Distribuição discreta de Bernoulli, Geométrica e Binomial

Ensaio de Bernoulli e Distribuição de Bernoulli:

Um evento é classificado como um **Ensaio de Bernoulli** é quando é avaliado apenas como sucesso ou fracasso, onde sucesso ou fracasso pode ser definido à vontade. Por exemplo, num lançamento de uma moeda, Cara pode ser considerado sucesso e Coroa Fracasso. Outro exemplo, num lançamento de um dado, os valores { 3 e 4 } podem ser considerados sucesso e os demais fracasso. Finalmente, se uma pessoa que entra por uma porta tem mais do que 1,70m de altura pode ser considerado sucesso e os demais valores, fracasso.

Usando-se a propriedade dos ensaios de Bernoulli, a **distribuição de Bernoulli** é definida como

$$P(X = \text{Sucesso}) = 1 - P(X = \text{Fracasso}) \text{ ou}$$

$$P(X = \text{Fracasso}) = 1 - P(X = \text{Sucesso})$$

A explicação é simples. Como os únicos resultados possíveis para um evento de Bernoulli são Sucesso ou Fracasso, então, a resposta a pergunta, "Qual a probabilidade de se ter um sucesso ou um fracasso num ensaio", sempre será 1 (100%). Assim, considerando-se que "p" é a probabilidade de sucesso e q a probabilidade de fracasso, temos:

$$P(X=\text{Sucesso ou Fracasso}) = p + q = 1$$

ou seja,

$$p + q = 1$$

Logo, ($q = 1 - p$) e ($p = 1 - q$). E portanto, pode-se escrever:

$$P(X=\text{Sucesso}) = p = 1 - q$$

e

$$P(X = \text{Fracasso}) = q = 1 - p$$

Pode-se representar estas relações de forma sucinta através da expressão:

$$P(X=k) = p^k \times (1-p)^{(1-k)} \qquad \textbf{Distribuição de Bernoulli}$$

onde p é a probabilidade de sucesso e K=1 em caso de sucesso e K=0 em caso de fracasso.

Problema 1:

Qual a probabilidade de se conseguir tirar 6 num lance de dados apenas num terceiro lançamento?

Antes de mais nada, é importante evidenciar coisas óbvias mas importantes:

1) foram realizados três lançamentos

2) sendo que apenas no terceiro houve sucesso, e que portanto

3) se tirou valores diferentes de 6 nos dois primeiros lançamentos.

Vamos chamar de Sucesso tirar 6 e de Fracasso, não tirar 6. Sistematizando os dados do problema para se aplicar os conceitos de probabilidade:

Espaço amostral: { 1, 2, 3, 4, 5, 6 }

Cardinalidade (n) do espaço amostral: n(EA) = 6

Evento de sucesso (S): { 6 }

Cardinalidade (n) do evento de sucesso: n(S) = 1

Evento de fracasso(F): { 1, 2, 3, 4, 5 }

Cardinalidade (n) do evento de fracasso: n(F) = 5

Obviamente n(F) tem de ser 5 porque o fracasso é complementar ao sucesso e portanto n(F) = n(EA) - n(S).

Nota: O evento de sucesso poderia ser definido diferentemente no exemplo anterior, como por exemplo: sucesso poderia ser quando o valor do lançamento for múltiplo de 3. Neste caso, o evento de sucesso teria cardinalidade 2 (n ({3, 6}) e o evento de fracasso teria cardinalidade 4, dado por n(F) = n(EA) - n(S) = 6 - 2 = 4, mas cuja cardinalidade também pode ser calculada, neste caso onde existem poucos elementos, apenas contando os elementos do subconjunto correspondente ao evento de Fracasso = { 1, 2, 4, 5 }.

