Universidade de São Paulo
Instituto De Ciências Matemáticas e de Computação

Prof. Jorge Bazán

Material adicional - Modelos probabilísticos

Estatística I - SME0320

São Carlos
2023/01

Capítulo 1

MODELOS PROBABILÍSTICOS

1.1 DISTRIBUIÇÃO HIPERGEOMÉTRICA

De uma população com N elementos, nos quais M são de interesse, pega-se uma amostra de n elementos. A variável aleatória X é definida como

$$X = \text{número de elementos de interesse na amostra.}$$

A função de probabilidade de X é dada por:

$$f(x) = P(X = x) = \frac{\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}, \tag{1.1}$$

$$x = \{\max\{0, n+M-N\}, \ldots, \min\{M, n\}\}$$

Os parâmetros desta distribuição podem assumir os seguintes valores

$$N = 1, 2, \ldots \qquad M = 0, 1, \ldots, N \qquad n = 1, 2, \ldots, N$$

A notação a ser usada é $X \sim H(N, M, n)$. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = n\frac{M}{N} \qquad \text{e} \qquad Var(X) = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Exemplo 1: Suponha que temos o seguinte experimento aleatório: retiram-se três bolas ao azar, sem reposição, de uma urna composta por cinco bolas vermelhas e três cinzas (ver Figura 1.1).

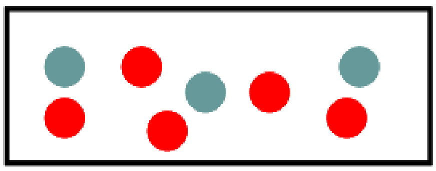


Figura 1.1: Urna com 5 bolas vermelhas e 3 cinzas.

Seja X = número de bolas cinzas extraídas, os valores dos parâmetros são:

$$N = 8 \qquad M = 5 \qquad n = 3.$$

Assim, segue que o suporte de X é dado por

$$x = \{\max\{0, 3+5-8\}, \ldots, \min\{5, 3\}\} = \{0, 1, 2, 3\}.$$

Neste caso, podemos deduzir que

$$P(X = x) = \frac{C^3_{3-x} C^5_x}{C^8_3}, \qquad x = 0, 1, 2, 3.$$

Exemplo 2: Há um total de N empresas que estão em condição de inadimplência. Um inspetor tem uma base de dados com 10 registros de empresas inadimplentes e um colega dele pega uma amostra, ao acaso, de 8 empresas inadimplentes e encontra que X delas já figuram na base de dados.

a) Para $N = 20$, verifique que X tem distribuição Hipergeométrica e identifique seus parâmetros.

b) Para um valor genérico de N, construa a função de probabilidade de X. Se resultou $X = 2$ e você conhece que N é igual a 40 ou 20, qual valor escolheria?

Solução:

a) Temos que $N = 20$, $M = 10$ e $n = 8$. Assim,

$$x = \{\max\{0, 8 + 10 - N\}, \ldots, \min\{10, 8\}\} = \{0, \ldots, 8\}$$

Logo,

$$P(X = x) = \frac{C^{20}_x C^{10}_{8-x}}{C^{20}_8}, \qquad x = 0, \ldots, 8.$$

b) Temos que

$$P(X = x) = \frac{C^{10}_x C^{N-10}_{8-x}}{C^N_8}.$$

São comparados ambos os casos:

– Se $N = 20$, então $P(X = 2) = \dfrac{C^{10}_2 C^{10}_6}{C^{20}_8}$.

– Se $N = 40$, então $P(X = 2) = \dfrac{C^{10}_2 C^{30}_6}{C^{40}_8}$.

Escolheremos o valor de N que possui maior probabilidade do evento acontecer $\{X = 2\}$.

1.2 MODELOS RELACIONADOS COM O PROCESSO BERNOULLI

1.2.1 DISTRIBUIÇÃO BERNOULLI

Temos um **experimento de Bernoulli** se, ao realizarmos um experimento, somente dois resultados são possíveis:

- $X = 1$ (**sucesso** com probabilidade p)
- $X = 0$ (**fracasso** com probabilidade $q = 1 - p$)

Exemplos:

- Atirar uma moeda e obter cara: $p = 1/2$.
- Escolher uma pessoa da população e esta pessoa é doente: $p = 1/1000$ (prevalência da enfermidade).
- Aplicar um tratamento a um doente e que este doente se cure: $p = 0,95$ (probabilidade de que o indivíduo se cure).

Como se aprecia em experimentos onde o resultado é dicotômico, uma variável é perfeitamente determinada conhecendo-se o *parâmetro p*.

Exemplo 3: Foi observado em **2000** acidentes de trânsito com impacto frontal e cujos motoristas **não possuíam** cintos de segurança, que **300** indivíduos ficaram com sequelas. Descreva o experimento usando conceitos de variáveis aleatórias.

Solução:
A noção frequentista de probabilidade nos permite aproximar a probabilidade de haver sequelas por meio de $300/2000 = 0,15 = 15\%$. Então a variável aleatória é

$$X = \text{“ter sequelas após um acidente sem cinto de segurança”},$$

que segue uma distribuição Bernoulli, onde:

- $X = 1$ tem probabilidade $p \approx 0,15$.
- $X = 0$ tem probabilidade $q \approx 0,85$.

Exemplo 4: Foi observado em **2000** acidentes de trânsito com impacto frontal e cujos motoristas **possuíam** cintos de segurança, que **10** indivíduos ficaram com sequelas. Descreva o experimento usando conceitos de variáveis aleatórias.

Solução:
A noção frequentista de probabilidade nos permite aproximar a probabilidade de haver sequelas por meio de $10/2000 = 0,005 = 0,5\%$. Então a variável aleatória é

$$X = \text{“ter sequelas após um acidente usando cinto de segurança”},$$

que segue uma distribuição Bernoulli, onde:

- $X = 1$ tem probabilidade $p \approx 0,005$.
- $X = 0$ tem probabilidade $q \approx 0,995$.

– Os exemplos 3 e 4 correspondem à distribuição Bernoulli para diferentes valores do parâmetro p.

1.2.2 DISTRIBUIÇÃO BINOMIAL

Em um processo Bernoulli definimos a variável aleatória

$$X = \text{Número de sucessos obtidos em } n \text{ ensaios.}$$

A função de probabilidade de X é dada por

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$x = 0, 1, \ldots, n$$

A notação a ser usada é $X \sim B(n, p)$. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = np \qquad \text{e} \qquad Var(X) = npq,$$

na qual $q = 1 - p$.

