

EP MAE0699 - Tópicos de probabilidades

Fabricio Kassardjian nusp:2234961
Robert Mota dos Santos nusp:9039927

15 de maio de 2019

Introdução

O trabalho se refere a estudar as curvas de distribuição para $T(v)$ e $C(v, w)$, onde \mathbf{T} representa o caminho mais curto de retorno ao vértice v e \mathbf{C} o caminho mais curto entre os vértices v e w . O modelo de Erdo-Rényi é utilizado para gerar os grafos aleatórios com n vértices e probabilidade p de ligação entre cada par de vértices.

Modelo 2

O modelo escolhido para o teste foi usando conexões não direcionadas e preguiçoso. O fato de ser preguiçoso implica que existem conexões para continuar no mesmo vértice. Além disso fica determinado que cada conexão só pode ser usada uma unica vez para cada caminho testado, assim evita-se que a distribuição \mathbf{T} tenha apenas valores 1 e 2. Como as conexões são aleatórias podem existir vértices isolados e também como não pode ser utilizado a mesma conexão para voltar podem existir valores de \mathbf{T} e \mathbf{C} que podemos considerar inf.

Metodologia

Estimação das distribuições

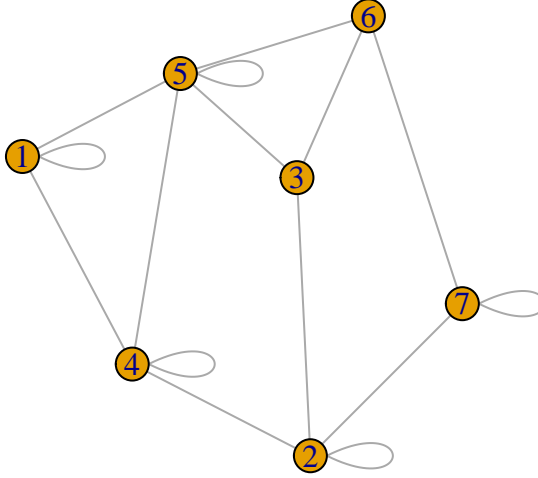
Para o teste primeiro inicia-se uma matriz A representando com **TRUE** quando a ligação entre os vértices está presente e **FALSE** quando não há ligação. A linha i da matriz representa o vértice de saída e a coluna j representa o vértice de chegada. Como o modelo é preguiçoso pode existir **TRUE** na diagonal principal da matriz, e pelo fato das conexões não serem direcionadas a matriz é simétrica.

Exemplo de matriz de conexões para 5 vértices com probabilidade de conexão 0.4:

```
n = 7
A = generateMatrix(n, 0.4)
print(1*A) #1* para deixar em formato numerico

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]  1   0   0   1   1   0   0
## [2,]  0   1   1   1   0   0   1
## [3,]  0   1   0   0   1   1   0
## [4,]  1   1   0   1   1   0   0
## [5,]  1   0   1   1   1   1   0
## [6,]  0   0   1   0   1   0   1
## [7,]  0   1   0   0   0   1   1

plot(graph_from_adjacency_matrix(A, mode = 'undirected', weighted = TRUE))
```



Para cada matriz gerada é testado para cada vértice o menor caminho de volta, usando uma busca em profundidade dos caminhos possíveis da matriz e armazenado o vetor com a contagem de cada valor para \mathbf{T} encontrado. O mesmo é feito para cada combinação de vértices possíveis para encontrar os valores de \mathbf{C} . Os caminhos podem ter tamanhos até n e iremos considerar o valor $n + 1$ como sendo infinito.

