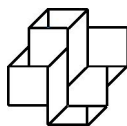


Laboratório Nacional de Computação Científica
Programa de Pós-Graduação em Modelagem Computacional
GA030 - Estatística

Variáveis Aleatória - LGN & TCL

João Pedro dos S. Rocha

Petrópolis-RJ
2024



Laboratório Nacional de Computação Científica
Programa de Pós-Graduação em Modelagem Computacional
GA030 - Estatística

Variáveis Aleatória - LGN & TCL

Trabalho apresentado como parte dos critérios de avaliação da
disciplina: GA030 - Estatística
Prof: Márcio Borges

Petrópolis-RJ
2024

Sumário

1. Descrição	4
2. Questões	4
2. 1. (a) Expectância e Variâncias Teóricas	4
2. 1. 1. Q	4
2. 1. 2. X	4
2. 1. 3. Y	4
2. 1. 4. T	4
2. 2. (b) Médias e variâncias amostrais	5
2. 3. (c) Histogramas	5
2. 4. (d) Construindo variáveis amostrais	6
2. 5. (e) + (f) Visualizando as distribuições construídas + Comparações	7
2. 5. 1. Médias Amostrais	7
2. 5. 2. Variâncias Amostrais	8
Bibliografia	11

1. Descrição

Neste documento estão descritos os resultados obtidos para os exercícios do arquivo trabGA030_2024.pdf disponibilizados em [1] a serem feitos utilizando os dados também disponibilizados em [1]. Serão apresentados os resultados item a item na mesma sequência do documento com as questões. Para a construção deste documento foi assumido que o leitor estará com uma cópia do documento que contém as questões (uma cópia do mesmo pode ser encontrada neste repositório na pasta docs).

2. Questões

Para referência, temos as seguintes variáveis: $Q \sim \mathcal{N}(0, 2)$, $X \sim \mathcal{U}(-1, 1)$, $Y \sim \mathcal{E}(\lambda = 0.05)$, $T \sim \mathcal{B}(15, 0.4)$. As questões que necessitam de apresentação de resultados para cada variável serão colocadas em sub-sessões separadas.

2. 1. (a) Expectância e Variâncias Teóricas

Para cada variável temos as expectâncias e variâncias teóricas especificadas abaixo, começando por Q .

2. 1. 1. Q

Para a distribuição $Q \sim \mathcal{N}(0, 2)$ temos que as expectâncias e variância teóricas são exatamente os parâmetros da distribuição pois a mesma é uma distribuição normal, ou seja, são respectivamente $\mu_Q = 0$ e $\sigma_Q^2 = 2$.

2. 1. 2. X

Para a variável $X \sim \mathcal{U}(-1, 1)$, temos a expectância e variância teórica dados por:

$$\mu_X = \frac{b + a}{2} = \frac{1 + (-1)}{2} = 0 \quad (1)$$

$$\sigma_X^2 = \frac{(b - a)^2}{12} = \frac{(1 - (-1))^2}{12} = \frac{4}{12} = \frac{1}{3} \quad (2)$$

Onde a é o limite inferior e b é o limite superior.

2. 1. 3. Y

A variável $Y \sim \mathcal{E}(\lambda = 0.05)$ é de uma distribuição exponencial, logo temos os valores teóricos:

$$\mu_Y = \frac{1}{\lambda} = \frac{1}{0.05} = \frac{1}{\frac{1}{20}} = 20 \quad (3)$$

$$\sigma_Y^2 = \frac{1}{\lambda^2} = \frac{1}{\left(\frac{1}{20}\right)^2} = \frac{1}{\frac{1}{400}} = 400 \quad (4)$$

2. 1. 4. T

A variável $T \sim \mathcal{B}(15, 0.40)$ é de uma distribuição binomial, logo temos os valores teóricos:

$$\mu_T = np = 15 \times 0.40 = 15 \times \frac{2}{5} = 6 \quad (5)$$

$$\sigma_T^2 = np(1-p) = 6 \left(1 - \frac{2}{5}\right) = 6 \times \frac{3}{5} = \frac{18}{5} = 3.6 \quad (6)$$

2. 2. (b) Médias e variâncias amostrais

Para este item os resultados foram calculados carregando os dados via a biblioteca numpy no Python e usando as funções embutidas para obter a média e variância amostral. Os resultados com 5 casas decimais foram colocados na Tabela 1.

