Métodos Estocásticos da Engenharia I

Capítulo 3 - Principais Modelos Discretos

Prof. Magno Silvério Campos

2024/2

Bibliografia

Essas notas de aulas foram baseadas nas seguintes obras:

1. CAMPOS, M. A.; RÊGO, L. C.; MENDONÇA, A. F. *Métodos Probabilísticos e Estatísticos com Aplicações em Engenharias e Ciências Exatas*. Rio de Janeiro: LTC, 2017.
2. CANCHO, V.G. *Notas de Aulas sobre Noções de Estatística e Probabilidade*. São Paulo: USP, 2010.
3. NETO, P. L. C; CYMBALISTA, M. *Probabilidades*. São Paulo: Editora Blucher, 2006.
4. HINES, W.W.; *et al. Probabilidade e Estatística na Engenharia*. 4. ed. Rio de Janeiro: LTC, 2006.
5. MENDES, F. C. T. *Probabilidade para Engenharias*. Rio de Janeiro: LTC, 2010.
6. MONTGOMERY, D.C.; RUNGER, G.C. *Estatística Aplicada e Probabilidade para Engenheiros*. 6. ed. Rio de Janeiro: LTC, 2016.
7. URBANO, J. *Estatística: uma nova abordagem*. Rio de Janeiro: Editora Ciência Moderna, 2010.
8. WALPOLE, R. E.; MYERS, R. H.; MYERS, S. L.; YE, K. *Probabilidade & Estatística para Engenharia e Ciências*. 8. ed. São Paulo: Pearson, 2009.

Aconselha-se pesquisá-las para se obter um maior aprofundamento e um melhor aproveitamento nos estudos.

Conteúdo Programático

1. Introdução;
2. Distribuição Uniforme Discreta (ou Inteira);
3. Ensaio e Distribuição de Bernoulli;
4. Distribuição Binomial;
5. Distribuição Polinomial ou Multinomial;
6. Distribuição Geométrica;
7. Distribuição de Pascal (ou Binomial Negativa);
8. Distribuição Hipergeométrica;
9. Distribuição Multi-hipergeométrica;
10. Distribuição de Poisson;
11. Distribuição de Pareto Discreta ou Zeta;
12. Distribuição Zipf.

Introdução

Definição

Algumas variáveis discretas geradas mediante processos de contagem podem ser associadas a funções de probabilidade que tenham um comportamento particular conhecido. Por exemplo, quando se estuda o número de artigos defeituosos em um lote ou quando se estuda o número de pessoas que chegam a um estabelecimento comercial num certo período de tempo, entre outros.

Nesses casos, é possível estudar o comportamento de tais variáveis através de funções de probabilidade particulares em cada caso. Nessa seção, são apresentadas algumas das principais funções de probabilidade ou distribuições de probabilidade, que podem ser utilizadas para analisar variáveis, tais como as descritas anteriormente.

Distibuição Uniforme Discreta (ou Inteira)

Definição: Uma variável aleatória X tem uma distribuição uniforme discreta se cada um dos n valores em sua faixa ($a \leq x \leq b$), isto é, $x_1, x_2, \ldots, x_n$ tiver igual probabilidade. Então,

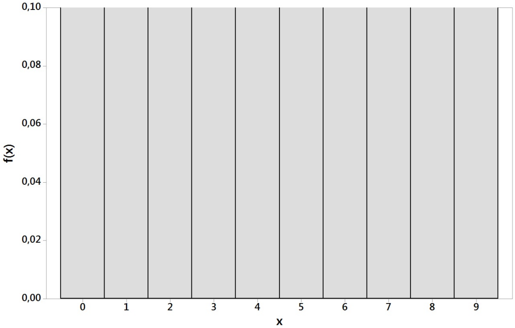
$$f(x_i) = 1/n.$$

Adotaremos a seguinte notação: $X \sim UD(a, b)$

Exemplo - [Montgomery e Runger(2016)]

O 1^0 dígito de um número serial de uma peça é igualmente provável de ser qualquer um dos dígitos 0 a 9. Se uma peça for selecionada de uma grande batelada e X for o primeiro dígito do número serial, então $X \sim (0, 9)$. Isto é,

$$f(x) = 1/10 \;\; \forall \;\; x \in \Re_x = [0, 1, \ldots, 9]$$



Se $X \sim UD(a,b)$, a média e a variância de X são dadas por:

Média

$$\mu = E(X) = \frac{a+b}{2}$$

Variância

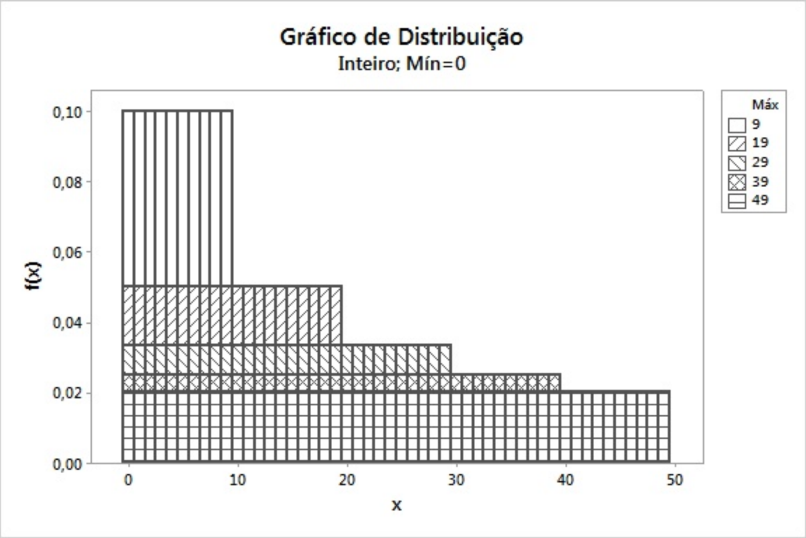
$$\sigma^2 = Var(X) = \frac{(b-a+1)^2 - 1}{12}$$