Exemplo 5: As peças que saem de uma linha de produção são classificadas como defeituosas **(D)** ou não defeituosas **(N)**, independentemente. A probabilidade de uma peça apresentar defeito é p e não muda. Isso implica que, se a população for finita, as observações serão feitas com reposição.
É definida uma variável aleatória X igual ao número de peças defeituosas em 3 analisadas; então $X = 0, 1, 2, 3$, e, em uma linha de produção de 3 peças, temos

$$P(X = x) = C_x^3 \, p^x (1-p)^{3-x}$$
$$x = 0, 1, 2, 3$$

Ver Figura 1.2 que mostra a distribuição de probabilidade.

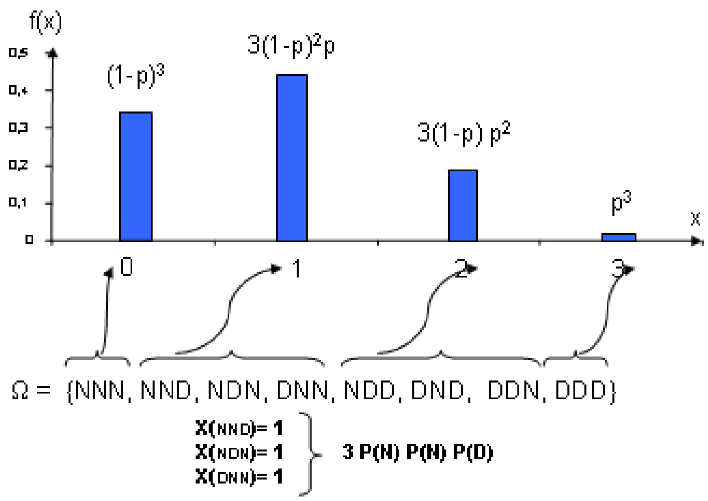


Figura 1.2: Função de distribuição de probabilidade do número de peças defeituosas em um modelo Binomial.

1.2.3 DISTRIBUIÇÃO GEOMÉTRICA

Em um processo Bernoulli definimos:

$$X = \text{Número de repetições necessárias para obter o primeiro sucesso.}$$

A função de probabilidade de X é dada por

$$f(x) = P(X = x) = (1-p)^{x-1} p$$
$$x = 1, 2, \ldots$$

A notação a ser usada é $X \sim G(p)$. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = \frac{1}{p} \qquad \text{e} \qquad Var(X) = \frac{q}{p^2},$$

na qual $q = 1 - p$.

Exemplo 6: Um dado é lançado quantas vezes até que um número maior que quatro ocorra. Neste caso, X é o número de lançamentos necessários.

Solução:
A: ocorre um número maior que quatro.
p: probabilidade de que ocorra o evento $A = 2/6$.
$1 - p$: probabilidade de que não ocorra o evento $A = 4/6$.
X: número de lançamentos até obter um número maior que 4, $R_X = 1, 2, 3, \ldots$

$$P(X = x) = p(A^C A^C \ldots A) = (1-p)^{x-1} p$$

1.2.4 DISTRIBUIÇÃO BINOMIAL NEGATIVA

Em um processo Bernoulli, definimos a variável

$$X = \text{Número de repetições necessárias até obter o } r\text{−ésimo sucesso.}$$

Neste caso, temos que a função de probabilidade de X é dada por

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

$$x = r, r+1, r+2, \ldots$$

A notação a ser usada é $X \sim BN(p)$. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = \frac{r}{p} \qquad \text{e} \qquad Var(X) = \frac{rq}{p^2},$$

na qual $q = 1 - p$.

Exemplo 7: Um dado é jogado tantas vezes até que um número maior que quatro ocorra três vezes. Neste caso, X é o número de lançamentos necessários.

Solução:
A: ocorrer um número maior que quatro.
p: probabilidade de que ocorra o evento $A = 2/6$.
$1 - p$: probabilidade de que não ocorra o evento $A = 4/6$.
X = número de lançamentos até obter três vezes um número maior que 4, $R_X = \{3, 4, 5, \ldots\},\ r = 3$

$$P(X = x) = p(A^C A^C \ldots AAA) = C_{3-1}^{x-1}(1-p)^{x-3}p^3$$

Exemplo 8: Em um determinado setor do comércio, 95% dos estabelecimentos comerciais pagam seus impostos em dia e de forma independente.

(a) Determine a probabilidade de que pelo menos 6 dos 8 estabelecimentos analisados paguem em dia seus impostos.

(b) Determine a probabilidade de que o número de estabelecimentos inspecionados, até o primeiro que pague seus impostos a tempo, esteja entre 5 (mínimo) e 15 (máximo).

(c) Determine a probabilidade de que o número de estabelecimentos inspecionados, até que o quarto pague seus importos a tempo, esteja entre 5 (mínimo) e 7 (máximo).

Solução:

(a) Seja X = Número de estabelecimentos que pagam em dia seus impostos dos 8 analisados. Então, podemos deduzir que o modelo adequado para esta v.a. é a distribuição Binomial com parâmetros $n = 8$ e

$$p = P(\text{sucesso}) = P(\text{que um estabelecimento pague seus impostos em dia}) = 0,95.$$

Assim, podemos escrever que

$$X \sim B(8,\ 0,95).$$

Portanto, a função de probabilidade de X é dada por:

$$f_X(x) = P(X = x) = C_x^8\, 0,95^x\, 0,05^{8-x}, \qquad x = 0, 1, \ldots, 8.$$

Finalmente, temos que

$$\begin{aligned}P(X \geq 6) &= P(X = 6) + P(X = 7) + P(X = 8)\\ &= f_X(6) + f_X(7) + f_X(8)\\ &= C_6^8\, 0.95^6\, 0.05^2 + C_7^8\, 0.95^7\, 0.05^1 + C_8^8\, 0.95^8\, 0.05^0\end{aligned}$$

(b) Seja Y = Número de estabelecimentos inspecionados até encontrar o primeiro que paga em dia seus impostos. Podemos deduzir que

$$Y \sim G(0,95).$$

Portanto, a função de probabilidade de Y é dada por

$$f_Y(y) = P(Y = y) = 0,95\,(0,05)^{y-1}, \qquad y = 1, 2, 3, \ldots.$$

Dessa forma, a probabilidade a ser calculada é $P(5 \leq Y \leq 15)$.

