Disciplina: Estatística Computacional e Otimização

Prof.: Luiz Duczmal

Lista: 1

Data de entrega: 06/05/24

Estudante: Pedro Mateus Moraes de Almeida

Matrícula: 2023668250

(1) Faça a simulação de dados do jogo WAR, com n lançamentos dos 6 dados vermelhos e amarelos, e estime as probabilidades p1, p2, p3 e p4 de cada um dos quatro possíveis resultados, conforme visto em aula. Repita ainda m vezes cada conjunto de n simulações de modo a construir empiricamente intervalos de 90% de confiança para cada um dos parâmetros p1, p2, p3 e p4. Compare os intervalos de confiança assim obtidos empiricamente com os intervalos de confiança obtidos teoricamente.

p1 (Vermelho ganha): 22.3% a 26.8%
p2 (Amarelo ganha): 22.3% a 27.0%
p3 (Empate): 48.3% a 53.6%
p4 (Todos os dados iguais): 0.0% a 0.0% (consistente com a raridade do evento)

O p1 e p2 mostram uma variação relativamente pequena, sugerindo que as estimativas são razoavelmente precisas e estáveis ao longo das repetições.

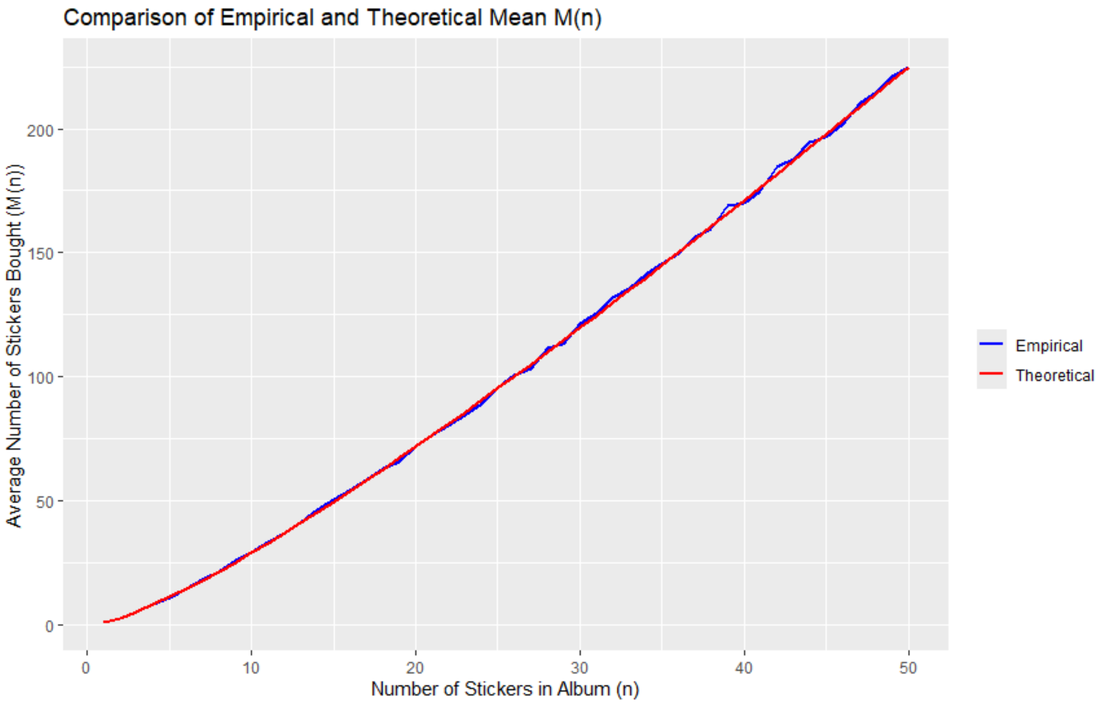
As simulações mostram que a probabilidade de empate (p3) é significativamente maior do que a de qualquer lado ganhar individualmente, o que pode ser explicado pela alta probabilidade de ambos os lados terem pelo menos um dado com o mesmo valor máximo em várias rodadas.

A ausência de qualquer evento onde todos os dados são iguais (p4 = 0.0%) é consistente um evento raro de acontece, considerando o grande número de combinações possíveis com 6 dados de 6 faces cada.

(2) Faça uma simulação do preenchimento de um album de n figurinhas, conforme visto em aula. Obtenha uma caracterização empírica da variável aleatória M(n) que indica o número de figurinhas compradas para se conseguir preencher um álbum de n figurinhas. Através de simulações para vários valores de n, descubra um valor aproximado para a média de M(n) em função de n, e compare com o valor teórico deduzido em sala.

