Trabalho de Variáveis Aleatórias

Ivan Cunha Vieira

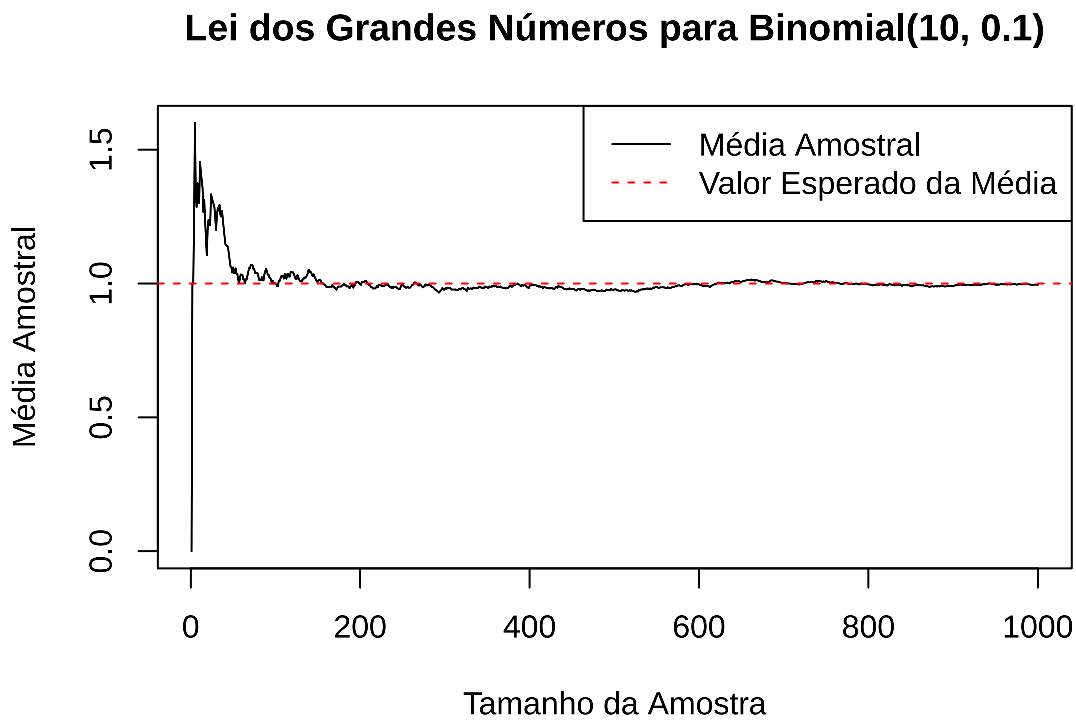
Contents

Questão 1

A Lei dos Grandes Números (LGN) é um dos resultados mais importantes na teoria estatística. Tal resultado permite afirmar que a média amostral tende para a média populacional, quando o tamanho da amostra é suficientemente grande. Sugira um programa, em código R, que permita mostrar computacionalmente a LGN para valores pseudo-aleatórios de uma variável aleatória binomial com parâmetros n = 10 e p = 0.1.

Resposta:

```r
# Definindo os parâmetros da distribuição binomial
n <- 10
p <- 0.1

# Número de amostras a gerar
num_amostras <- 1000

# Vetor que irá conter as médias obtidas
seq_medias <- NULL

# Definir seed para reprodutibilidade
set.seed(123)

# Armazenar os números aleatórios gerados
y <- rbinom(num_amostras, size = n, prob = p)

#
for (i in 1:num_amostras){
  seq_medias[i] <- mean(y[1:i])
}

# Plotar as médias amostrais em função do tamanho da amostra
plot(1:num_amostras, seq_medias, type = "l",
     main = "Lei dos Grandes Números para Binomial(10, 0.1)",
     xlab = "Tamanho da Amostra", ylab = "Média Amostral")

# Adicionar a linha horizontal no valor esperado da média teórica
abline(h = n * p, col = "red", lty = 2)
legend("topright", legend = c("Média Amostral", "Valor Esperado da Média"),
       col = c("black", "red"), lty = c(1, 2))
```


Lei dos Grandes Números para Binomial(10, 0.1)
Média Amostral
Valor Esperado da Média
Média Amostral
0.0
0.5
1.0
1.5
0
200
400
600
800
1000
Tamanho da Amostra

Questão 2

Seja X uma variável aleatória em um espaço de probabilidade (Ω, A, P) e suponha que X = U(0, 1). Obtenha a distribuição de Y = -ln(X).