(c) Seja Z = Número de estabelecimentos inspecionados até encontrar o quarto que paga em dia seus impostos. Podemos deduzir que

$$Z \sim BN(4,\ 0,95).$$

Portanto, a função de probabilidade de Z é dada por

$$f_Z(z) = P(Z = z) = \binom{z-1}{3} 0,95^3 (0,05)^{z-3}, \qquad z = 3, 4, 5, \ldots.$$

Dessa forma, a probabilidade a ser calculada é $P(5 \leq Z \leq 7)$.

1.3 MODELOS RELACIONADOS COM O PROCESSO POISSON

1.3.1 DEFINIÇÃO DE PROCESSO POISSON

É um processo composto de eventos discretos que são independentes no espaço e no tempo. Diz-se que um conjunto de eventos discretos é gerado por um processo de Poisson de taxa ω se, para qualquer intervalo I (geralmente tempo) de comprimento suficientemente pequeno $h > 0$:

(i) P(ocorrência de um evento em I) $\approx \omega h$,

(ii) P(ocorrência de 2 ou mais eventos em I) ≈ 0,

(iii) A ocorrência de eventos em intervalos disjuntos do tipo I são independentes.

Exemplos:

- Chegada de aviões a um aeroporto por dia
- Chegada de clientes ao banco por dia
- Ingressos a intranet em uma hora
- Pessoas atendidas em um Refeitório Central entre as 1pm e 2 pm
- Subidas de uma inversão na bolsa de valores por dia
- Número de acidentes de trabalho por semana
- Chegada de caminhões a um armazém por dia
- Subidas de preços do petróleo por semana

1.3.2 VARIÁVEIS QUE SE PODEM DEFINIR

Considerando o exemplo de "Chegada de aviões a um aeroporto por dia" é possível definir as seguintes variáveis aleatórias:

- Número de aviões que chegam a um aeroporto por dia
- Tempo decorrido até a chegada do primeiro avião
- Tempo decorrido até a chegada do terceiro avião

A primeira corresponde a um modelo Poisson denotada por $P(\lambda = \omega t)$, a segunda a um modelo Exponencial denotada por $Exp(\beta = \omega)$ e a terceira associada a um modelo Gama denotada por $\Gamma(\alpha, \beta = \omega)$. Note que o modelo Exponencial é um caso particular do modelo Gama, ou seja,

$$Exp(\beta = \omega) = \Gamma(\alpha = 1, \beta = \omega).$$

1.3.3 DISTRIBUIÇÃO POISSON

Se o número de eventos esperados λ, num intervalo de extensão t, é $\lambda = \omega t$ (ω é a taxa de eventos por unidade de t), então definimos a variável

$$X = \text{número de eventos que ocorrem num intervalo de extensão } t,$$

cuja função de probabilidade de X é dada por

$$f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$x = 0, 1, 2, 3, \ldots.$$

A notação a ser usada é $X \sim P(\lambda)$. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = \lambda \qquad \text{e} \qquad Var(X) = \lambda.$$

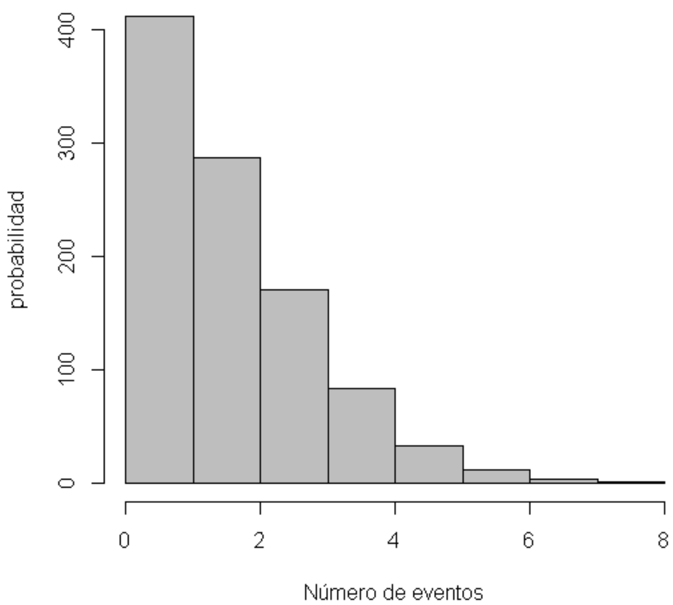


Figura 1.3: Distribuição de Poisson.

1.3.4 DISTRIBUIÇÃO GAMA e EXPONENCIAL

Se o número de eventos esperados λ, num intervalo de extensão t, é $\lambda = \omega t$ (ω é a média de ocorrências de eventos por unidade de t), então definimos a variável aleatória contínua

$$Y = \text{tempo até que ocorra o primeiro evento},$$

cuja a função de probabilidade de Y é dada por

$$f_Y(y) = \beta e^{-\beta y} \qquad y > 0, \text{ com } \beta = \omega.$$

Este modelo é denominado distribuição Exponencial e a notação a ser usada é $Y \sim Exp(\beta)$. O valor esperado e a variância são dados, respectivamente, por

$$E(Y) = \frac{1}{\beta} \qquad \text{e} \qquad Var(Y) = \frac{1}{\beta^2}.$$

Se definimos a variável aleatória contínua

$$W = \text{tempo até que ocorra o } \alpha\text{-ésimo evento},$$

então a função de probabilidade de W é dada por

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} e^{-\beta w} \qquad w > 0, \text{ com } \beta = \omega.$$

Este modelo é denominado distribuição Gama e a notação a ser usada é $W \sim \Gamma(\alpha, \beta)$. O valor esperado e a variância são dados, respectivamente por

$$E(W) = \frac{\alpha}{\beta} \qquad \text{e} \qquad Var(W) = \frac{\alpha}{\beta^2}.$$

