Laboratório Nacional de Computação Científica
UNIDADE DE PESQUISA DO MCTI

Distribuições de probabilidade

Modelos Compartimentais em Epidemiologia e Inferência Bayesiana

Gustavo Libotte e Regina Almeida

10 de janeiro de 2022

Distribuições de probabilidade discretas
Distribuição binomial

- Essa distribuição de probabilidade descreve processos nos quais um evento pode ter apenas **um de dois resultados possíveis**.
- Os exemplos incluem jogar uma moeda, detectar algo em uma verificação de segurança ou ganhar na loteria.
- Seja p a probabilidade de um evento, o qual denominamos de "sucesso"; $1 - p$ é a probabilidade do outro evento ("falha").
- Se a tentativa for repetida independentemente n vezes, estamos interessados na probabilidade de obter exatamente r sucessos, que podemos rotular de $P\left(r \mid p, n\right)$.
- Suponha que as primeiras r tentativas sejam bem-sucedidas e o restante, $n - r$, sejam todas falhas.
- Como as tentativas são **independentes**, a probabilidade dessa sequência é apenas o produto das probabilidades, $p^r\left(1-p\right)^{n-r}$.

Distribuições de probabilidade discretas
Distribuição binomial

- Esta é a probabilidade de apenas uma sequência particular de tentativas. O número de sequências únicas com essa probabilidade é o número de maneiras de selecionar r a partir de n (sem substituição), que é ${}_nC_r$.
- O número de maneiras de selecionar r objetos de um conjunto de n sem levar em conta a ordem de seleção é chamado de **número de combinações** ${}_nC_r$,

$$ {}_nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} $$

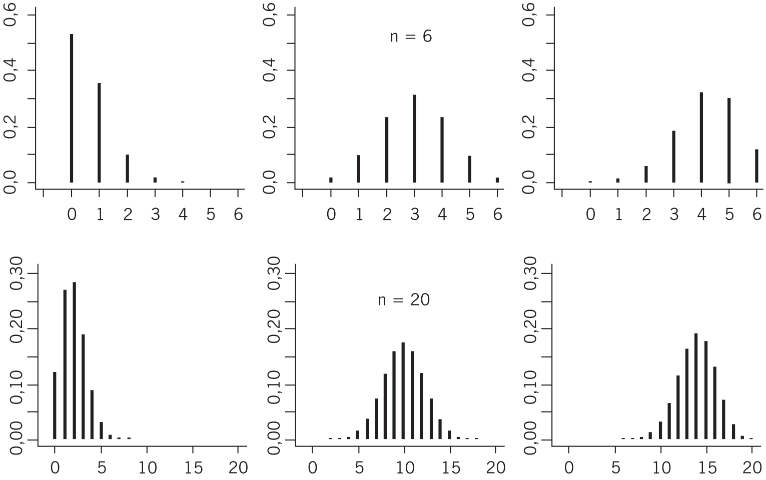
- O valor ${}_nC_r$ é também chamado de **coeficiente binomial**.
- A distribuição binomial é dada por

$$ P(r \mid p, n) = {}_nC_r p^r (1-p)^{n-r} = \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} $$

- Esta é a probabilidade de obter r sucessos de n tentativas se a probabilidade de um sucesso for p, onde é claro $0 \leq r \leq n$.

Distribuições de probabilidade discretas

Distribuição binomial



p = 0,1 p = 0,5 p = 0,7

Distribuições de probabilidade discretas
Distribuição binomial

- **Exemplo**: Suponha que a probabilidade de falha de uma estrutura devido a terremotos seja estimada em 10^{-5} por ano. Supondo que a vida útil do projeto da estrutura seja de 50 anos e a probabilidade de falha em cada ano permaneça constante e independente durante sua vida útil, então a probabilidade de nenhuma falha pode ser estimado usando a distribuição binomial como

$$
\begin{aligned}
P(\text{não haver falha em 50 anos}) &= P\left(0 \mid 10^{-5}, 50\right) \\
&= \binom{50}{0}\left(10^{-5}\right)^{0}\left(1-10^{-5}\right)^{50-0} = \frac{50!}{0!(50-0)!}\left(1-10^{-5}\right)^{50} \\
&\approx 0{,}99950 \\
P(\text{haver falha em 50 anos}) &= 1 - P(\text{não haver falha em 50 anos}) \\
&= 1 - 0{,}99950 = 0{,}00050.
\end{aligned}
$$

