Teorema Central do Limite - Um Apanhado

Resumo

O Teorema Central do Limite (TCL) é um dos mais importantes dentro da Estatística e desempenha papel de destaque na análise de dados, inferências e elaboração de planos de amostras; na análise de erros em mensurações científicas é sempre invocado como princípio de partida a respeito da distribuição de erros. De modo bastante simplista, este teorema assegura que quaisquer das estatísticas de quaisquer momentos centrais de uma variável aleatória independente e com reposição irão convergir para uma distribuição normal padrão, $N(0,1)$, para um número cada vez maior de medidas da estatística daquele momento central. Apesar de sua importância para a Estatística, o TCL é comumente apenas citado nas literaturas introdutórias sem qualquer demonstração rigorosa; eis aqui, portanto, a proposta de apresentá-lo por completo, fornecendo ao leitor uma demonstração deste teorema de papel tão central. Além do mais, após sua apresentação rigorosa, ser também apresentada uma discussão a respeito de cuidados com o seu uso e os limites de sua aplicação.

1 O TCL em experimentos computacionais

Tome um experimento estatístico bastante simples, onde tomada uma sequência das estatísticas X_i de uma classe de variáveis aleatórias X, portanto todas com a mesma distribuição, é calculada a média – ou a variância, ou outro momento qualquer –, sendo este processo repetido um número n de vezes. Se forem colecionados todos os resultados das médias obtidas com os sucessivos sorteios das amostras – ou as variâncias, ou quaisquer outros momentos –, como seria a distribuição deste conjunto? Esta distribuição dependeria da distribuição inicial da variável aleatória X?

Tal experimento é bastante simples de ser realizado, em especial através de simulações computacionais. O código apresentado abaixo trata do experimento proposto escrito em R, onde é escolhida uma variável aleatória com distribuição uniforme, sendo com ela realizados sorteios de amostras com $n = \{2, 3, 5, 8, 12, 19, 25, 36, 45\}$ elementos, sendo calculada a média para cada tipo de amostra e armazenada num conjunto de 90 resultados que é utilizado para verificar a forma da distribuição das médias para cada escolha de n.

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
```

```
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha
for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  mean( runif( n=n[i,1])) # Calcula a média
         ↪  para o conjunto sorteado e a armazena na matriz,
         ↪  distribuição uniforme
        X$x2[j,i] =  mean( runif( n=n[i,2])) # Calcula a média
         ↪  para o conjunto sorteado e a armazena na matriz,
         ↪  distribuição uniforme
        X$x3[j,i] =  mean( runif( n=n[i,3])) # Calcula a média
         ↪  para o conjunto sorteado e a armazena na matriz,
         ↪  distribuição uniforme
    }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dunif(x),xlim=c(0,1),xlab="x") # Plota a distribuição
 ↪  uniforme
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=2") # Plota o primeiro
 ↪  histograma
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=3") # Plota o segundo
 ↪  histograma
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=5") # Plota o terceiro
 ↪  histograma
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=8") # Plota o quarto
 ↪  histograma
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=12") # Plota o quinto
 ↪  histograma
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=19") # Plota o sexto
 ↪  histograma
```