Exemplo - [Montgomery e Runger(2016)]

Seja a variável aleatória X o número das 48 linhas telefônicas que estão em uso em certo tempo. Considere que X seja uma variável aleatória discreta uniforme, com uma faixa de 0 a 48. Então, $X \sim UD(0,48)$ e

$$\mu = E(X) = \frac{0+48}{2} = 24 \;\; e \;\; \sigma^2 = Var(X) = \frac{(48-0+1)^2 - 1}{12} = 200$$

Alguns esboços da distribuição Uniforme Discreta (ou Inteira) para valores escolhidos para a e b são ilustrados no gráfico abaixo.



Ensaio e Distribuição de Bernoulli

Ensaio de Bernoulli

Há muitos experimentos que têm somente dois resultados possíveis, chamados de *sucesso* (S) e *fracasso* (F). Logo, o espaço amostral para esse tipo de experimento é $\Omega = \{S, F\}$. Por exemplo, ao lançar uma moeda, obtém-se somente dois resultados possíveis, cara (H - *Heads*) ou coroa (T - *Tails*). Chama-se de sucesso, o evento de interesse.

No exemplo, caso o interesse seja "cara", obtém-se um sucesso quando no ensaio ocorre cara. Caso contrário, obtém-se um fracasso. Um experimento com essa característica é chamado de experimento ou **ensaio de Bernoulli.**

Distribuição de Bernoulli

Distribuição de Bernoulli

Seja a variável aleatória X, definida como o número de sucessos num ensaio de Bernoulli. Então, o contradomínio de X é dado por $R_X = \{1, 0\}$. Isto é,

$$X = \begin{cases} 1, & \text{se o resultado é do ensaio é sucesso} \\ 0, & \text{se o resultado é fracasso.} \end{cases}$$

A variável aleatória assim definida chama-se *variável aleatória de Bernoulli.*

Sejam $P(S) = p$ e $P(F) = q = 1 - p$ as probabilidade de sucesso e fracasso, respectivamente. A distribuição de probabilidade da variável aleatória X de Bernoulli, é chamada de **Distribuição de Bernoulli,** e é dada por

x	0	1
$f(x) = P[X = x]$	$q = 1 - p$	p

A Distribuição de Bernoulli pode também ser expressa como

$$f(x) = P[X = x] = \begin{cases} p^x(1-p)^{1-x}, & se\ x = 0, 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Notação

$X \sim bernoulli(p)$, com $M_X(t) = (1 - p) + pe^t$.

Média e Variância

$$\begin{aligned} \mu_X &= E(X) = 0 \times q + 1 \times p = p. \\ \sigma^2 &= Var(X) = E(X^2) - \mu^2 = 0^2 \times q + 1^2 \times p - p^2 = p(1-p) \end{aligned}$$

Distribuição Binomial

Experimento Binomial

Existem muitos problemas, nos quais o experimento consiste em n ensaios (ou experimentos) de Bernoulli $\varepsilon_1, \ldots, \varepsilon_n$. Dizemos que uma sequência de ensaios de Bernoulli forma um processo de Bernoulli ou **experimento Binomial** quando satisfizer às seguintes condições:

(i) Cada ensaio tem somente dois resultados possíveis S ou F.

(ii) Os ensaios são independentes, isto é, o resultado (sucesso ou fracasso) de qualquer ensaio é independente do resultado de qualquer outro ensaio.

(iii) A probabilidade de sucesso, p, permanece constante de ensaio em ensaio. Logo, a probabilidade de fracasso $q = 1 - p$ também permanece constante.

Exemplo - [Cancho(2010)]

Suponha um experimento onde uma moeda é lançada três vezes. Suponha também que p seja a probabilidade de sair coroa (T) em cada lançamento. Seja X a variável aleatória que representa o número de coroas obtidas ao final dos três lançamentos. Achar a distribuição de probabilidade de X.

Solução: O espaço amostral para experimento de lançar uma moeda três vezes é:

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Seja X_i ($i = 1, 2, 3$) a variável aleatória de Bernoulli que representa o número coroas no lançamento i. Então a variável

$$X = X_1 + X_2 + X_3,$$

representa o número de coroas nos 3 lançamentos da moeda. Pode-se mostrar que $X_i \sim bernoulli(p)$.

w_i	$P(\{w_i\})$	$X_1(w_i)$	$X_2(w_i)$	$X_3(w_i)$	$X(w_i) = X_1(w_i) + X_2(w_i) + X_2(w_i)$
HHH	$(1-p)^3$	0	0	0	0
HHT	$(1-p)^2p$	0	0	1	1
HTH	$(1-p)^2p$	0	1	0	1
THH	$(1-p)^2p$	1	0	0	1
HTT	$(1-p)p^2$	0	1	1	2
THT	$(1-p)p^2$	1	0	1	2
TTH	$(1-p)p^2$	1	1	0	2
TTT	p^3	1	1	1	3