- **Mais um exemplo**: Geralmente, em cerca de 80% dos chamados que um certo técnico em computação recebe para resolver panes nos computadores de clientes ele constata que o problema decorreu da presença de algum vírus.
- Suponha que, em um determinado dia, esse técnico vai visitar seis desses clientes cujos computadores necessitam de conserto.
- Admita também que os seis clientes não se comunicam por meio de computador (o que garante a independência da existência de vírus em cada computador). Calcule a probabilidade de que:
  - a) Pelo menos quatro entre os seis computadores estejam com vírus.
  - b) No máximo dois dentre eles estejam com vírus.
  - c) Todos os seis estejam com vírus.

Distribuições de probabilidade discretas
Distribuição binomial

- Considere:
  - Sucesso = "o defeito no computador é devido a presença de vírus" $(p = P(\text{ sucesso }) = 0,80)$.
  - X = número de computadores com vírus entre os 6 a serem consertados.
- a) Então $X \sim \text{Bin}(4 \mid 0{,}8, 6)$.
- $P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) =$

$$= \left(\begin{array}{c} 6 \\ 4 \end{array}\right) 0,8^4 \times 0,2^2 + \left(\begin{array}{c} 6 \\ 5 \end{array}\right) 0,8^5 \times 0,2 + \left(\begin{array}{c} 6 \\ 6 \end{array}\right) 0,8^6 = 0,90112.$$

- Isso significa que é bem alta a probabilidade de pelo menos quatro entre os seis computadores estarem com vírus.
- b) $\mathrm{P}(\mathrm{X} \leq 2) = \mathrm{P}(\mathrm{X} = 0) + \mathrm{P}(\mathrm{X} = 1) + \mathrm{P}(\mathrm{X} = 2) =$

$$= \left(\begin{array}{c} 6 \\ 0 \end{array}\right) 0,2^6 + \left(\begin{array}{c} 6 \\ 1 \end{array}\right) 0,8 \times 0,2^5 + \left(\begin{array}{c} 6 \\ 2 \end{array}\right) 0,8^2 \times 0,2^4 = 0,01696.$$

- Este valor indica que é baixíssima a probabilidade de que no máximo dois deles estejam com vírus.
- c) $\mathrm{P}(\mathrm{X} = 6) = 0,8^6 = 0,26214$. Finalmente, não é tão pequena a probabilidade de que todos estejam com vírus.

Distribuições de probabilidade discretas
Distribuição de Poisson

- A distribuição binomial descreve eventos em que há um evento definido ocorrendo que tem um resultado bidirecional: é um "sucesso" ou um "fracasso"; algo acontece ou não.
- No entanto, muitos processos naturais são unilaterais, isto é, só são identificáveis por terem acontecido.
- Os exemplos são **quedas de raios** e **emissão de partículas** α de uma fonte radioativa.
- Não podemos contar não-eventos porque **não podemos identificar uma sequência de eventos** onde algo deveria acontecer ou não.
- Suponha que, em média, esses eventos ocorram a uma taxa λ, de modo que λ é o **número esperado de eventos em algum intervalo de tempo**. Gostaríamos de encontrar a probabilidade de obtermos r eventos neste intervalo.

