Problema do colecionador de cupons

O **problema do coletor de cupons** está no campo de estudo da probabilidade. Ele descreve a seguinte situação: uma empresa coloca n cupons diferentes, de forma aleatória, nas caixas de sua linha de cereais. Toda caixa recebe apenas um cupom, dos n tipos de cupons disponíveis. A pergunta que o problema faz é: Qual é a probabilidade de que mais de *t* caixas precisam ser compradas, para que o colecionador consiga coletar todos os *n* cupons? A análise matemática deste problema mostra que o valor de tentativas necessárias esperado cresce na ordem de n log(n), e vamos mostrar o porquê.

Primeiro, é estabelecida uma variável aleatória $W_{n,k}$, que representa o número de caixas que o colecionador precisa comprar até conseguir completar sua coleção com todos os *n* cupons. Seja Z_i a quantidade de caixas de cereal necessárias para $i = 1, 2, ..., n$, até que o número de cupons distintos pelo colecionador aumente de i − 1 para i.

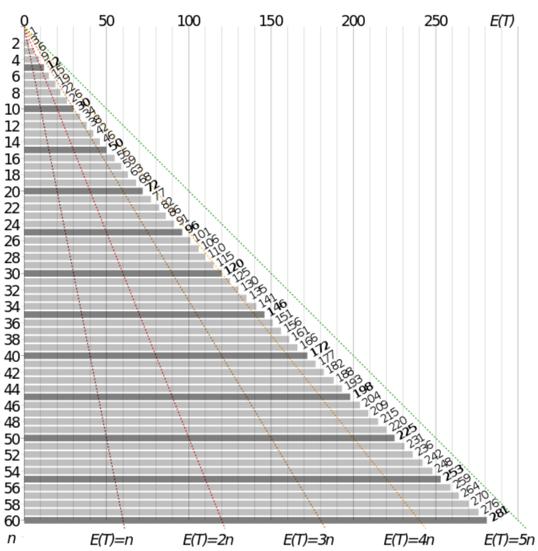
Assim, quando o colecionador tem i−1 cupons diferentes, a probabilidade p_i de que um cupom obtido de forma aleatória seja um dos que ele ainda não tem, será igual a 1 − (i − 1)/n. Dessa forma, cada Z_i é uma variável geométrica aleatória com probabilidade de sucesso p_i. Logo, a esperança de Z_i é E[Z_i] = n/(n−i+1).

E como $W_{n,k} = Z_1 + Z_2 + ... + Z_k$, obtemos a seguinte solução: $E[W_{n,k}] = \sum_{i=1}^{k} \frac{n}{n-i+1}$

Para k=n, temos: $E[W_{n,n}] = \sum_{i=1}^{k} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i}$

Sendo o número harmônico $H(n) = \sum_{k=1}^{n} \frac{1}{k} = log\, n + c$, (sendo c uma constante) substituindo na equação acima, obtemos enfim a solução final para o problema:

$$E[W_{n,n}] = n\, log\, n + nc = n\, log\, n + \theta(n)$$



De acordo com a hierarquia de funções, temos que n log n ≻ nc, logo o comportamento que predomina na função é n log n, e portanto ela tem custo assintótico **Θ(n log n).** Ao lado, temos um gráfico do número de cupons n, versus o número esperado de tentativa necessárias para coletar todos eles, E (T).

Observação: apesar de muita pesquisa, não consegui encontrar outras soluções popularizadas do problema proposto.