```
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=25") # Plota o sétimo
 ↪  histograma
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=36") # Plota o oitavo
 ↪  histograma
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=45") # Plota o nono
 ↪  histograma
```

O resultado obtido com o código acima tem a característica apresentada na figura (1); nele pode ser verificado que com o aumento de n os histogramas se agrupam mais no entorno do valor teórico da média para a distribuição – amplitudes menores nas abscissas – e a forma da distribuição se assemelha cada vez mais a um sino.

Um código muito similar ao anterior pode ser empregado para se fazer o mesmo[1], mas agora para as variâncias, sendo o resultado apresentado na figura (2). Nele são observadas as mesmas características de antes, apenas com a diferença de que as distribuições iniciais são nitidamente assimétricas, mas perdendo esta característica com o aumento de n.

Em ambos casos fica evidente que a distribuição associada à estatística dos momentos da distribuição das variáveis aleatórias uniformes não é uma outra distribuição uniforme, mas sim uma distribuição simétrica em forma de sino que assim se apresenta mais claro quanto maior é o número n de elementos no sorteio.

Alguém poderia questionar que o fenômeno observado é caso isolado e está conectado com a distribuição simétrica das variáveis iniciais, i.e. uniforme. Então o mesmo experimento pode ser realizado para outros tipos de variáveis aleatórias, como uma variável com distribuição exponencial, com distribuição χ^2 ou com distribuição β, todas elas parametrizadas de forma não simétrica para verificar se há algum fundamento pensar numa conexão específica. Os resultados para todos estes casos estão apresentados nas figuras (3-8) e os códigos das simulações estão no apêndice deste trabalho para aqueles interessados em reproduzir o experimento.

Em todos os casos apresentados é observado o mesmo fenômeno; mesmo que alguma assimetria se revele nos primeiros histogramas, ela logo é desfeita com o aumento de n e toma uma forma simétrica tipo sino com seu pico se fechando cada vez mais nos valores teóricos dos momentos relacionados nos experimentos.

Algo assim não é mera obra do casuísmo, há algo mais geral ditando este comportamento de convergência nos diferentes casos, que trata justamente do Teorema Central do Limite (TCL). Ao longo da demonstração poderão ser verificadas as condições necessárias para haver a convergência, bem como que a distribuição de convergência é sempre a mesma em todos os casos e corresponde justamente a uma distribuição normal, $N(0, 1)$.

[1] O código correspondente se encontra no apêndice deste trabalho. Uma vez serem os diversos códigos quase idênticos, não cabendo poluir o texto principal com tais repetições quase idênticas.

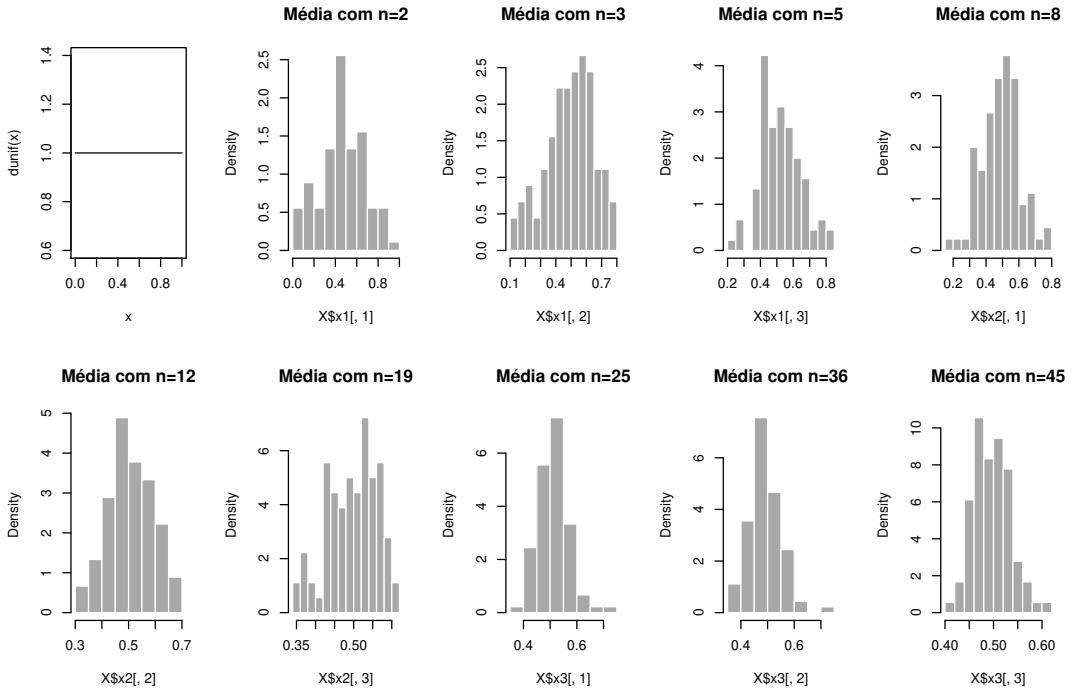

Figura 1: Resultado das médias da simulação computacional para variáveis aleatórias de distribuição uniforme.

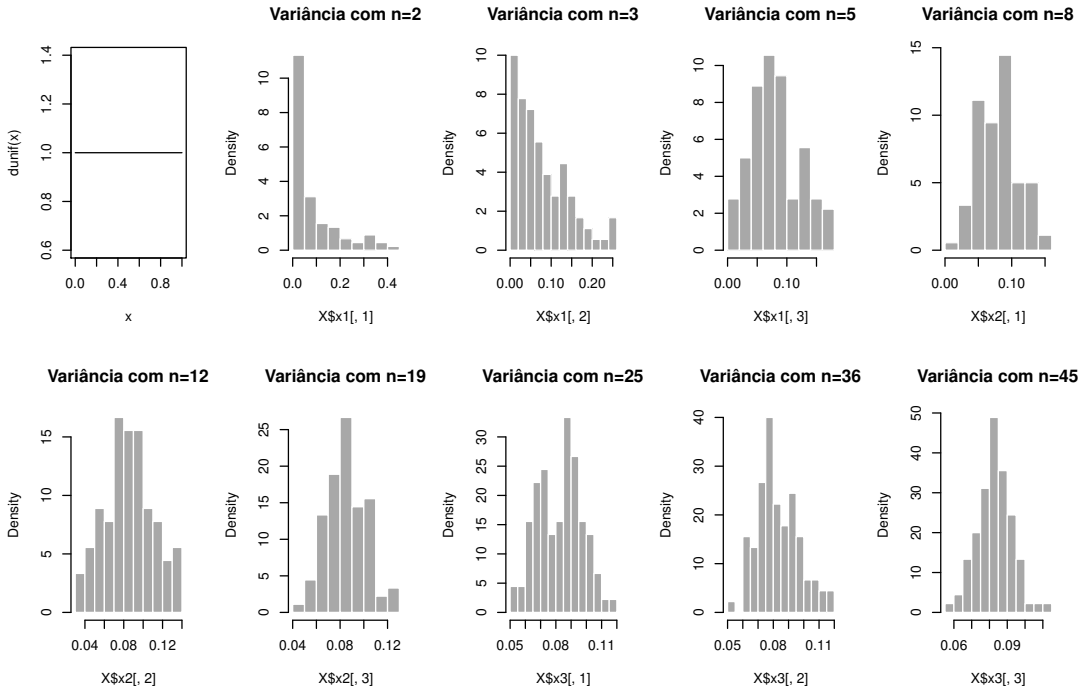

Figura 2: Resultado das variâncias da simulação computacional para variáveis aleatórias de distribuição uniforme.

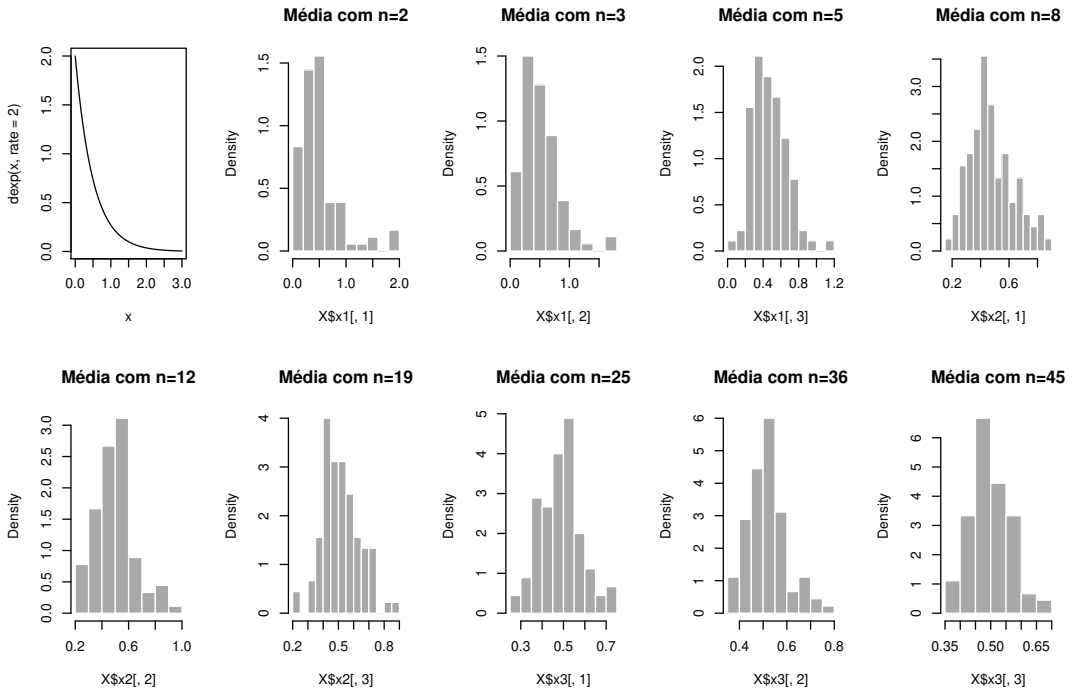

Figura 3: Resultado das distribuições para as médias para uma distribuição exponencial.

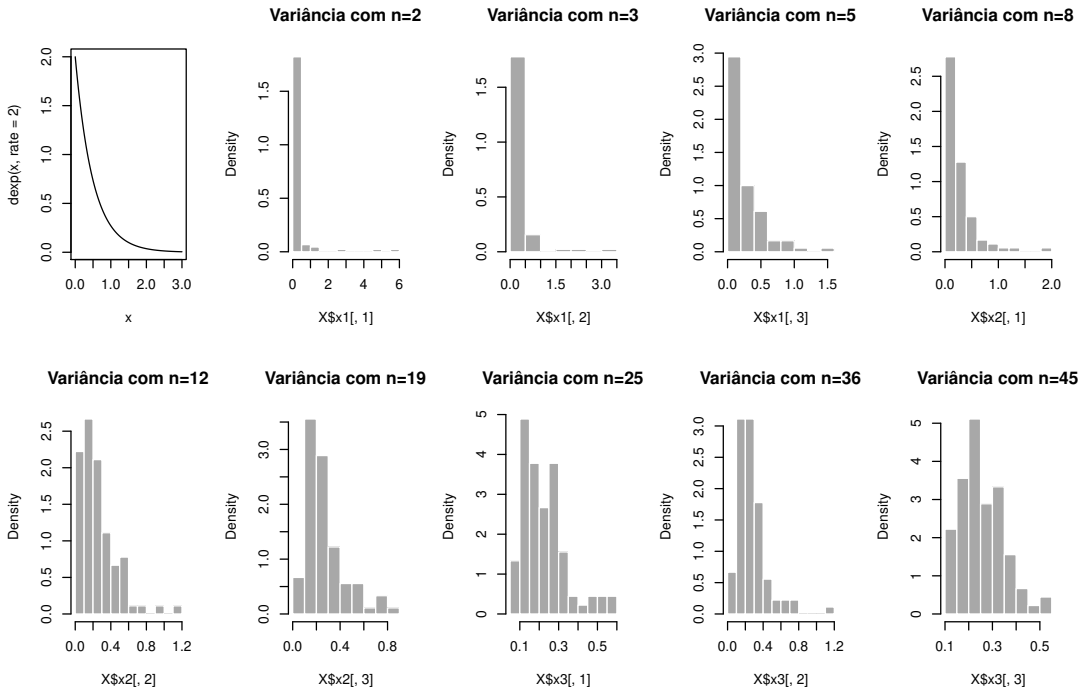

Figura 4: Resultado das distribuições para as variâncias para uma distribuição exponencial.

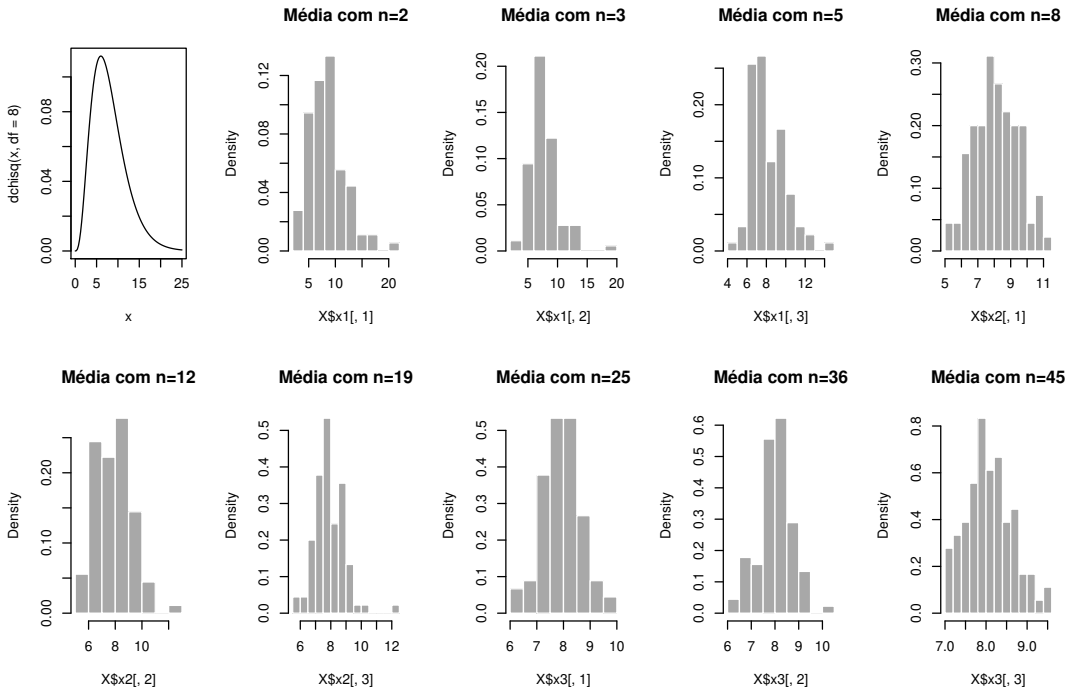

Figura 5: Resultado das distribuições para as médias para uma distribuição χ^2.

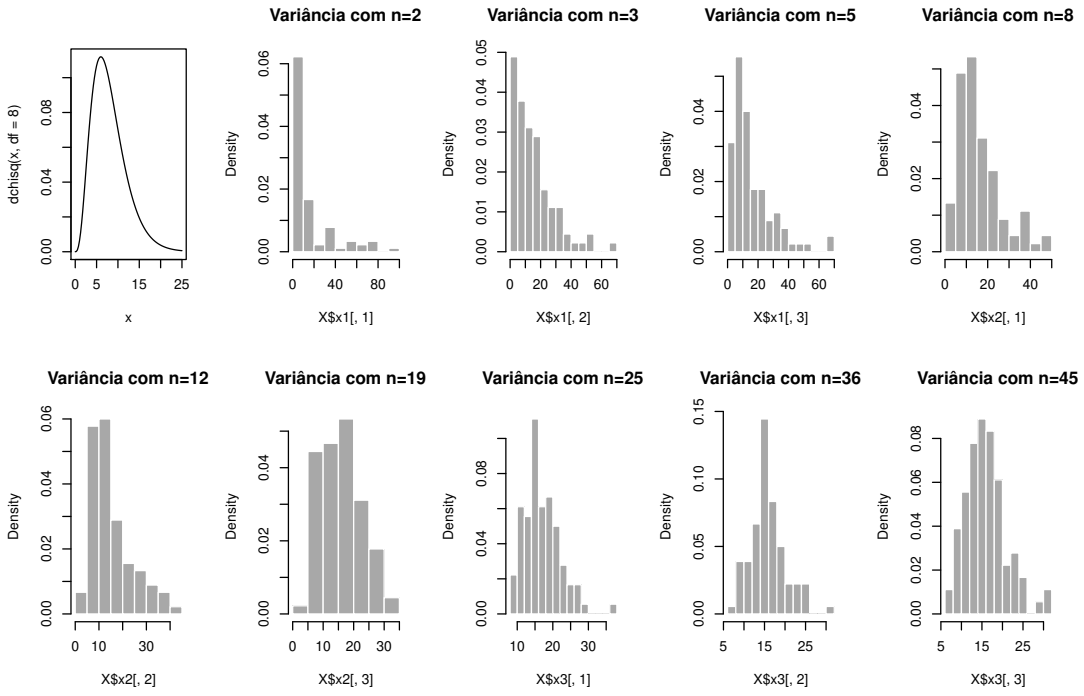

Figura 6: Resultado das distribuições para as variâncias para uma distribuição χ^2.

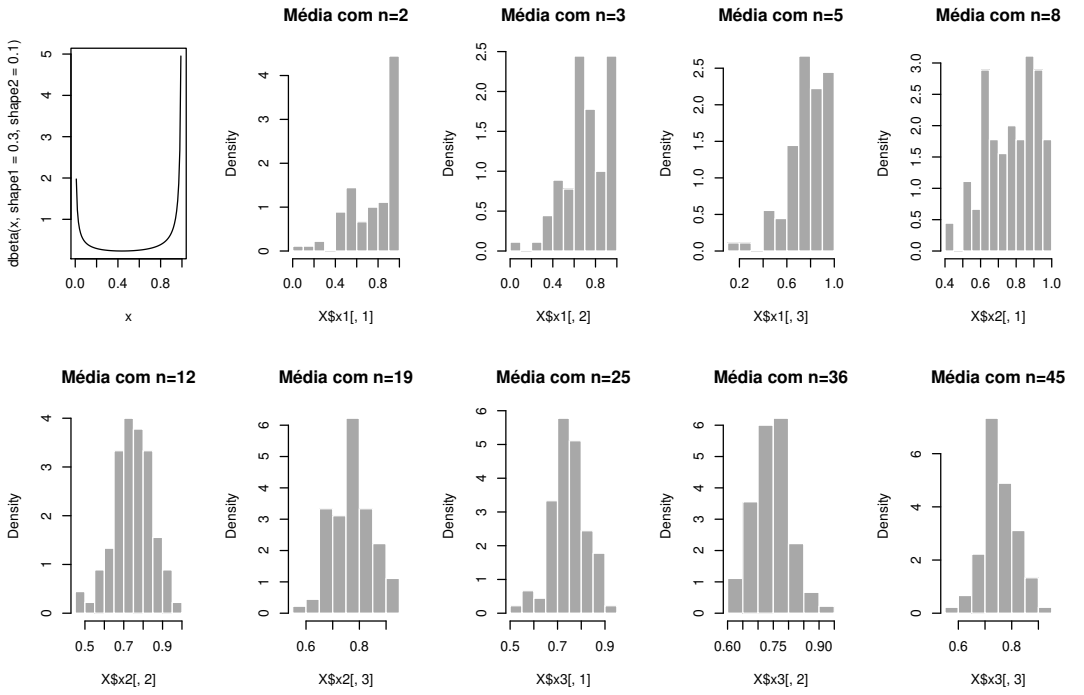

Figura 7: Resultado das distribuições para as médias para uma distribuição β.

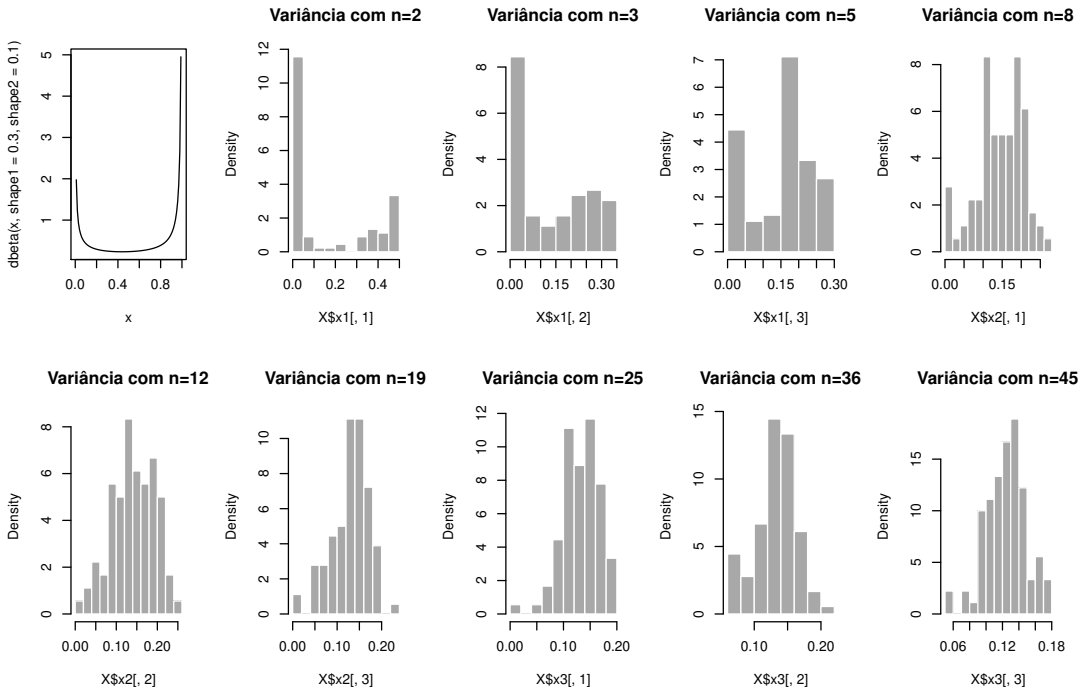

Figura 8: Resultado das distribuições para as variâncias para uma distribuição β.

2 Função característica

Antes de partir para a demonstração do TCL se faz necessário discutir um pouco sobre a função característica de uma distribuição, uma vez que este conceito será bastante utilizado.

Tome uma variável aleatória independente X com distribuição de probabilidade $f(x)$, cuja média é dada por μ e sua variância por σ^2. Tal conjunto de informações pode ser utilizado para definir a variável aleatória padrão $Z = \frac{X-\mu}{\sigma}$, que centra a média de X na origem; assim, por definição, o *m-éssimo* momento central da distribuição é dado por

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^m\right] = \int \left(\frac{x-\mu}{\sigma}\right)^m f(x)dx. \tag{1}$$

Onde a média trata do primeiro momento central e é sempre nula, pois

$$\begin{aligned} E\left[\left(\tfrac{X-\mu}{\sigma}\right)\right] &= \int \left(\tfrac{x-\mu}{\sigma}\right) f(x)dx = \tfrac{1}{\sigma}\int x f(x)dx - \tfrac{\mu}{\sigma}\int f(x)dx \\ &= \tfrac{1}{\sigma}\mu - \tfrac{\mu}{\sigma} = 0. \end{aligned} \tag{2}$$

Bem como a variância central é sempre igual a 1 e trata do segundo momento da distribuição, uma fez que

$$\begin{aligned} E\left[\left(\tfrac{X-\mu}{\sigma}\right)^2\right] &= \int \left(\tfrac{x-\mu}{\sigma}\right)^2 f(x)dx = \tfrac{1}{\sigma^2}\left(\int (x^2 - 2\mu x + \mu^2) f(x)dx\right) \\ &= \tfrac{1}{\sigma^2}\left(\int x^2 f(x)dx - 2\mu\int x f(x)dx + \mu^2\int f(x)dx\right) \\ &= \tfrac{1}{\sigma^2}\left(\int x^2 f(x)dx - \mu^2\right) = \tfrac{1}{\sigma^2}(E[X^2] - E[X]^2) = 1. \end{aligned} \tag{3}$$

Para simplificar a notação, adote a variável aleatória padrão Z e se atenta para o fato da equidade da medida $f(x)dx = f(z)dz = dF$ que implica na relação $f(z) = \sigma f(x)$, cujo primeiro momento é nulo e a variância é 1 independente da distribuição $f(x)$, conforme mostrado. A função característica de Z é definida por

$$E\left[e^{iZt}\right] = F(t) = \int e^{izt} f(z)dz; \tag{4}$$

onde qualquer momento m desejado da distribuição pode ser diretamente calculado pela derivada em ordem m da função $F(t)$ em $t = 0$, i.e.

$$E[Z^m] = \frac{1}{i^m}\left.\frac{d^{(m)}F}{dt^{(m)}}\right|_{t=0}. \tag{5}$$

Não será feita qualquer demonstração sobre a condição de unicidade da relação entre $\Rightarrow$ dada uma distribuição $f(z)$ haver um único conjunto de momentos relacionado a ela e $\Leftarrow$ dado um conjunto de momentos haver apenas uma única distribuição $f(z)$ relacionada a ele. Isto porque tal tarefa escaparia do objetivo principal deste trabalho.

Todavia, perceba um fato muito importante que se extrai das equações (4) e (5): um conjunto de infinitos momentos define uma única distribuição e uma única distribuição define apenas um único conjunto de infinitos momentos, podendo esta identi-

dade ser representada pela unicidade da função característica, a qual liga o conjunto dos momentos à distribuição e vice-versa.

Posto isso, importante se torna saber qual a função característica relacionada com a distribuição normal $N(0,1)$ que corresponde a uma distribuição gaussiana na variável Z anteriormente definida; assim

$$N(t) = E[e^{itZ}] \quad = \frac{1}{\sqrt{2\pi}} \int e^{itz} e^{-\frac{z^2}{2}} dz = e^{-\frac{t^2}{2}}; \tag{6}$$

ou seja, a função característica do tipo normal/gaussiana está relacionada unicamente a uma distribuição do tipo normal/gaussiana. Este último resultado é o mais importante e será bastante utilizado.

3 Caso específico - TCL de DeMoivre e Laplace

Para constar nesta parte, é dita que uma variável aleatória é **independente** quando a realização de uma medida não implica qualquer influência **funcional** nas demais. E uma variável **identicamente distribuída** diz respeito à permanência de sua distribuição, e aqui incluídos os seus parâmetros característicos também, sem qualquer tipo de influência; desse modo, construir uma cadeia de variáveis aleatórias identicamente distribuídas é dizer que a distribuição de probabilidade permanece sempre a mesma em todas as realizações.

Um bom contra-exemplo é o sistema formado por n bolas pretas e m bolas brancas, todos misturados e sem recomposição após uma medida; a toda realização a proporção de bolas pretas e de bolas brancas é alterada e, portanto, os parâmetros da distribuição também, sendo este o mais clássico exemplo de variáveis não identicamente distribuídas, mesmo todas elas possuindo a mesma classe de distribuição.

Para começar a apresentação da demonstração do teorema, e até para seguir sua história no caminho da generalização maior possível, será nesta seção apresentada a demonstração do Teorema Central do Limite **para uma variável aleatória independente e identicamente distribuída** X **que possui distribuição binomial de probabilidade** com parâmetros n e p, onde n representa o número de elementos da população e p a probabilidade de ser obtido um único sucesso numa única medida.

Dado o sistema acima, considere a sequência de m sorteios aleatórios de X, formando a variável aleatória $S_m = X_1 + X_2 + \ldots + X_m$, ocorrendo sempre a recomposição da população após a realização de uma observação[2]. Da distribuição binomial são conhecidos os resultados $\mu_m = mp$ para a média de m elementos obtidos num sorteio e, também, $\sigma_m^2 = mp(1-p)$ para a variância do mesmo conjunto. Logo a variável padrão para cada grupo sorteado é assim escrita:

$$Z_m = \frac{S_m - mp}{\sqrt{mp(1-p)}}; \tag{7}$$

[2]Note que com a condição de recomposição da população após a realização de um sorteio a condição $m \leq n$ não é necessária, pois por construção podem ser reunidas variáveis S_m com um número m qualquer em razão de sempre haver uma população de n elementos para ser um deles sorteado. A bem da verdade, bastaria uma única e exclusiva variável de Bermoulli para se montar o sistema aqui tratado. Portanto, a construção apresentada é meramente para melhor didática e aproximação a um processo mais familiar.

cuja função característica assim resulta:

$$E[e^{itZ}] = E\left[exp\left(\frac{S_m - mp}{\sqrt{mp(1-p)}}\right)it\right] = exp\left(-\frac{itmp}{\sqrt{mp(1-p)}}\right)E\left[exp\left(\frac{itS_m}{\sqrt{mp(1-p)}}\right)\right]. \quad (8)$$

No entanto, para distribuições discretas de probabilidades, como a distribuição binomial, é do conhecimento que

$$E[X] = \sum_{j=0}^{m} x_j f(x_j); \quad (9)$$

logo a equação (8) pode ser colocada de forma mais direta, sendo

$$\begin{aligned} E[e^{itZ}] &= exp\left(-\tfrac{itmp}{\sqrt{mp(1-p)}}\right)\textstyle\sum_{x=0}^{m} exp\left(\tfrac{itx}{\sqrt{mp(1-p)}}\right)\binom{m}{x}p^x(1-p)^{m-x} \\ &= exp\left(-\tfrac{itmp}{\sqrt{mp(1-p)}}\right)\textstyle\sum_{x=0}^{m}\binom{m}{x}\left(pe^{\left(\frac{it}{\sqrt{mp(1-p)}}\right)}\right)^x(1-p)^{m-x} \\ &= exp\left(-\tfrac{itmp}{\sqrt{mp(1-p)}}\right)\left(1-p+pe^{\left(\frac{it}{\sqrt{mp(1-p)}}\right)}\right)^m \\ &= \left((1-p)e^{\left(-\frac{itp}{\sqrt{mp(1-p)}}\right)} + pe^{\left(\frac{it(1-p)}{\sqrt{mp(1-p)}}\right)}\right)^m. \end{aligned} \quad (10)$$