O contradomínio da variável X é: $R_X = \{0, 1, 2, 3\}$. Logo,

$$
\begin{aligned}
P[X=0] &= P(\{HHH\}] = (1-p)(1-p)(1-p) = (1-p)^3 \\
P[X=1] &= P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) = 3p(1-p)^2 \\
P[X=2] &= P(\{HTT\}) + P(\{THT\}) + P(\{TTH\}) = 3p^2(1-p) \\
P[X=3] &= P(\{TTT\}) = p^3
\end{aligned}
$$

A distribuição de probabilidades da variável aleatória X é dada por

x	0	1	2	3
$f(x) = P[X = x]$	$1(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

O comportamento de X fica completamente determinado pela função

$$f(x) = \begin{cases} \binom{3}{x} p^x (1-p)^{3-x}, & x = 0, 1, 2, 3 \\ 0, & \text{caso contrário} \end{cases}$$

onde $\binom{3}{x} = \frac{3!}{x!(3-x)!}$.

Observe que as probabilidades correspondem aos termos do desenvolvimento em Binômio de Newton de $[p + (1-p)]^3$, o que justifica o nome **Distribuição Binomial** escolhido para esse modelo.

Relembrando: Para as constantes a e b, a expansão binomial é dada por

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^k \cdot b^{n-k}$$

Distribuição Binomial

Considere a repetição de n ensaios de Bernoulli independentes, todos com a mesma probabilidade de sucesso p. A VA que conta o **número total de sucessos** nos n ensaios de Bernoulli, é denominada de variável aleatória Binomial com parâmetros n e p e sua função de probabilidade é dada por

$$f(x) = P[X = x] = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & \text{caso contrário} \end{cases}$$

onde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ representa o coeficiente Binomial.

Notação

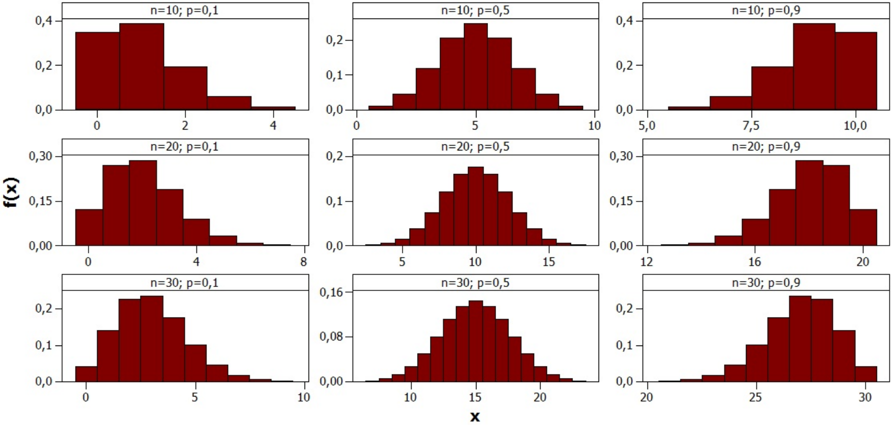
$X \sim B(n, p)$, com $M_X(t) = [pe^t + (1-p)]^n$.

Média e Variância

(a) $\mu = E(X) = np$

(b) $\sigma^2 = Var(X) = np(1-p)$

Exemplos de distribuições binomiais para valores selecionados de n e p



Exemplo 1 - [Montgomery e Runger(2016)]

Cada amostra de ar tem 10% de chance de conter um determinado poluente orgânico. Considere que as amostras sejam independentes com relação à presença de poluente. Encontre a probabilidade de que nas próximas 18 amostras exatamente 2 contenham o poluente. Além disso, determine o número médio e a variância das amostras que contém o poluente.

Exemplo 2 - [Cancho(2010)]

Suponha que os nascimentos de menino e menina sejam igualmente prováveis e que o nascimento de qualquer criança não afeta a probabilidade do sexo do próximo nascimento. Então, determine a probabilidade de que nasçam:

(a) Exatamente 4 meninos em 10 nascimentos.

(b) Pelo menos, 4 meninos em 10 nascimentos.

(c) No máximo, um menino em 10 nascimentos.

Exemplo 3 - [Cancho(2010)]

Um processo seletivo é composto por uma prova de múltipla escolha, constituída de 10 questões, cada uma com 4 alternativas. Suponha que os candidatos que irão fazer a prova não estudaram para a mesma. O selecionador estabeleceu que para ser aprovado, o candidato deve acertar pelo menos 6 questões. Qual é a probabilidade do candidato ser aprovado nesse processo?

Distribuição Polinomial ou Multinomial

A distribuição Polinomial ou Multinomial é uma generalização da distribuição Binomial, onde o espaço amostral Ω é repartido em k eventos mutuamente exclusivos $B_1, B_2, \ldots, B_k$, respectivamente com probabilidades $p_1, p_2, \ldots, p_k$, tal que $p_1 + p_2 + \ldots + p_k = 1$.

Seja considerar n repeticões independentes do mesmo experimento, sendo que $p_i \;\; (i = 1, 2, \ldots, k)$ permanecem constantes durante as repetições.

Também, sejam $X_1, X_2, \ldots, X_k$ as x_i ocorrências de $B_1, B_2, \ldots, B_k$, respectivamente, tal que $x_1 + x_2 + \ldots + x_k = n$.