- Faça uma análise descritiva;
- Qual a distribuição de Y ? Veja de forma simulada e na forma teórica;
- Com base na distribuição da variável aleatória (v.a.) Y do exercício acima, implemente uma função em R que gere observações de Y.

Resposta:

Temos que X é uma v.a. com distribuição uniforme (0,1) e Y uma v.a. com distribuição -ln(X). Dessa forma, nós podemos utilizar de transformações para calcular a probabilidade de Y, como a seguir:

$$P(Y \leq y) = P(-\ln(X) \leq y)$$

Procurando simplificar a inequação, encontramos:

$$-\ln(X) \leq y \Rightarrow \ln(X) \geq -y \Rightarrow X \geq e^{-y}$$

Dessa forma:

$$P(Y \leq y) = P(X \geq e^{-y}) = 1 - P(X < e^{-y})$$

Encontramos a $Fx(x)$ e a $Fy(y)$, como demonstrado:

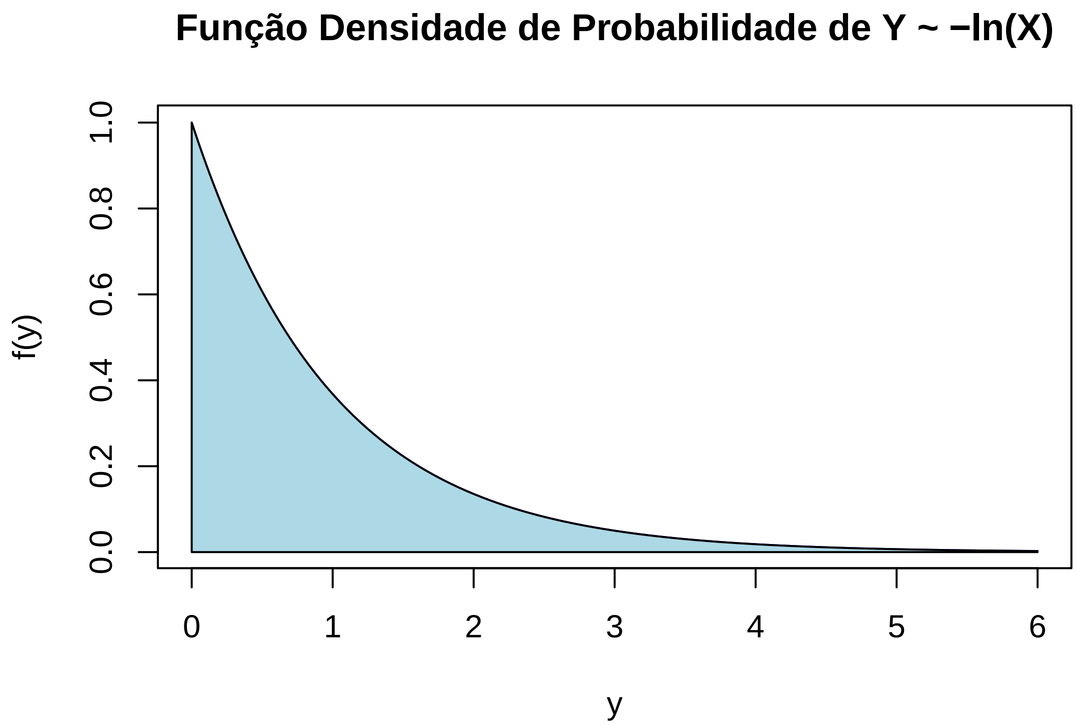
$$P(X < e^{-y}) = F_X(e^{-y}) = e^{-y} P(Y \leq y) = 1 - P(X < e^{-y}) = 1 - F_X(e^{-y}) = 1 - e^{-y} = F_Y(y)$$

