

AGA 0505 - Análise de Dados em Astronomia

2. Distribuições de Probabilidades

Laerte Sodré Jr.

1o. semestre, 2024

aula de hoje:

1. distribuições de probabilidades
2. a distribuição normal ou gaussiana
3. o teorema do limite central
4. distribuições discretas
5. distribuições contínuas
6. distribuições bivariadas
7. a distribuição gaussiana bivariada

É provável que coisas improváveis aconteçam

Aristóteles

o que são probabilidades?

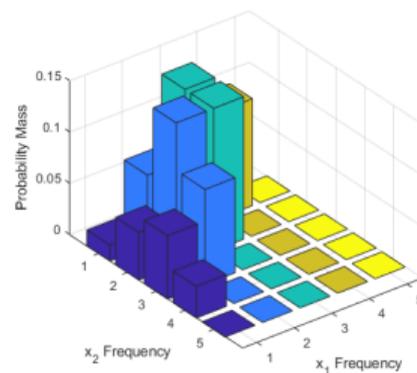
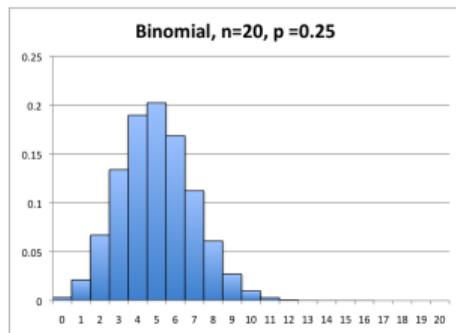
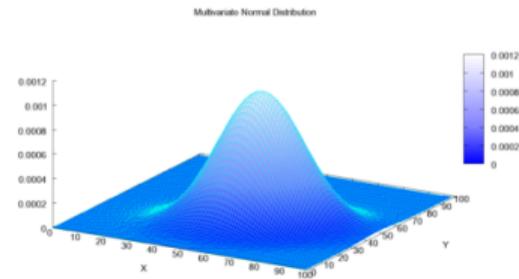
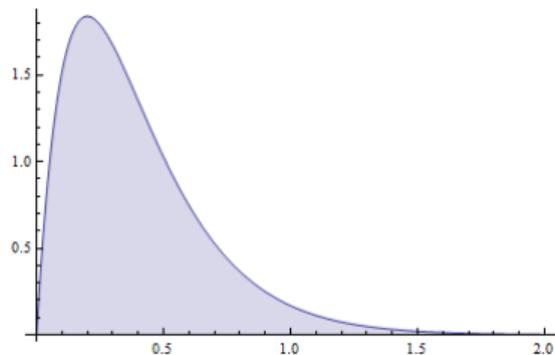
- probabilidade como medida da plausibilidade (chatGPT):
 - a probabilidade pode ser vista como uma medida da incerteza associada a uma determinada proposição ou evento, em que valores mais altos de probabilidade indicam que é mais plausível ou mais provável que a proposição ou evento seja verdadeiro, enquanto valores mais baixos indicam que é menos provável
 - exemplo: um meteorologista diz que há 70% de chance de chuva amanhã: com base na análise das condições climáticas atuais, é plausível que haja chuva, mas ainda há uma certa incerteza sobre isso; a probabilidade de 70% sugere que há uma maior chance de chuva do que de não chuva, mas ainda existe uma possibilidade significativa de que não chova
 - podemos considerar probabilidade como medida de frequência como um caso particular de medida de plausibilidade

distribuições de probabilidades

- vamos descrever medidas/observações x como uma variável aleatória, contínua ou discreta
- variável aleatória: variável cujos valores são resultados de um processo aleatório ou estocástico, obedecendo a uma certa distribuição de probabilidades, $P(x)$
- notação: $x \sim P(x)$
 x é uma variável aleatória com distribuição que obedece a $P(x)$
- $P(x)$ pode ser uma função discreta ou contínua
- $P(x)$ contínua: função de distribuição ou densidade de probabilidades (FDP ou PDF):
 $P(x)dx$: número entre 0 e 1 que mede o grau de plausibilidade de que uma certa variável x esteja entre x e $x + dx$
- no caso de variáveis discretas a distribuição de probabilidades é chamada de função de massa de probabilidades (FMPs)
- as FDPs/FMPs podem ser multivariadas, i.e., funções de várias variáveis: $P(x, y, z)$
- uma distribuição de probabilidades pode ser condicional, depender de outras variáveis:
Ex.: $P(x|\mu, \sigma)$: a distribuição de probabilidades de x depende das quantidades μ e σ
- $P(x|...)$: a distribuição depende de tudo que está a direita da barra |