Nestas condições, a probabilidade de B_1 ocorrer x_1 vezes, B_2 ocorrer x_2 vezes, ..., e B_k ocorrer x_k vezes é dada por:

$$\begin{aligned} f(x_1, x_2, \ldots, x_k) &= P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) \\ &= \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \end{aligned}$$

sendo $\sum_{i=1}^{k} x_i = n$ e $\sum_{i=1}^{k} p_i = p$.

Notação

$X \sim M(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k)$

Média e Variância

(a) $\mu_i = E(X_i) = np_i$

(b) $\sigma_i^2 = Var(X_i) = np_i(1 - p_i)$

Observação

Vale lembrar que o número de maneiras de se arranjar n objetos, x_1 dos quais são de uma espécie, x_2 são de uma segunda espécie, ... e x_k são de uma $k-$ésima espécie, é dado pelo coeficiente multinomial $\frac{n!}{x_1!x_2!\cdots x_k!}$, que corresponde a uma permutação com repetição.

Exemplo 1 - [Walpole e outros(2009)]

Para um certo aeroporto que possui três pistas de decolagem/aterrissagem, sabe-se que, em um cenário ideal, as probabilidades de que as pistas individuais sejam acessadas pela chegada aleatória de voos são: 20% para a pista 1, 15% para a pista 2 e 65% para a pista 3.

Qual é a probabilidade de que 6 aviões, chegando aleatoriamente ao aeroporto, sejam distribuídos da seguinte maneira: 2 aviões aterrissem na pista 1, 1 avião aterrisse na pista 2 e 3 aviões aterrissem na pista 3?

Exemplo 2 - [Urbano(2010)]

As causas mais frequentres de paralisação de uma certa linha de produção, e respectivos percentuais, são:

- Falhas mecânicas : 20%;
- Erros de operação: 30%;
- Material impróprio: 35%;
- Falta de energia: 15%.

Durante certo período ocorreram 5 paralisações. Qual a probabilidade de que tenham sido 2 devidas a erros de operação e 3 devidas a material impróprio?

Distribuição Geométrica

Em uma série de ensaios de Bernoulli (tentativas independentes, com probabilidade constante p de sucesso), seja a variável aleatória X o número de tentativas **até que o primeiro sucesso ocorra**. Então, X é uma variável aleatória geométrica, com parâmetro $0 < p < 1$ e

$$f(x) = P[X = x] = \begin{cases} (1-p)^{x-1} \cdot p, & x = 1, 2, \ldots, \\ 0, & \text{caso contrário} \end{cases}$$

Notação

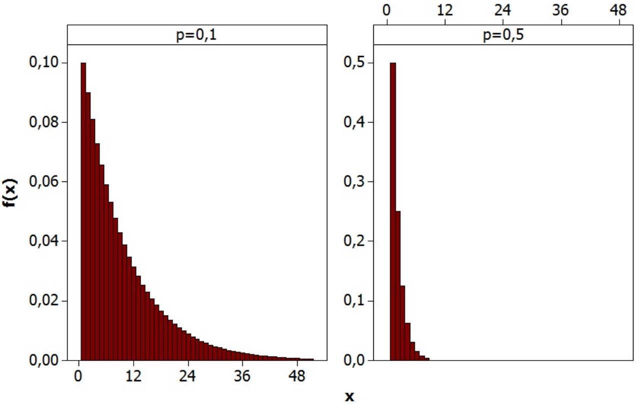
$X \sim Geom(p)$, com $M_X(t) = \frac{pe^t}{1-qe^t}$.

Média e Variância

(a) $\mu = E(X) = \frac{1}{p}$

(b) $\sigma^2 = Var(X) = \frac{1-p}{p^2}$

Exemplos de funções de probabilidade para variáveis aleatórias geométricas são mostrados na figura a seguir:



Note que a altura da linha em x é $(1-p)$ vezes a altura da linha em $(x-1)$. Ou seja, as probabilidades diminuem em uma progressão geométrica. A distribuição tem esse nome por causa desse resultado.

Exemplo - [Montgomery e Runger(2016)]

A probabilidade de uma pastilha de freio conter uma partícula grande de contaminação é 0,01. Se for considerado que as pastilhas são independentes, qual será a probabilidade de que exatamente 125 pastilhas necessitem ser analisadas para que uma partícula grande seja detectada pela primeira vez?

Propriedade da falta de memória da Distribuição Geométrica

Uma propriedade interessante e útil da Distribuição Geométrica é que ela não tem memória, isto é,

$$P(X > s + x | X > s) = P(X > x).$$

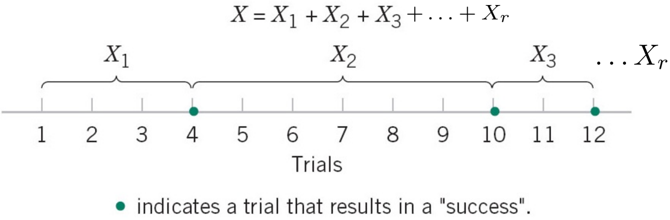
A Distribuição Geométrica é a única distribuição discreta que tem essa *propriedade de falta de memória.*

Exemplo - [Montgomery e Runger(2016)]

No exemplo anterior, suponha que já tenhamos analisado 90 pastilhas sem detectar uma partícula grande. Qual é a probabilidade de que pelo menos outras 20 pastilhas adicionais precisem ser analisadas para encontrar a primeira imperfeição?