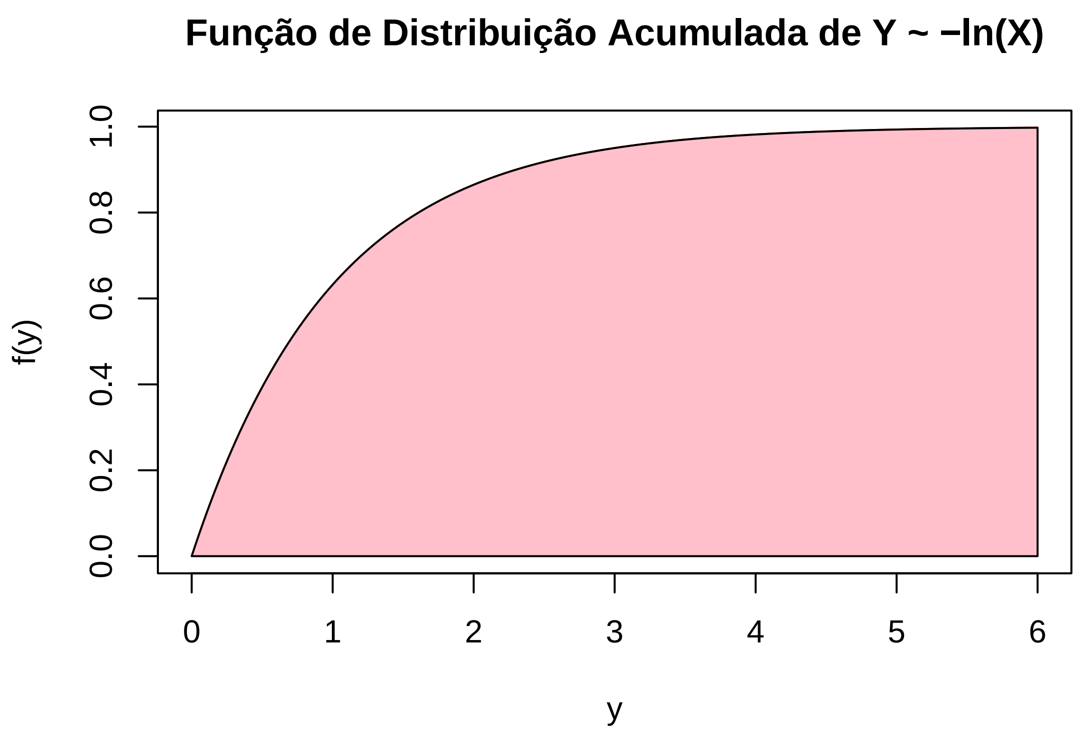
Tomando a derivada:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(1 - e^{-y}) = e^{-y} \text{ , para } y > 0$$

E, dessa forma, encontramos a função densidade de probabilidade (f.d.p) de Y, que segue uma distribuição exponencial com parâmetro $\lambda = 1$ ($Y \sim Exp(1)$).

Abaixo, a representação gráfica das funções de distribuição acumulada e de densidade de probabilidade de Y:





Implementando uma função geradora da distribuição de Y:

```r
gerar_Y <- function(n) {
  # Gera n observações de uma variável aleatória Y ~ Exp(1)
  y_values <- rexp(n, rate = 1)
  return(y_values)
}

# Definir seed para reprodutibilidade
set.seed(123)

# Exemplo de uso: gerar 10 observações de Y
obs <- gerar_Y(10)
print(obs)
```

```
##  [1] 0.84345726 0.57661027 1.32905487 0.03157736 0.05621098 0.31650122
##  [7] 0.31422729 0.14526680 2.72623646 0.02915345
```