Continuando na solução do Problema 1:

Como cada lançamento não afeta em nada o resultado do outro lançamento, então os eventos de cada lançamento são independentes entre si. Assim, a probabilidade de só se ter um sucesso no terceiro lançamento (lançamento 3) pode ser descrita como:

$$P(\text{Problema1}) = P(F) \times P(F) \times P(S) = P(F)^2 \times P(S)$$

Agora, se alterássemos o Problema 1 para que apenas no quinto lançamento se obtivesse um número 6, a probabilidade disto acontecer neste novo problema (Problema 2) seria de:

$$P(\text{Problema2}) = P(F) \times P(F) \times P(F) \times P(F) \times P(S) = P(F)^4 \times P(S)$$

Antes de prosseguir a leitura, tenha certeza de que entendeu a razão porque a expressão acima é verdadeira.

Hummm.... Note que o expoente em P(F) representa o número de fracassos que precede ao sucesso (e tem de ser assim, porque o expoente E é apenas uma forma compacta de dizer que houve E fracassos prévios que, como eventos independentes, devem ser multiplicados), e este número é sempre 1 a menos que o "momento" (ordem do lançamento) em que se obteve sucesso, digamos, o momento N.

Distribuição Geométrica

Assim, pode-se generalizar a resposta para um problema onde um evento de sucesso só ocorre na N-ésima tentativa (sendo precedido por n-1 fracassos), criando-se a seguinte FÓRMULA que SUMARIZA o raciocínio anterior:

$$P(X = n) = P(F)^{(n-1)} \times P(S) \qquad \textbf{Distribuição Geométrica}$$

Sendo que P(S) é a probabilidade de sucesso e P(F) a probabilidade de fracasso.

Como os eventos são Ensaios de Bernoulli, ou seja, ou representam sucessos ou fracassos, temos:

$$P(EA) = P(S) + P(F) \quad <=> \quad 1 = P(S) + P(F)$$

Porque a UNIÃO do evento de sucesso com o evento de fracasso é todo o espaço amostral.

Assim, a probabilidade de fracasso é dada por:

$$P(F) = P(EA) - P(S) \;<=>\; P(F) = 1 - P(S).$$

E, pela mesma razão, a probabilidade de sucesso também é dada por:

$$P(S) = P(EA) - P(F) \quad <=> \quad P(S) = 1 - P(F)$$

Modelo Binomial

Vamos revisitar o Problema 1 . Para simplificar a escrita em vez de escrever P(F) e P(S) escreveremos apenas F e S para representar, respectivamente, a probabilidade de fracasso e de sucesso.

Assim qualquer solução para o Problema 1 (sucesso apenas no N-ésimo evento) tem sempre a forma:

$$P(X = n) = F \times F \times F \times F \ldots \times F \times S = FFFF\ldots FS = F^{(n-1)} S$$

onde o número de efes (F) que precedem S é sempre N-1.

Vamos agora criar um novo problema, afrouxando um dos requerimentos do problema 1. Porém antes tente responder a seguinte questão:

Informalmente, as chances de se ter um único sucesso em três lançamentos são maiores, iguais ou menores do que se ter um sucesso apenas no último lançamento?

Problema 3:

Qual a probabilidade de se conseguir tirar 6 num lance de dados apenas uma vez em três lançamentos?

De novo evidenciando coisas óbvias mas importantes:

1) foram realizados três lançamentos

2) sendo que houve apenas um sucesso, e que portanto

3) se tirou valores diferentes de 6 em dois lançamentos.

Qual a diferença deste problema em relação ao problema 1?

A única diferença é que o lançamento de sucesso não precisa mais ter sido conseguido no último lançamento, ou seja, a POSIÇÃO do sucesso pode ocorrer em qualquer dos três lançamentos .

Assim, se a posição de sucesso fosse a última posição, teríamos uma solução representada pelas letras abaixo e, ao lado são exibidos os subconjuntos do espaço amostral que contém os valores possíveis para os resultados dos lançamentos correspondentes a cada uma das posições.