distribuições de probabilidades

exemplos:



distribuições de probabilidades

- clássicas
 - uniforme
 - binomial
 - normal
 - Poisson
 - exponencial

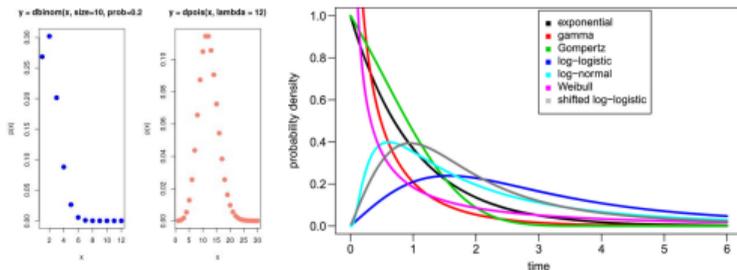
- fauna estatística
 - β , Cauchy, χ^2 , F, Γ , t, Weibull,...

- astrofísica
 - lei de potência
 - Schecter
 - log-normal
 - ...

dois tipos de distribuições:

- discretas

- contínuas



a distribuição normal ou gaussiana

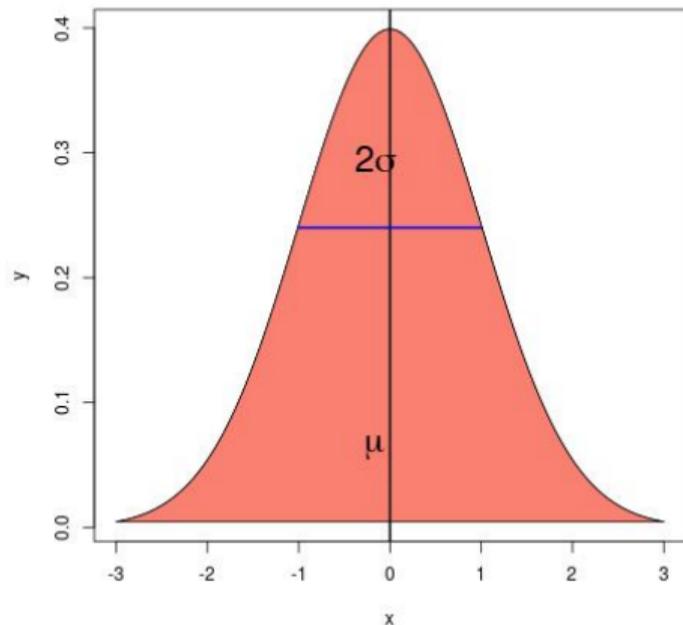
- distribuição gaussiana:

$$P(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

2 parâmetros:

- média μ
- desvio padrão σ (ou variância σ^2)

$p = 1/\sigma^2$ é denominado *precisão*



o teorema do limite central

- Teorema do Limite Central: em condições bem gerais, fazer médias produz uma distribuição gaussiana, *independentemente da distribuição a partir da qual os dados são gerados*
 - seja $\{x_1, x_2, \dots, x_n\}$ um conjunto de variáveis aleatórias amostradas de uma mesma distribuição de probabilidades $P(x)$, que tem média μ e variância σ^2

- para n tendendo a infinito, a média dos $\{x_i\}$ tende a

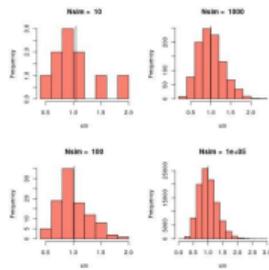
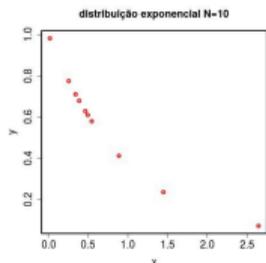
$$\frac{\sum_i x_i}{n} \longrightarrow \mu$$