Distribuição de Pascal (ou Binomial Negativa)

Observe a figura a seguir:



Fonte: Adaptado de Montgomery & Runger (2016), p.72

Seja X_1 o número de tentativas requeridas para obter o primeiro sucesso; Seja X_2 o número de tentativas requeridas para obter o segundo sucesso; Seja X_3 o número de tentativas requeridas para obter o terceiro sucesso, e assim por diante. Então, o número total de tentativas requeridas para obter r sucessos é $X = X_1 + X_2 + X_3 + \ldots + X_r$.

Devido à propriedade de falta de memória da Distribuição Geométrica, cada uma das variáveis aleatórias $X_1, X_2, \ldots, X_r$ tem uma distribuição geométrica, com o mesmo valor de p.

Consequentemente, uma variável aleatória binomial negativa pode ser interpretada como a soma de r variáveis aleatórias geométricas, conforme ilustrado na figura acima.

Comparação

Variável aleatória binomial $\rightarrow$ o número de tentativas é predeterminado e o número de sucessos é aleatório;

Variável aleatória binomial negativa $\rightarrow$ o número de sucessos é predeterminado e o número de tentativas é aleatório.

Distribuição de Pascal (ou Binomial Negativa)

Em uma série de tentativas de Bernoulli (tentativas independentes, com probabilidade constante p de sucesso), seja a variável aleatória X o número de tentativas até que r sucessos ocorram. Então X é uma **variável de Pascal ou variável aleatória binomial negativa** com parâmetros $0 < p < 1$ e $r = 1, 2, 3, \ldots$, e

$$f(x) = P[X = x] = \begin{cases} \binom{x-1}{r-1}(1-p)^{x-r} \cdot p^r, & x = r, r+1, r+2 \ldots \\ 0, & \text{caso contrário} \end{cases}$$

Pelo fato de no mínimo r tentativas serem requeridas para obter r sucessos, a faixa de X é de r a ∞.

Notação

$X \sim Pascal(p, r)$, com $M_X(t) = (\frac{pe^t}{1-qe^t})^r$.

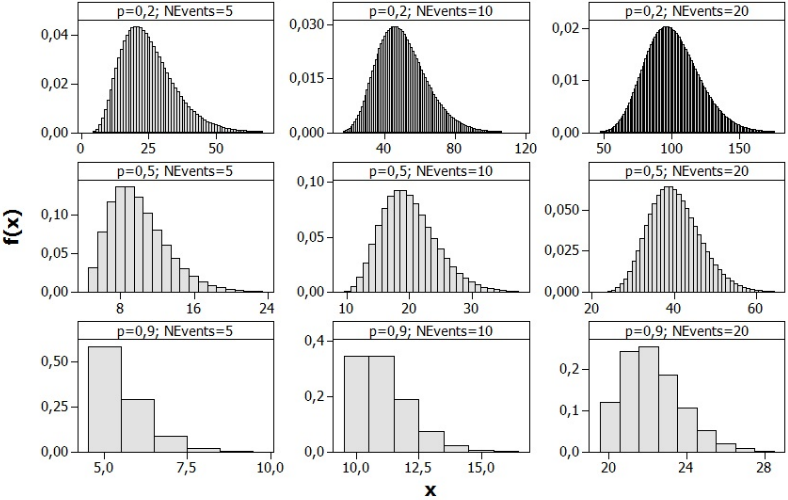
Média e Variância

$\mu = E(X) = \frac{r}{p}$ e $\sigma^2 = Var(X) = \frac{r(1-p)}{p^2}$

Observação

No caso especial em que $r = 1$, uma variável aleatória binomial negativa é uma variável aleatória geométrica.

Distribuições binomiais negativas selecionadas são ilustradas a seguir:


p=0,2; NEvents=5
p=0,2; NEvents=10
p=0,2; NEvents=20
p=0,5; NEvents=5
p=0,5; NEvents=10
p=0,5; NEvents=20
p=0,9; NEvents=5
p=0,9; NEvents=10
p=0,9; NEvents=20
f(x)
x

Exemplo - [Montgomery e Runger(2016)]

Um site da internet contém três servidores idênticos. Somente um deles é usado para operar o site; os outros dois são sobressalentes que podem ser ativados no caso do sistema principal falhar. A probabilidade de uma falha no computador principal (ou em qualquer sistema sobressalente ativado) é de 0,0005.

- Supondo que cada solicitação represente uma tentativa independente, qual será o tempo médio para a falha de todos os três servidores?

- Qual é a probabilidade de todos os três servidores falharem em, no máximo, 5 solicitações?

Aplicações da Distribuição Binomial numa Amostra

O sorteio de uma amostra de n elementos de uma população pode ser considerado como um experimento que consiste de n ensaios (ou experimentos) de Bernoulli. Os n ensaios serão independentes nos seguintes casos:

(a) Quando os elementos da amostra são sorteados com ou sem reposição de uma população infinita. Obviamente, o resultado de um sorteio qualquer é independente do outro sorteio e a proporção p de sucessos ($P(S) = p$) permanece constante em cada sorteio. Então, é aplicável a distribuição Binomial.

(b) Quando os elementos da amostra são sorteados com reposição de uma população finita. Suponha que a população tenha N elementos, dos quais M são de certa classe que temos interesse. Define-se, assim, a variável X : número de elementos da classe de interesse na amostra de tamanho n.