F F S ~~~~ {1,2,3,4,5} { 1,2,3,4,5} {6} **Distribuição Geométrica**

Isso significa que os resultados de três lançamentos poderiam ser (1, 1, 6), (1, 2, 6) ... (4, 4, 6) ... e por aí vai. Qualquer organização em que as duas primeiras posições tivessem valores do subconjunto {1,2,3,4,5} (evento de Fracasso) e que se tenha 6 na última posição é uma sequência possível de resultados de lançamentos.

Agora, se o sucesso estivesse na primeira posição teríamos as letras que estão associadas a possíveis resultados de lançamentos:

S F F ~~~~ {6} {1,2,3,4,5} { 1,2,3,4,5}

Já se o sucesso ocupasse a posição intermediária teríamos:

F S F ~~~~ {1,2,3,4,5} {6} { 1,2,3,4,5}

Então qual será a probabilidade associada ao problema 3?

Pense:

1) As soluções FFS, FSF e SFF tem o mesmo valor de probabilidade?

2) As soluções FFS, FSF e SFF podem ocorrer ao mesmo tempo? Se combinam ao mesmo tempo ou são soluções alternativas?

Respondendo:

1) Como a ordem dos fatores não altera o resultado, todos os três produtos, FFS, FSF e SFF são iguais! E como são iguais, pode-se organizá-los numa ordem comum, por exemplo, colocando sempre o fracasso no início. Logo, cada solução parcial para o problema 3 sempre é igual a FFS

2) E como cada solução parcial é alternativa, a solução do problema 3 é o resultado da UNIÃO de todas as soluções alternativas, o que, em termos de probabilidade, leva a SOMA das probabilidades.

.Assim, a resposta para o problema 3, com 3 lançamentos e 1 sucesso é dada pela expressão abaixo:

$$P(3,1) = FFS + FSF + SFF = FFS + FFS + FFS = 3 \times FFS = 3 \times F^{(3-1)} \times S$$

$$P(3,1) = 3 \times F^{(3-1)} \times S$$

Problema 4:

Qual a probabilidade de se conseguir tirar 6 num lance de dados apenas uma vez em CINCO lançamentos?

A lógica da solução para 1 sucesso em 5 lançamentos é a mesma utilizada no problema 3, e portanto podemos a solução é dada por:

$$P(5,1) = FFFFS + FFFSF + FFSFF + FSFFF + SFFFF =$$

$$FFFFS + FFFFS + FFFFS + FFFFS + FFFFS =$$

$$5 \times FFFFS = 5 \times F^{(5-1)} \times S$$

$$P(5,1) = 5 \times F^{(5-1)} \times S$$

Pensando um pouco, podemos ver que, quando ocorre 1 sucesso em 5 lançamentos, mandatoriamente, por serem ensaios de Bernoulli, tem de ocorrer 4 fracassos, que são representados no resultado pelo expoente (5-1), ou seja, NF = 5 - NS. Isso ocorre porque, como cada lançamento é um ensaio de Bernoulli, o total de lançamentos é a soma de todos os sucessos e fracassos, ou seja,

N = NS + NF ~~~~ (No Problema 4: 5 = 1 + 4)

O que leva, por uma simples manipulação algébrica, às óbvias expressões:

NF = N - NS e NS = N - NF

As duas expressões são relevantes porque às vezes um problema informa dois dos três elementos das expressões, e deve-se calcular o elemento omitido. Por exemplo, um problema pode informar o número de ensaios (lançamentos) e o número de fracassos. No problema 4, foi informado o número de ensaios (lançamentos = 5) e o número de sucessos (1), o que leva a expressão abaixo para o cálculo do número de fracassos.