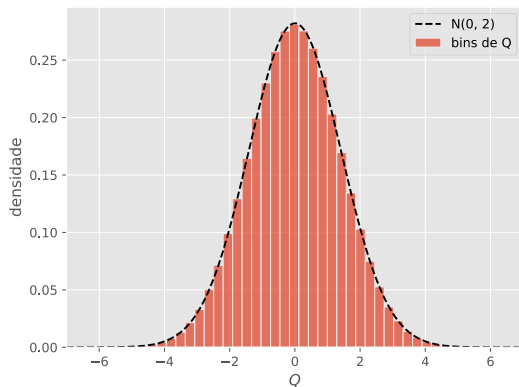
	$\bar{\mu}$	$\bar{\sigma}$	μ	σ	$\Delta\mu$	$\Delta\sigma$
Q	-0.00009	1.99961	0.00000	2.00000	0.00009	0.00039
X	-0.00023	0.33325	0.00000	0.33333	0.00023	0.00008
Y	20.00756	400.36956	20.00000	400.00000	0.00756	0.36956
T	5.99979	3.60221	6.00000	3.60000	0.00021	0.00221

Tabela 1: Resultado amostral contra teórico

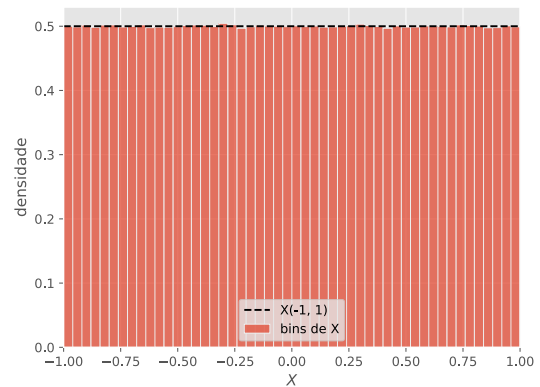
Podemos observar que as estimativas amostrais ficaram bastante próximas como esperado, com diferenças apenas a partir da terceira casa decimal com exceção da variável Y , que apresentou uma estimativa com diferença na primeira casa, no entanto esta é uma variável exponencial cuja variância é de 400, comparando via a diferença relativa percentual temos apenas 0.092%.

2. 3. (c) Histogramas

Montando os histogramas das variáveis aleatórias (com 50 bins nas distribuições contínuas), que podem ser vistos na Figura 1 e Figura 2, podemos ver uma excelente concordância com as distribuições analíticas de origem.

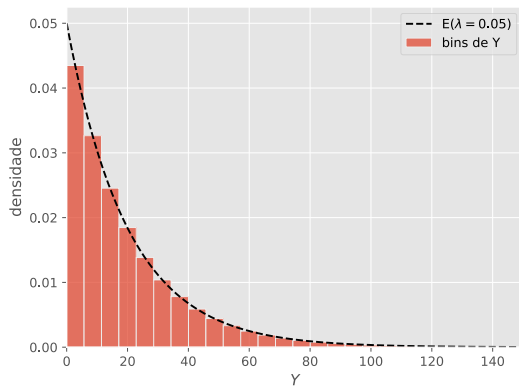


(a) Histograma da variável Q

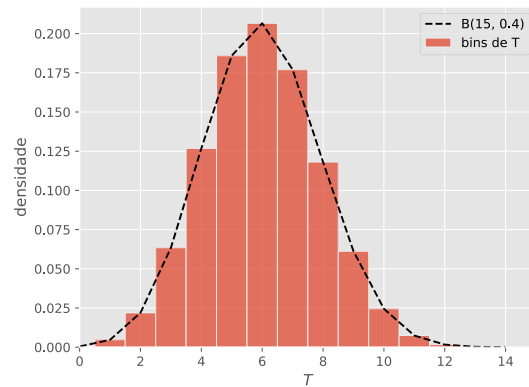


(b) Histograma da variável X

Figura 1: Histogramas das variáveis Q e X



(a) Histograma da variável Y



(b) Histograma da variável T

Figura 2: Histogramas das variáveis Y e T

2. 4. (d) Construindo variáveis amostrais

Para construir as variáveis média amostral e variância amostral foram usados os seguintes códigos em Python usando a biblioteca numpy. Para a média amostral:

```
sizes = [5, 10, 50]
N = 10_000

for var, rv in data.items():

    for s, ax in zip(sizes, axes):
        sample_var_rv = (
            np.random.choice(rv, size=(N, s), replace=True)
            .mean(axis=1)
        )

# Continuação com plotagem #
```