Os sorteios individuais são ensaios de Bernoulli, onde um elemento da classe de nosso interesse corresponde a um “sucesso” e o experimento de tomar uma amostra de tamanho n, com reposição, consiste nos n ensaios independentes de Bernoulli, onde $p = P(sucesso) = \frac{M}{N}$; isto é, X tem distribuição binomial,

$$f(x) = \binom{n}{x} \left[\frac{M}{N}\right]^{x} \left[1 - \frac{M}{N}\right]^{n-x}, \quad x = 1, \ldots, n$$

Distribuição Hipergeométrica

Suponha uma população finita com N elementos , dividida em duas classes. Uma classe com M elementos (sucessos), $(M < N)$, e a outra com $N - M$ elementos (fracassos).

Por exemplo, no caso particular de N peças produzidas, podem ser consideradas as classes: M peças defeituosas e (N-M) peças não defeituosas.

Considere agora, o seguinte experimento:

Uma amostra aleatória de tamanho n $(n < N)$, sem reposição, é sorteada da população finita de N elementos. Uma variável aleatória X pode ser definida da seguinte forma:
X: Número de elementos com a característica de interesse (sucessos) na amostra de tamanho n.

A variável aleatória assim definida chama-se variável aleatória Hipergeométrica e sua função de probabilidade é:

$$f(x) = P(X = x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = 0, 1, \ldots, \min\{n, M\} \\ 0, & \text{caso contrário} \end{cases}$$

Notação

$X \sim H(N, M, n)$

Média e Variância

Se $X \sim H(N, M, n)$, então

(a) $E(X) = n\frac{M}{N}$

(b) $Var(X) = n\frac{M}{N}(1 - \frac{M}{N})(\frac{N-n}{N-1})$

Exemplo 1 - [Hines e outros(2006)]

Em um setor de inspeção de qualidade de uma empresa, lotes de eixos são recebidos periodicamente. Os lotes contém 100 unidades, e o seguinte plano de amostragem de aceitação é usado.

Seleciona-se uma amostra aleatória de 10 unidades, sem reposição. O lote é aceito se a amostragem tiver, no máximo, um eixo defeituoso.

Suponha que um lote seja recebido e que é 5% defeituoso. Qual é a probabilidade de que seja aceito?

Exemplo 2 - [Montgomery e Runger(2016)]

Uma batelada de peças contém 100 peças do processo A e 200 peças de um outro processo, B. Se 4 peças forem selecionadas, ao acaso e sem reposição:

- Qual será a probabilidade de que elas sejam todas provenientes do processo A?

- Qual é a probabilidade de duas ou mais peças na amostra serem provenientes do processo A?

- Qual é a probabilidade de no mínimo uma peça na amostra ser proveniente do processo A?

- Determine o número médio esperado de peças na amostra provenientes do processo A, bem como seu desvio-padrão.

Distribuição Binomial como aproximação da Distribuição Hipergeométrica

Fato:

A Distribuição Binomial pode ser usada como limite da Distribuição Hipergeométrica quando n for suficientemente pequeno em relação a N. Isto é, Se $X \sim H(N, M, n)$ e $f = \frac{n}{N} < 0,10$, então $X \sim B(n, \frac{M}{N})$.

Exemplo - [Cancho(2010)]

Foram colocadas em uma caixa, 100 peças, 40 das quais foram fabricadas pela indústria A e as outras, pela indústria B. Retiradas, sem reposição, 8 peças, qual é a probabilidade de que sejam 4 da indústria B?

Distribuição Multi-Hipergeométrica

Seja uma população finita de N elementos, dos quais, M_i $(i = 1, 2, \ldots, k)$ têm determinadas características.

Seja também, selecionar n elementos $(n \leq N)$ sem reposição.

Assim, podemos associar a esse experimento uma variável aleatória multidimensional $\mathbf{X} = \{X_1, X_2 \ldots, X_k\}$, com cada dimensão indicando o número de vezes que ocorre a correspondente característica nas n extrações.

A distribuição de probabilidades para essa variável multidimensional é conhecida como distribuição Multi-Hipergeométrica, e é uma generalização da distribuição Hipergeométrica.

Sua distribuição de probabilidades é dada pela seguinte função:

$$\begin{aligned} f(x_1, x_2, \ldots, x_k) &= P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) \\ &= \frac{\binom{M_1}{x_1} \cdot \binom{M_2}{x_2} \cdot \ldots \cdot \binom{M_k}{x_k}}{\binom{N}{n}}, \end{aligned}$$

onde $\sum_{i=1}^{k} M_i = N$ e $\sum_{i=1}^{k} x_i = n$.

Exemplo - Adaptado de [Neto e Cymbalista(2006)]

Em um processo de seleção para trainees de uma empresa, há 5 engenheiros, 4 economistas e 7 administradores. São chamados para entrevista, sem reposição, os oito que mais se destacaram na fase de dinâmica de grupos. Qual é a probabilidade de que, entre os chamados, haja 3 engenheiros, 1 economista e 4 administradores?

Distribuição de Poisson

Processo de Poisson

Muitos problemas consistem em observar a ocorrência de eventos discretos num intervalo contínuo (unidade de medida), como por exemplo, o número de manchas (falhas) por unidade de medida (digamos $1m^2$) no esmaltado de uma superfície metálica. Ao se definir a variável aleatória X : número de manchas em um metro quadrado dessa superfície, o contradomínio é $\Re_X = \{0, 1, 2, \ldots, \}$

Outro exemplo é contar o número de chamadas que chegam a uma central telefônica de uma empresa num intervalo de tempo de 2 horas. É um evento discreto, visto que o tempo de chegada de qualquer delas é um ponto isolado num período de 2 horas.