Distribuições de probabilidade discretas
Distribuição de Poisson

- Podemos descrever isso como o **limite de um processo binomial**. Se tomarmos a série temporal em n divisões, então no limite, conforme as divisões se tornam muito pequenas, de modo que n fica grande, podemos escrever $p = \lambda/n$. A distribuição binomial é

$$P(r \mid \lambda/n, n) = \frac{n!}{r!(n-r)!} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \tag{1}$$

- Agora levamos n para infinito para produzir um *continuum* de eventos. Como $n \to \infty$ com r finito,

$$\frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-r+1) \to n^r \tag{2}$$

- porque cada termo tende para n e

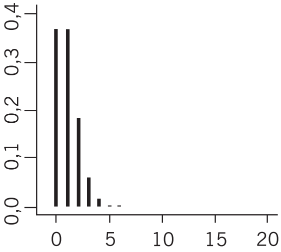
$$\left(1 - \frac{\lambda}{n}\right)^{n-r} \to \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda} \tag{3}$$

[1] A Equação (3) é uma definição da **constante de Euler** (e).

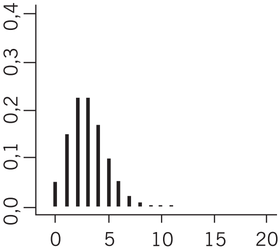
Distribuições de probabilidade discretas
Distribuição de Poisson

- Inserindo os termos das Eqs. (2) e (3) na Eq. (1), n^r cancela e terminamos com a **distribuição de Poisson**
$$P(r \mid \lambda) = \frac{e^{-\lambda}\lambda^r}{r!}$$
onde $\lambda > 0$ e $r \geq 0$.
- Esta é a probabilidade de se obter r eventos se o número médio esperado for λ, ou seja, $\mathbb{E}(r) = \lambda$.
- Uma propriedade importante desta distribuição é que sua variância é $\mathrm{Var}(r) = \lambda$, ou seja, **igual à média**.
- Além disso:
$$\begin{aligned} P(r+1 \mid \lambda) &= \frac{e^{-\lambda}\lambda^{r+1}}{(r+1)!} \\ &= \frac{e^{-\lambda}\lambda^r}{r!}\frac{\lambda}{r+1} \\ &= P(r \mid \lambda)\frac{\lambda}{r+1} \end{aligned}$$

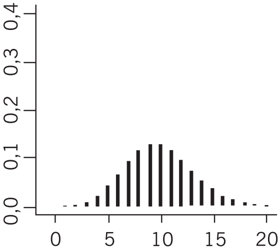
Distribuições de probabilidade discretas
Distribuição de Poisson



$\lambda = 1$



$\lambda = 3$



$\lambda = 10$

- **Exemplo**: Admita que o número de falhas no recapeamento de um fio condutor de eletricidade obedece a uma distribuição de Poisson e que em média há duas falhas por metro.
- Qual a probabilidade de que:
  - a) em um determinado metro de fio o recapeamento apresente três falhas?
  - b) em sete metros de fio sejam encontradas no máximo 10 falhas?

Solução:

- a) Seja X a variável aleatória que representa o número de falhas num dado metro de fio. Então $X \sim \text{Poisson}\,(3 \mid 2)$. Assim, $P(X = 3 \mid 2) = \frac{e^{-2}2^3}{3!} = 0,1804$.
- b) Nesse caso a unidade de longitude considerada é sete metros. Logo, para essa nova situação a taxa média por unidade é $\lambda = 2 \times 7 = 14$.

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- A **distribuição normal** geralmente ocorre em aplicações práticas, por conta do seu papel no *teorema central do limite*, que afirma que a soma de variáveis aleatórias de distribuição arbitrária segue assintoticamente uma distribuição normal quando o tamanho da amostra se torna suficientemente grande [1].
- A variável aleatória X, que tome todos os valores reais $-\infty < x < \infty$, tem uma **distribuição normal** (ou **distribuição Gaussiana**) se a sua função densidade de probabilidade for da forma

$$f_X\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right), \tag{4}$$

onde $\mu_X,\ \sigma_X > 0$ são os parâmetros de posição e dispersão da variável aleatória.

[1] A. F. Karr.
Probability.
Springer New York, New York, 1 edition, 1993.