python

E para a variância amostral:

```
for var, rv in data.items():

    for s, ax in zip(sizes, axes):
        sample_var_rv = (
            np.random.choice(rv, size=(N, s), replace=True)
            .var(axis=1, ddof=1)
        )

# Continuação com plotagem #
```

python

Onde a variável data é um dicionário (tabela hash) contendo os dados relativos à cada variável relacionados ao nome da variável. O código roda o núcleo que sorteia 10 mil amostras (`np.choice`) e tira a estatística (`{mean, var}(axis=1)`) para cada variável no dicionário

data e para cada tamanho de amostra na lista `sizes` com reposição. Depois segue com o código de plotagem, que pode ser achado por completo no caderno jupyter que acompanha o relatório.

2. 5. (e) + (f) Visualizando as distribuições construídas + Comparações

As variáveis foram construídas 3 a 3 para cada variável aleatória de origem. Vamos começar listando as médias amostrais.

2. 5. 1. Médias Amostrais

Começando com a variável Q podemos ver que para a mesma as médias amostrais aparecem naturalmente como uma normal com variância $\frac{\sigma^2}{n}$ como esperado por conta da lei dos grandes números e teorema central do limite.

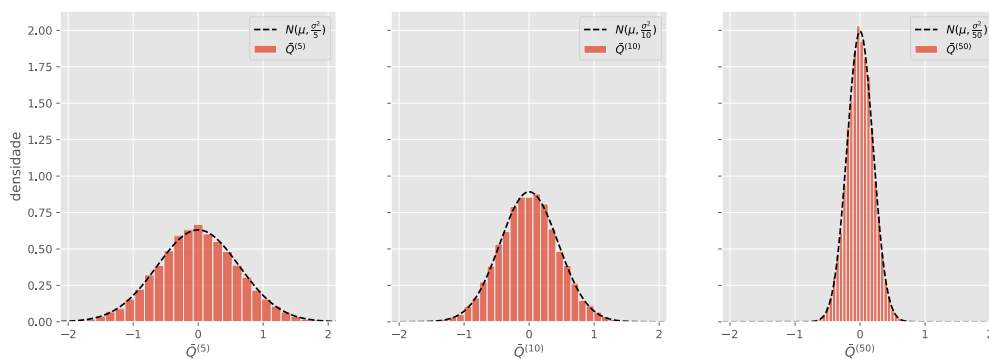


Figura 3: Média amostral de Q

Para a variável X também temos uma aderência rápida a uma normal, com apenas um pouco mais de ruído perceptível à olho.

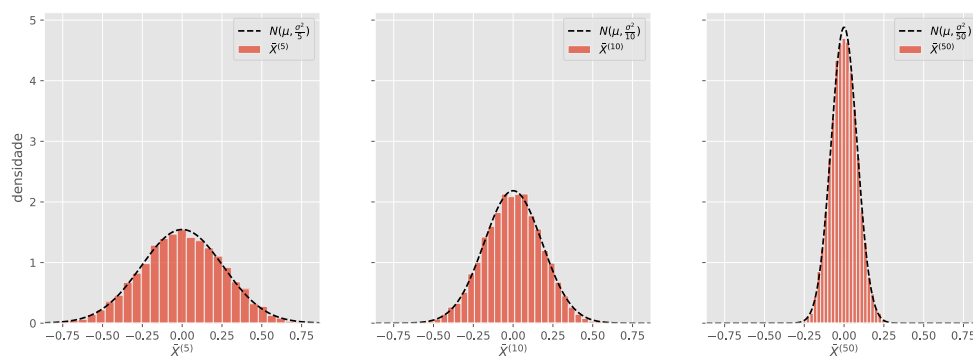


Figura 4: Média amostral de X

Para a variável Y podemos ver uma aderência menor no início, porém com o aumento do tamanho da amostra é possível ver a convergência para uma normal.