Mas se na última equação (10) forem as exponenciais expandidas em séries de Taylor e tais séries somadas com os fatores multiplicativos existentes na expressão, i.e. $(1-p)$ e p, chegar-se-á ao simples resultado

$$E[e^{itZ}] = \left(1 - \frac{t^2}{2m} + \ldots\right)^m = \left(1 + \frac{-\frac{t^2}{2}}{m} + \ldots\right)^m; \quad (11)$$

onde o limite para m muito grande pode ser tomado e que leva ao esperado[3]

$$\lim_{m\to\infty} E[e^{itZ}] = \lim_{m\to\infty}\left(1 + \frac{-\frac{t^2}{2}}{m} + \ldots\right)^m = e^{-\frac{t^2}{2}}; \quad (12)$$

desse modo ficando demonstrado que a função característica na presente situação converge para uma função característica normal, ou em outras palavras, a estatística obtida para quaisquer momentos centrais convergirá para uma distribuição do tipo normal uma vez que apenas esta distribuição detém como função característica uma função gaussiana.

Tal demonstração é bastante limitada, pois nela foram consideradas variáveis aleatórias independentes e identicamente distribuídas cuja distribuição inicial é a bino-

[3] Que por construção do problema sempre pode ser feito uma vez estar ocorrendo reposição da população original de n elementos, disponibilizando assim a formação de conjuntos de dimensões quaisquer para a realização S_m.

mial. Na próxima seção será enfraquecida a condição sobre uma específica distribuição, mas serão ainda mantidas as condições das variáveis serem independentes e identicamente distribuídas.

De todo modo, já é visto que o fenômeno de convergência ilustrado inicialmente não trata de mera casualidade, mas de um fenômeno de convergência para grandes números.

4 TCL para quaisquer distribuições de variáveis independentes e identicamente distribuídas

Para a demonstração do TCL em variáveis aleatórias independentes e identicamente distribuídas sem qualquer conhecimento prévio da distribuição associada, considere um conjunto de n elementos de uma amostra de variáveis aleatórias X_i com média igual a μ e variância igual a σ^2, sendo esta amostra representada pela variável aleatória $S_n = X_1 + X_2 + \ldots + X_n$, onde, por construção, se sabe que a média é $n\mu$ e a variância $n\sigma^2$ em razão da condição de variáveis identicamente distribuídas.

Nestas condições é defina a variável padrão Z e se busca mostrar que a sua função característica converge para uma distribuição normal, em outras palavras,

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1). \tag{13}$$

Tomando a função característica para a variável padrão Z definida acima e fazendo pequena manipulação algébrica, se obtém

$$\begin{aligned} \Psi_{Z_n}(t) = E[e^{itZ_n}] \quad &= E[e^{it\frac{S_n - n\mu}{\sigma\sqrt{n}}}] = E[e^{i\frac{t}{\sqrt{n}}\frac{X_1-\mu}{\sigma}} e^{i\frac{t}{\sqrt{n}}\frac{X_2-\mu}{\sigma}} \ldots e^{i\frac{t}{\sqrt{n}}\frac{X_n-\mu}{\sigma}}] \\ &= E[e^{i\frac{t}{\sqrt{n}}\frac{X_1-\mu}{\sigma}}]E[e^{i\frac{t}{\sqrt{n}}\frac{X_2-\mu}{\sigma}}]\ldots E[e^{i\frac{t}{\sqrt{n}}\frac{X_n-\mu}{\sigma}}] = E[e^{i\frac{t}{\sqrt{n}}\frac{X-\mu}{\sigma}}]^n; \end{aligned} \tag{14}$$

isto porque todas as variáveis X_i são independentes e identicamente distribuídas.

Agora, olhando somente a a expressão $\psi(\frac{t}{\sqrt{n}}) = E[e^{i\frac{t}{\sqrt{n}}\frac{X-\mu}{\sigma}}]$, que também por construção é sabido possuir primeira e segunda derivadas em razão da existência de média e de variância para a variável X; é possível escrevê-la como uma expansão de Taylor até primeira ordem no entorno de $\frac{t}{\sqrt{n}} = 0$ e mais um resto de Lagrange com $0 < \alpha(\frac{t}{\sqrt{n}}) \leq \frac{t}{\sqrt{n}}$, do seguinte modo

$$\psi\left(\frac{t}{\sqrt{n}}\right) = \psi(0) + \psi'(0)\frac{t}{\sqrt{n}} + \frac{1}{2}\psi''\left(\alpha\left(\frac{t}{\sqrt{n}}\right)\right)\frac{t^2}{n}; \tag{15}$$

que pode ser um pouco mais trabalhada com a soma e subtração dos termos $\frac{t^2}{2n}\psi''(0)$, resultando em

$$\begin{aligned} \psi\left(\tfrac{t}{\sqrt{n}}\right) \quad &= \psi(0) + \psi'(0)\tfrac{t}{\sqrt{n}} + \tfrac{t^2}{2n}\psi''(0) + \tfrac{t^2}{2n}[\psi''\left(\alpha\left(\tfrac{t}{\sqrt{n}}\right)\right) - \psi''(0)] \\ &= \psi(0) + \psi'(0)\tfrac{t}{\sqrt{n}} + \tfrac{t^2}{2n}\psi''(0) + \tfrac{t^2}{2n}e\left(\tfrac{t}{\sqrt{n}}\right). \end{aligned} \tag{16}$$

Todavia, por ser conhecida a variável aleatória em ψ como aquela padrão construída pelos parâmetros da variável X, é sabido que $\psi(0) = 1$; $\psi'(0) = iE[\frac{X-\mu}{\sigma}] = 0$, do mesmo modo que $\psi''(0) = i^2E[\left(\frac{X-\mu}{\sigma}\right)^2] = -1$. Substituindo estes resultados na expressão (16), chega-se ao resultado:

$$\psi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n}\left[1 - e\left(\frac{t}{\sqrt{n}}\right)\right]. \tag{17}$$

Com ele, pode se concluir que a expressão (14) tem a forma mais simples e escrita como

$$\Psi_{Z_n}(t) = E[e^{itZ_n}] = \left[1 - \frac{t^2}{2n}\left[1 - e\left(\frac{t}{\sqrt{n}}\right)\right]\right]^n. \tag{18}$$

A questão final para a demonstração do teorema é calcular agora qual o limite para n tendendo ao infinito na equação (18). Em primeiro lugar deve ser notado que a função $e(\frac{t}{\sqrt{n}}) = \psi''\left(\alpha\left(\frac{t}{\sqrt{n}}\right)\right) - \psi''(0)$ tende ao limite $\to 0$ quando n tende ao infinito, uma vez que, neste limite $\lim_{n\to\infty}\psi''\left(\alpha\left(\frac{t}{\sqrt{n}}\right)\right) = \psi''(0)$. Desse modo, resta apenas o último limite a ser verificado, o qual trata daquele

$$\lim_{n\to\infty}\Psi_{Z_n}(t) = \lim_{n\to\infty}\left[1 + \frac{-\frac{t^2}{2}}{n}\right]^n = e^{-\frac{t^2}{2}} \tag{19}$$

ou seja, no limite de $n \to \infty$, a função característica $\Psi_{Z_n}(t)$ tende a mesma função característica da distribuição gaussiana/normal; o que conclui a demonstração do teorema, qualquer que seja a distribuição associada à variável aleatória inicial X.

Outra vez fica claro que o processo de serem eleitas amostras de um conjunto de variáveis aleatórias populacionais independentes e identicamente distribuídas e sobre elas ser produzida uma estatística; gera tal estatística um espaço métrico caracterizado por uma medida do tipo gaussiana/normal, sendo ela tão clara quanto maiores as estatísticas produzidas e independente da medida inicial característica da variável populacional.

Na próxima seção será apresentada a demonstração mais geral até o presente do TCL, onde a condição de variável identicamente distribuída é derrubada.

5 Teorema Central do Limite de Lindeberg

A demonstração de Lindeberg do TCL é até o presente a mais geral e forte, sendo válida a convergência para uma amostra de variáveis aleatórias X_i, com média igual a $\mu_i = E[X_i]$ e variância $\sigma_i^2 = E[X_i^2] - E[X_i]^2$, não sendo requerida a condição de variáveis identicamente distribuídas, i.e. podendo serem as distribuições diferentes entre elas.

As únicas condições necessárias são que as variáveis aleatórias sejam independentes e que a condição de Lindeberg seja satisfeita para qualquer ϵ positivo, i.e.

$$\lim_{n\to\infty}\left[\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 f_k(x)dx\right] = 0, \quad \forall\epsilon > 0; \tag{20}$$

onde $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$, tratando da soma de todas as variâncias.

Dentro do regime, para $n \to \infty$, no qual a participação individual de cada σ_k^2 na soma s_n^2 seja ínfima; mesmo para o maior valor dele, uma vez serem as distribuições diferentes e cada dar peso específico à variância. Ou em linguagem matemática

$$\max_{1 \leq K \leq n} \frac{\sigma_k^2}{s_n^2} \xrightarrow{n \to \infty} 0. \tag{21}$$

Se tais condições são satisfeitas e $S_n = \sum_{i=1}^{n} X_i$, então é válida a convergência seguinte

$$\frac{S_n - E[S_n]}{s_n} \xrightarrow{D} N(0,1); \tag{22}$$

ou de modo mais preciso

$$E[exp\left(it\frac{S_n - E[S_n]}{s_n}\right)] \xrightarrow{n \to \infty} e^{-\frac{t^2}{2}}. \tag{23}$$

E, pelas definições de S_n e a independência das variáveis X_i, a equação (23) de partida para a demonstração pode ser reescrita de forma mais explícita como

$$\prod_{k=1}^{n} E[exp\left(it\frac{X_k - \mu_k}{s_n}\right)] \xrightarrow{n \to \infty} e^{-\frac{t^2}{2}}; \tag{24}$$

onde pode ser notado que, por não haver mais a condição de distribuições idênticas para as variáveis, a última expressão se escreve como um produto de cada elemento e não mais uma potência n de termos idênticos.

Para a demonstração da convergência (24), fixe inicialmente um único termo k, cuja expressão é do tipo $\exp(itx)$ e tem a seguinte expansão de Taylor:

$$e^{itx} = \sum_{n=0}^{\infty} \frac{(itx)^n}{n!}; \tag{25}$$

podendo ela ser escrita tanto em segunda ordem quanto em terceira de modo equivalente como

$$\begin{aligned} e^{itx} &= 1 + itx + \theta_1(x)\tfrac{(tx)^2}{2}, && |\theta_1(x)| < 1; \\ e^{itx} &= 1 + itx - \tfrac{(tx)^2}{2} + \theta_2(x)\tfrac{(tx)^3}{6}, && |\theta_2(x)| < 1; \end{aligned} \tag{26}$$

ou de modo mais compacto com a expressão

$$e^{itx} = 1 + itx - \frac{(tx)^2}{2} + err(x). \tag{27}$$

Onde foi introduzida a função erro $err(x)$ definida por duas sentenças a depender do módulo de x da seguinte forma

$$err(x) = \begin{cases} (1+\theta_1(x))\frac{(tx)^2}{2}, & \text{com } |x| > \epsilon; \\ \\ \theta_2(x)\frac{(tx)^3}{6}, & \text{com } |x| \leq \epsilon. \end{cases} \tag{28}$$

Tomando apenas um único elemento do produto em (24) e aplicando a expressão (27) se chega exatamente ao seguinte resultado

$$\begin{aligned} E[\exp(it\tfrac{X_k-\mu_k}{s_n})] &= \int f_k(x)dx + \int it\tfrac{x-\mu_k}{s_n} f_k(x)dx - \int \tfrac{t^2}{2}\left(\tfrac{x-\mu_k}{s_n}\right)^2 f_k(x)dx \\ &\quad + \int err\left(\tfrac{x-\mu_k}{s_n}\right) f_k(x)dx; \end{aligned} \tag{29}$$

onde

$$err\left(\frac{x-\mu_k}{s_n}\right) = \left[1+\theta_1\left(\frac{x-\mu_k}{s_n}\right)\right]\frac{t^2}{2}\left(\frac{x-\mu_k}{s_n}\right)^2 + \theta_2\left(\frac{x-\mu_k}{s_n}\right)\frac{t^3}{6}\left(\frac{x-\mu_k}{s_n}\right)^3, \tag{30}$$

valendo a primeira expressão quando $|x-\mu_k| > \epsilon s_n$ e a segunda quando $|x-\mu_k| \leq \epsilon s_n$.

Agora a equação (29) é reescrita com a introdução explícita da função erro, o que resulta