Questão 3

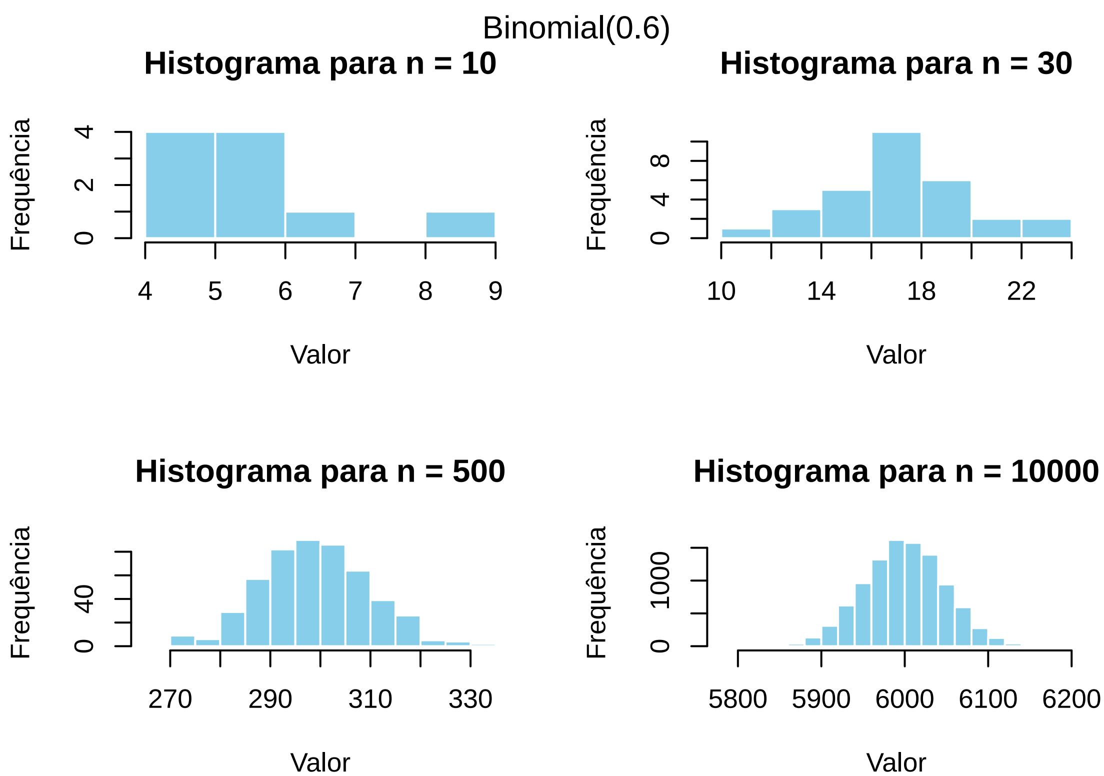
Gere X1, X2, . . . , Xn variáveis aleatórias, n = 10, 30, 500, 10000, das distribuições a seguir:

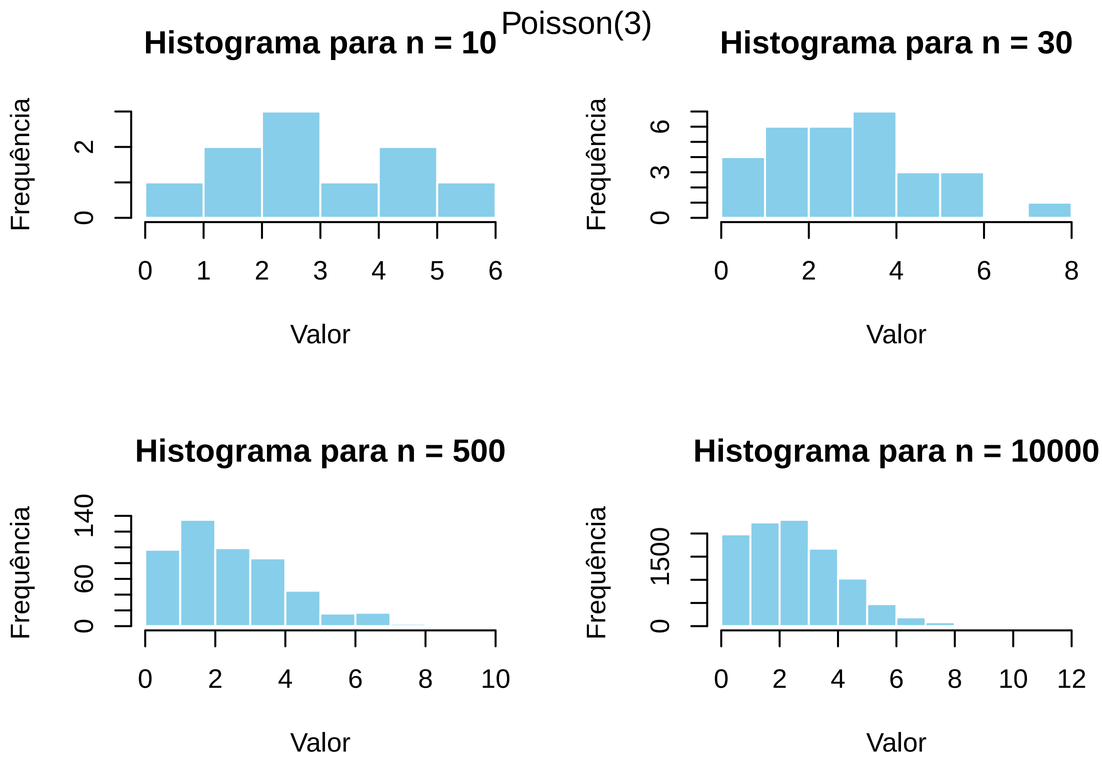
- Binomial(n, 0.6)
- Poisson(3)
- $X^2(6)$
- F(5,3)