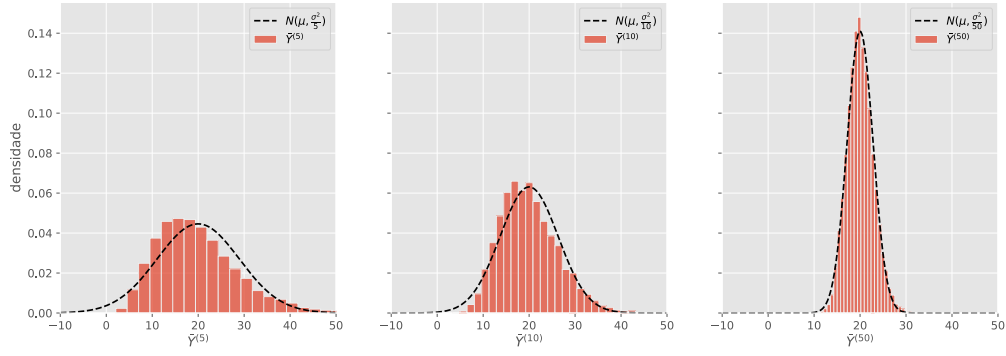


Figura 5: Média amostral de Y

Na variável T podemos ver artefatos na variável média amostral, isso acontece por conta da variável de origem ser uma binomial, ou seja, uma variável discreta. No entanto, da mesma forma podemos ver a variável aderindo a uma normal com o aumento do tamanho da amostra.

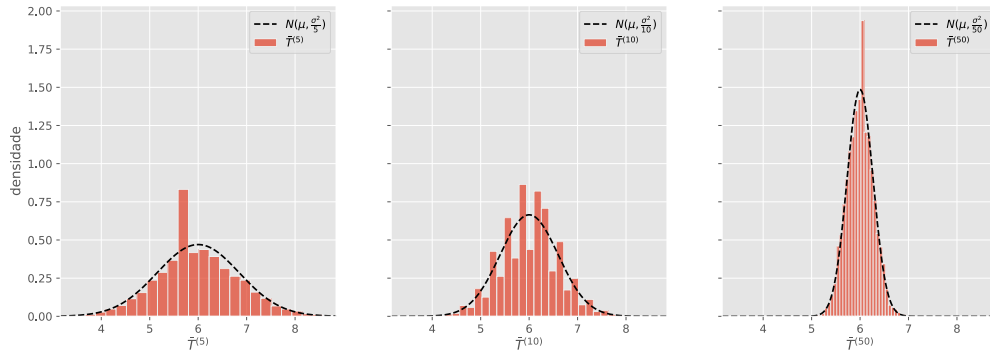


Figura 6: Média amostral de T

Assim podemos ver a aderência de todas as variáveis a uma normal $N(\mu, \frac{\sigma}{n})$ em todos os casos. Vejamos agora as variâncias.

2. 5. 2. Variâncias Amostrais

Com a variável variância amostral podemos ver o comportamento específico das amostras de Q aderindo a uma variável χ^2 com o aumento da amostra quando multiplicado pelo fator $\frac{n-1}{\sigma}$ dado que Q é uma normal. Como podemos ver na Figura 7.

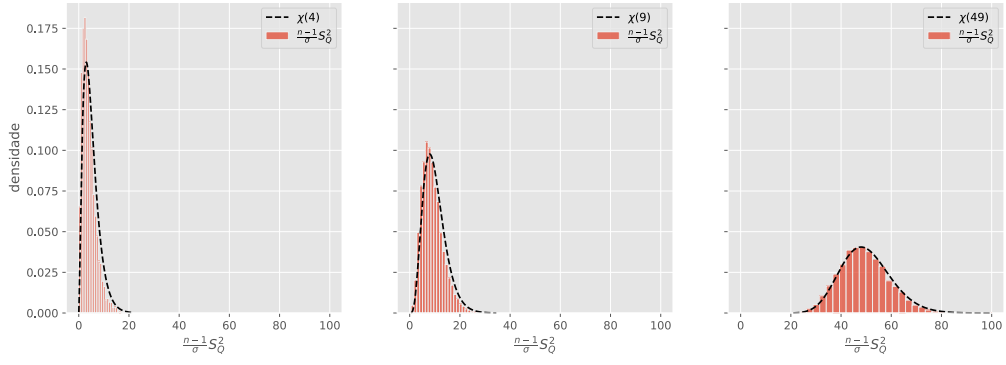


Figura 7: Variância amostral da variável Q . Com χ^2 sobreposta.