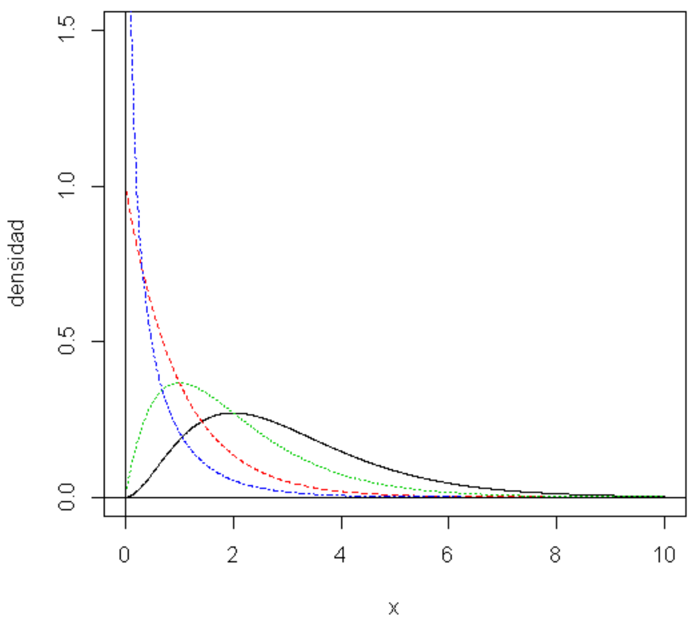


Figura 1.4: Distribuição Gama para $\lambda = 2$ e diferentes valores de α, caso exponencial quando $\alpha = 1$.

Exemplo 9: Suponha que o número de acidentes que acontecem por mês num instituto tecnológico segue a distribuição Poisson, de maneira que a probabilidade de que ocorram 2 acidentes seja igual a 2/3 da probabilidade de que ocorra 1 acidente.

(a) Calcular a probabilidade de que não ocorram acidentes em 3 meses.

(b) Calcular o esperado de acidentes por ano.

Solução:

(a) Seja X = Número de acidentes que acontecem por mês.

$$X \sim P(\omega)$$

Temos que,

$$P(X=2) = \frac{2}{3}P(X=1) \quad \Rightarrow \quad \frac{e^{-\omega}\omega^2}{2!} = \frac{2}{3}\,\frac{e^{-\omega}\omega}{1!} \quad \Rightarrow \quad \omega = \frac{4}{3}$$

Logo, espera-se $\omega = 4/3$ de acidentes por mês.

Suponha que há uma mudança na unidade de tempo. Definimos a variável

$$Y = \text{Número de acidentes em 3 meses.}$$

Portanto,

$$Y \sim P(\lambda = 3 \cdot \omega = 4)$$

Finalmente calculamos

$$P(Y=0) = \frac{e^{-4}4^0}{0!} = e^{-4}.$$

(b) Seja Z = Número de acidentes que acontecem por ano. Temos que

$$Z \sim P(\lambda^* = 12 \cdot \omega = 16)$$

Logo,

$$E[Z] = \lambda^* = 16$$

Exemplo 10: Suponha que num terminal rodoviário chegue, ao acaso, uma média de 8 ônibus por hora.

(a) Determine a distribuição de probabilidade do número de ônibus que chegam por hora no terminal.

(b) Se tal terminal pode atender apenas 10 ônibus por hora, qual é a probabilidade de que numa hora determinada devam retornar os ônibus não atendidos?

(c) Calcule o esperado de ônibus atendidos entre 10 e 12 da noite.

Solução: Seja X = Número de ônibus que chegam no terminal por hora. Temos que $E(X) = 8$.

(a) $X \sim P(\lambda) \quad \Rightarrow \quad E(X) = \lambda = 8$.

(b) $P(X = x) = \dfrac{e^{-8}8^x}{x!}, \quad$ com $x = 0, 1, 2, \ldots$.

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{x=0}^{10} P(X = x) = 1 - \sum_{x=0}^{10} \frac{e^{-8}8^x}{x!}.$$

Exemplo 11: A distribuição da duração em meses de uma lampada é Exponencial com parâmetro β. Qual é o valor de β se conhecemos que há uma probabilidade de $0,7$ de que um destes objetos tenha uma duração menor do que 6 meses?

Solução: Seja X = tempo de vida da lampada. Assim

$$X \sim Exp(\beta).$$

Podemos demostrar que a função de distribuição acumulada de X é dada por:

$$F(x) = 1 - e^{-\beta x}$$

Por condição do problema, temos que $P(X < 6) = 0,7$, então

$$\begin{aligned} P(X < 6) = F(6) = 0,7 \quad &\Rightarrow \quad 1 - e^{-6\beta} = 0,7 \\ &\Rightarrow \quad e^{-6\beta} = 0,3 \\ &\Rightarrow \quad -6\beta = \ln(0,3) \\ &\Rightarrow \quad \beta = 0,2 \end{aligned}$$

Exemplo 12: Suponha que o tempo necessário para uma criança de 8 anos montar um quebra-cabeça de 100 peças tem uma distribuição exponencial com média de 40 minutos.

(a) Calcular a probabilidade de que o quebra-cabeça seja montado em menos de meia hora.

(b) Se não montou o quebra-cabeça em meia hora, qual é a probabilidade de que monte no período de 20 minutos a mais?

Solução: Seja X = tempo necessário para criança montar o quebra-cabeça, em minutos. Assim,

$$X \sim Exp(\beta)$$

(a) $P(X < 30) = F(30) = 1 - e^{-30/40} = 0,528$.

(b)

$$\begin{aligned} P(X < 50 \mid X > 30) &= \frac{P(30 < X < 50)}{P(X > 30)} = \frac{F(50) - F(30)}{1 - F(30)} \\ &= \frac{0,714 - 0,528}{0,472} = 0,394 \end{aligned}$$

Exemplo 13: A soma do preço de um livro escolar de uma editora sobe, de acordo a um processo de Poisson, à razão de duas vezes por ano.

(a) Que tempo você espera que aconteça entre duas subidas?

(b) Que tempo você espera que aconteça entre três subidas?

(c) Qual a probabilidade de que as previsões anteriores se vejam superadas?