$$\begin{aligned} E[\exp(it\tfrac{X_k-\mu_k}{s_n})] &= 1 + itE\left[\tfrac{X_k-\mu_k}{s_n}\right] - \tfrac{t^2}{2}E\left[\left(\tfrac{X_k-\mu_k}{s_n}\right)^2\right] \\ &\quad + \tfrac{t^2}{2}\int_{|x-\mu_k|>\epsilon s_n}\left[1+\theta_1\left(\tfrac{x-\mu_k}{s_n}\right)\right]\left(\tfrac{x-\mu_k}{s_n}\right)^2 f_k(x)dx \\ &\quad + \tfrac{t^3}{6}\int_{|x-\mu_k|\leq\epsilon s_n}\theta_2\left(\tfrac{x-\mu_k}{s_n}\right)\left(\tfrac{x-\mu_k}{s_n}\right)^3 f_k(x)dx; \end{aligned} \tag{31}$$

sendo ainda possível simplificá-la lembrando que $E[X_k] = \mu_k$ e que $E[(X_k-\mu_k)^2] = \sigma_k$, resultando então

$$\begin{aligned} E[\exp(it\tfrac{X_k-\mu_k}{s_n})] &= 1 - \frac{t^2\sigma_k^2}{2s_n^2} \\ &\quad + \frac{t^2}{2s_n^2}\int_{|x-\mu_k|>\epsilon s_n}\left[1+\theta_1\left(\tfrac{x-\mu_k}{s_n}\right)\right](x-\mu_k)^2 f_k(x)dx \\ &\quad + \frac{t^3}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}\theta_2\left(\tfrac{x-\mu_k}{s_n}\right)\left(\tfrac{x-\mu_k}{s_n}\right)(x-\mu_k)^2 f_k(x)dx \\ &= 1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}. \end{aligned} \tag{32}$$

A expressão $e_{n,k}$ deve ser analisada em maior detalhe e, para isso, note que a condição $|\theta_1(x)| < 1$ em (26) permite concluir ser verdadeira a desigualdade $0 < [1+\theta_1(x)] < 2$, com um supremo que não excede a 2. Outro fato extraído de (26) e de (28) é que $|X_k-\mu_k| \leq \epsilon s_n$ implica diretamente que $\left|\frac{X_k-\mu_k}{s_n}\right| \leq \epsilon$; bem como $|\theta_2(x)| < 1$ implica na validade da desigualdade $-\epsilon < \epsilon\theta_2(x) < \epsilon$ para $\epsilon > 0$. Trazendo estas

observações para a análise da expressão $|e_{n,k}|$, se verifica

$$\begin{aligned}
|e_{n,k}| = \quad & \frac{t^2}{2s_n^2}\int_{|x-\mu_k|>\epsilon s_n}\left|1+\theta_1\left(\frac{x-\mu_k}{s_n}\right)\right|(x-\mu_k)^2 f_k(x)dx \\
& + \frac{|t^3|}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}\left|\theta_2\left(\frac{x-\mu_k}{s_n}\right)\right|\left|\frac{x-\mu_k}{s_n}\right|(x-\mu_k)^2 f_k(x)dx \\
& \leq \frac{t^2}{2s_n^2}\int_{|x-\mu_k|>\epsilon s_n}\mathbf{2}(x-\mu_k)^2 f_k(x)dx + \frac{|t^3|}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}\epsilon(x-\mu_k)^2 f_k(x)dx \\
& = \frac{t^2}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 f_k(x)dx + \frac{\epsilon|t^3|}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}(x-\mu_k)^2 f_k(x)dx
\end{aligned} \tag{33}$$

Agora, se somadas todas as contribuições de k para o módulo da expressão do erro, o resultado para a desigualdade será aquele

$$\begin{aligned}
\sum_{k=1}^{n}|e_{n,k}| \leq \quad & \underbrace{t^2\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 f_k(x)dx}_{\text{termo Lindeberg}} \\
& + \underbrace{\frac{\epsilon|t^3|}{6}\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|\leq\epsilon s_n}(x-\mu_k)^2 f_k(x)dx}_{\text{termo de soma das variâncias}}.
\end{aligned} \tag{34}$$

E como apresentado na equação anterior, a primeira parte trata justamente do termo de Lindeberg apresentado em (20) que, **se satisfeito**, terá seu limite igual a zero para $n \to \infty$. Já o segundo termo da equação convergirá sempre para o intervalo $[0,1]$ a depender da escolha de ϵ, sendo portanto limitado, para qualquer n **e se** $\epsilon \to 0$, i.e. toda a parte convergirá para zero em razão do termo multiplicativo depender explicitamente de ϵ. Portanto

$$\sum_{k=1}^{n}|e_{n,k}| \xrightarrow{n\to\infty} 0. \tag{35}$$

Uma vez apresentado este fato, tome o resultado (32) em (24) que resulta em

$$\prod_{k=1}^{n}E[exp\left(it\frac{X_k-\mu_k}{s_n}\right)] = \prod_{k=1}^{n}\left(1-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}\right); \tag{36}$$

restando apenas mostrar que para $n \to \infty$ a equação (36) tende para o resultado $e^{-\frac{t^2}{2}}$ e, assim, verificada a convergência da distribuição da estatística para grandes números para uma distribuição gaussiana/normal, o que encerra a demonstração.

No entanto a passagem de (36) para o resultado (24) não é tão trivial, requerendo uma demonstração específica. Então, considere neste momento o seguinte:

Lema 5.1 *Considere $c_{n,k}$ uma sequência de números complexos cuja série $\sum_{k=1}^{n}c_{n,k}$ converge para um número complexo c limitado quando n tende ao infinito. Bem como, é ainda satisfeita a condição $\max_{1\leq k\leq n}|c_{n,k}| \to 0$ quando $n \to \infty$. Nestas condições*

se tem

$$\prod_{k=1}^{n}(1+c_{n,k}) \xrightarrow{n\to\infty} e^{c}. \tag{37}$$

Prova 5.1.1 *Como se está no limite $n\to\infty$ e a condição $\max_{1\leq k\leq n}|c_{n,k}|\to 0$ é assegurada, a equação (37) pode ser reescrita pela continuidade analítica da função exponencial como*

$$\lim_{n\to\infty}\prod_{k=1}^{n}(1+c_{n,k}) = \lim_{n\to\infty}\prod_{k=1}^{n}e^{c_{n,k}} = \lim_{n\to\infty}e^{\sum_{k=1}^{n}c_{n,k}} = e^{c}; \tag{38}$$

uma vez que a somatória no presente limite converge para o número limitado c. Encerrando, assim, a demonstração do lema (5.1).

Retornando para a equação (36), os números $c_{n,k}$ do lema (5.1) tratam na situação de $c_{n,k} = -\frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}$ e, assim sendo,

$$\begin{aligned}\sum_{k=1}^{n}c_{n,k} &= \sum_{k=1}^{n}\left(-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}\right) = \sum_{k=1}^{n}-\frac{t^2\sigma_k^2}{2s_n^2} + \sum_{k=1}^{n}e_{n,k}\\ &= -\frac{t^2}{2}\sum_{k=1}^{n}\frac{\sigma_k^2}{s_n^2} + \sum_{k=1}^{n}e_{n,k} \xrightarrow{n\to\infty} -\frac{t^2}{2};\end{aligned} \tag{39}$$

isto porque (35) já mostra o limite do segundo termo e o primeiro termo é por construção de s_n^2 que resulta, assim, no valor 1.

É necessário mostrar ainda que $\max_{1\leq k\leq n}|c_{n,k}|\to 0$ na situação concreta para se justificar o resultado do lema (5.1). Sendo assim,

$$\begin{aligned}\max_{1\leq k\leq n}|c_{n,k}| &= \max_{1\leq k\leq n}\left|-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}\right|\\ &\leq \max_{1\leq k\leq n}\frac{t^2\sigma_k^2}{2s_n^2} + \max_{1\leq k\leq n}|e_{n,k}|\\ &= \frac{t^2}{2}\underbrace{\max_{1\leq k\leq n}\frac{\sigma_k^2}{s_n^2}}_{\text{Lindeberg}} + \max_{1\leq k\leq n}|e_{n,k}| \xrightarrow{n\to\infty} 0;\end{aligned} \tag{40}$$

uma vez que a condição de Lindeberg assegura a convergência para zero do primeiro termo, equação (21), e o segundo termo já foi mostrado em (35) que converge a zero. Logo, o resultado do lema (5.1) é aplicado na presente situação e conclui-se que

$$\lim_{n\to\infty}\prod_{k=1}^{n}\left(1-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}\right) = e^{-\frac{t^2}{2}}, \tag{41}$$

o que encerra a demonstração do Teorema Central do Limite na versão de Lindeberg e assim mostrando que para grandes números uma estatística de variáveis aleatórias independentes mas não identicamente distribuídas que satisfaçam a condição de Lindeberg converge para uma distribuição do tipo normal/gaussiana.

Tal demonstração é muito importante em Teoria de Erros, uma vez que a totalidade de erros acumulada sobre uma medida tem inúmeras origens distintas e, portanto, distribuições diferentes. O TCL na versão de Lindeberg nesta situação diz que, mesmo assim, a distribuição do erro associada à medida segue uma distribuição normal/gaussiana.

Para fechar esta seção é necessário apresentar a demonstração da condição de Lindeberg em (20) e em seu regime em (21), que apenas foram colocados e utilizados na demonstração do TCL.

5.1 Demonstração da condição de Lindeberg

Tome a própria definição de s_n^2, i.e.

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2; \tag{42}$$

onde é assumida a diferença entre σ_j^2 em razão da possibilidade de diferentes distribuições associadas às variáveis aleatórias.

Sendo assim, como estratégia de uniformizar as diferentes distribuições, para cada σ_j^2 tome **uma distribuição uniforme com média na origem e parâmetros** $-k$ **e** k tal que $\sigma_{j_k}^2 = \frac{k^2}{3}$, ou mais explícito:

$$s_n^2 = \sum_{j=1}^{n} \frac{k_j^2}{3}, \text{ para } k_j \in \mathbb{R}_*^+. \tag{43}$$

É clara a existência de n elementos no conjunto de $k_j \in \mathbb{R}_*^+$, podendo eles serem repetidos ou não, além de ordenáveis no corpo dos números naturais; e sendo verdade no limite de $n \to \infty$ que

$$s_n^2 \leq \sum_{k=1}^{n} \frac{k^2}{3}, \text{ para } k \in \mathbb{N}. \tag{44}$$

Desse modo, dividindo a equação (44) por n^3, ela passa a ser

$$\frac{s_n^2}{n^3} \leq \frac{1}{n^{2+1}} \sum_{k=1}^{n} \frac{k^2}{3} = \frac{1}{3}\left(\frac{1}{n^{2+1}} \sum_{k=1}^{n} k^2 \right), \tag{45}$$

cujo limite para $n \to \infty$ da expressão entre parênteses deve ser estudado.

Lema 5.2 *Para* $\lambda \in \mathbb{R}_*^+$ *é válido que*

$$\lim_{n\to\infty} \frac{1}{n^{\lambda+1}} \sum_{k=1}^{n} k^\lambda = \frac{1}{\lambda+1}. \tag{46}$$

Prova 5.2.1 *Dados* k^λ *e* x^λ *com as seguintes condições:*

- $x^\lambda \leq k^\lambda$, *onde* $k-1 \leq x \leq k$ *e, ainda,*

$$\int_{k-1}^{k} x^\lambda dx \leq \int_{k-1}^{k} k^\lambda dx = k^\lambda;$$

- $x^\lambda \geq k^\lambda$, *onde* $k \leq x \leq k+1$ *e, ainda,*

$$\int_{k}^{k+1} x^\lambda dx \geq \int_{k}^{k+1} k^\lambda dx = k^\lambda;$$

que permitem escrever a desigualdade

$$\int_{k-1}^{k} x^\lambda dx \leq k^\lambda \leq \int_{k}^{k+1} x^\lambda dx. \tag{47}$$

Empreendendo agora a somatória dos elementos de (47) com k variando de 1 até n, a desigualdade resultante será

$$\int_{0}^{n} x^\lambda dx \leq \sum_{k=1}^{n} k^\lambda \leq \int_{1}^{n+1} x^\lambda dx, \tag{48}$$

que, por fim, resultará na desigualdade

$$\frac{(n+1)^{\lambda+1}-1}{\lambda+1} \leq \sum_{k=1}^{n} k^\lambda \leq \frac{(n+1)^{\lambda+1}}{\lambda+1}. \tag{49}$$

Observe na primeira metade da desigualdade (49) que $[(n+1)^{\lambda+1}-1] \leq n^{\lambda+1}$*; então, de pose deste fato e dividindo toda a desigualdade por* $n^{\lambda+1}$*, chega-se ao intervalo de interesse*

$$\frac{1}{\lambda+1} \leq \frac{1}{n^{\lambda+1}} \sum_{k=1}^{n} k^\lambda \leq \left(\frac{n+1}{n}\right)^{\lambda+1} \frac{1}{\lambda+1}. \tag{50}$$

Nesta última desigualdade já se observa claramente o limite inferior, cabendo apenas ser analisado o limite superior, o qual resulta

$$\lim_{n\to\infty}\left[\left(\frac{n+1}{n}\right)^{\lambda+1} \frac{1}{\lambda+1}\right] = \frac{1}{\lambda+1}. \tag{51}$$

Como o limite superior converge para o limite inferior, então fica demonstrado o lema (5.2).

Com o resultado apresentado pelo lema (5.2), o limite para a equação (45) logo fica determinado

$$\lim_{n\to\infty} \frac{s_n^2}{n^3} \leq \lim_{n\to\infty} \frac{1}{3}\left(\frac{1}{n^{2+1}} \sum_{k=1}^{n} k^2\right) = \frac{1}{9}. \tag{52}$$

E este último resultado é importante para verificar que

$$\lim_{n\to\infty}\frac{s_n^2}{n^2}=\lim_{n\to\infty}\frac{s_n^2}{n^3}n=\underbrace{\lim_{n\to\infty}\frac{s_n^2}{n^3}}_{limitado}\lim_{n\to\infty}n=\infty, \tag{53}$$