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- A **função de distribuição acumulada** correspondente à distribuição normal é dada por

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \mathrm{d}x \,.$$

- Calcular o **valor esperado** de uma variável aleatória X com distribuição normal significa obter o resultado de

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \mathrm{d}x \,. \tag{5}$$

- Fazendo a transformação linear estabelecida por $Z = (X - \mu_X)/\sigma_X$ e observando que $\mathrm{d}x = \sigma_X\, \mathrm{d}z$, a expressão anterior toma a forma

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma_X z + \mu_X)\, \mathrm{e}^{-z^2/2}\, \mathrm{d}z \\
&= \sigma_X \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\mathrm{e}^{-z^2/2}\, \mathrm{d}z + \mu_X \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2}\, \mathrm{d}z \,.
\end{aligned}$$

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- Nessa soma de integrais, o termo da esquerda possui integrando $g(z)$ que satisfaz a propriedade $g(z) = -g(z)$, isto é, $g(z)$ é uma **função ímpar** e o cálculo da sua integral (que tem intervalos simétricos) resulta em zero.
- A segunda integral é igual a um, pois representa a área total abaixo da curva que descreve a função densidade de probabilidade, sem considerar o termo μ_X.
- Portanto, a Eq. (5) equivale a $\mathbb{E}[X] = \mu_X$.
- Agora, considere

$$\mathbb{E}\left[X^2\right] = \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \mathrm{d}x \,.$$

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- Por meio da mesma transformações realizada no caso anterior, a expressão de $\mathbb{E}\left[X^2\right]$ assume a forma

$$\mathbb{E}\left[X^2\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\sigma_X z + \mu_X\right)^2 \mathrm{e}^{-z^2/2}\, \mathrm{d}z$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma_X^2 z^2 \mathrm{e}^{-z^2/2}\, \mathrm{d}z + 2\mu_X\sigma_X \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\mathrm{e}^{-z^2/2}\, \mathrm{d}z + \mu_X^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2}\, \mathrm{d}z\,.$$

- Nesse caso, para calcular a primeira integral, é possível proceder com a técnica de **integração por partes**, por meio das escolhas de $u = z$ e $\mathrm{d}v = z\mathrm{e}^{-z^2/2}\,\mathrm{d}z$. Dessa forma, o seu resultado (**cujo desenvolvimento não será detalhado aqui**) será igual a um.
- No caso da resolução da segunda e terceira integrais desse somatório, pode-se notar que os seus resultados já foram encontrados durante a análise de $\mathbb{E}\left[X\right]$, realizada anteriormente.
- Utilizando os mesmos argumentos já citados, essas integrais valem zero e um, respectivamente.

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- Logo, $\mathbb{E}\left[X^2\right] = \sigma_X^2 + \mu_X^2$.
- Recuperando a formulação da variância,

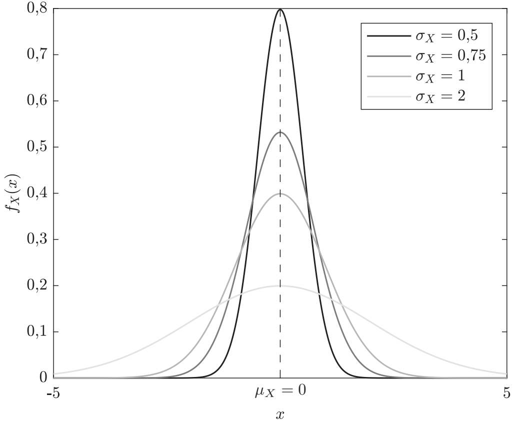
$$\mathrm{Var}\left[X\right] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 ,$$