Solução: Definimos as seguintes variáveis

- X = Subidas de preço por ano, $X \sim P(\omega = 2)$
- Y = Tempo até a primeira subida, $Y \sim Exp(\omega = 2)$
- Z = Tempo até a α-ésima subida $Z \sim \Gamma(\alpha, \omega = 2)$

Assim, temos que

(a) Tempo entre duas subidas: $E(Y) = \dfrac{1}{\omega} = \dfrac{1}{2}$ (6 meses)

(b) Tempo entre três subidas: $E(Z) = \dfrac{\alpha}{\omega} = \dfrac{2}{2}$ (1 ano)

(c) $P(Y > 0,5) = 1 - F_Y(0,5)$, $\quad P(Z > 1) = 1 - F_Z(1)$

1.4 MODELOS UNIFORME e BETA

1.4.1 DISTRIBUIÇÃO UNIFORME

Uma variável aleatória X segue uma distribuição Uniforme se sua função de densidade é dada por

$$f(x) = \frac{1}{b-a}, \qquad \text{com } a, b \in \mathbb{R} \text{ e } a < x < b.$$

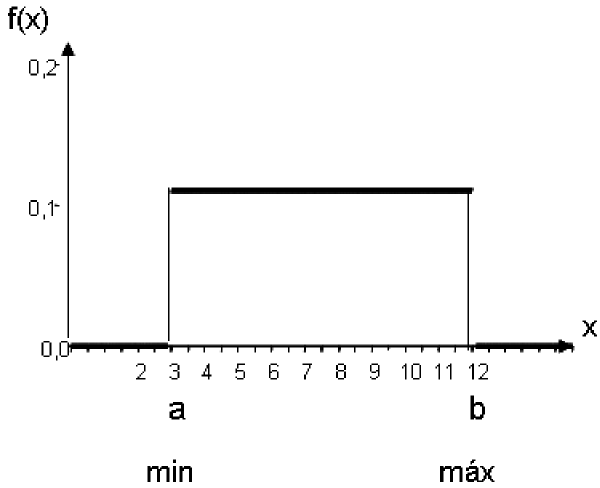


Figura 1.5: Função densidade Uniforme

Consideraremos a notação $X \sim U(a, b)$. O valor esperado e a variância são dados, respectivamente, por

$$E\,(X) = \frac{a+b}{2} \qquad \text{e} \qquad Var\,(X) = \frac{(b-a)^2}{12}.$$

Finalmente, a função de distribuição acumulada é dada por

$$F(x) = \frac{x-a}{b-a}.$$

1.4.2 DISTRIBUIÇÃO BETA

Uma variável aleatória X segue uma distribuição Beta se sua função de densidade é dada por

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\; x^{\alpha-1}\,(1-x)^{\beta-1} \qquad 0 \le x \le 1$$

$$= B(\alpha, \beta)\; x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 \le x \le 1$$

Consideraremos a notação $X \sim Beta(\alpha, \beta)$. O valor esperado e a variância são dados, respectivamente, por

$$E\,(X) = \frac{\alpha}{\alpha+\beta} \qquad \text{e} \qquad Var\,(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

na qual as funções Beta e Gama são definidas como:

– **Beta:**

$$B(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^1 x^{r-1}(1-x)^{s-1}dx$$

– **Gama:**

$$\Gamma(n) = \int_0^\infty y^{n-1}e^{-y}dy \qquad n > 0$$

Exemplos de densidades da distribuição Beta são apresentadas na Figura 1.6.

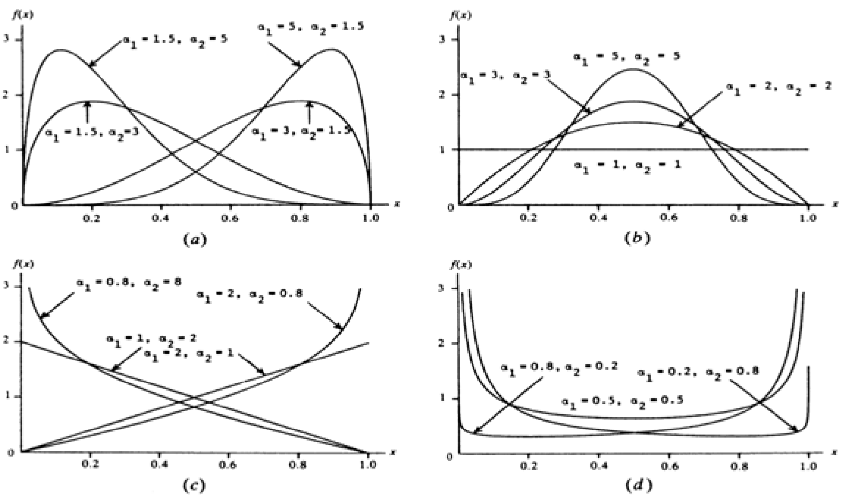


Figura 1.6: Exemplos da distribuição Beta.

A distribuição uniforme é adequada para descrever uma variável que assume seus valores uniformes ou indistintamente num intervalo de extremos finitos. No caso da distribuição Beta, aplica-se também para modelar uma variável aleatória que assume seus valores num intervalo de extremos finitos, mas sem que seja uniforme. Uma aplicação comum de ambos modelos é modelar proporções no intervalo $(0, 1)$.

Exemplo 14: Seja X uma variável aleatória distribuída uniformemente no intervalo $[a, b]$, $a < b$, com média igual a $1,5$ e variância $49/12$, calcular a probabilidade de que X seja menor que 1.

Exemplo 15: Estudos conduzidos sobre a proporção de professores com um nível adequado em matemática, a nível nacional, indicaram que esta proporção está em torno de $2/10$ e com uma dispersão média de $2/15$. Há interesse em determinar um modelo probabilístico para descrever X, a proporção de professores com nível adequado em matemática, e obter a probabilidade de que tal proporção exceda a $0,25$.

Exemplo 16: Um pesquisador educacional propõe que a proporção de estudantes que ingressam em uma universidade pode ser modelada usando a seguinte função densidade

$$f(y) = 6y(1-y), \qquad 0 \leq y \leq 1$$

Um segundo pesquisador propõe usar outro modelo, dado por

$$f(x) = 1, \qquad 0 \leq x \leq 1$$

Determine quanto é a proporção esperada e a variância da proporção de ingressantes à universidade, segundo os modelos propostos, e compare em qual desses modelos existe maior variabilidade.

1.5 MODELO NORMAL

1.5.1 DISTRIBUIÇÃO NORMAL

Uma variável aleatória X segue uma distribuição Normal se sua função densidade é dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{com } -\infty < x < \infty, \quad \mu \in \boldsymbol{R}, \sigma > 0$$

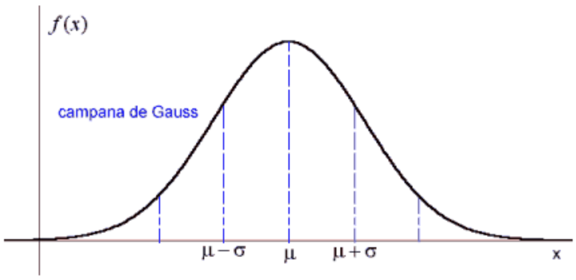


Figura 1.7: Função densidade normal (μ, σ^2).