NF = 5 - 1 = 4

Com isso em mente, vamos expressar a solução do problema 5 utilizando expoentes para os 2 fatores.

$$P(5,1) = 5 \times F^{(5-1)} \times S^{(1)}$$

Neste ponto, pode-se fazer uma pequena generalização. Note que o expoente do fator Fracasso (F) sempre será exatamente igual a quantidade de fracassos porque cada possível sequência de lançamentos tem, por exigência do problema, um número bem definido de fracassos. O mesmo ocorre com o expoente do fator Sucesso (S). Assim, já é possível generalizar os expoentes dos fatores F e S da expressão como:

$$P(5,1) = 5 \times F^{NF} \times S^{NS}$$

onde, como definido anteriormente, NF é o número de Fracassos e NS é o número de Sucessos.

Problema 5:

Qual a probabilidade de se conseguir tirar 6 num lance de dados QUATRO vezes em CINCO lançamentos?

P(5,4) = ???

Agora ao invés de 1, queremos 4 sucessos em 5 lançamentos. Antes de mais nada, calcula-se o número de fracassos usando a expressão NF = N - NS, ou seja, NF = 5 - 4 = 1. Usando a mesma lógica dos Problemas 3 e 4, teríamos:

$$P(5,4) = FSSSS + SFSSS + SSFSS + SSSFS + SSSSF =$$

$$FSSSS + FSSSS + FSSSS + FSSSS + FSSSS =$$

$$5 \times FSSSS = 5 \times F \times SSSS = 5 \times F^{(5-4)} \times S^{(4)}$$

Resumidamente,

$$P(5,4) = 5 \times F^{(5-4)} \times S^{(4)}$$

Um padrão começa a aparecer com relação ao multiplicador e aos expoentes, não é mesmo? No entanto, como até agora os problemas referem-se apenas a 1 sucesso ou a 1 fracasso, ainda não é conveniente tentar uma generalização. É necessário criar problemas mais amplos, como o problema a seguir .

Problema 6:

Qual a probabilidade de se conseguir tirar 6 num lance de dados DUAS vezes em CINCO lançamentos?

P(5,2) = ???

Antes de mais nada, calcula-se o número de fracassos usando a expressão NF = N - NS, ou seja, NF = 5 - 2 = 3. Ao usar a mesma lógica dos Problemas 3 e 4, percebemos que agora existem 2 posições possíveis para o lançamento de sucesso ocorrer. Adaptando as expressões dos Problemas 3 e 4 a esta nova realidade temos:

$$P(5,2) =$$

$$FFFSS + FFSFS + FSFFS + SFFFS$$

$$+ FFSSF + FSFSF + SFFSF$$

+ FSSFF + SFSFF

SSFFF

Reordenando os fatores para FFFSS porque a ordem dos fatores não altera o resultado, temos:

P(5,2) =

FFFSS + FFFSS + FFFSS + FFFSS

+ FFFSS + FFFSS + FFFSS

+ FFFSS + FFFSS

FFFSS

Que é igual a:

$$P(5,2) = 10 \times FFFSS = 10 \times FFF \times SS = 10 \times F^{(5-2)} \times S^{(2)}$$

$$P(5,2) = 10 \times F^{(5-2)} \times S^{(2)}$$

Comparando a solução do problema 6 com a solução dos problemas anteriores, vemos que, enquanto os expoentes seguem um padrão claro do tipo

$$P(5,2) = 10 \times F^{NF} \times S^{NS}$$

onde, NF = N - NS, ou seja,

$$P(5,2) = 10 \times F^{N - NS} \times S^{NS}$$

O fator multiplicador não é simplesmente N, como nos problemas anteriores. Por que?

Bom, entendendo que este fator multiplicativo corresponde a QUANTAS VEZES o padrão composto por produtos de F e de S, FS aparece na solução, temos de entender a forma como estes produtos são produzidos, para poder usar algum MÉTODO DE CONTAGEM, como arranjos, permutações ou combinações, que evite ter de gerar exaustivamente todos os produtos para depois contá-los.