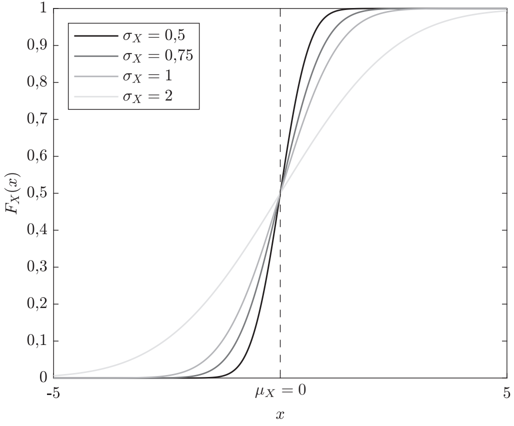
  e comparando-a com os resultados obtidos nessa análise, é possível concluir que $\mathrm{Var}\left[X\right] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 = \sigma_X^2 + \mu_X^2 - \mu_X^2 = \sigma_X^2$.
- Deste modo, **verifica-se que** os dois parâmetros, μ_X e σ_X^2, que caracterizam a distribuição normal, são a **média** e a **variância** da variável aleatória X, respectivamente.

Variável aleatória com distribuição normal

Portanto, diz-se que uma variável aleatória X tem distribuição normal, denotada como $X \sim \mathcal{N}\left(\mu_X,\ \sigma_X^2\right)$, com média $\mu_X \in \mathbb{R}$, tal que $-\infty < \mu_X < \infty$, e variância $\sigma_X^2 \in \mathbb{R}_+^*$, se a sua densidade de probabilidades for dada pela Eq. (4).

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)





Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- A distribuição Gaussiana pode ser **normalizada** pela transformação da variável $Z = (X - \mu_X) / \sigma_X$, utilizada anteriormente.
- Assumindo uma nova variável normal padrão Z, com média igual a zero e desvio-padrão igual a um, a função densidade de probabilidade da Eq. (4) assume a forma

$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) . \tag{6}$$

- Naturalmente, a função de distribuição acumulada também está sujeita à transformação na variável aleatória, passando a ser representada como

$$\Phi_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}\theta^2\right) \mathrm{d}\theta .$$

- Especificamente nesse caso em que $Z \sim \mathcal{N}(0,\ 1)$, é dito que Z possui **distribuição normal padrão**.

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- **Exemplo**: A carga de ruptura de cabos de aço de 8mm, usados em guinchos e produzidos por uma certa Companhia têm uma distribuição Normal com média de 2.210 kg e desvio padrão de 25 kg.
- A especificação mínima para a dita carga é de 2.180 kg.
- Cabos com carga de ruptura entre 2.130 e 2.180 kg ainda podem ser comercializados, porém a um preço menor, mas se tiverem carga de ruptura inferior a 2.130kg devem ser descartados.
    - Qual a porcentagem de cabos que satisfazem a especificação?
    - Qual a porcentagem de cabos que, mesmo não satisfazendo a especificação, poderiam ser vendidos?
    - Qual a porcentagem de cabos que deveriam ser descartados?

Distribuições de probabilidade contínuas
Distribuição normal (Gaussiana)

- Seja X a variável aleatória que representa a carga de ruptura, em kg, dos cabos de aço. Então, $X \sim \mathcal{N}\left(2210, 25^2\right)$.
- a) Os cabos satisfarão a especificação se $X > 2180$.

$$\mathrm{P}(\mathrm{X} > 2180) = 1 - \Phi\left(\frac{2180 - 2210}{25}\right) = 1 - \Phi(-1,2) = 0,8849$$

- Aproximadamente $88,5\%$ dos cabos produzidos satisfazem as especificações.
- b)

$$P(2130 < X < 2180) = \Phi\left(\frac{2180 - 2210}{25}\right) - \Phi\left(\frac{2130 - 2210}{25}\right) = \Phi(-1,2) - \Phi(-3,2)$$

$$= \Phi(3,2) - \Phi(1,2) = 0,9993 - 0,8849 = 0,1144.$$

- Aproximadamente $11,4\%$ dos cabos podem ser vendidos a um preço inferior.
- c) $\mathrm{P}(\mathrm{X} < 2130) = \Phi\left(\frac{2130-2210}{25}\right) = \Phi(-3,2) = 1 - \Phi(3,2) = 1 - 0,9993 = 0,0007$
- Portanto, deveriam ser descartados aproximadamente $0,07\%$ dos cabos produzidos.