Exemplo de valores de $T(v)$ para cada vértice de \mathbf{A}

```
for(i in 1:n) {
  Ti = findPath(i,i,A,0,n+1)
  cat(sprintf("T(%d) = %d\n",i, Ti))
}
```

```
## T(1) = 1
## T(2) = 1
## T(3) = 3
## T(4) = 1
## T(5) = 1
## T(6) = 3
## T(7) = 1
```

Será gerado para cada tamanho de $n \in \{6, 7, 8, 9, 10, 11, 12\}$ uma amostra de 200 matrizes e feita uma contagem para cada valor de \mathbf{T} encontrado. A distribuição é estimada tirando a média da contagem por $n * 200$. Assim:

$$\hat{P}(T = k) = \frac{1}{n * 200} \sum_{i=1}^{n*200} \mathbb{1}_{(T=k)}$$

Para a distribuição de \mathbf{C} é usado processo similar mas como temos as combinações entre os pares serão

estimados $n * n$ valores para cada matriz, assim:

$$\hat{P}(C = k) = \frac{1}{n^2 * 200} \sum_{i=1}^{n^2 * 200} \mathbb{1}_{(C=k)}$$

Tamanho da amostra

Para determinar um tamanho bom de amostra para a aproximação da estimação, fixamos $n = 8$ e geramos a distribuição e o gráfico para alguns tamanhos de amostra ($N \in 25, 50, 75, 100, 150, 200, 250, \dots, 750, 800$). Depois calculamos a soma das diferenças ao quadrado entre os valores de cada distribuição e colocamos em um gráfico. No gráfico pode ser verificado se houve convergência e com que tamanho de amostra podemos considerar a convergência.

$$erro = \sum_{i=1}^{n+1} (P(T = k) - P'(T = k))^2$$

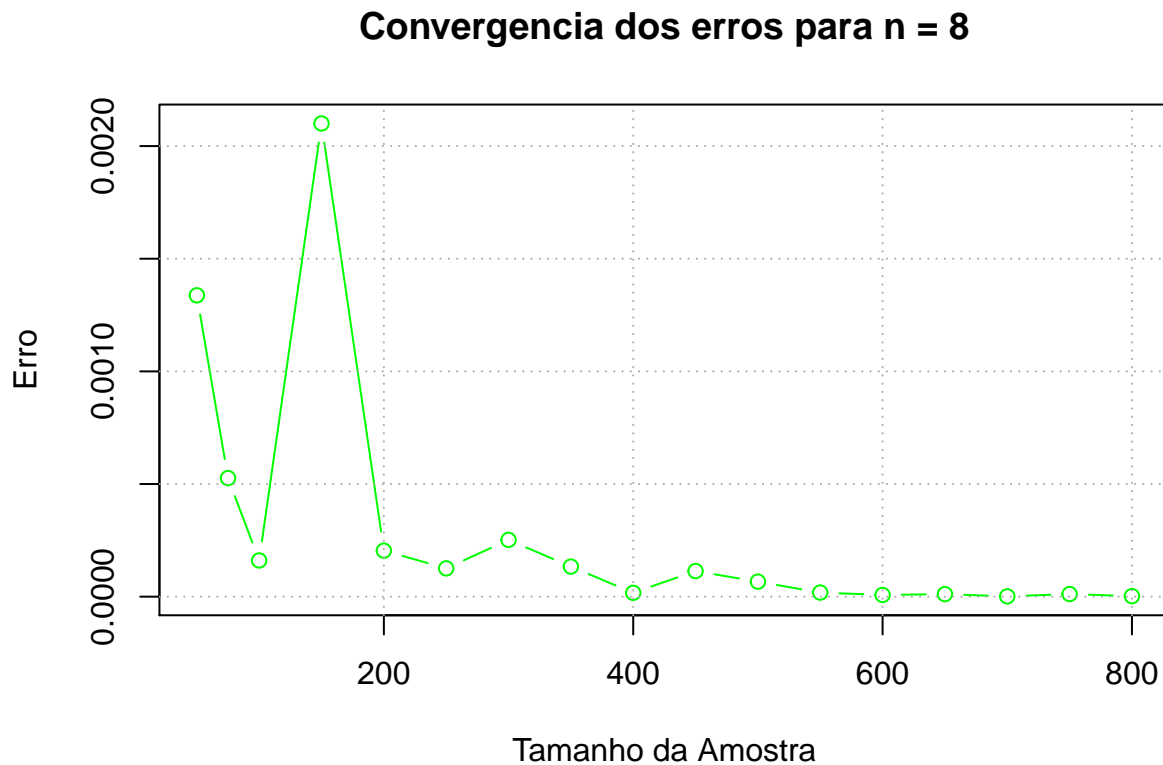
onde $P(T = k)$ é a probabilidade para o tamanho de amostra atual e $P'(T = k)$ a probabilidade da amostra anterior.