e sua variância a

$$Var(\{x_i\}) \longrightarrow \frac{\sigma^2}{n},$$

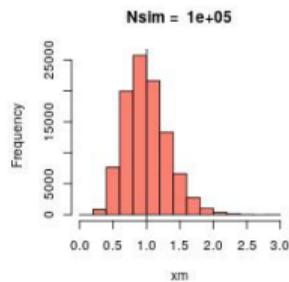
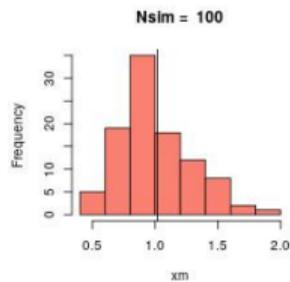
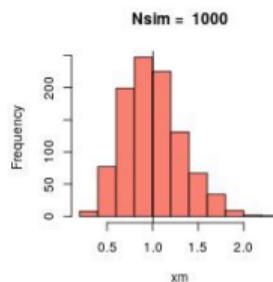
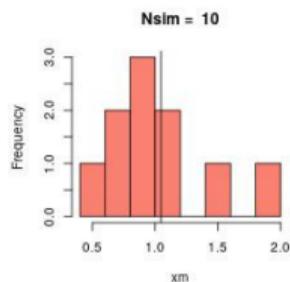
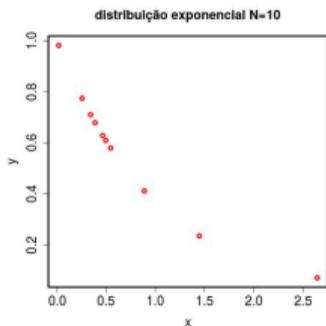
independentemente da forma de $P(x)$

N.B.: σ/\sqrt{n} : o “erro da média”



o teorema do limite central

- Teorema do Limite Central:
 - exemplo: *valores médios* em N_{sim} simulações de 10 pontos cada uma, com uma distribuição exponencial
 - conforme N_{sim} aumenta, a distribuição dos *valores médios* fica cada vez mais parecida com uma gaussiana



- uma consequência: os erros de médias de dados vão tender a ser 'gaussianos'

distribuições discretas

- espaço amostral constituído por variáveis aleatórias discretas:

$$S = \{x_1, x_2, \dots, x_k\}$$

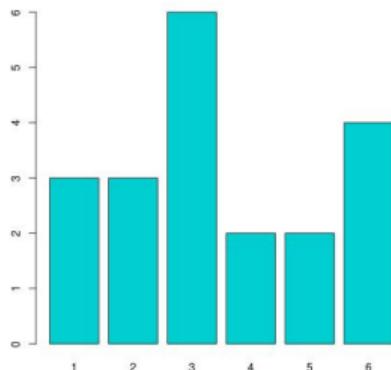
- exemplo: a distribuição uniforme discreta

$$P(x) = 1/k,$$

onde k é a dimensão do espaço amostral
moeda: $k=2$
dado de 6 faces: $k=6$

- média da população: $\mu = \sum_{l=1}^k lP(l) = \frac{k+1}{2}$
- variância da da população: $\sigma^2 = \sum_{l=1}^k (l - \mu)^2 P(l) = \frac{k^2-1}{12}$

- para um dado de 6 faces: $k=6$
 $\mu = 3.5$ e $\sigma = (35/12)^{1/2} \simeq 1.71$
- exemplo: amostra com simulação de 20 jogadas de um dado:
3 6 3 2 2 6 3 5 4 6 6 1 2 3 5 3 3 1 4 1
- média da amostra: $\bar{x} = \sum_{x=1}^{20} x_i P(x_i) = 3.45$
- desvio padrão da amostra:
 $\hat{\sigma}^2 = V^{1/2} = \left[\sum_{x=1}^{20} (x_i - \bar{x})^2 P(x_i) \right]^{1/2} \simeq 1.73$



distribuição binomial

aplica-se quando há apenas dois resultados possíveis:
sucesso/falha, detecção/não-deteção, cara/coroa

- p : probabilidade de sucesso em cada tentativa (é a mesma para todas as tentativas)
- as tentativas são independentes entre si
- probabilidade de n sucessos em N tentativas:

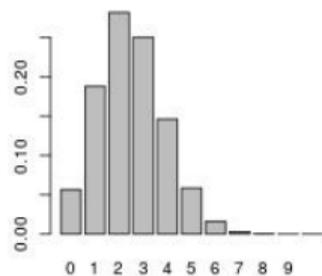
$$P(n|p, N) = \binom{N}{n} p^n (1-p)^{N-n} = \frac{N(N-1)\dots(N-n+1)}{n(n-1)\dots 1} p^n (1-p)^{N-n} = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

- média e variância:

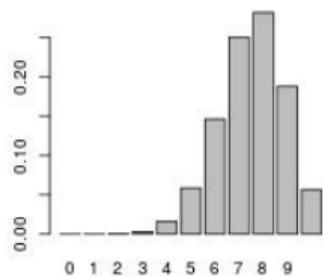
$$\bar{n} = \sum_{n=0}^N nP(n) = Np \quad \sigma^2 = \sum_{n=0}^N (n - \bar{n})^2 P(n) = Np(1-p)$$

distribuição binomial

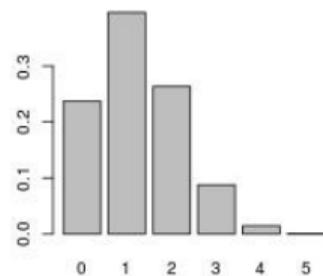
N,p: 10 0.25



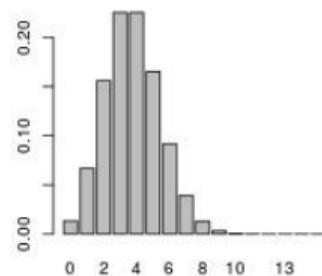
N,p: 10 0.75



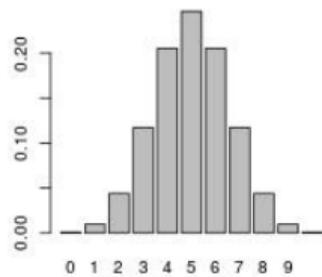
N,p: 5 0.25



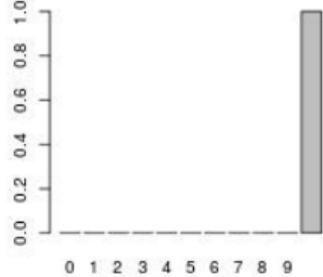
N,p: 15 0.25



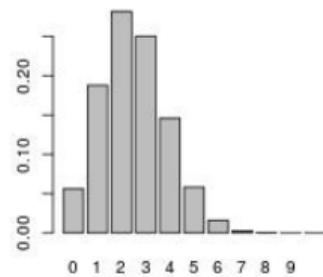
N,p: 10 0.5



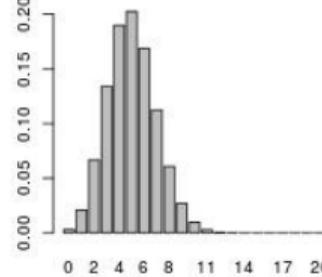
N,p: 10 1



N,p: 10 0.25



N,p: 20 0.25



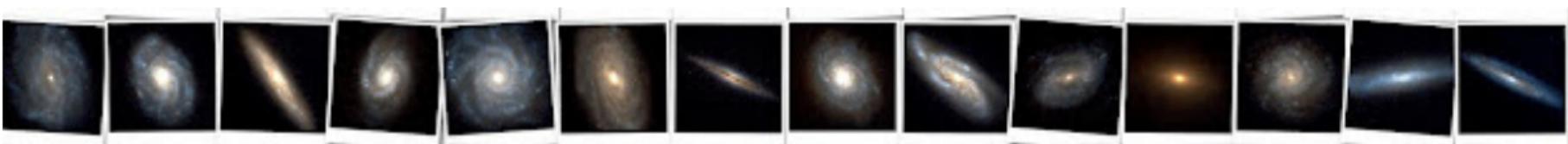
a distribuição binomial

- exemplo: em uma certa região do céu espera-se o mesmo número de estrelas e galáxias em um certo intervalo de magnitudes
 - considerando-se 10 objetos, qual é a probabilidade de se ter 8 ou mais galáxias?
 - no caso: $N = 10$ e $p = 1/2$
 - probabilidade de $n \geq 8$ sucessos em N tentativas:

$$\begin{aligned}P(n \geq 8|p, N) &= \sum_{n=8}^N \binom{N}{n} p^n (1-p)^{N-n} = \\ &= \frac{1}{2^{10}} \sum_{n=8}^N \binom{N}{n} = \frac{56}{124} \simeq 0.055\end{aligned}$$

- assim, a probabilidade de $n \geq 8$ galáxias em $N = 10$ objetos é $\sim 5.5\%$

a distribuição binomial



- exemplo: Um estudante tem que classificar 12 galáxias em 5 tipos diferentes. Suponha que todos os tipos tenham a mesma probabilidade. Se o estudante classificar as galáxias ao acaso, qual é a probabilidade de se obter até 4 classificações corretas?
 - como os 5 tipos têm a mesma probabilidade, a probabilidade de cada um é $p = 1/5 = 0.2$
 - a probabilidade de se obter ao acaso $n \leq 4$ classificações corretas para $N = 12$ galáxias é dada por

$$P(n \leq 4 | p, N) = \sum_{n=0}^4 \binom{N}{n} p^n (1-p)^{N-n} \simeq 0.92744$$