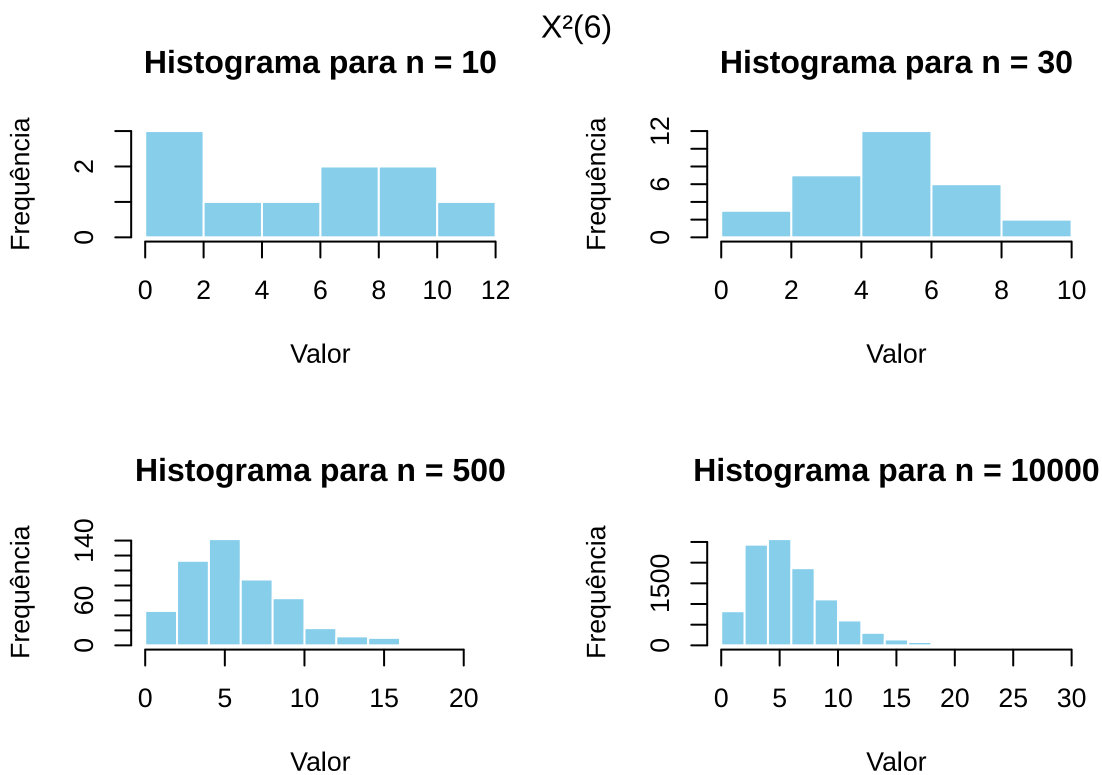
Diante disso: [i.] Considere o Teorema Central do Limite, ou seja, pegue as variáveis acima e verifique, com o qqplot (qqnorm) se de fato as variáveis acima convergem para a normal. Se sim, diga para qual normal (parâmetros) ocorre a convergência. Para responder sobre a convergência utilize apenas n = 10000 e diga se coincide com a teoria. [ii.] Aplique um teste de ajuste (aderência) e conclua sobre o que afirma no item anterior.