Teste de aderência

Simulação

Tamanho da amostra

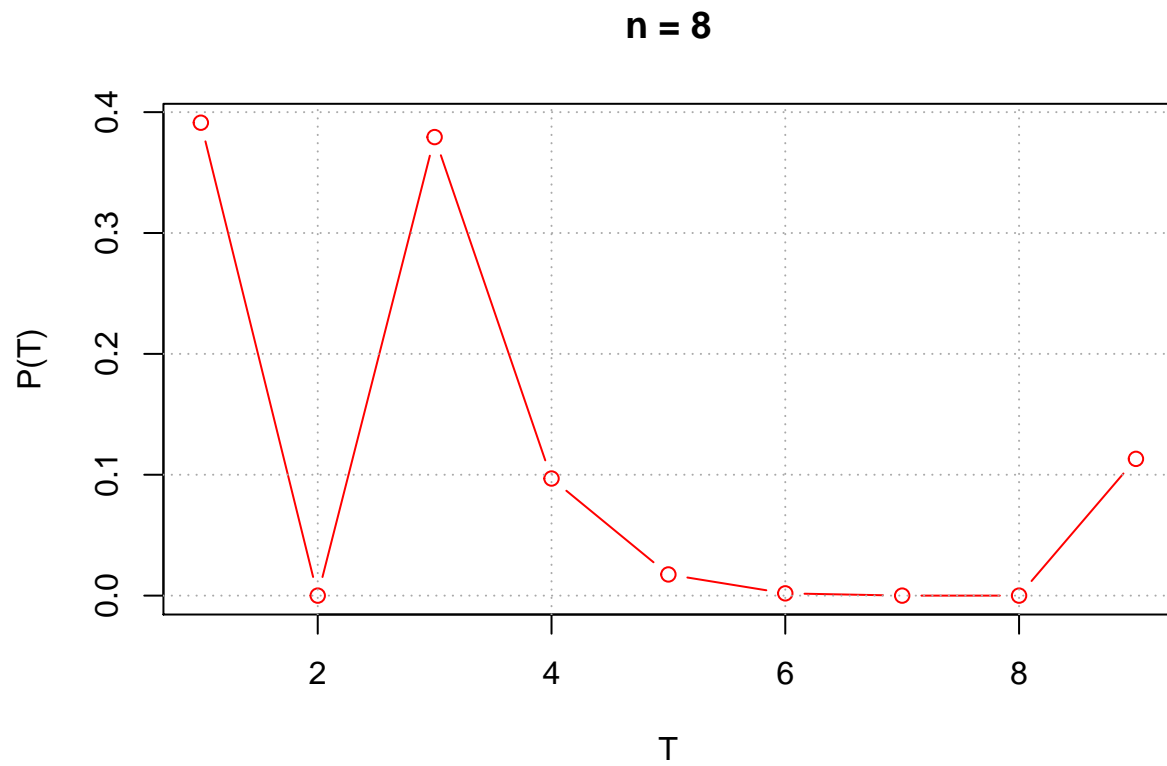
Gráfico para vários tamanhos de amostra com $p = 0.4$:

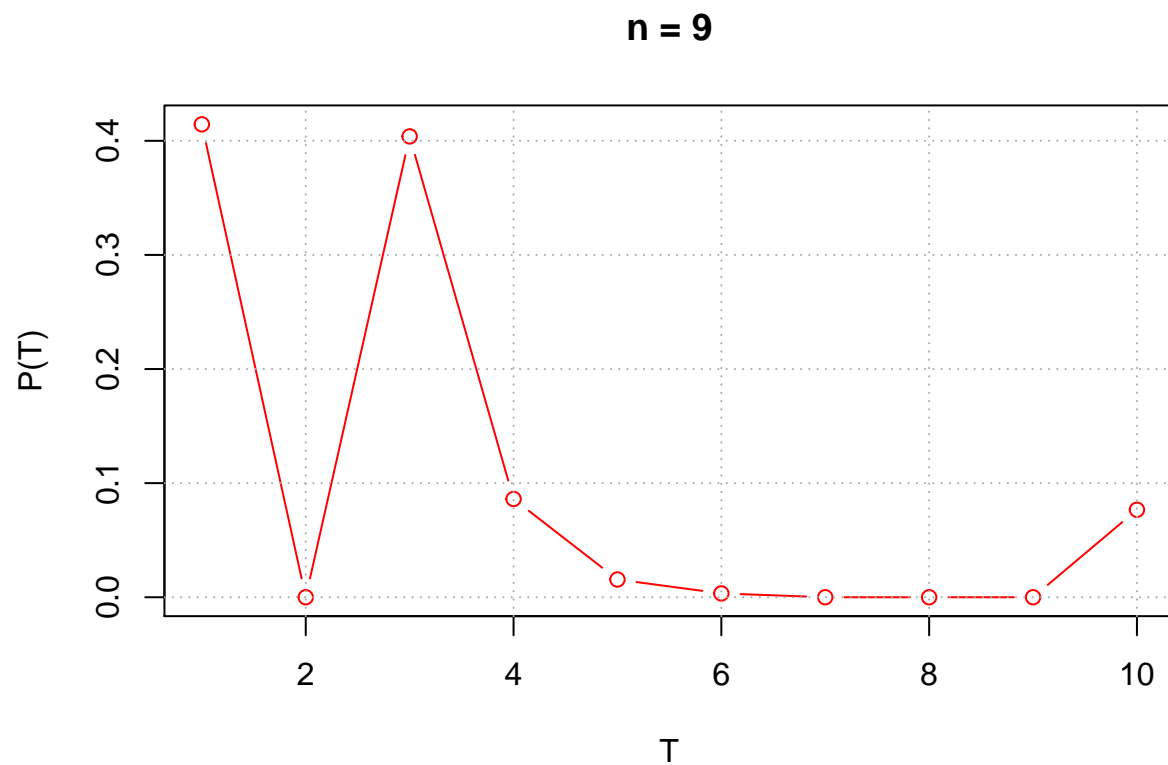


Pelo gráfico podemos considerar uma amostra com tamanho 200 razoável para as estimações de distribuição.