Notação: $X \sim N(\mu, \sigma^2)$.

É chamada também distribuição Gaussiana ou sino de Gauss em homenagem a K. Gauss. É uma distribuição do tipo contínua. Seu comportamento é caraterizado pela Curva Normal, que é o gráfico de sua densidade em forma de sino (é simétrica e mesocúrtica) mostrada na Figura 1.7. Os parâmetros da distribuição são a média μ e o desvio padrão σ. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = \mu \qquad \text{e} \qquad Var(X) = \sigma^2.$$

Se uma variável aleatória X tem distribuição Normal, para calcular a probabilidade de que X esteja entre dois valores a e b, temos que calcular a área abaixo da curva entre a e b. Esta área se calcula mediante integração, porém não é possível obter uma solução analítica, então utilizamos tabelas ou programas para seu cálculo. Em geral, a função de distribuição acumulada $F(x)$ não tem expressão analítica.

1.5.2 DISTRIBUIÇÃO NORMAL PADRÃO

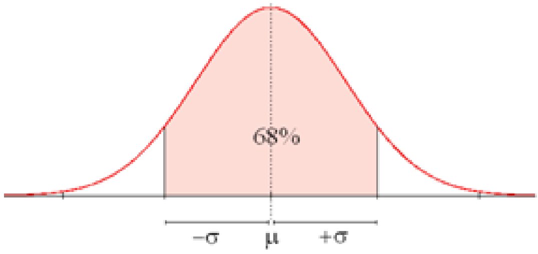
Seja $Z \sim N(0, 1)$, denominada distribuição normal padrão. Sua função densidade é dada por

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty.$$

Neste caso $\mu = 0$ e $\sigma = 1$, e $F(z)$ podemos obter por meio de tabelas.

Padronização: Se $X \sim N(\mu, \sigma^2)$, então $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

Transformação: Se $X \sim N(\mu, \sigma^2)$, então $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$



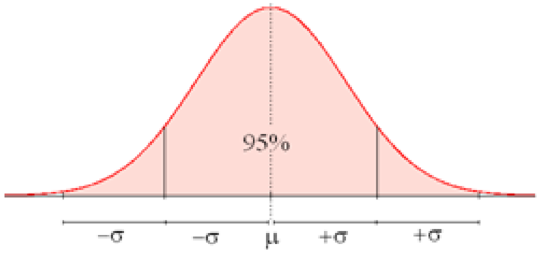


Figura 1.8: Probabilidade de curvas Normais.

Probabilidades significativas: As seguintes probabilidades são obtidas por meio de tabelas:

$$1\sigma : P(\mu - \sigma \leq x \leq \mu + \sigma) = 0,683$$
$$2\sigma : P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0,955$$
$$3\sigma : P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0,9971$$

apresentadas na Figura 1.8.

Soma de normais independentes: Se $X_1, \ldots, X_n$ são v.a. independentes, na qual

$$X_j \sim N(\mu_j, \sigma_j^2),\ j = 1, 2, \ldots, n$$
$$\Rightarrow S = \sum_{i=1}^{n} X_j \sim N\left(\sum_{i=1}^{n} \mu_j, \sum_{i=1}^{n} \sigma_j^2\right)$$

No caso especial, se $X_j \sim N(\mu, \sigma^2),\ j = 1, 2, \ldots, n$

$$\Rightarrow S = \sum_{i=1}^{n} X_j \sim N(n\mu, n\sigma^2)$$
$$\Rightarrow \overline{X} = \frac{\sum_{i=1}^{n} X_j}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Teorema Central do Limite (TCL): Se $X_1, \ldots, X_n$ são v.a. independentes, na qual $X_j \sim f(\mu, \sigma^2),\ j = 1, 2, \ldots, n$, f uma distribuição qualquer, não necessariamente Normal, tal que $E(X_j) = \mu$, $Var(X_j) = \sigma^2$, e n suficientemente grande

$$\Rightarrow\ \overline{X}\ =\ \frac{\sum_{j=1}^{n} X_j}{n}\ \overset{aprox}{\sim}\ N\left(\mu, \frac{\sigma^2}{n}\right)$$

O TCL é um dos teoremas mais usados porque nos diz que a soma de variáveis independentes de qualquer distribuição pode ser aproximar pela distribuição Normal. Devido a isto, por exemplo, as distribuições Poisson e Binomial podem ser aproximadas por uma Normal para um número suficientemente grande de eventos. Também a adição de números gerados por qualquer outra distribuição pode ser aproximada por uma distribuição aproximadamente Normal.

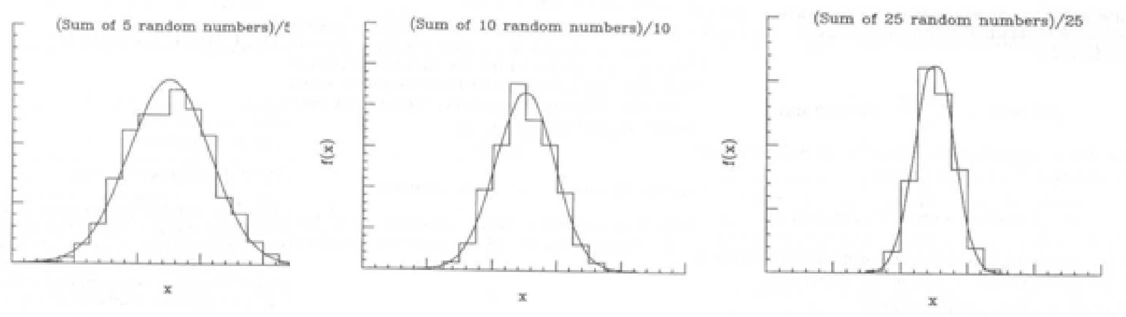


Figura 1.9: Aproximação da distribuição Binomial pela Normal.

1.5.3 USO DA TABELA NORMAL

– Ver tabela no Anexo A.

Exemplo 17: Qual a probabilidade de que um valor de Z esteja entre 0 e $2,03$?

(i) Buscar na tabela a área correspondente a $Z \leq 2,03$.

(ii) Neste caso, $P(Z \leq 2,03) = 0,9788$.