Os eventos discretos gerados num intervalo contínuo (unidade: comprimento, área, volume, tempo, etc.) formam um processo de Poisson com parâmetro λ se satisfazem às seguintes propriedades:

1. O número médio de ocorrência dos eventos numa unidade de medida (comprimento, área, volume, tempo, etc.) é conhecido e igual a λ.
2. A ocorrência de um evento numa unidade de medida t não afeta a ocorrência ou a não ocorrência em outra unidade de medida t contígua. Isto é, a ocorrência dos eventos em unidades de medida contíguas são independentes.

Uma variável discreta X tem Distribuição de Poisson com parâmetro μ, se sua função de probabilidade é dada por

$$f(x) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, \dots, \tag{1}$$

onde

X :	número de eventos discretos em t unidades de medida.
λ :	é a média de eventos discretos em uma unidade de medida.
t :	número de unidades de medida.
$\mu = \lambda t$:	é a média de eventos discretos em t unidades de medida.

Notação

$X \sim P(\mu)$, com $M_X(t) = e^{\mu(e^t - 1)}$.

Média e Variância

(a) $E(X) = \mu$

(b) $Var(X) = \mu$

Exemplo 1 - [Cancho(2010)]

Suponha que a central telefônica de um empresa de grande porte receba, em média, 3 chamadas a cada 4 minutos. Qual é a probabilidade de que a central recepcione 2 ou menos chamadas em um intervalo de 2 minutos?

Exemplo 2 - [Montgomery e Runger(2016)]

Falhas ocorrem ao acaso ao longo do comprimento de um fio delgado de cobre. Suponha que o número de falhas siga uma Distribuição de Poisson, com uma média de 2,3 falhas por milímetro.

- Determine a probabilidade de existirem exatamente 2 falhas em um milímetro de fio.

- Determine a probabilidade de 10 falhas em 5 milímetros de fio.

- Determine a probabilidade de existir no mínimo, 1 falha em dois milímetros de fio.

Distribuição de Poisson com aproximação da distribuição Binomial

A Distribuição de Poisson pode ser utilizada para aproximar probabilidades de uma Distribuição Binomial quando n é suficientemente grande $(n \to \infty)$ e p é muito pequeno $(p \to 0)$.

Nesse caso, faz-se $X \sim B(n, p)$ *ser aproximada por* $X \sim P(np)$.

Na prática, considera-se que a aproximação é aceitável se

$$p < 0,1$$

e n grande. Nesse caso, considera-se que $X \sim P(np)$.

Exemplo - [Hines e outros(2006)]

A probabilidade de que um rebite particular na superfície da asa de uma nova aeronave seja defeituoso é 0,001. Há 4000 rebites na asa. Qual é a probabilidade de que sejam instalados não mais de seis rebites defeituosos?

Propriedade Reprodutiva da Distribuição de Poisson

Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes , com Distribuição de Poisson com parâmetros $\mu_1, \ldots, \mu_n$, respectivamente, então a variável aleatória

$$Y = X_1 + \cdots + X_n,$$

tem Distribuição de Poisson com parâmetros $\mu = \mu_1 + \cdots + \mu_n$.

Exemplo - [Cancho(2010)]

Em uma fábrica, foram registradas em três semanas, a média de acidentes: 2,5 na primeira semana, 2 na segunda semana e 1,5 na terceira semana. Suponha que o número de acidentes por semana segue um processo de Poisson. Qual é a probabilidade de que haja 4 acidentes nas três semanas?

Distribuição de Pareto Discreta (ou Zeta)

Algumas variáveis aleatórias discretas apresentam distribuição com caudas pesadas (*heavy tails*), isto é, têm probabilidades grandes nas caudas. Alguns exemplos: *número de consumidores afetados por um blackout, tamanhos de arquivos transferidos via web, número de links em um site, entre outros.*

O comportamento estocástico dessas variáveis pode ser modelado por uma distribuição de Pareto, da seguinte maneira:

$$f(x) = P(X = x) = \frac{C}{x^{1+\alpha}}, \quad x = 1, 2, 3, \ldots, N,$$

onde C é uma constante > 0, $\alpha > 0$ e N pode ser finito ou infinito.

Observa-se que as probabilidades decrescem de acordo com uma potência de x. Por isto, ela é chamada de uma distribuição de lei de potência (*power law*). Ela não cai com uma rapidez exponencial como é o caso de uma Poisson ou de uma Geométrica.

Comparando:

- A distribuição de Poisson tem cauda curta, valores com probabilidades significativas estão concentrados em uma faixa estreita em torno de sua média $E(X) = \mu$;
- Já a distribuição de Pareto tem cauda longa, gerando facilmente valores muito grandes, ordens de grandeza maiores que $E(X)$; e,
- Por sua vez, a distribuição Geométrica é um caso intermediário.