a distribuição de Poisson

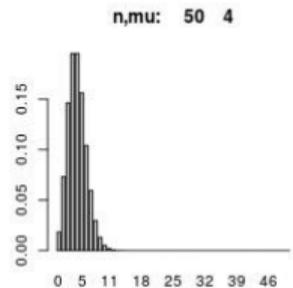
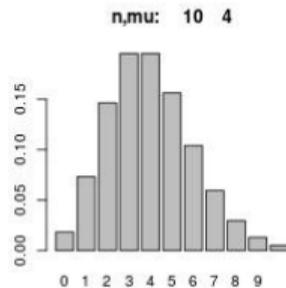
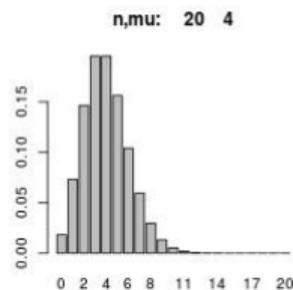
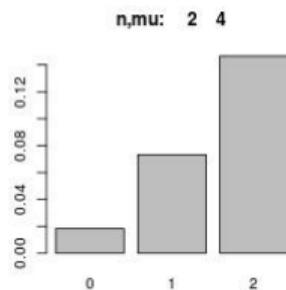
- probabilidade de n eventos ocorrerem num certo intervalo de tempo, ou em certa região do espaço, se estes eventos ocorrem independentemente e com uma média fixa μ

$$P(n|\mu) = \frac{\mu^n}{n!} e^{-\mu}$$

- média e variância:

$$\bar{n} = \sum_{n=0}^N nP(n) = \mu$$

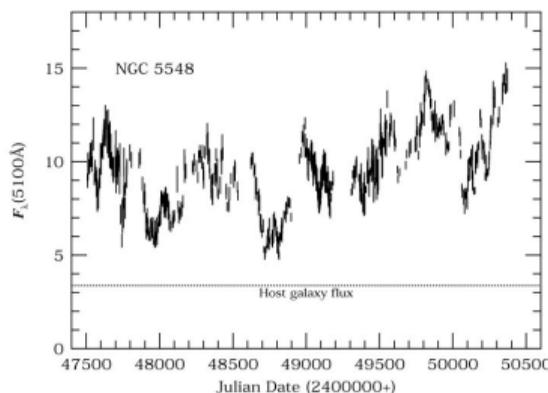
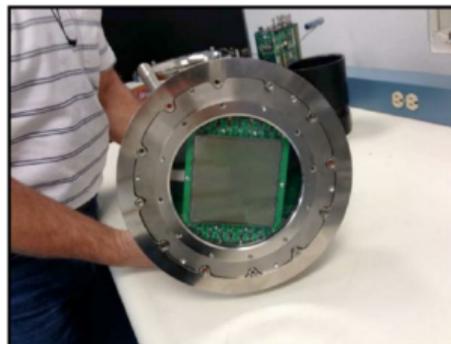
$$\sigma^2 = \sum_{n=0}^N (n - \bar{n})^2 P(n) = \mu$$



a distribuição de Poisson

aplicações:

- contagens em detectores
- densidade de estrelas/galáxias
- explosões de supernovas
- variabilidade de AGNs
- ...



a distribuição de Poisson

exemplo:

- em um detector, o número esperado de fótons de uma fonte é de 4.5 por segundo; qual é a probabilidade de se detectar 6 fótons em 2 segundos?
 - taxa de fótons: $\lambda = 4.5/\text{sec}$
 - número esperado no intervalo de tempo t : $\mu = \lambda t = 4.5 \times 2 = 9$
 - probabilidade:

$$P(n|\mu) = \frac{\mu^n}{n!} e^{-\mu} = \frac{9^6}{6!} e^{-9} \simeq 0.09$$

distribuições contínuas: a distribuição exponencial

- descreve o intervalo de tempo entre eventos que obedecem a um processo poissoniano, isto é, que ocorrem continuamente e independentemente, com uma taxa média constante, $\lambda > 0$:

$$P(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$

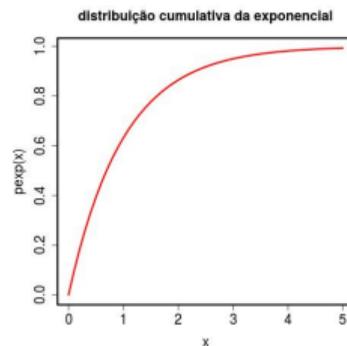
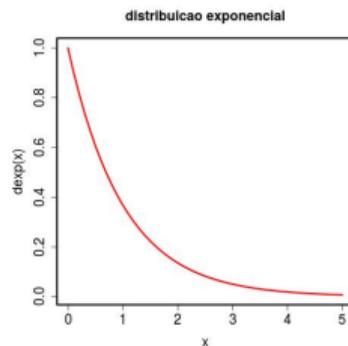
- média e variância:

$$\bar{x} = \frac{1}{\lambda}$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