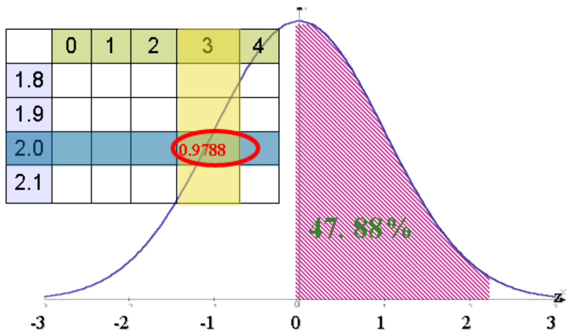


Figura 1.10: Uso da Tabela Normal.

(iii) Como queremos

$$P(0 \leq Z \leq 2,03) = P(Z \leq 2,03) - P(Z \leq 0) = 0,9788 - 0,5 = 0.4788$$

Exemplo 18: Seja X é uma população Normal com média $\mu = 70$ e $\sigma = 10$. Calcule as seguintes probabilidades:

(a) $P(X < 60)$

(b) $P(X > 95)$

(c) $P(50 < X < 80)$

Solução:

(a)

$$\begin{aligned}
P(X<60) &= P\left(\frac{x-\mu}{\sigma} < \frac{60-70}{10}\right) = P\left(Z < \frac{60-70}{10}\right) \\
&= P(Z<-1) = F(-1) \quad \text{Usar tabela Normal} \\
&= 0,1587
\end{aligned}$$

(b)

$$\begin{aligned}
P(X>95) &= 1-P(X\leq 95) \\
&= 1-P\left(\frac{x-\mu}{\sigma} \leq \frac{95-70}{10}\right) = 1-P\left(Z \leq \frac{95-70}{10}\right) \\
&= 1-P(Z\leq 2,5) = 1-F(2,5) \quad \text{Usar tabela Normal} \\
&= 1-0,9938 = 0,0062
\end{aligned}$$

(c)

$$\begin{aligned}
P(50<X<80) &= P\left(\frac{50-70}{10} < \frac{x-\mu}{\sigma} < \frac{80-70}{10}\right) \\
&= P\left(\frac{50-70}{10} < Z < \frac{95-70}{10}\right) \\
&= P\left(-2<Z<1\right) = F(1)-F(-2) \quad \text{Usar tabela Normal} \\
&= 0,8413-0,0228 = 0,8185
\end{aligned}$$

Exemplo 19: Numa operação financeira, a taxa de rentabilidade, R, é considerada como uma variável aleatória com distribuição Normal de média $0,05$ e desvio padrão $0,25$. Suponha que se fazem duas operações independentes com estas características, sendo uma de 10 u.m. e a outra de 20 u.m.

(a) Calcule a probabilidade de que a taxa de rentabilidade, R, associada a esta operação financeira seja superior a 0.3.

(b) Em qual inversão o rendimento final pode ser menor ou igual a 15 u.m.?

(c) Calcule a probabilidade da soma dos rendimentos finais ser pelo menos 30 u.m.

Solução: Temos que, $R \sim N(0.05, 0.25^2)$. Assim,

(a) $P(R>0,3) = P\left(\frac{R-\mu}{\sigma} > \frac{0,3-0,05}{0,25}\right) = P(Z>1) = 0,159.$

(b) Definimos RF_j como o rendimento final da inversão j-ésima, $j=1,2$. Então,

$$RF_1 = 10+10R \qquad \text{e} \qquad RF_2 = 20+20R$$

Portanto,

$$\begin{aligned}
P(RF_1\leq 15) = P(R\leq 0,5) &= \left(Z \leq \frac{0,5-0,05}{0,25}\right) \\
&= P(Z\leq 1,8) = 0,964
\end{aligned}$$

$$\begin{aligned}
P(RF_2\leq 15) = P(R\leq -0,25) &= P\left(Z \leq \frac{-0,25-0,05}{0,25}\right) \\
&= P(Z\leq -1,2) = 0,115
\end{aligned}$$

(c) Seja $RF = RF_1 + RF_2$ a soma dos rendimentos finais, assim $RF = 30 + 30R$. Portanto

$$\begin{aligned} RF &\sim N(30 + 30 \times 0,05, 30^2 \times 0,25^2) \\ &\sim N(31,5,\ 7,5^2) \end{aligned}$$

Logo, pode-se calcular

$$\begin{aligned} P(RF \geq 30) &= P\left(Z \geq \frac{30 - 31,5}{7,5}\right) \\ &= P(Z \geq -0,2) = 1 - P(Z < -0,2) \\ &= 1 - 0,4207 = 0,5793 \end{aligned}$$

1.5.4 DISTRIBUIÇÃO WEIBULL

Uma variável aleatória X segue uma distribuição Weibull se sua função densidade é dada por

$$f(x) = \alpha\beta x^{\alpha-1} e^{-\beta x^{\alpha}}, \qquad x > 0.$$

A notação a ser usada é $X \sim W(\alpha, \beta)$. O valor esperado e a variância são dados, respectivamente, por

$$E(X) = \frac{\Gamma\left(1 + \dfrac{1}{\alpha}\right)}{\beta^{1/\alpha}} \qquad \text{e} \qquad Var(X) = \frac{\Gamma\left(1 + \dfrac{2}{\alpha}\right) - \Gamma^2\left(1 + \dfrac{1}{\alpha}\right)}{\beta^{2/\alpha}}.$$

A função de distribuição acumulada é dada pela seguinte expressão

$$F(x) = 1 - e^{-\beta x^{\alpha}}.$$

Adicionalmente, temos as seguintes propriedades:

- Se $X \sim Exp(\beta) \ \Rightarrow \ X^{1/\alpha} \sim W(\alpha, \beta)$.
- Se $X \sim W(\alpha, \beta) \ \Rightarrow \ X^{\alpha} \sim Exp(\beta)$.

Exemplo 20: Deseja-se estudar a resistência de um material contra tremores de terra, X. O sistema de avaliação de materiais considera que o material é resistente se X excede o nível ou valor de V. Quão grande deve ser V para que a probabilidade de resistência do material seja de 95%, considerando o modelo probabilístico Weibull? A média e variância da resistência, segundo estudos anteriores, foi de 3 e 1, respectivamente.

