiores para $\theta = 100$ em amostras menores. A segunda simulação confirmou a consistência dos valores teóricos e simulados de $REQM$ e vício, nos mostrando que quando $n = 2\theta$ o estimador $\hat{\theta}_{EMV}$ já alcança uma grande acurácia nos valores de θ .

Análises visuais mostraram que para θ ímpar, as distribuições diferem ligeiramente devido à natureza fracionária de $\frac{\theta}{2}$. As probabilidades de $P(Y_n = \theta)$ estabilizam para valores maiores de θ quando $n(\theta)$. O estudo validou a robustez do estimador $\hat{\theta}_{EMV}$ e destacou a importância do tamanho da amostra e dos valores extremos para garantir a precisão das estimativas e evitar erros numéricos.

$\hat{\theta}_{MM}$ VS $\hat{\theta}_{EMV}$

Os resultados se apresentam favoráveis para $\hat{\theta}_{EMV}$.

6 Informações Computacionais

- Hardware
 - Processador: i3-9100f, 3.50GHz, 4 núcleos e 4 threads
 - Memória RAM: 8GB, DDR4
- Software

- SO: Windows 11 Pro, Versão 23H2, 64 bits
- Linguagem de Programação: R Versão 4.3.1 R Core Team (2023)
- Ambiente de Desenvolvimento Integrado (IDE): RStudio Versão 2024.4.2.764 Posit team (2024)
- Pacotes:
 - * kableExtra Versão 1.4.0 Zhu (2024)
 - * purrr Versão 1.0.2 Wickham & Henry (2023)
 - * ggplot2 Versão 3.5.0 Wickham (2016)
 - * cowplot Versão 1.1.3 Wilke (2024)
- Tempo de Execução do código: 35.20 segundos
- Semente solta
- Código utilizado: (<https://drive.google.com/file/d/19ZSDCFZ-nDtv0KuFsqJHs7H4vcE2gFi6/view?usp=sharing>)

Referências Bibliográficas

Morettin, P. A. & Bussab, W. O. (2017), *Estatística básica*, Saraiva Educação SA.

Posit team (2024), *RStudio: Integrated Development Environment for R*, Posit Software, PBC, Boston, MA.

URL: <http://www.posit.co/>

R Core Team (2023), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

URL: <https://www.R-project.org/>

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

URL: <https://ggplot2.tidyverse.org>

Wickham, H. & Henry, L. (2023), *purrr: Functional Programming Tools*. R package version 1.0.2.

URL: <https://CRAN.R-project.org/package=purrr>

Wikipedia contributors (2024), ‘Degenerate distribution — Wikipedia, the free encyclopedia’. [Online; accessed 14-July-2024].

URL: https://en.wikipedia.org/w/index.php?title=Degenerate_distribution&oldid=1202936736

Wilke, C. O. (2024), *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 1.1.3.

URL: <https://CRAN.R-project.org/package=cowplot>

Zhu, H. (2024), *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.4.0.

URL: <https://CRAN.R-project.org/package=kableExtra>