Já no caso das outras variáveis não temos garantias teóricas sobre a distribuição da variância amostral. Portanto não podemos sobrepor uma distribuição usando apenas informações teóricas.

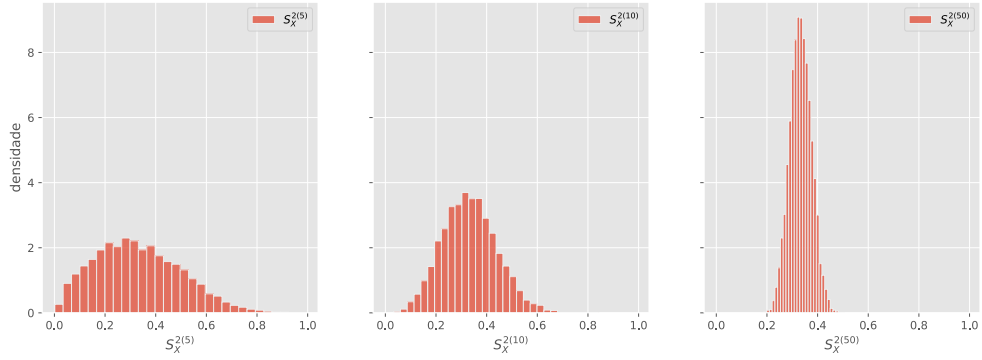


Figura 8: Variância amostral da variável X .

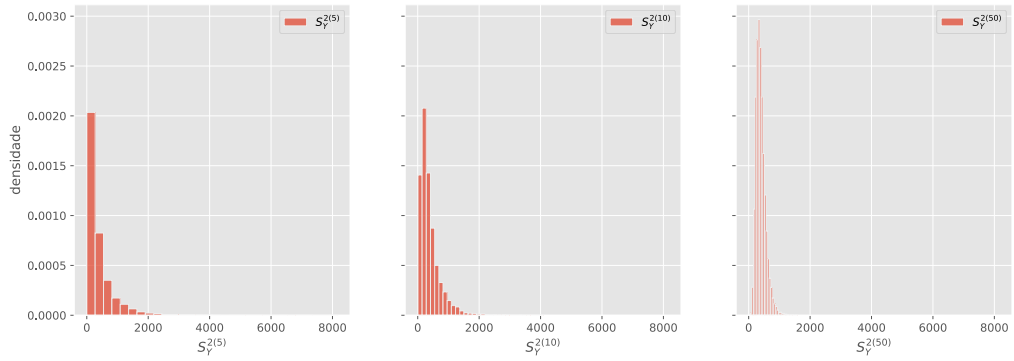


Figura 9: Variância amostral da variável Y .

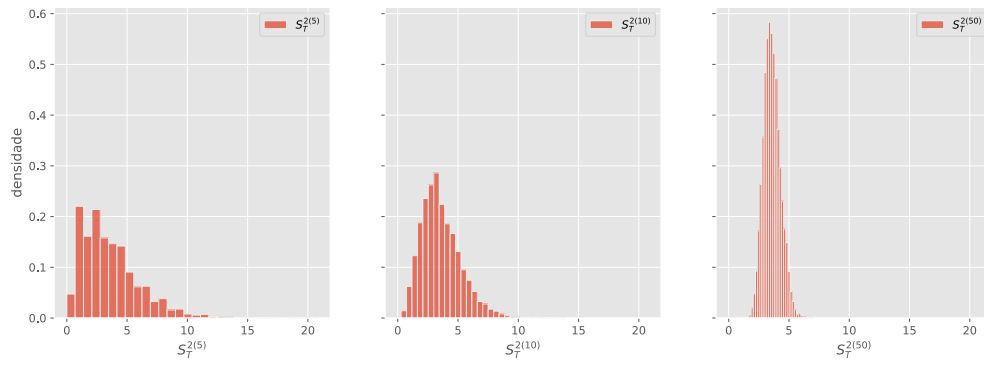


Figura 10: Variância amostral da variável T .

Apesar da falta de garantias teóricas podemos observar uma tendência geral das variáveis a ficar com uma aparência de distribuição normal, mesmo para a variável discreta, o que indica uma ação do teorema central do limite.

Bibliografia

- [1] M. Borges, «GA-030 - Estatística». [Em linha]. Disponível em: <https://lncc.br/~mrborges/>