Solução: Por condição do problema, temos que

$$E(X) = \frac{\Gamma\left(1 + \dfrac{1}{\alpha}\right)}{\beta^{\frac{1}{\alpha}}} = 3 \qquad \text{e} \qquad Var(X) = \frac{\Gamma\left(1 + \dfrac{2}{\alpha}\right) - \left(\Gamma\left(1 + \dfrac{1}{\alpha}\right)\right)^2}{\beta^{\frac{2}{\alpha}}} = 1$$

Da equação com respeito a esperança, $E(X)$, podemos isolar β, ou seja

$$\beta = \left(\frac{\Gamma(1 + \frac{1}{\alpha})}{3}\right)^{\alpha}.$$

Substituindo na equação da variância, $Var(X)$, temos

$$3^2 \, \frac{\Gamma(1 + \frac{2}{\alpha}) - \Gamma(1 + \frac{1}{\alpha})^2}{\Gamma(1 + \frac{1}{\alpha})^2} = 1$$

Finalmente, devemos encontrar o valor de α que resolve esta equação não linear. Para encontrar esta solução, podemos usar um programa numérico, por exemplo R, Matlab ou Excel, usando vários métodos.

Esses programas realizam os cálculos de forma iterativa até que a solução apresente um erro mínimo. Usando estes programas, temos que o valor de $\alpha = 3,3035$ e, portanto, $\beta = 0,0185$. Com estes valores podemos calcular V, tal que $P(X > V) = 0,95$. Temos que

$$P(X > V) = 1 - F(V) = 1 - \left(1 - e^{-0,0185V^{3,3035}}\right) = e^{-0,0185V^{3,3035}}$$

Agora temos que isolar V na seguinte equação

$$e^{-0,0185V^{3,3035}} = 0,95$$

Neste caso, obtemos

$$V = \left(-\frac{ln(0,95)}{0,0185}\right)^{1/3,3035} = 1,36$$

1.6 DISTRIBUIÇÕES DISCRETAS E CONTÍNUAS MAIS IMPORTANTES (RESUMO)

- **Processo ou ensaios Bernoulli:** experimentos aleatórios independentes (com substituição) que correspondem a dois resultados possíveis: $X = 1$ (sucesso) ou $X = 0$ (fracasso), na qual $P(X = 1) = p$ e $P(X = 0) = q = 1 - p$.

- Na distribuição Hipergeométrica os ensaios são sem substituição.

- **Processo Poisson:** Diz-se que um conjunto de eventos discretos é gerado por um processo de Poisson de taxa ω se, para qualquer intervalo I (geralmente de tempo) de comprimento suficientemente pequeno $h > 0$:

  (i) $P(\text{ocorrência de um evento en } I) = \omega h$,

  (ii) $P(\text{ocorrência de 2 ou mais eventos em } I) = 0$,

  (iii) A ocorrência de eventos em intervalos disjuntos do tipo I são independentes.

- **Processo Bernoulli:** $B(n,p)$, $\mathrm{B}N(r,p) \longrightarrow G(p)$

- **Processo Poisson:** $P(\lambda)$, $\Gamma(\alpha,\beta) \longrightarrow Exp(\beta)$

- **Normal:** $N(\mu,\sigma^2)$

- **Beta-uniforme:** $Beta(\alpha,\beta) \longrightarrow U(0,1)$

Tabela 1.1: Distribuições Comuns.

Nome	**Notação:** $X \sim$	**Definição**	**Parâmetro(s)**
Binomial	$B(n, p)$	Número de sucessos em n ensaios	$p = P(\text{sucesso})$
Geométrica	$G(p)$	Número de ensaios até conseguir o primeiro sucesso	$p = P(\text{sucesso})$
Binomial negativa	$\mathrm{B}N(r, p)$	Número de ensaios até conseguir o r-ésimo sucesso	$p = P(\text{sucesso})$
Hipergeométrica	$H(N, M, n)$	Número de elementos do tipo A na amostra n	N: população, M: do tipo A, n: amostra
Poisson	$P(\lambda)$	Número de eventos em $[0, t]$ gerados por um processo	$\lambda = \omega\, t$, ω: taxa, t: unidade (tempo)
Exponencial	$Exp(\beta)$	Tempo transcorrido até a ocorrência do 1º evento	$\beta = \omega$
Gama	$\Gamma(\alpha, \beta)$	Tempo transcorrido até a ocorrência do α-ésimo evento	$\beta = \omega$ e α o número de sucessos
Weibull	$W(\alpha, \beta)$	Distribuição assimétrica contínua positiva	α, β
Normal	$N(\mu, \sigma^2)$	Distribuição simétrica contínua na reta	μ: média, σ^2: variância
Uniforme	$U(\alpha, \beta)$	Distribuição contínua, uniforme no intervalo $[\alpha,\ \beta]$	α: limite inferior do intervalo, β: limite superior do intervalo
Beta	$Beta(\alpha, \beta)$	Proporção	α, β

Tabela 1.2: Distribuições de probabilidade, suporte, média e variância das distribuições comuns.

Notação: $X \sim$	Distribuição de probabilidade	Suporte	$\mu_x = E(X)$	$V(X) = \sigma_X^2$
$B(n,p)$	$C_X^n q^{n-x} p^x$	$x = 0, 2, \cdots, n$	np	npq
$G(p)$	$q^{x-1}p$	$x = 1, 2, 3, \cdots$	$1/p$	q/p^2
B$N(r,p)$	$C_{r-1}^{x-1} q^{x-r} p^r$	$r, r+1.r+2, \cdots$	r/p	rq/p^2
$H(N,M,n)$	$\dfrac{C_x^M C_{n-x}^{N-M}}{C_n^N}$	$x = 0, 2, \cdots, n$	$n\dfrac{M}{N}$	$n(\frac{M}{N})(\frac{N-M}{N})(\frac{N-n}{N-1})$
$P(\lambda)$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$	$x = 0, 1, 2, 3, \cdots$	λ	λ
$Exp(\beta)$	$\beta e^{-\beta X}$	$x > 0$	$1/\beta$	$1/\beta^2$
$\Gamma(\alpha,\beta)$	$\dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$	$x > 0$	α/β	α/β^2
$W(\alpha,\beta)$	$\alpha\beta x^{\alpha-1} e^{-\beta x^\alpha}$	$x > 0$	$\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\beta^{1/\alpha}}$	$\frac{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)}{\beta^{2/\alpha}}$
$N(\mu,\sigma^2)$	$\dfrac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$	$-\infty < x < \infty$	μ	σ^2
$U(\alpha,\beta)$	$\dfrac{1}{\beta-\alpha}$	$\alpha < x < \beta$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$
$Beta(\alpha,\beta)$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 < x < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$