Outras distribuições de probabilidade
Distribuição Gama

- Dizemos que X tem uma distribuição **gama** com parâmetros α e β e escrevemos $X \sim Ga(\alpha, \beta)$, se X tem função de densidade

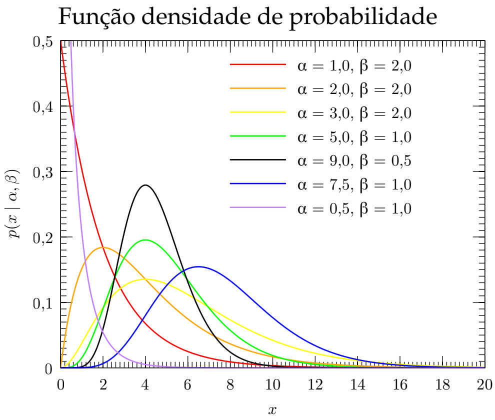
$$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0$$

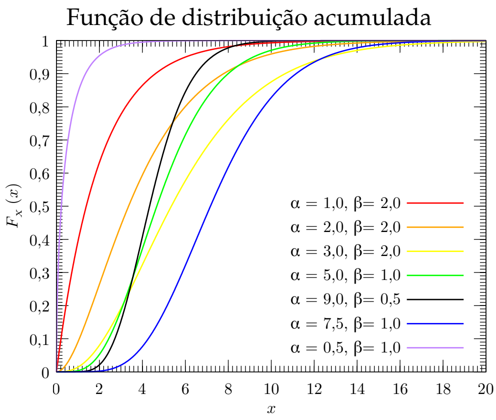
  para valores $\alpha, \beta > 0$. Aqui, $\Gamma(z)$ é a função gama, definida para $z > 0$ por

$$\begin{aligned} \Gamma(z) &= \int_0^\infty t^{z-1} \exp(-t) \mathrm{d}t \\ &= (z-1)! \end{aligned}$$

- O parâmetro α é referido como o parâmetro de **forma** e β como o parâmetro de **escala**.
- A **média** e a **variância** de uma distribuição gama são α/β e α/β^2, respectivamente.

Outras distribuições de probabilidade
Distribuição Gama





Outras distribuições de probabilidade
Distribuição Beta

- Dizemos que X tem uma distribuição **beta** com parâmetros $a, b > 0$ e escrevemos $X \sim Be(a, b)$, se X tem função de densidade

$$p(x \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

  para $0 < x < 1$.

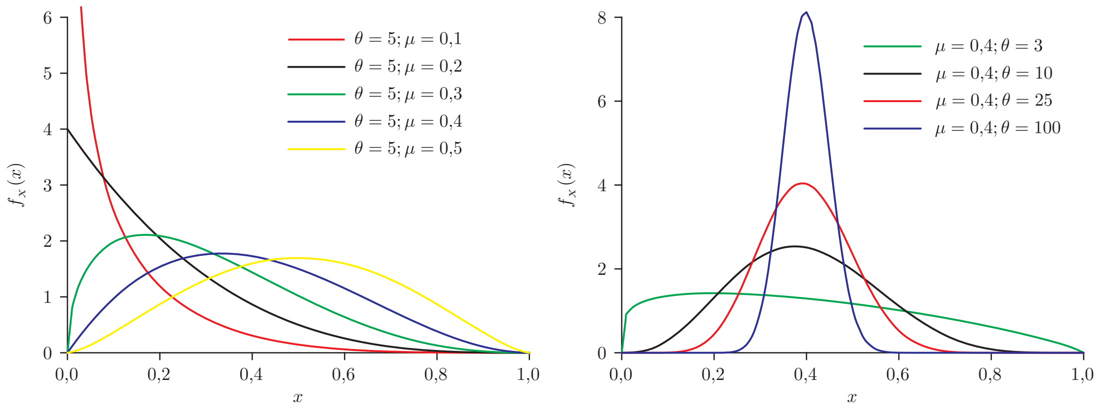
- A **média** e a **variância** da distribuição beta são, respectivamente,

$$\mu = \frac{a}{a+b} \qquad \sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mu(1-\mu)}{\theta+1}$$

  onde $\theta = a + b$.