- distribuição cumulativa:

$$F(x|\lambda) = \begin{cases} 1 - e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$



distribuições contínuas: a distribuição exponencial

exemplo:

- a taxa de supernovas média em uma galáxia como a Via Láctea é de 1 em 50 anos
- a última supernova conhecida que explodiu na Via Láctea foi a de Kepler, em 1604
- qual é a probabilidade de se esperar mais de 400 anos para se observar uma nova supernova?
- nesse caso,

$$\begin{aligned} \int_{t_0}^{\infty} \lambda e^{-\lambda t} dt &= e^{-\lambda t_0} = \\ &= e^{-400/50} = e^{-8} \simeq 0.00033546 \end{aligned}$$

ou, $\sim 0.03\%$

- a distribuição exponencial não tem memória: o passado não afeta a probabilidade futura

a distribuição uniforme contínua

- distribuição uniforme (entre 0 e X):

$$P(x|X) = \begin{cases} \frac{1}{X}, & \text{se } 0 \leq x \leq X \\ 0, & \text{de outro modo} \end{cases}$$

- média

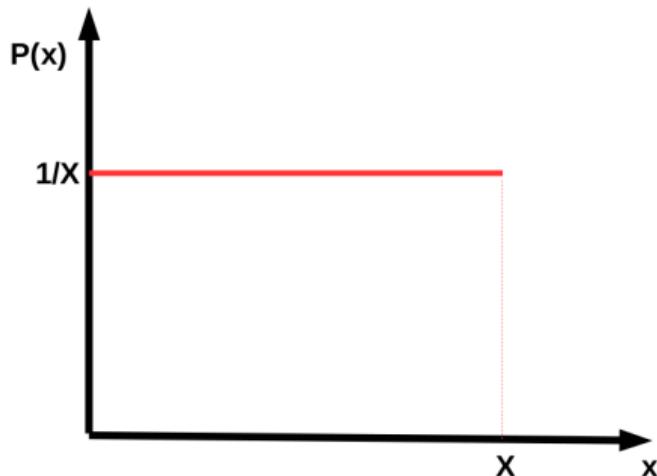
$$E(x) = \int_0^X x \frac{1}{X} dx = \frac{X}{2}$$

- variância

$$\sigma^2 = E(x^2) - E(x)^2 = \frac{X^2}{3} - \frac{X^2}{4} = \frac{X^2}{12}$$

- distribuição cumulativa:

$$F(x) = \int_0^x \frac{1}{X} dx = \frac{x}{X}$$



a versátil distribuição beta

- definida no intervalo $0 < x < 1$

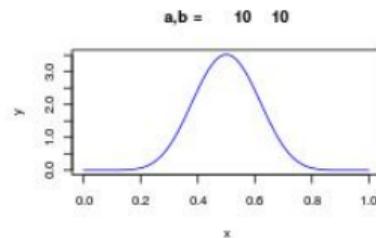
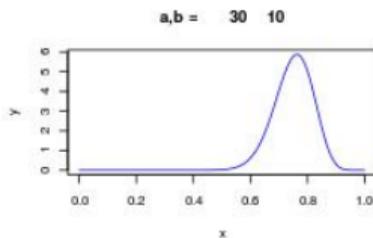
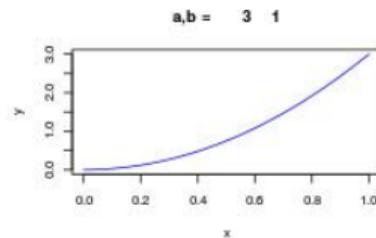
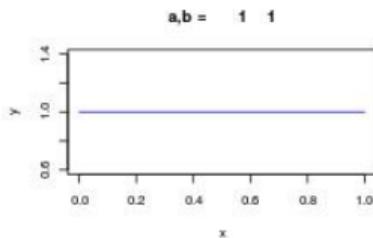
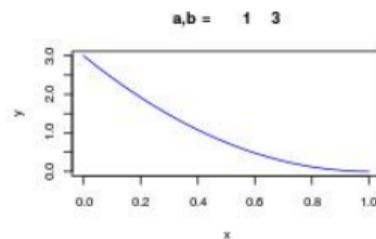
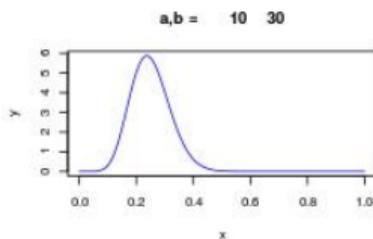
$$P(x|a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1 - x)^{b-1} \quad (a, b > 0)$$

função gama:

$$\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} dz \quad \Gamma(n) = (n - 1)!$$