ou de modo equivalente

$$\lim_{n\to\infty}\frac{n^2}{s_n^2}=0. \tag{54}$$

Com o resultado deste último limite e diante da construção elaborada em (44), é claramente observado que o elemento de maior variância na soma s_n^2 é da ordem n^2. Sendo assim, fica claro que

$$\lim_{n\to\infty}\max_{1\leq K\leq n}\frac{\sigma_k^2}{s_n^2}=\lim_{n\to\infty}\frac{n^2}{3s_n^2}=0; \tag{55}$$

demonstrando assim o regime (21) enunciado na condição de Lindeberg.

Pois bem, considere agora as seguintes igualdades e desigualdades

$$\begin{aligned}
\frac{\sigma_k^2}{s_n^2} &= \frac{1}{s_n^2}\int(x-\mu_k)^2 f_k(x)dx \\
&= \frac{1}{s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}(x-\mu_k)^2 f_k(x)dx+\frac{1}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 f_k(x)dx \\
&\leq \frac{1}{s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}(\epsilon s_n)^2 f_k(x)dx+\frac{1}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 f_k(x)dx \\
&\leq \epsilon^2+\frac{1}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 f_k(x)dx \\
&\leq \epsilon^2+\sum_{j=1}^{n}\frac{1}{s_n^2}\int_{|x-\mu_j|>\epsilon s_n}(x-\mu_j)^2 f_j(x)dx;
\end{aligned} \tag{56}$$

como a conclusão (55) diz que até para o maior σ_k^2 a razão tende a zero para $n\to\infty$ e ϵ pode ser aproximado de zero o quanto queira, então de (56) é visto diretamente que

$$\lim_{n\to\infty}\sum_{j=1}^{n}\frac{1}{s_n^2}\int_{|x-\mu_j|>\epsilon s_n}(x-\mu_j)^2 f_j(x)dx=0; \tag{57}$$

finalizando, assim, a demonstração da condição de Lindeberg apresentada em (20).

6 Discussões sobre aplicações do TCL

Ao longo das demonstrações do Teorema Central do Limite foi recorrente a observação da convergência para a distribuição normal paulatinamente com o aumento da estatística, i.e. de n, e nos exemplos práticos apresentados é evidente que tal conver-

gência não cumpre o mesmo ritmo para todos ao longo dos processos. Não basta tão simplesmente que algumas condições sejam satisfeitas, é preciso também que haja um número de processos razoáveis para convergir, portanto.

Logo, em aplicações práticas onde o número de processos é em regra bastante limitado, há sempre a questão da convergência dos observáveis e sua estabilidade antes do TCL ser invocado.

Esta questão precisa ser levantada sempre antes da invocação ao TCL em situações onde o número de processos não seja abundante e, quando não possível de fazê-lo, as consequências pela sua suposição serem avaliadas. Não raro o próprio leitor já deparou, por exemplo, com testes de hipóteses com o emprego direto de uma distribuição normal sem qualquer reflexão se (a) é aquela que melhor descreve a dispersão dos dados, se (b) a variabilidade observada nos dados é somente flutuações provocadas por erros de medidas ou (c) quais as consequências por se supor o TCL na avaliação seja para validar, seja para refutar hipóteses.

Outro ponto importante e constante em todas as demonstrações é a necessidade da independência funcional das variáveis aleatórias, sem ela o TCL não pode ser invocado. E nas situações práticas do mundo, infelizmente, esta condição é facilmente violada. Por exemplo, se o sistema apresentado na demonstração de DeMoivre e Laplace não contiver reposição, a cada observável tomado não apenas o tamanho da população se alteraria, como também a frequência de sucessos; ou seja, as variáveis aleatórias teriam agora dependência funcional em relação ao que ocorreu nos observáveis passados o que invalida o TCL. A violação da reposição, no geral, cria vínculos entre as variáveis e, portanto, dependência entre elas. A simples existência de dinâmica do sistema estudado pode representar a existência de dependências entre as variáveis, isto porque se $X(t)$ e $Y(t)$ são duas variáveis aleatórias dependentes do tempo, sempre será possível escrever localmente t em função de uma ou outra e, portanto, $X(Y)$ que explicita a dependência funcional. Em muitas situações práticas de aplicação do TCL o que é avaliado é a quase independência entre as variáveis, com uma população tão grande em relação aos observáveis tomados ou o seu quase regime estático, quando reconhecida a existência de dinâmica das variáveis.

E, mesmo no caso de se trabalhar com variáveis não identicamente distribuídas, deve ser observada na demonstração da versão de Lindeberg quão forte é a necessidade de um número elevado de processos para assegurar o regime de validade e, portanto, a convergência. Caso o número de processos seja limitado na realidade prática, uma inspeção será requerida antes da invocação do TCL.

Portanto, apesar do TCL ser um resultado comum a uma imensa diversidade de composições, o que o torna bastante útil, sua invocação requer cuidadosas reflexões e até verificações prévias.

7 Fechando com alguns contra-experimentos

Se no início deste trabalho foram apresentados exemplos de convergência da média e da variância em situações dentro dos limites de validade do TCL, nada melhor agora do que fechar com alguns contra-experimentos nos quais sabidamente algumas das condições de convergência foram violadas a fim de ilustrar a falta de aplicabilidade

do teorema nestas situações.

Contra exemplos são ideais para o amadurecimento dos conceitos trabalhados, podendo o leitor apreciar as consequências reais de quando quaisquer condições são violadas ou fragilizadas. E, com o madurecimento dos conceitos, o amadurecimento do espírito crítico que é tão importante para as tomadas de decisões por parte do pesquisador.

7.1 Dependência dinâmica sobre o sistema estudado

O primeiro experimento fictício apresentado trata de uma sequência de variáveis aleatórias do tipo uniforme, com média igual a zero e parâmetro k variando deterministicamente e monotonicamente com a evolução dos processos de medida, sendo ele incrementado a cada novo sorteio de amostras dentro da população.

Um paralelo entre este exemplo fictício e a realidade poderia ser, por exemplo, uma pesquisa na área de psicologia humana onde se procura medir certa característica psíquica do indivíduo com sua mente desarmada/distraída sob aquela característica, i.e. sem o indivíduo compreender aquilo que se procura com a pesquisa. Entretanto, em razão das limitações de recursos humanos e materiais, a pesquisa não pode ser executada em um único e exclusivo processo, requerendo assim diversas investidas de campo em momentos distintos. Ocorre que, por curiosidade da população investigada, informações e especulações são trocadas entre aqueles que participaram e aqueles que ainda não; e tudo a respeito de alguma provável orientação da pesquisa. Mesmo que os próximos indivíduos pesquisados não saibam precisamente aquilo que está sendo medido, alguns deles trazem especulações e crenças que, sob a influência de tais, não alteram o valor médio da medida na amostra, mas expande a variabilidade em razão do aumento paulatino de indivíduos na pesquisa que estão sob a influência de alguma crença especulatória. O aumento monotônico do parâmetro k está, assim, associado ao aumento do número de indivíduos nas investidas de campo que estejam sob a influência de alguma crença. Uma questão, portanto, bastante comum nesta área de estudo.

Nestas condições o leitor percebe que trata o caso de uma sequência de variáveis aleatórias **aparentemente** independentes e não identicamente distribuídas em razão das variações monotônicas do parâmetro da distribuição uniforme. Além disso, pode o leitor verificar que tal contexto satisfaz o regime de Lindeberg em (21), sendo ele proporcional a $1/j$, onde j trata do número de investidas de campo. Portanto, por construção, aparenta ser uma situação típica de possível uso do TCL para testes de hipóteses ou análise de dados. No entanto, não o é.

Isto porque o incremento dado ao parâmetro k a cada processo é equivalente a atribuir um tipo de dependência temporal às variáveis aleatórias, existindo um parâmetro característico para cada processo. E o efeito último disso é provocar dependência funcional entre as variáveis aleatórias, o que invalida o TCL.

Os resultados das simulações computacionais podem ser observados nas figuras (9-10), onde é nitidamente percebida uma distribuição com aspecto de sino no entorno da média igual a zero com uma sistemática tendência de acumulação com o aumento do número das amostras. Todavia, quando observado o resultado para a variância, claramente vê que a distribuição não detém uma forma de sino, ainda que

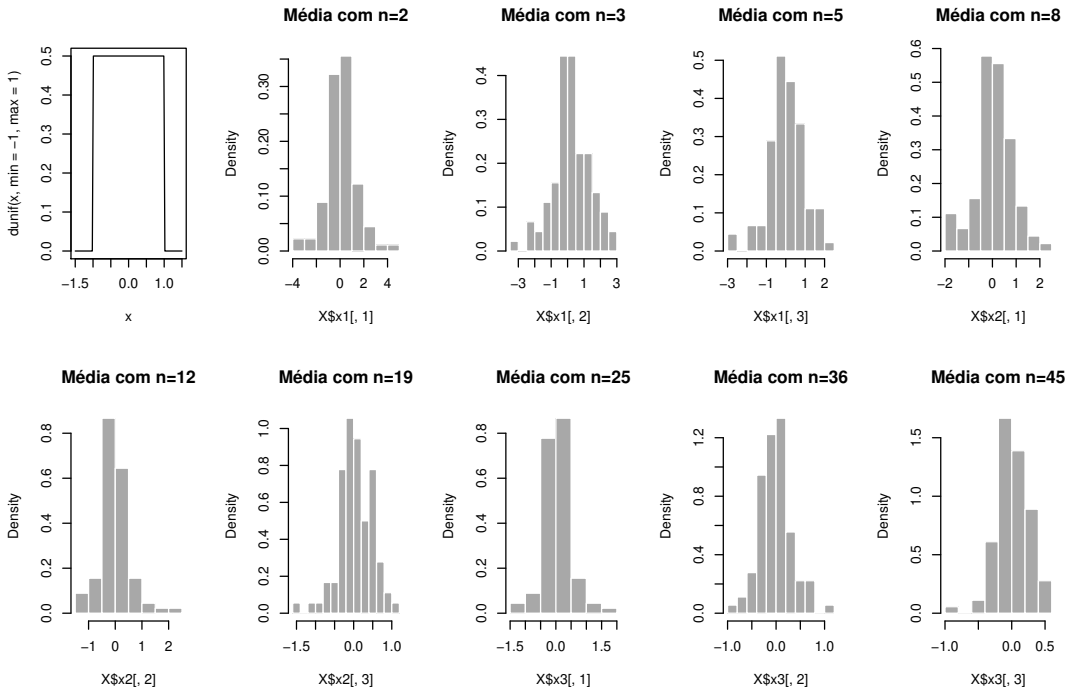

Figura 9: Resultado das distribuições para as médias para uma distribuição uniforme.

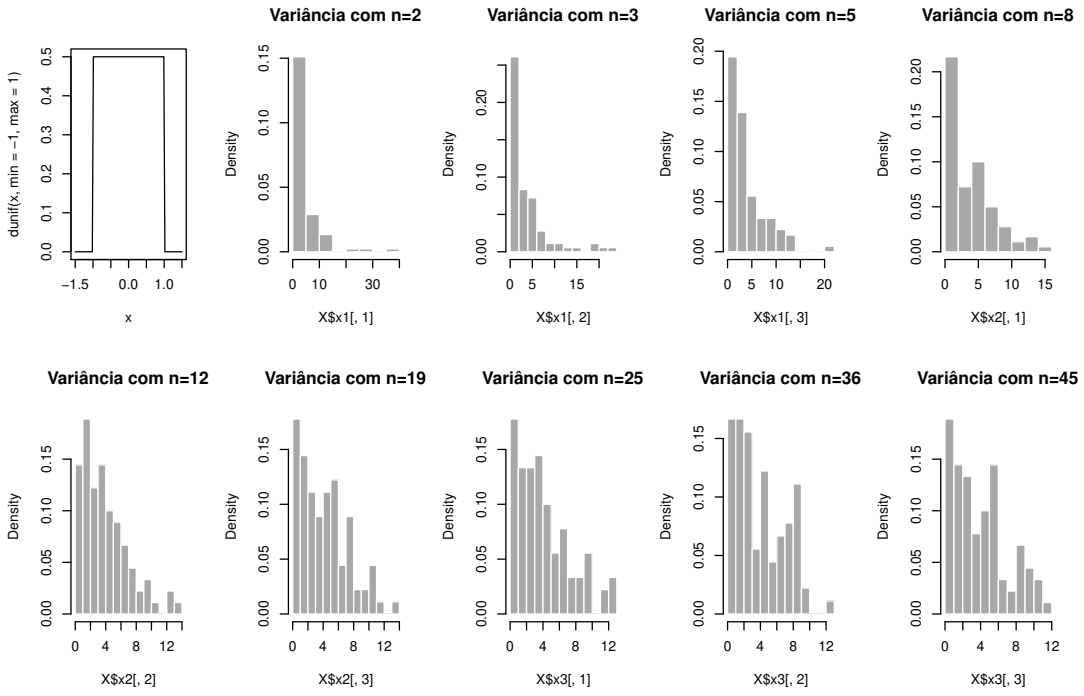

Figura 10: Resultado das distribuições para as variâncias para uma distribuição uniforme.

se aumente o tamanho das amostras, nem mesmo aparenta existir convergência de seu intervalo noutro mais definido. Algo bastante diferente daquilo apreciado anteriormente nas figuras (1-2). Logo, constata-se que a distribuição para qual o sistema está convergindo não pode ser aquela de uma normal/gaussiana.

A violação provocada neste contra-exemplo foi bastante sutil, pois num contexto real o pesquisador não dispõe tão facilmente da informação quanto ao processo de influência especulativa sobre a população; o que fatalmente poderá o levar acreditar na independência das variáveis aleatórias e, pior, apenas com a observação visual da distribuição das médias ele poderá ainda crer na validade do TCL; dificilmente ele se preocupará em observar a distribuição das variâncias.

Os códigos em linguagem R do contra-experimento aqui apresentado encontram-se no apêndice.

Por outro lado, caso o pesquisador adote uma linha bastante rigorosa de análise, a inconformidade dos seus dados com o modelo estatístico previamente vislumbrado para a pesquisa estará nesta situação denunciando a influência de qualquer fenômeno não considerado até então no trabalho. Sua missão será, então, procurar descobrir o que esteja influenciando o sistema de estudo. E o conhecimento sobre as condições de convergência do TCL pode, neste caso, funcionar como um bom guia para cercar as possíveis fontes do problema.

7.2 Sucessivas retiradas de amostras sem reposição da população original

O próximo contra-experimento trata da situação na qual a condição de reposição da população é quebrada; isto gera a criação de vínculo entre os parâmetros da distribuição da população e as realizações, i.e. a cada realização, os parâmetros da distribuição são alterados.