- A **distribuição uniforme** em $[0, 1]$ é um membro da família beta, quando $a = 1$ e $b = 1$.

Outras distribuições de probabilidade
Distribuição Beta



Outras distribuições de probabilidade
Distribuição Binomial negativa

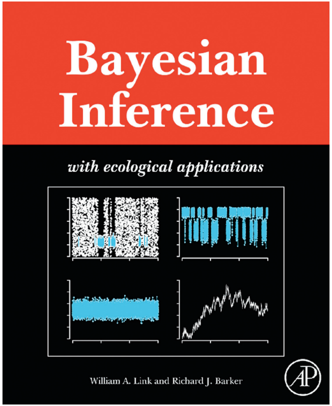
- Notamos que a distribuição de Poisson é frequentemente escolhida como ponto de partida para a análise de dados de **contagem**.
- A distribuição de Poisson é limitada pelo fato de que se $X \sim P(\lambda)$, então $\mathbb{E}(X) = \mathrm{Var}(X)$.
- Os dados de contagem são normalmente superdispersos em relação à distribuição de Poisson: **sua variância é maior do que sua média**.
- Uma **solução** é modelar a média λ da distribuição de Poisson como uma **variável aleatória**.
- A função de distribuição **binomial negativa** é dada por

$$\Pr(X = k) = \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)} \left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{1}{1+\beta}\right)^{k} \qquad \text{para } k = 0, 1, \ldots$$

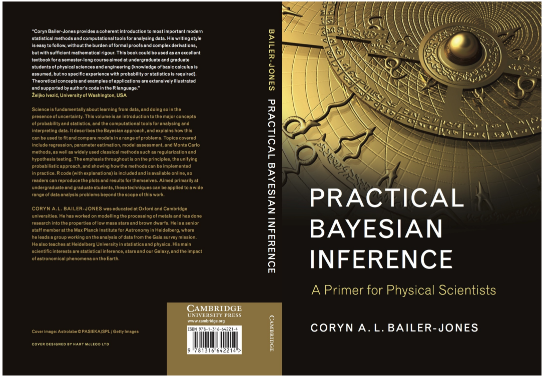
- A distribuição binomial negativa tem média e variância, respectivamente, iguais a

$$\mathbb{E}(X) = \alpha/\beta \qquad \mathrm{Var}(X) = \frac{\alpha(1+\beta)}{\beta^2}$$

Mais distribuições de probabilidade



[1] W. A. Link and R. J. Barker.
Bayesian Inference: with ecological applications.
Academic Press, London, 1 edition, 2010.



[1] C. A. L. Bailer-Jones.
Practical Bayesian Inference: a Primer for Physical Scientists.
Cambridge University Press, Cambridge, 1 edition, 2017.

Atividade prática

- Seja X uma variável aleatória com distribuição **normal**. Assuma que os parâmetros desta distribuição são desconhecidos.
- Em um determinado evento aleatório, obteve-se as seguintes amostras de X:

i	0	1	2	3	4	5	6	7	8
x_i	-2.0000	-1.5000	-1.0000	-0.5000	0	0.5000	1.0000	1.5000	2.0000
$f_X(x_i)$	0.0663	0.1093	0.1613	0.2130	0.2516	0.2660	0.2516	0.2130	0.1613

- Como podemos descobrir quais os parâmetros (média e desvio-padrão) da distribuição normal que melhor **aproximam** estes valores?
- Seja $y \sim \mathcal{N}\left(\mu, \sigma^2\right)$, para $0 \leq \mu \leq 10$ e $1 \leq \sigma \leq 10$, onde $y_i = y\left(x_i\right)$. Obtenha $y^* \sim \mathcal{N}\left(\mu^*, {\sigma^*}^2\right)$ tal que $\left(\mu^*, {\sigma^*}^2\right) = \underset{\mu,\sigma}{\operatorname{argmin}} \sum_{i=0}^{8} \left(y_i - f_X(x_i)\right)^2$.