- propriedades:

- $E(x) = a/(a + b)$
- $var(x) = ab/[(a + b)^2(a + b + 1)]$
- $moda(x) = (a - 1)/(a + b - 2)$



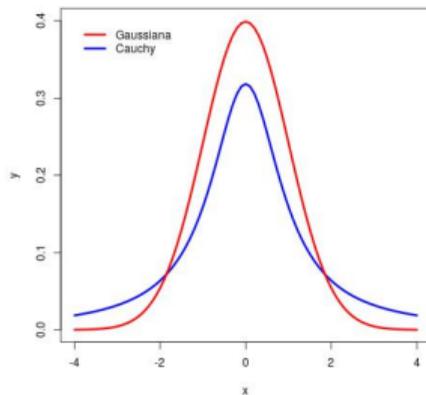
a distribuição de Cauchy

- distribuição de Cauchy ou lorentziana:

$$P(x|\mu, \gamma) = \frac{\gamma}{\pi[\gamma^2 + (x - \mu)^2]}$$

- parâmetro de localização: μ
parâmetro de escala: $\gamma > 0$
- como $P(x) \propto 1/x^2$ para $x \gg 1$, a média, variância e desvio padrão não existem (“caudas pesadas”)
- dada uma amostra $\{x_i\}$ com distribuição de Cauchy, μ e γ devem ser estimados com métodos robustos (como medianas e quartis)

- se x e y são duas variáveis gaussianas independentes amostradas de $N(0, 1)$, então $z = x/y$ distribui-se como Cauchy com $\mu = 0$ e $\gamma = 1$



distribuição em lei de potência

muito comum em Astronomia:

- exemplos: contagens de galáxias, função de massa inicial de Salpeter
- também muito comum em outras áreas: flutuações no mercado econômico, taxa de crescimento de empresas, distribuição de salários, etc
- $N(> L)$: número de objetos ou eventos com uma dada propriedade medida (e.g. a luminosidade) maior do que um certo valor L

- se a distribuição for uma lei de potência com expoente γ , temos:

$$N(> L) \propto L^{\gamma+1} \quad (\text{forma integral})$$

ou

$$dN \propto (\gamma + 1)L^\gamma dL \quad (\text{forma diferencial})$$

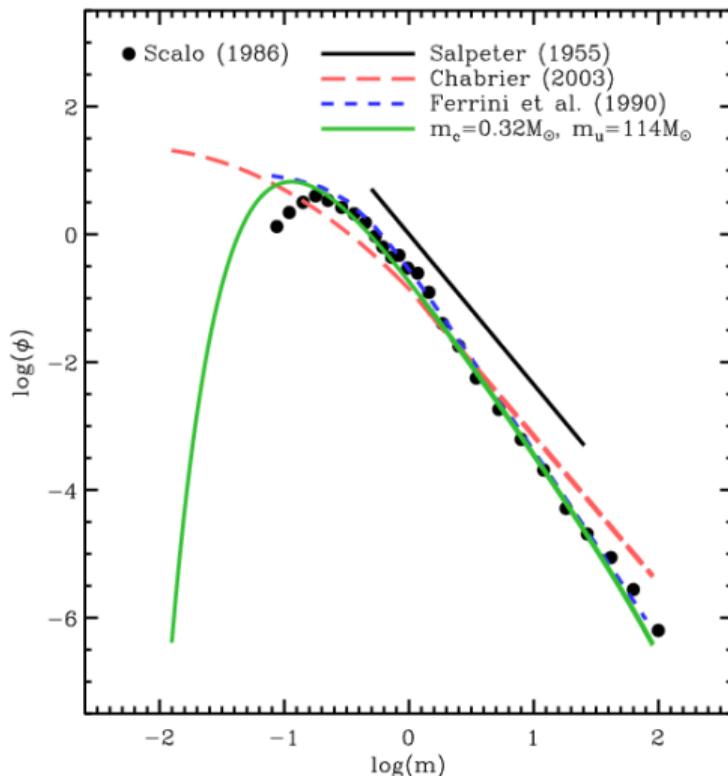
- *distribuição independente de escala*:
se $f(x) = x^\gamma$, então
 $f(ax) = a^\gamma x^\gamma = \text{cte } x^\gamma = \text{cte } f(x)$
a *forma* da função não muda com uma mudança de escala da variável

distribuição em lei de potência

- tem média e variância infinitas, a menos que limites sejam definidos
- isso ocorre na maioria dos casos concretos: podem existir limitações físicas em ambos os 'lados' da distribuição
- exemplo de distribuição em lei de potência: a função de massa inicial de Salpeter (1955)

$$\xi(M) = \xi_0 M^{-2.35}$$

$\xi(M)dM$: número de estrelas nascidas com massas entre M e $M + dM$



distribuições bivariadas

- vamos considerar agora distribuições de duas ou mais variáveis
- dadas 2 variáveis x e y , a distribuição de probabilidades conjunta de x e y é $P(x, y)$:

- $P(x, y) \geq 0$

- $\int P(x, y) dx dy = 1$ ou $\sum_{x, y} P(x, y) = 1$

- As distribuições $P(x)$ e $P(y)$ são as **distribuições marginais** de $P(x, y)$:

$$P(x) = \int P(x, y) dy \quad \text{ou} \quad P(x) = \sum_y P(x, y)$$

$$P(y) = \int P(x, y) dx \quad \text{ou} \quad P(y) = \sum_x P(x, y)$$

- distribuição cumulativa conjunta (não é muito usada):

$$F(x', y') = P(x < x', y < y')$$

valor esperado e correlação

- **valor esperado** de uma função $f(x, y)$:

$$E(f) = \int_x \int_y f(x, y)P(x, y)dxdy$$

ou

$$E(f) = \sum_x \sum_y f(x, y)P(x, y)$$

- **covariância** entre x e y

$$Cov(x, y) = E[(x - E(x))(y - E(y))] = E(xy) - E(x)E(y) \equiv \sigma_{xy}$$

- o **coeficiente de correlação de Pearson** é um número entre -1 e +1 definido como

$$\rho = Corr(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

σ_x^2, σ_y^2 : valores esperados da variância de x e y

a distribuição gaussiana bivariada

- gaussiana bivariada:

$$P(x, y | \mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{Z^2}{2(1-\rho^2)}\right)$$

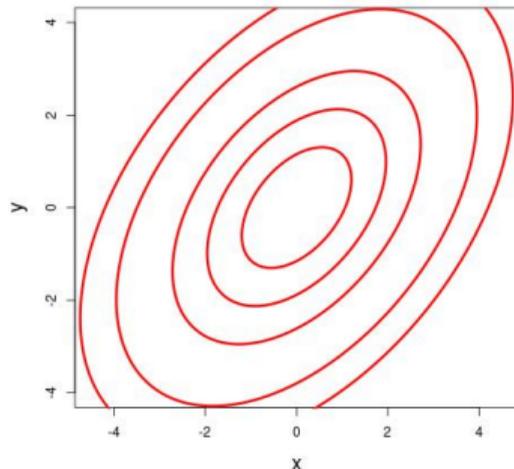
onde

$$Z^2 = \frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} - 2\rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y}$$

- coeficiente de correlação:

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

- contornos $P(x, y) = \text{cte}$: elipses



distribuição gaussiana bivariada em notação vetorial

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}$$

notação vetorial:

- vetor com as variáveis:

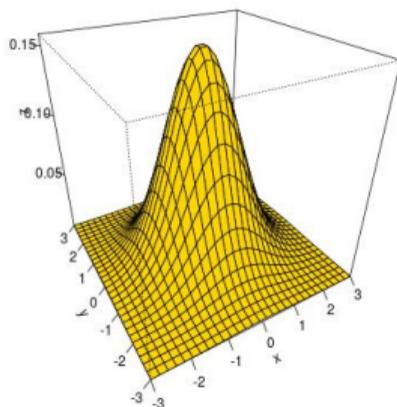
$$\mathbf{x} = (x, y)^T$$

- vetor com as médias:

$$\boldsymbol{\mu} = (\mu_x, \mu_y)^T$$

- matriz de covariância:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$



então,

$$P(x, y) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\} \sim N(\boldsymbol{\mu}, \Sigma)$$

a distribuição gaussiana bivariada

Note que:

- $-1 \leq \rho \leq +1$
- x e y são independentes apenas se $\rho = 0$
- as distribuições marginais são

$$P(x) \sim N(\mu_x, \sigma_x) \quad e \quad P(y) \sim N(\mu_y, \sigma_y)$$

- covariância entre x e y :

$$\text{Cov}(x, y) = \rho$$

- distribuições condicionais
(ver completando_o_quadrado.pdf):

$$P(y|x) \sim N(\mu_{y|x}, \sigma_{y|x}),$$

onde

$$\mu_{y|x} = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x)$$

$$\sigma_{y|x} = \sigma_y \sqrt{1 - \rho^2}$$