Distribuições para T

Valores estimados e gráficos da distribuição, usando 200 amostras, e $p = 0.4$



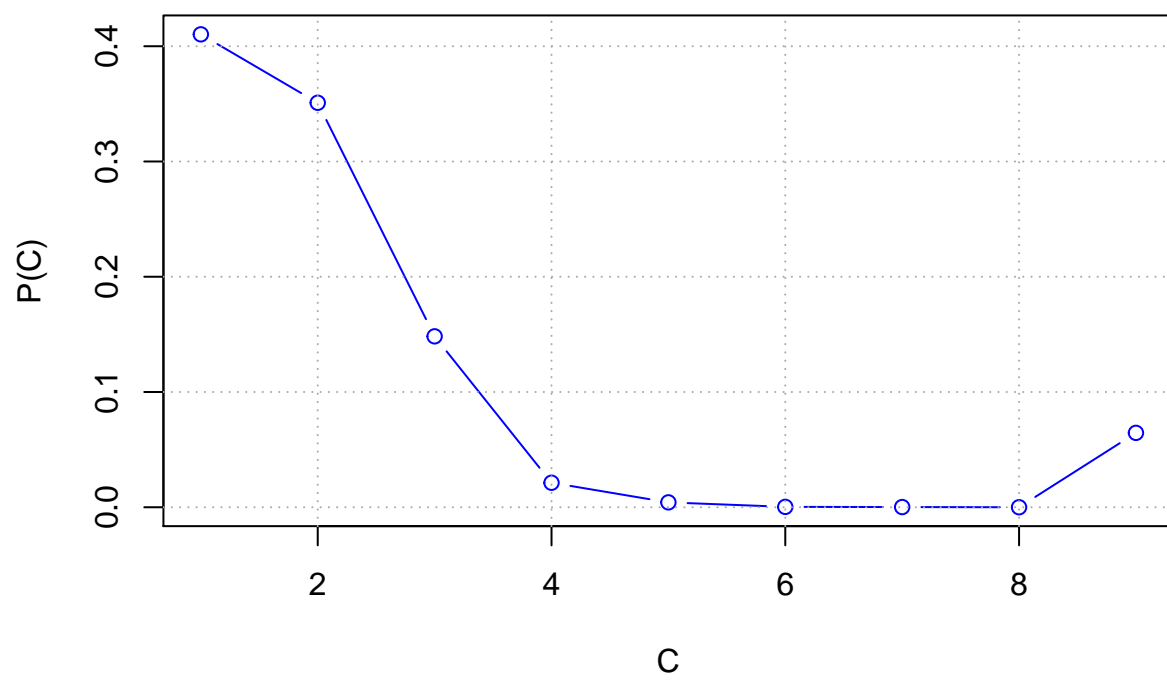


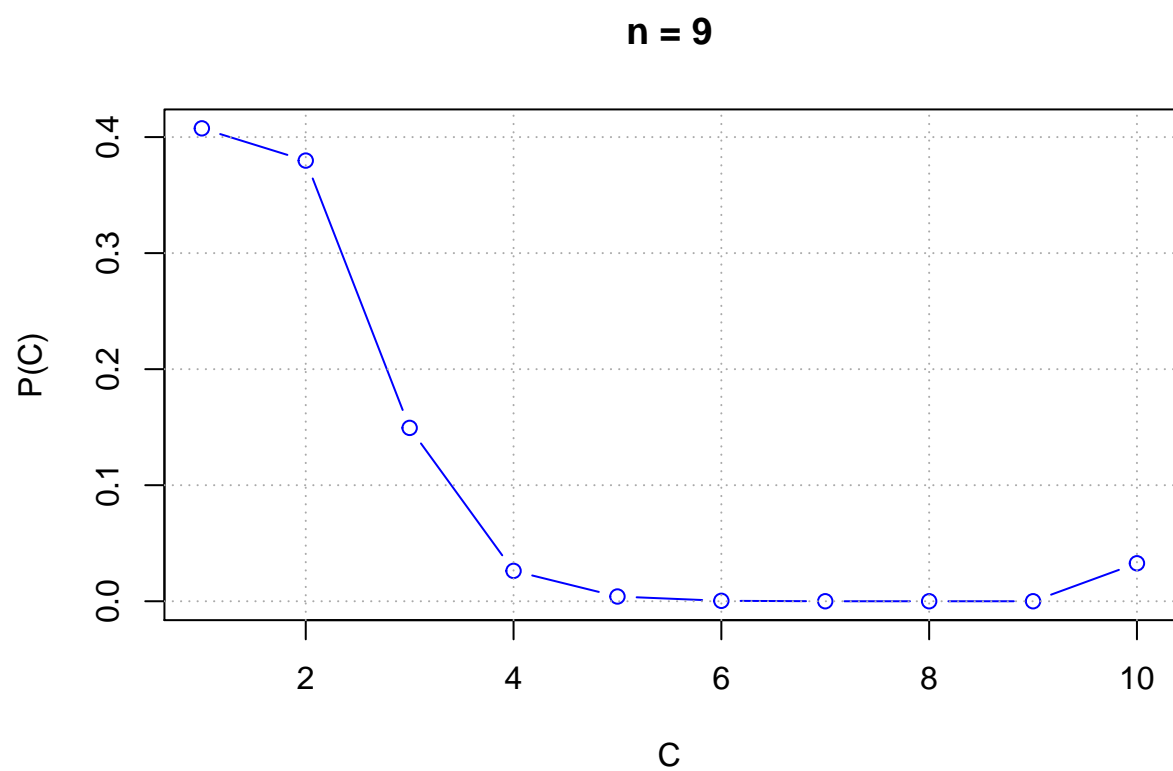
Teste de aderência para T

Distribuições para C

Valores estimados e gráficos da distribuição. Como a quantidade de testes aumenta em n vezes para a distribuição C, reduzimos a quantidade de matrizes geradas pela metade, e portanto usamos uma amostra de tamanho 100 com $p = 0.4$.

n = 8





Teste de aderência para C