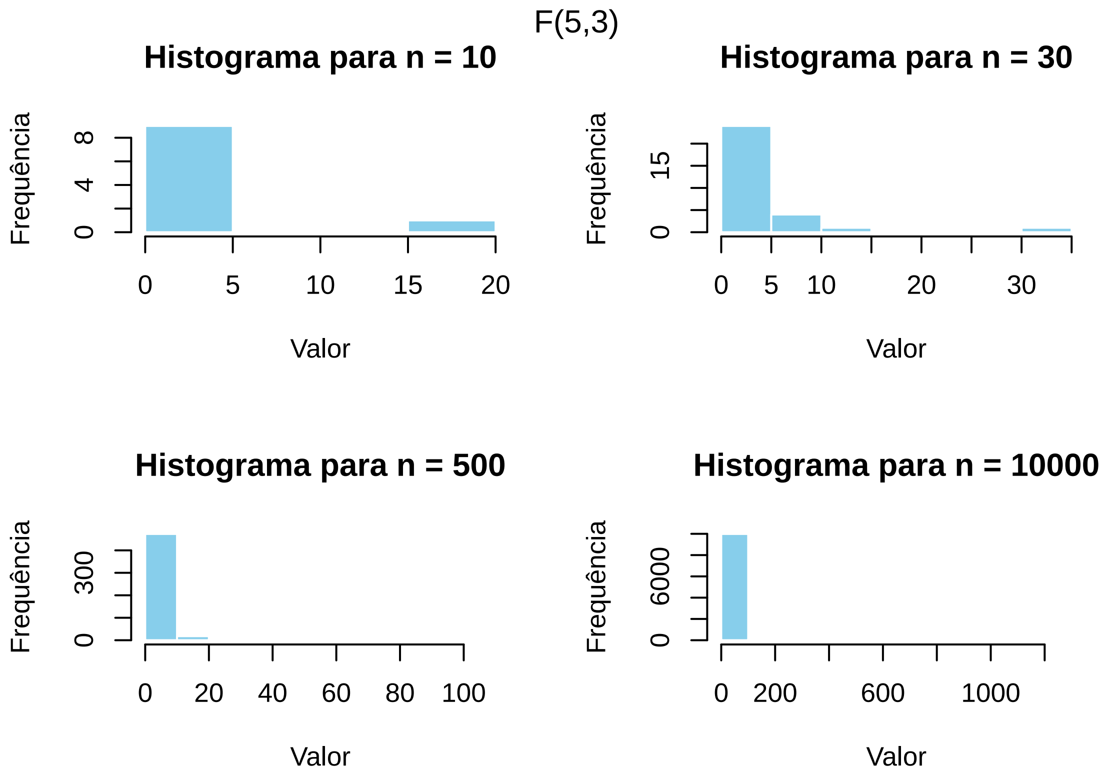
Resposta

Plotando os gráficos para verificar as formas das distribuições, na respectiva ordem do exercício:

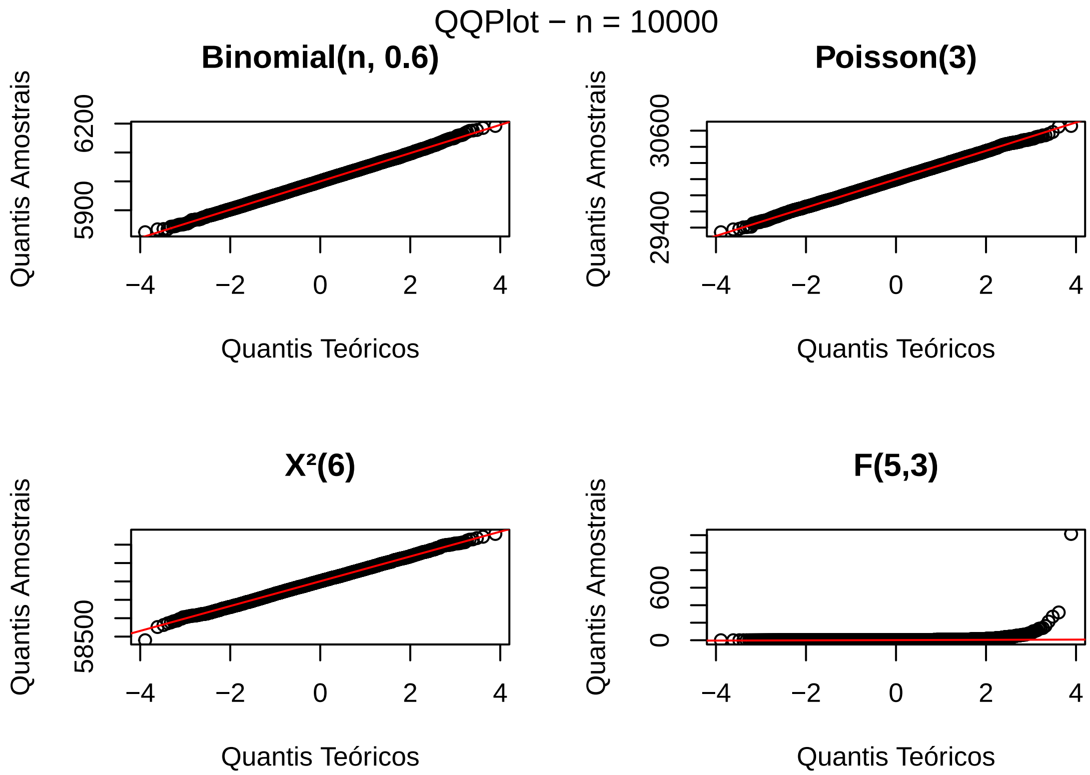



Poisson(3)
Histograma para n = 10
Histograma para n = 30
Histograma para n = 500
Histograma para n = 10000
Frequência
Valor


X²(6)
Histograma para n = 10
Histograma para n = 30
Histograma para n = 500
Histograma para n = 10000
Frequência
Valor



Com o objetivo de verificar se as variáveis geradas acima convergem para uma normal, utilizaremos a função qqnorm (que produz uma normal dos valores atribuídos à ela) e qqline (que adiciona uma linha dos valores teóricos da normal), de tal forma que apenas será empregue o valor de n = 10000.



De acordo com a visualização dos gráficos, é perceptível uma semelhança no comportamento das variáveis aleatórias das distribuições Poisson, Binomial e X^2, enquanto a distribuição F apresenta comportamento diferente, sendo a única que não aparente interceptar totalmente a linha dos valores teóricos (vermelha).

Ao fim, para complementar a visualização, um teste de ajuste (teste de Anderson-Darling) será feito para termos valores numéricos que serão utilizados para confirmação dos resultados:

```
## $Dist_Binomial
##
##  Anderson-Darling normality test
##
## data:  amostra_binomial
## A = 0.30885, p-value = 0.5586
##
##
## $Dist_Poisson
##
##  Anderson-Darling normality test
##
## data:  amostra_poisson
## A = 0.20212, p-value = 0.8795
##
##
```

```
## $Dist_Chisq
##
##  Anderson-Darling normality test
##
## data:  amostra_chisq
## A = 0.61764, p-value = 0.108
##
##
## $Dist_F
##
##  Anderson-Darling normality test
##
## data:  amostra_f
## A = 2660.2, p-value < 2.2e-16
```

Após os testes, é possível observar que o único p-valor que se apresentou como significativamente baixo foi no teste feito para a distribuição F, demonstrando, estatisticamente, que essa distribuição não converge para uma normal. Os demais testes apresentaram p-valores significativamente altos, o que comprova que essas distribuições convergem para uma normal. Os resultados dos testes evidenciam, assim, semelhança com o que foi visto nos gráficos.

Dessa forma, levando em consideração apenas o que é dito pelo Teorema Central do Limite e o que foi apresentado nos gráficos e nos testes, estes são os resultados de cada distribuição e suas aproximações da normal, caso convergissem:

- Para a Binomial(10000, 0.6), a distribuição converge para uma normal. De forma teórica, os valores dos parâmetros deveriam ser $\mu = np = 10000 * 0.6 = 6000$ e $\sigma^2 = np(1-p) = 6000*0.4 = 2400$. Na simulação, a distribuição expressou proximidade com a teórica, convergindo para uma normal com parâmetros:

```
## [1] "média = 5999.9497 e variância = 2366.00637054705"
```

- Para a Poisson(3), a distribuição converge para uma normal. De forma teórica, os valores dos parâmetros deveriam ser $\mu = n\lambda = 10000 * 3 = 30000$ e $\sigma^2 = n\lambda = 30000$. Na simulação, a distribuição expressou proximidade com a teórica, convergindo para uma normal com parâmetros:

```
## [1] "média = 29999.5636 e variância =  30134.8816432043"
```

- Para a X²(6), a distribuição converge para uma normal. De forma teórica, os valores dos parâmetros deveriam ser $\mu = nk = 10000 * 6 = 60000$ e $\sigma^2 = 2nk = 120000$. Na simulação, a distribuição expressou proximidade com a teórica, convergindo para uma normal com parâmetros:

```
## [1] "média = 60002.4245619013 e variância =  118003.827413903"
```

- Para a F(5,3), a distribuição não converge para uma normal.