Pode num primeiro instante o leitor se equivocar e imaginar ser esta situação perfeitamente análoga ao contexto de Lindeberg, correspondendo as variações dos parâmetros à existência de distribuições não identicamente distribuídas. Mas aqui não é o que ocorre, uma vez que a criação de um vínculo acarreta uma dependência funcional, ou mais rigoroso, acarreta uma dependência de **um funcional** haja vista que a configuração presente da população é função de todo um conjunto de possíveis histórias. Em resumo, a configuração da população é dependente da história conhecida através das realizações da variáveis aleatórias passadas, eis que a independência das variáveis aleatórias não é afirmada e, portanto, não é válida a convergência de Lindeberg.

O presente contra-experimento é bastante rico em reflexões, pois ele trás uma limitação do mundo real na aquisição de informação sobre um sistema de estudo. Como regra em diversos trabalhos das áreas de biologia, saúde, humanidades e economia, a reposição da população é colocada como uma idealização para trazer simplicidade ao problema; uma vez que a dinâmica de condução da pesquisa é selecionar elementos distintos daquela população sem repetição.

Caso se assuma uma população tão grande comparada com o número de elementos selecionados nos levantamentos de campo, a influência desempenhada pelo vínculo da não reposição se torna diminuta e, assim, pode a situação real convergir

para o ideal. Ainda assim, dentro do problema concreto estudado, deve o pesquisador se questionar o quão grande deve ser sua população para se dizer que pode ele trabalhar com uma realidade idealizada.

Aos interessados em explorar a questão de idealização versus realidade, segue no apêndice o código em linguagem R deste experimento, podendo testá-lo através da convergência do TCL configurações distintas dos parâmetros do problema. E, assim, verificar quais ordens de grandeza necessita para o tamanho de sua população e os parâmetros do seu levantamento de campo a fim de justificar o emprego de um tratamento estatístico idealizado.

A população alvo dos levantamentos pode ser considerada como aquela formada por dois tipos distintos de elementos, por exemplo: um conjunto de dados de cor branca e outro conjunto de dados de cor preta, sendo a característica branca tida como especial. A cada levantamento são criados n grupos de k elementos em cada sem qualquer reposição da população nestes sorteios; nesta série de n grupos são apurados resultados de médias, variâncias e outros momentos centrais através das realizações de dados de cor branca. Então, para se ter a informação de qual a distribuição seguida pelas estatísticas dos levantamentos, são colecionados c ciclos de levantamentos. Caso se esteja sob o domínio de convergência do TCL, é esperada a verificação de distribuições estáveis do tipo gaussianas. E, pela construção do problema, é do conhecimento prévio que tal convergência apenas ocorreria para uma população grande suficiente para a influência do vínculo de não reposição ser negligenciável.

Um paralelo real deste mesmo problema poderia ser proposto na área médica, onde um conjunto de características definiria uma população preferencialmente alvo de certo problema médico – o dado de cor branca –, existindo c diversos grupos espalhados por todo o planeta que estivessem dentro de sua área de atuação coletando k pessoas distintas dentro da população previamente definida para a criação ao longo do tempo de condução do trabalho de n grupos padronizados a fim de ser verificada a frequência de aparecimento do problema médico – número de dados de cor branca –, sendo, por cada equipe médica, formados n grupos ao longo do tempo a fim de terminar uma estatística característica para aquela localidade. Por fim, reunidas todas as estatísticas locais obtidas pelas equipes médicas, ter-se-ia a distribuição mundial da frequência de aparecimento do problema médico dentro da população alvo. Veja aqui que trata exatamente no mesmo problema proposto no parágrafo anterior e onde o problema da não reposição é intrínseco.

Tal tipo de sorteio, i.e. aquele que seleciona k elementos de um único grupo, tem distribuição de probabilidade conhecida e trata de uma hipergeométrica. Desse modo, para o sorteio seguinte, basta atualizar seus parâmetros segundo o resultado realizado anteriormente, e assim sucessivamente; literalmente consumindo com a população.

O resultado do experimento para uma população formada por 43.000 dados de cor branca, 200.000 dados de cor preta, 30 elementos sorteados em cada grupo, número variável de grupos de diversas situações e 90 equipes; que tem na situação de maior consumo da população o índice de 50% com $n = 45$ e o menor de 2,2% com $n = 2$, está apresentado nas figuras (11-12).

Na figura (11) é primeiramente observado que não há o que ser falado quanto à

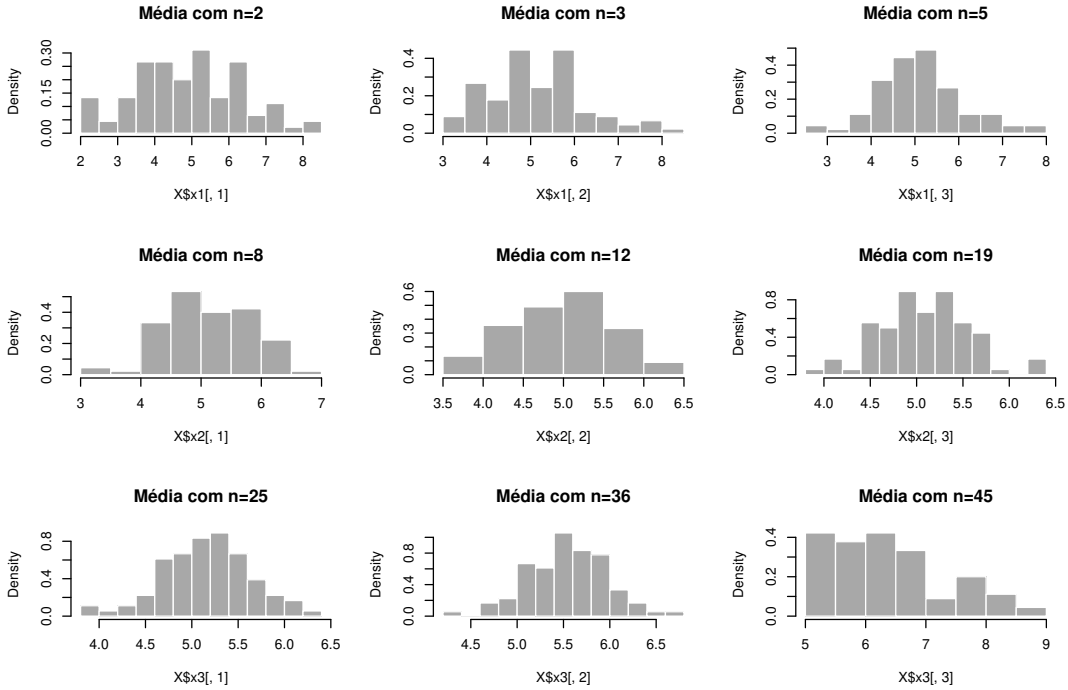

Figura 11: Resultado das distribuições para as médias. Usados 43000 dados brancos, 200000 dados pretos, 30 elementos sorteados por grupo e 90 equipes realizando o mesmo procedimento, o número de grupos trabalhados por equipe encontra-se destacado em cada gráfico.

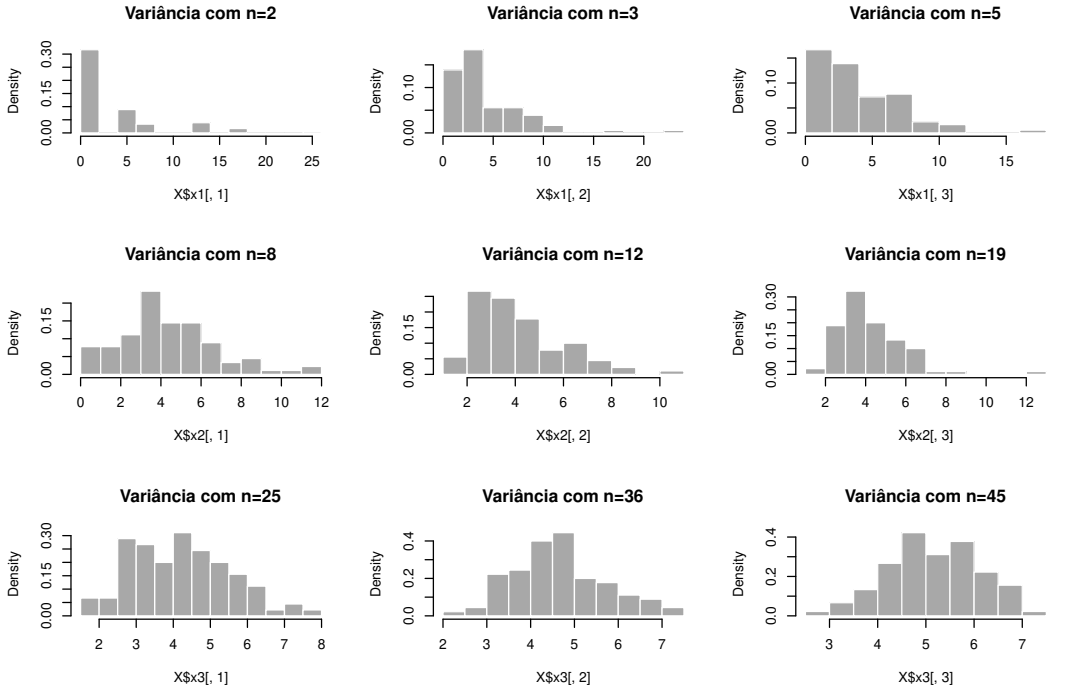

Figura 12: Resultado das distribuições para as variâncias nas mesmas condições anteriores do experimento.

convergência com o aumento de n; logo, refinar as estatísticas nesta situação não levará a qualquer convergência tipo TCL. O mesmo se aplicando para a figura (12). E a razão para isso é bastante óbvia: qualquer refinamento da estatística implica maior consumo da população e, quanto maior o seu consumo, maior é também é a sua aleatoriedade com a variedade de histórias possíveis de terem ocorrido, em outras palavras, a informação a respeito da população inicial é apenas mais severamente degrada com as incursões.

Neste exemplo fica bastante claro que situações do mundo real podem colocar sutis barreiras na aquisição de informação, pois investir em mais dados da população aqui trabalhada não necessariamente implicará obter uma melhor informação, até o contrário, podendo levar a uma situação de ignorância compatível com as investidas menos ambiciosas.

Em resumo, não há como invocar o TCL para trabalhar com problemas de um sistema como o aqui descrito, pelo menos não com os parâmetros utilizados.

Para finalizar os comentários sobre os resultados, observe na figura (11) um fato bastante interessante: a distribuição inicial se apresenta disforme e com o aumento de n vai adquirindo uma forma mais centrada num mesmo valor estático para a média e no entorno dele vai assumindo contornos de sino tipo gaussiano, mas perdendo esta característica após certo patamar de n; e algo similar parece ocorrer para a variância também. Casando os dois resultados, é por volta de $n = 36$ que se tem um resultado com valores de média e variância mais estabilizados e a distribuição mais próxima a uma gaussiana. Apesar de não ser o objetivo principal deste trabalho, este fato merece ser indicado aqui, pois, ainda que não se possa invocar o TCL neste caso, aparenta existir uma configuração de parâmetros do trabalho de pesquisa que possibilita uma informação mais bem definida do que as demais configurações. E isto é peculiar em sistemas que se modificam com o ato de medição, como aquele aqui apresentado.

Resta agora verificar qual o resultado para o mesmo sistema com um número bem maior de sua população e, para isso, foram tomados 4.300.000 dados brancos e 20.000.000 de dados pretos, com a manutenção dos demais parâmetros da simulação anterior. E os resultados são apresentados nas figuras (13-14).

Como esperado, os resultados são agora mais compatíveis com aquilo observado no TCL, indicando que houve uma melhor convergência com o crescimento da população; no entanto, em relação aos resultados do início deste trabalho, pode ser observado que a convergência é ainda sofrível.

Isto ocorre justamente porque as mudanças na população passam agora a serem diminutas, estando ela quase no regime de reposição. Na configuração de maior incursão, por exemplo, são 0,5% de consumo da população para $n = 45$. Quanto menor o percentual de consumo da população, mais próximo se estará da condição ideal de reposição de elementos e menor a influência de dependência entre as variáveis; portanto, mais próximo do regime de validade do TCL.

Mas note neste exemplo o quão foi alto o custo de tal aproximação ao regime de validade do TCL: de uma população de 243.000 elementos, saltou-se para outra de 24.300.000 elementos e ainda com uma convergência bastante fraca. Lembrando que em questões médicas, econômicas, biológicas e humanas a realidade impõe um limite supremo claro, i.e. o número de habitantes no planeta da ordem de 8 bilhões,

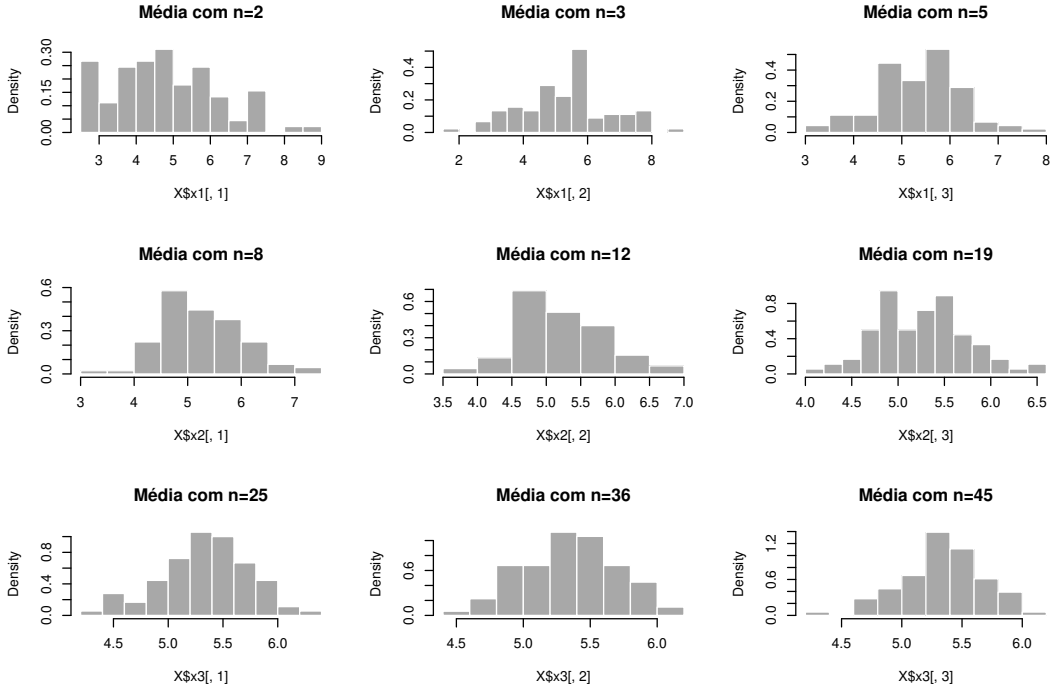

Figura 13: Resultado das distribuições para as médias. Usados 4.300.000 dados brancos, 20.000.000 dados pretos, 30 elementos sorteados por grupo e 90 equipes realizando o mesmo procedimento, o número de grupos trabalhados por equipe encontra-se destacado em cada gráfico.

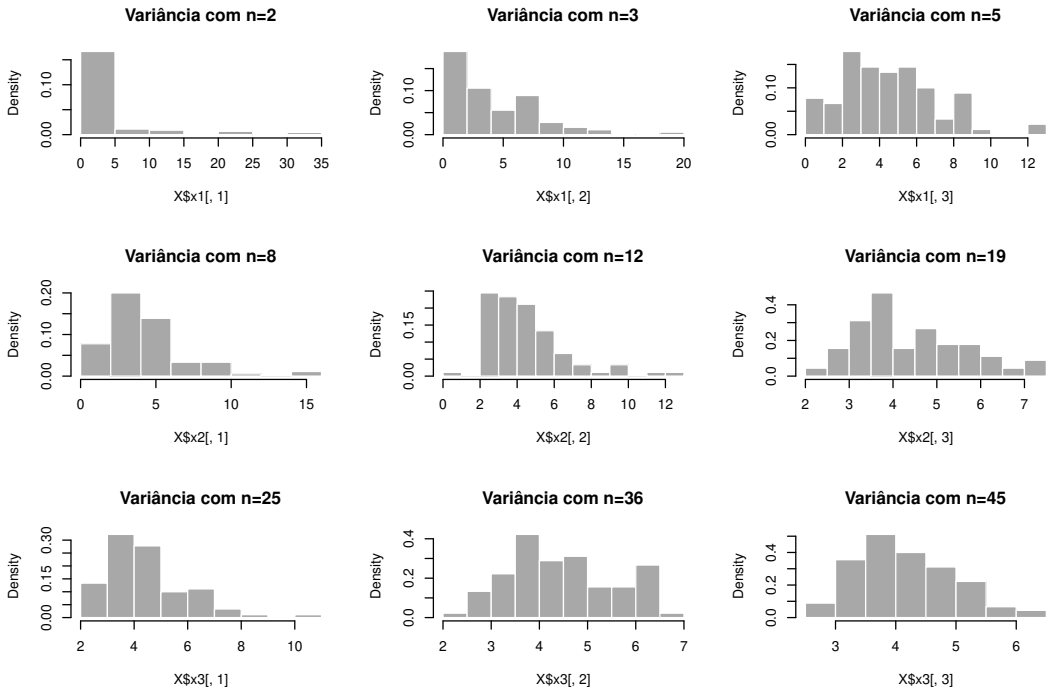

Figura 14: Resultado das distribuições para as variâncias nas mesmas condições anteriores do experimento.

a idealização máxima da realidade num problema real pode vir a ser um erro.

Com isso, fica ilustrado também que diante dos problemas reais deve o pesquisador definir antes aquilo que, na situação concreta, seja um *"grande número"* para justificar o emprego de distribuições gaussianas mediante TCL; em especial nas áreas de saúde, humanas, econômicas e biológicas, onde o espaço amostral real pode ser um tanto limitado para o uso de certos resultados matemáticos.

Referências

[1] R. B. Alves. Teorema central do limite para martingais. Master's thesis, Pontificia Universidade Catolica do Rio de Janeiro, www.maxwell.vrac.puc-rio.br:32327, 2017.

[2] W. FELLER. *An introduction to Probability Theory and its applications*, volume I. 1968.

[3] H. FISCHER. *A history of the Central Limit Theorem from classical to modern*. New York: Springer, 2011.

[4] T. Fischetti. *Data Analysis with R*. Packtpub, second edition, 2018.

A Códigos em R para as simulações

A.1 Variâncias para distribuição uniforme

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  var( runif( n=n[i,1])) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição uniforme
        X$x2[j,i] =  var( runif( n=n[i,2])) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição uniforme
        X$x3[j,i] =  var( runif( n=n[i,3]))  # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição uniforme
    }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dunif(x),xlim=c(0,1),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=3")
```

```
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=45")
```