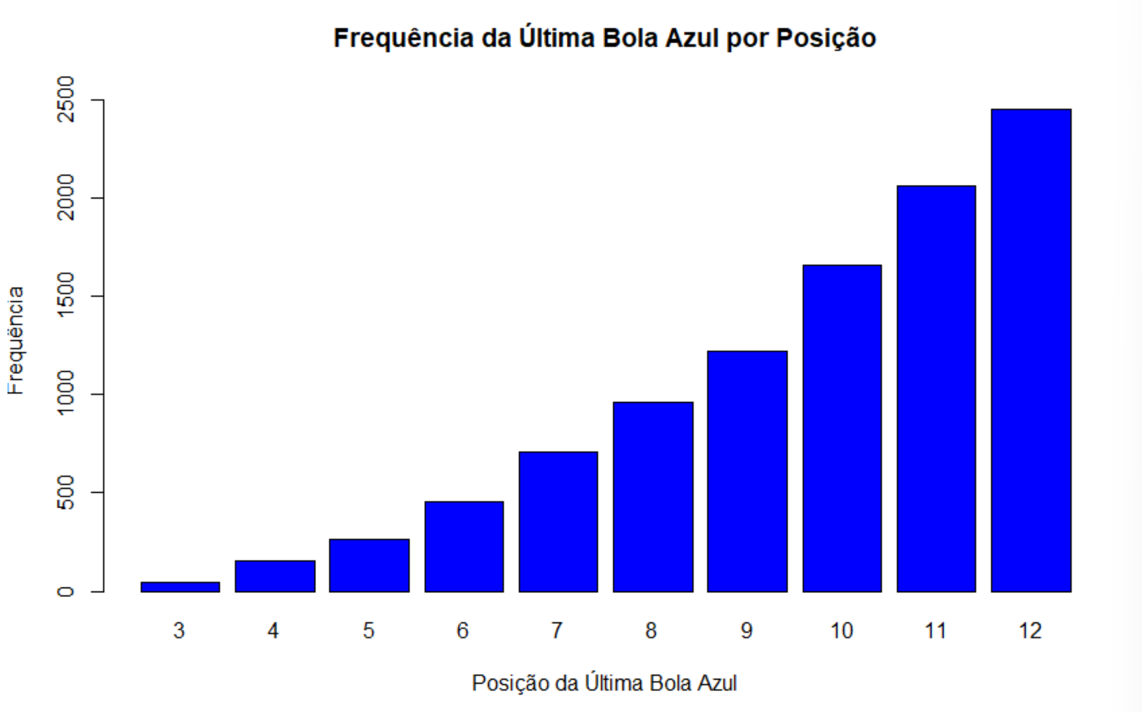
O primeiro histograma (obtido na sala de aula) mostra a distribuição das simulações para n=50n=50. O pico da distribuição está em torno de 225, que está próximo da média calculada de 225,1855, confirmando que a média das simulações é representativa da distribuição.

Quando utilizo como comparação os resultados desenvolvido, obtive 224.917 como resultado das simulações desenvolvidas, valor muito próximo do que se espera teoricamente. Abaixo podemos visualizar graficamente a linha do teórica que é constante as simulações possuem pequenas distorções.



Portanto, os resultados empíricos estão em ótima concordância com os valores teóricos, indicando que o modelo teórico do número harmônico é uma boa aproximação para este problema.

(3) Faça um código em R para simular a retirada aleatória, sem reposição, de uma caixa com 3 bolas azuis, 4 bolas brancas, e 5 bolas cinzas, e conte quantas vezes a última bola azul a ser retirada é a i-ésima bola, para i=1,2,...,12.



Posições mais próximas ao início (3, 4, 5) têm frequências significativamente menores, indicando que é menos provável que a última bola azul seja retirada tão cedo no processo.

A frequência da última bola azul aumenta conforme se aproxima do final das 12 posições. Isso é esperado, pois quanto mais bolas são retiradas sem encontrar todas as azuis, maior a chance de que uma das poucas bolas restantes seja azul.

A probabilidade de a última bola azul ser uma das últimas bolas a ser retirada aumenta progressivamente (maior frequência da última bola azul é a 12ª, com 2.454 ocorrências.). Isso reflete a lógica de que, à medida que o número de bolas diminui, as chances de retirar uma bola azul específica (quando apenas uma resta) tornam-se maiores. Esse é um comportamento esperado para situações onde não é feita a reposição.

(4) Faça um código em R para simular o problema das entrevistas, conforme visto em aula. Através de simulações para vários valores de n (n=10, n=20, n=30, e n=40), descubra uma expressão aproximada para o número ótimo de candidatos que são descartados inicialmente que maximiza a probabilidade de se escolher o melhor candidato. Encontre uma expressão aproximada para este número em função de n.

Fazendo uma análise do resultado do "problema do secretário" podemos tirar algumas conclusões, mas primeiro vamos aos resultados:

- Para n=10n=10, k ótimo é 3 com uma probabilidade máxima de 0.408.
- Para n=20n=20, k ótimo é 6 com uma probabilidade máxima de 0.399.
- Para n=30n=30, k ótimo é 11 com uma probabilidade máxima de 0.406.
- Para n=40n=40, k ótimo é 16 com uma probabilidade máxima de 0.383.

Como podemos notar após as simulações k cresce quase de forma linear e que os resultados apesar de serem diferentes, são muito próximos. Dessa forma podemos criar uma função n abaixo:

Kotimo =[$f(n) \approx 0{,}399*n$]

É uma função aproximada tendo em vista os resultados apresentados.