Analisando o Problema 6, podemos dizer que os produtos F e S representam **TODAS AS FORMAS POSSÍVEIS DE SE OBTER** 2 sucessos em 5 lançamentos. Dito de outra forma, de quantas formas diferentes podemos colocar 2 elementos indistintos entre si (S) em 5 posições (posições na sequência de lançamentos)? Ou ainda, quantas são as COMBINAÇÕES de 5 elementos tomados 2 a 2?

A resposta de todas as perguntas acima é justamente o conceito de combinação da Análise Combinatória e é representado por:

$$C(N,K) = A(N,K) / P(K)$$

onde A(X,Y) é o arranjo de X em Y e P(X) são permutações de X.

Assim, a solução do Problema 6 pode ser expressa de forma mais sucinta como:

$$P(5,2) = C(5,2) \times F^{(5-2)} \times S^{(2)}$$

Com base neste raciocínio, podemos enfim generalizar a solução e deduzir o modelo Binomial.

Fazendo N = n e NS = k, temos:

$$P(n, k) = C(n,k) \times F^{(n-k)} \times S^{k}$$

E lembrando que F = (1 - S) por que os eventos são ensaios de Bernoulli, podemos escrever:

$$P(n, k) = C(n,k) \times (1-S)^{(n-k)} \times S^{k}$$

Finalmente, se chamarmos S, que é a probabilidade de Sucesso, de "p", temos a FÓRMULA da **Distribuição Binomial** que SUMARIZA todo o raciocínio anterior:

$$P(n, k) = C(n,k) \times (1-p)^{(n-k)} \times p^{k}$$

Porém, a expressão acima anda não está no formato padrão, P(X=?). Fazendo este pequeno ajuste, e percebendo que n não aparece como parâmetro da fórmula, temos:

$$P(X = k) = C(n,k) \times (1-p)^{(n-k)} \times p^{k}$$ **Distribuição Binomial**

Distribuição Hipergeométrica

A Distribuição Hipergeométrica é utilizada para resolver problemas onde se espera saber a probabilidade de um evento ocorrer a partir de um subconjunto de um determinado espaço amostral E quando se sabe detalhes de como o espaço amostral E está organizado em partes.

Por exemplo, seja um espaço amostral "E" composto por um saco "N" (20) bolas divididas em 6 bolas vermelhas, 10 bolas verdes e 4 bolas azuis. Pode-se usar a distribuição hipergeométrica se quisermos saber, qual a probabilidade de se obter "k" (digamos 3) bolas de um determinado tipo (digamos bolas vermelhas) quando se retiram "n" (digamos 8) bolas do saco de uma vez, ou uma de cada vez mas sem recolocá-las no saco. Note que essas "n" bolas são um subconjunto do espaço amostral "E".

A boa maneira de "decorar" a fórmula da Distribuição Hipergeométrica é a seguinte:

1) Usando a distribuição binomial para lembrar de uma estrutura conhecida, tem-se:

$$P(X = k) = C(n,k) \times (1\text{-}p)^{(n\text{-}k)} \times p^{k}$$

2) Fixa-se apenas nos expoentes dos fatores:

K (N - K)

Sendo que no caso do exemplo,

N = 20 (6 bolas vermelhas + 10 verdes + 4 azuis) e

K = 6 (porque o problema quer saber sobre bolas vermelhas)

N - K = 14

3) Entendendo que K e (N-K) seriam expoentes para **todo o espaço amostral E**, é necessário **adaptá-los proporcionalmente para o subconjunto de E** no qual o evento deve ocorrer, correspondente à retirada simultânea de bolas (ou sequencial sem reposição) que serve de base para o evento de interesse.

Seja:

n = 8 número de bolas que foram retiradas do saco.

k = 3 número de bolas vermelhas das que foram retiradas do saco

Assim, a fórmula correspondente a Distribuição Hipergeométrica é:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Que no caso do exemplo é:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{6}{3}\binom{20-6}{8-3}}{\binom{20}{8}} = \frac{\binom{6}{3}\binom{14}{5}}{\binom{20}{8}}$$