A.2 Média e variância para a distribuição exponencial

Código para a média:

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  mean( rexp( n=n[i,1], rate=2)) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição exponencial
```

```
        X$x2[j,i] =  mean( rexp( n=n[i,2], rate=2)) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição exponencial
        X$x3[j,i] =  mean( rexp( n=n[i,3], rate=2)) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição exponencial
    }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dexp(x, rate=2),xlim=c(0,3),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=45")
```

Código para a variância:

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
```

```
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  var( rexp( n=n[i,1], rate=2)) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição exponencial
        X$x2[j,i] =  var( rexp( n=n[i,2], rate=2)) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição exponencial
        X$x3[j,i] =  var( rexp( n=n[i,3], rate=2))  # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição exponencial
    }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dexp(x, rate=2),xlim=c(0,3),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=45")
```

A.3 Média e variância para a distribuição χ^2

Código para a média:

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  mean( rchisq( n=n[i,1], df=8)) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição X^2
        X$x2[j,i] =  mean( rchisq( n=n[i,2], df=8)) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição X^2
        X$x3[j,i] =  mean( rchisq( n=n[i,3], df=8)) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição X^2
    }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dchisq(x, df=8),xlim=c(0,25),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=3")
```

```
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=45")
```

Código para a variância:

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  var( rchisq( n=n[i,1], df=8)) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição X^2
        X$x2[j,i] =  var( rchisq( n=n[i,2], df=8)) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição X^2
```

```
        X$x3[j,i] =  var( rchisq( n=n[i,3], df=8))  # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz, distribuição X^2
        }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dchisq(x, df=8),xlim=c(0,25),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=45")
```

A.4 Média e variância para a distribuição β

Código para a média:

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
```

```
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  mean( rbeta( n=n[i,1], shape1=.3,
         ↪  shape2=.1)) # Calcula a média para o conjunto
         ↪  sorteado e a armazena na matriz, distribuição beta
        X$x2[j,i] =  mean( rbeta( n=n[i,2], shape1=.3,
         ↪  shape2=.1)) # Calcula a média para o conjunto
         ↪  sorteado e a armazena na matriz, distribuição beta
        X$x3[j,i] =  mean( rbeta( n=n[i,3], shape1=.3,
         ↪  shape2=.1)) # Calcula a média para o conjunto
         ↪  sorteado e a armazena na matriz, distribuição beta
    }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dbeta(x,  shape1=.3, shape2=.1),xlim=c(0,1),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=45")
```

Código para a variância:

```
# Experimento:

n=cbind(c(2, 3, 5), # números de elementos do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de elementos do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de elementos do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  var( rbeta( n=n[i,1], shape1=.3,
         ↪  shape2=.1)) # Calcula a variância para o conjunto
         ↪  sorteado e a armazena na matriz, distribuição beta
        X$x2[j,i] =  var( rbeta( n=n[i,2], shape1=.3,
         ↪  shape2=.1)) # Calcula a variância para o conjunto
         ↪  sorteado e a armazena na matriz, distribuição beta
        X$x3[j,i] =  var( rbeta( n=n[i,3], shape1=.3,
         ↪  shape2=.1))  # Calcula a variância para o conjunto
         ↪  sorteado e a armazena na matriz, distribuição beta
 }
}

# Plotar resultados:

par(mfrow=c(2,5))
curve(dbeta(x,  shape1=.3, shape2=.1),xlim=c(0,1),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=5")
```

```
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=45")
```

B Códigos em R para os contra-exemplos

B.1 Médias e variâncias para distribuição uniforme

Código para a média:

```
#Contra experimento com a distribuição uniforme: média
# Experimento:
n=cbind(c(2, 3, 5), # números de sorteios do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de sorteios do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de sorteios do sétimo ao nono
         ↪  experimentos
X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    k<-1.0 # Parâmetro k inicial
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  mean( runif( n=n[i,1], min=-1.0*k, max=k))
         ↪  # Calcula a média para o conjunto sorteado e a
         ↪  armazena na matriz, distribuição exponencial
```

```
        X$x2[j,i] =  mean( runif( n=n[i,2],  min=-1.0*k, max=k))
         ↪  # Calcula a média para o conjunto sorteado e a
         ↪  armazena na matriz, distribuição exponencial
        X$x3[j,i] =  mean( runif( n=n[i,3],   min=-1.0*k,
         ↪  max=k)) # Calcula a média para o conjunto sorteado e
         ↪  a armazena na matriz, distribuição exponencial
        k<-k+0.05 # Incrementos atribuídos à k ao longo das
         ↪  pesquisas de campo
    }
}
# Plotar resultados:
par(mfrow=c(2,5))
curve(dunif(x, min=-1.0, max=1.0), xlim=c(-1.5,1.5),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=45")
```

Código para a variância:

```
#Contra experimento com a distribuição uniforme: variância
# Experimento:
n=cbind(c(2, 3, 5), # números de sorteios do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de sorteios do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de sorteios do sétimo ao nono
         ↪  experimentos
X<-list(
```

```
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  variâncias do grupo sorteado da terceira linha

for ( i in 1:3){  # varrendo cada linha de experimentos
    k<-1.0 # Parâmetro k inicial
    for ( j in 1: 90){  # Registrando o resultado do conjunto
     ↪  sorteado na matriz
        X$x1[j,i] =  var( runif( n=n[i,1], min=-1.0*k, max=k)) #
         ↪  Calcula a variância para o conjunto sorteado e a
         ↪  armazena na matriz, distribuição uniforme
        X$x2[j,i] =  var( runif( n=n[i,2],  min=-1.0*k, max=k))
         ↪  # Calcula a variância para o conjunto sorteado e a
         ↪  armazena na matriz, distribuição uniforme
        X$x3[j,i] =  var( runif( n=n[i,3],   min=-1.0*k, max=k))
         ↪  # Calcula a variância para o conjunto sorteado e a
         ↪  armazena na matriz, distribuição uniforme
        k<-k+0.05 # Incrementos atribuídos à k ao longo das
         ↪  pesquisas de campo
    }
}
# Plotar resultados:
par(mfrow=c(2,5))
curve(dunif(x, min=-1.0, max=1.0), xlim=c(-1.5,1.5),xlab="x")
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=36")
```

```
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=45")
```

B.2 Médias e variâncias para sistema sem reposição - hipergeométrica

Código para a média:

```
#Contra experimento com a distribuição hipergeométrica sem
 ↪  reposição

#### PARÂMETROS DA POPULAÇÃO
pop.brancos=43000
pop.pretos=200000
############################

#### PARÂMETROS DO LEVANTAMENTO
lev.equipes=90
lev.elementos=30
##############################

#### Função que gera o sorteio reduzindo a população

sorteio=function(brancos, pretos, dimensao,iteracao, ciclos){
        # brancos - número de dados brancos
        # pretos - número de dados pretos
        # dimensao - tamanho do conjunto de elementos extraídos
         ↪  num sorteio
        # iteracao - número de elementos retirados em cada
         ↪  sorteio
        # ciclos - número de vezes que o processo será repetido

        if((brancos+pretos)<(iteracao*dimensao*ciclos)){ #
         ↪  Verificação de número de elementos suficientes para
         ↪  efetuar todos os sorteios
                stop("Não há condições de efetuar todos os
                 ↪  sorteios. Número de elementos inferior ao
                 ↪  necessário")
                exit()
        }
        amostra<-rep(0, dimensao) # Vetor que armazena os
         ↪  resultados dos sorteios
```

```
        for(i in 1:dimensao){ # Realiza ciclos de sorteios com
         ↪  modificação na população e atualizando seus
         ↪  parâmetros
                amostra[i]<-rhyper(1, brancos, pretos, iteracao)
                brancos= brancos-amostra[i]
                pretos= pretos-(dimensao-amostra[i])
        }


        resultados<-c(brancos, pretos, amostra)

        return(resultados)
} ####### fim da função sorteio
################################


n=cbind(c(2, 3, 5), # números de sorteios do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de sorteios do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de sorteios do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha

# Definindo o número de elementos na população
total.brancos<-pop.brancos
total.pretos<-pop.pretos

# Definindo o número de retiradas num mesmo processo
itera<-lev.elementos

# Definindo a matriz inicial de parâmetros
PARAMETROS<-list(
        brancos=matrix(total.brancos, nrow=3, ncol=3),
        pretos=matrix(total.pretos, nrow=3, ncol=3))

# Iniciar a simulação
```

```
for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1:lev.equipes){  # Registrando o resultado do
     ↪  conjunto sorteado na matriz
        result=sorteio(brancos=PARAMETROS$brancos[i,1],
         ↪  pretos=PARAMETROS$pretos[i,1], dimensao=n[i,1],
         ↪  itera, ciclos=(lev.equipes-j+1))
        X$x1[j,i] =  mean(result[3:(n[i,1]+2)]) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz
        PARAMETROS$brancos[i,1]=result[1]
        PARAMETROS$pretos[i,1]=result[2]

        result=sorteio(brancos=PARAMETROS$brancos[i,2],
         ↪  pretos=PARAMETROS$pretos[i,2], dimensao=n[i,2],
         ↪  itera, ciclos=(lev.equipes-j+1))
        X$x2[j,i] =  mean(result[3:(n[i,2]+2)]) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz
        PARAMETROS$brancos[i,2]=result[1]
        PARAMETROS$pretos[i,2]=result[2]

        result=sorteio(brancos=PARAMETROS$brancos[i,3],
         ↪  pretos=PARAMETROS$pretos[i,3], dimensao=n[i,3],
         ↪  itera, ciclos=(lev.equipes-j+1))
        X$x3[j,i] =  mean(result[3:(n[i,3]+2)]) # Calcula a
         ↪  média para o conjunto sorteado e a armazena na
         ↪  matriz
        PARAMETROS$brancos[i,3]=result[1]
        PARAMETROS$pretos[i,3]=result[2]

 }
}

################## Plotar resultados:
par(mfrow=c(3,3))
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=5")
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=8")
```

```
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Média com n=45")
```

Código para a variância:

```
#Contra experimento com a distribuição hipergeométrica sem
 ↪  reposição

#### PARÂMETROS DA POPULAÇÃO
pop.brancos=43000
pop.pretos=200000
############################

#### PARÂMETROS DO LEVANTAMENTO
lev.equipes=90
lev.elementos=30
##############################

#### Função que gera o sorteio reduzindo a população

sorteio=function(brancos, pretos, dimensao,iteracao, ciclos){
        # brancos - número de dados brancos
        # pretos - número de dados pretos
        # dimensao - tamanho do conjunto de elementos extraídos
         ↪  num sorteio
        # iteracao - número de elementos retirados em cada
         ↪  sorteio
        # ciclos - número de vezes que o processo será repetido

        if((brancos+pretos)<(iteracao*dimensao*ciclos)){ #
         ↪  Verificação de número de elementos suficientes para
         ↪  efetuar todos os sorteios
                stop("Não há condições de efetuar todos os
                 ↪  sorteios. Número de elementos inferior ao
                 ↪  necessário")
```

```
                exit()
        }
        amostra<-rep(0, dimensao) # Vetor que armazena os
         ↪  resultados dos sorteios
        for(i in 1:dimensao){ # Realiza ciclos de sorteios com
         ↪  modificação na população e atualizando seus
         ↪  parâmetros
                amostra[i]<-rhyper(1, brancos, pretos, iteracao)
                brancos= brancos-amostra[i]
                pretos= pretos-(dimensao-amostra[i])
        }


        resultados<-c(brancos, pretos, amostra)

        return(resultados)
} ####### fim da função sorteio
################################

n=cbind(c(2, 3, 5), # números de sorteios do primeiro ao
 ↪  terceiro experimentos
        c(8,12, 19), # números de sorteios do quarto ao sexto
         ↪  experimentos
        c(25, 36, 45)) # números de sorteios do sétimo ao nono
         ↪  experimentos

X<-list(
    x1=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da primeira linha
    x2=matrix(NA,nrow=90,ncol=3), # Matriz que armazenara as
     ↪  médias do grupo sorteado da segunda linha
    x3=matrix(NA,nrow=90,ncol=3)) # Matriz que armazenara as
     ↪  médias do grupo sorteado da terceira linha

# Definindo o número de elementos na população
total.brancos<-pop.brancos
total.pretos<-pop.pretos

# Definindo o número de retiradas num mesmo processo
itera<-lev.elementos

# Definindo a matriz inicial de parâmetros
PARAMETROS<-list(
        brancos=matrix(total.brancos, nrow=3, ncol=3),
```

```
        pretos=matrix(total.pretos, nrow=3, ncol=3))

# Iniciar a simulação
for ( i in 1:3){  # varrendo cada linha de experimentos
    for ( j in 1:lev.equipes){  # Registrando o resultado do
     ↪  conjunto sorteado na matriz
        result=sorteio(brancos=PARAMETROS$brancos[i,1],
         ↪  pretos=PARAMETROS$pretos[i,1], dimensao=n[i,1],
         ↪  itera, ciclos=(lev.equipes-j+1))
        X$x1[j,i] =  var(result[3:(n[i,1]+2)]) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz
        PARAMETROS$brancos[i,1]=result[1]
        PARAMETROS$pretos[i,1]=result[2]

        result=sorteio(brancos=PARAMETROS$brancos[i,2],
         ↪  pretos=PARAMETROS$pretos[i,2], dimensao=n[i,2],
         ↪  itera, ciclos=(lev.equipes-j+1))
        X$x2[j,i] =  var(result[3:(n[i,2]+2)]) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz
        PARAMETROS$brancos[i,2]=result[1]
        PARAMETROS$pretos[i,2]=result[2]

        result=sorteio(brancos=PARAMETROS$brancos[i,3],
         ↪  pretos=PARAMETROS$pretos[i,3], dimensao=n[i,3],
         ↪  itera, ciclos=(lev.equipes-j+1))
        X$x3[j,i] =  var(result[3:(n[i,3]+2)]) # Calcula a
         ↪  variância para o conjunto sorteado e a armazena na
         ↪  matriz
        PARAMETROS$brancos[i,3]=result[1]
        PARAMETROS$pretos[i,3]=result[2]

 }
}

################### Plotar resultados:
par(mfrow=c(3,3))
hist(X$x1[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=2")
hist(X$x1[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=3")
hist(X$x1[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=5")
```

```
hist(X$x2[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=8")
hist(X$x2[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=12")
hist(X$x2[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=19")
hist(X$x3[,1], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=25")
hist(X$x3[,2], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=36")
hist(X$x3[,3], col="darkgray", border="white",freq = FALSE,
 ↪  breaks = 10, main="Variância com n=45")
```

ARTIGO-04-JAN-19.pdf

Documento número #de4fd699-6070-45c8-9224-10da6dc5e391

Assinaturas

Flávio Henrique Severino Oliveira Vieira
Assinou

Log

04 Jan 2019, 11:26:01	Operador com email lino.henry@gmail.com na Conta 3ecfc2a2-8dc0-4859-aeab-ae67fb0e14cc criou este documento número de4fd699-6070-45c8-9224-10da6dc5e391. Data limite para assinatura do documento: 03 de Fevereiro de 2019 (23:59). Finalização automática após a última assinatura: habilitada. Idioma: Português brasileiro.
04 Jan 2019, 11:26:50	Operador com email lino.henry@gmail.com na Conta 3ecfc2a2-8dc0-4859-aeab-ae67fb0e14cc adicionou à Lista de Assinatura: LINO.HENRY@GMAIL.COM, para assinar, com os pontos de autenticação: email (via token); Nome Completo; CPF; Data de Nascimento; endereço de IP.
04 Jan 2019, 11:29:28	Flávio Henrique Severino Oliveira Vieira assinou. Pontos de autenticação: email LINO.HENRY@GMAIL.COM (via token). CPF informado: 042.531.026-46. IP: 177.148.211.86. Componente de assinatura versão 1.22.4 disponibilizado em https://app.clicksign.com.
04 Jan 2019, 11:29:28	Processo de assinatura finalizado automaticamente. Motivo: finalização automática após a última assinatura habilitada. Processo de assinatura concluído para o documento número de4fd699-6070-45c8-9224-10da6dc5e391.

Hash do documento original (SHA256): 2f8d761d33c12684144855a5b0e6dc4d32f278773e3bcbea972cae7b15aa567a