Pode-se observar também que, ao se fixar um valor de α, o valor de C fica determinado, uma vez que a soma das probabilidades deve resultar 1, ou seja,

$$\begin{aligned} 1 &= P(X=1) + P(X=2) + P(X=3) + \ldots \\ &= C\left(\frac{1}{1^{1+\alpha}} + \frac{1}{2^{1+\alpha}} + \frac{1}{3^{1+\alpha}} + \ldots\right) \\ &= C\left(\sum_{x=1}^{N} \frac{1}{x^{1+\alpha}}\right), \end{aligned}$$

o que implica em

$$C = \frac{1}{\sum_{x=1}^{N} \frac{1}{x^{1+\alpha}}}.$$

Nota: quando $N = \infty$, esta constante está associada com a *função zeta de Riemann* definida como

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \text{que converge quando } \; s > 1.$$

Assim, quando $N = \infty$, temos

$$C = \frac{1}{\zeta(1+\alpha)} \quad \text{que converge quando } \; 1 + \alpha > 1 \; \rightarrow \; \alpha > 0.$$

Distribuição de Pareto Discreta (ou Zeta)

$$f(x) = P(X = x) = \frac{C}{x^{1+\alpha}}, \quad x = 1, 2, 3, \ldots, N,$$

onde C é uma constante > 0, $\alpha > 0$ e N pode ser finito ou infinito.

- No caso de N finito, tem-se $C = \frac{1}{\sum_{x=1}^{N} \frac{1}{x^{1+\alpha}}}$;
- Já no caso de $N = \infty$, tem-se
  $C = \frac{1}{\zeta(1+\alpha)}$ que converge quando $1 + \alpha > 1 \;\rightarrow\; \alpha > 0$.

Notação

$X \sim Pareto(N, \alpha)$

Média e Variância

1. Caso N finito:
   - $E(X) = C\sum_{x=1}^{N} \frac{1}{x^{\alpha}}$
   - $V(X) = C\sum_{x=1}^{N} \frac{1}{x^{\alpha-1}} - \left[C\sum_{x=1}^{N} \frac{1}{x^{\alpha}}\right]^2$
2. Caso $N = \infty$:
   - $E(X) = \frac{\zeta(\alpha)}{\zeta(1+\alpha)}, \quad \alpha > 1$
   - $V(X) = \frac{\zeta(\alpha-1)}{\zeta(1+\alpha)} - \left[\frac{\zeta(\alpha)}{\zeta(1+\alpha)}\right]^2, \quad \alpha > 2$

Exemplo 1

Alguns *sites* da internet apresentam poucos *links*, mas outros apresentam muitos *links*. Considere que o número de *links* possa ser modelado por uma distribuição de Pareto Discreta (ou Zeta). Considerando um valor de $\alpha = 0,5$, qual a probabilidade de se encontrar apenas um *link* em uma página qualquer? E de se encontrar páginas que contenham 50 *links*?

Exemplo 2

Alguns *blackouts* energéticos podem afetar poucas cidades de uma dada região. Porém, certas vezes, afetam um número grande de cidades. Considere que o número de cidades afetadas por *blackouts* possa ser modelado por uma distribuição de Pareto Discreta (ou Zeta). Considerando um valor de $\alpha = 0,5$ e que uma região contenha 5 cidades, qual a probabilidade de todas elas serem afetadas por um eventual *blackout*? Calcule também o número médio de cidades afetadas.

Distribuição Zipf

Um distribuição que também apresenta *caudas pesadas* é a distribuição Zipf, cujo nome é em homenagem a seu desenvolvedor, o americano George Kingsley Zipf (1902-1950).

Para essa situação, temos

$$f(x) = P(X = x) = \frac{x^{-\alpha}}{\zeta(\alpha)}, \;\; x = 1, 2, 3, \ldots \;\; \text{e} \;\; \alpha > 1$$

onde $\zeta(\alpha) = \sum_{x=1}^{\infty} \frac{1}{x^{\alpha}}$ é a *função zeta de Riemann*, sendo convergente quando $\alpha > 1$.

Notação

$X \sim Zipf(\alpha)$

Média e Variância

- $E(X) = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}, \quad \alpha > 2$
- $V(X) = \frac{1}{\zeta(\alpha)}\zeta(\alpha - 2) - \left[\frac{\zeta(\alpha-1)}{\zeta(\alpha)}\right]^2, \quad \alpha > 3$

Exemplo - Adaptado de [Campos e outros(2017)]

Os tamanhos de arquivos armazenados em um banco de dados seguem uma distribuição Zipf com parâmetro α quando estes tamanhos são medidos em Kilobytes.

1. Se os tamanhos dos arquivos de 1 KB são 10.000 vezes mais prováveis que tamanhos de arquivos de 1 MB, então qual o valor do parâmetro α?
2. Quanto mais prováveis são tamanhos de arquivos de 1MB em comparação com tamanhos de arquivos de 1 GB?

- Campos, M., outros, 2017. Métodos Probabilísticos e Estatísticos com Aplicações em Engenharias e Ciências Exatas. Rio de Janeiro: LTC.
- Cancho, V., 2010. Notas de aulas sobre noções de estatística e probabilidade - São Paulo: USP.
- Hines, W., outros, 2006. Probabilidade e Estatística na Engenharia. Rio de Janeiro: LTC.
- Montgomery, D., Runger, G., 2016. Estatística Aplicada e Probabilidade para Engenheiros. Rio de Janeiro: LTC.
- Neto, P., Cymbalista, M., 2006. Probabilidades. São Paulo: Blucher.
- Urbano, J., 2010. Estatística: uma nova abordagem. Rio de Janeiro: Editora Ciência Moderna.
- Walpole, R., outros, 2009. Probabilidade e Estatística para Engenharia e Ciências. São Paulo: Pearson.