

Métodos Quantitativos (e Qualitativos)

2. Probabilidades

Salvatore Cosentino
D.Mat U.Minho

24 nov 2020

Probabilidades

A **teoria das probabilidades** é uma teoria física que estuda as **regularidades** observadas nos fenómenos que são, por uma razão ou outra, **aleatórios**.

Como toda teoria física, é constituída por uma **estrutura matemática**

(a teoria da medida, e umas ideias e técnicas próprias, como a noção de independência, a entropia, ...)

e por umas **regras operacionais**, que fazem a ligação entre as **previsões** do modelo matemático e as **observações** das experiências sobre o pedaço de natureza que queremos compreender.

As previsões são **leis universais** que emergem em presença de **números grandes** de provas, ou de elementos, ...

Mecânica estatística & termodinâmica

O exemplo paradigmático é o caso da **mecânica estatística**.

O **estado microscópico** de um litro gás é descrito pelas posições e velocidades de cada uma das

$$\sim 10^{24} \quad \text{moléculas}$$

No entanto, o **estado macroscópico** é descrito por apenas

$$2$$

observáveis macroscópicos, escolhidos entre **temperatura**, **volume** e **pressão**.

As leis da **termodinâmica** podem ser deduzidas a partir da descrição microscópica, newtoniana ou quântica, do sistema usando ...

... algumas hipóteses razoáveis sobre as observações e sobre o modelo microscópico (a noção de observável macroscópico, o curto alcance das forças moleculares, a independência entre regiões distantes, a hipótese ergódica de Boltzmann, ...)

... e resultados chave da teoria das probabilidades, como a lei dos grandes números, o teorema limite central, ...

ALEA

O arquétipo de um fenómeno aleatório é o lançamento de um **dado** (ALEA, em latim).

Em princípio, a queda de um dado é descrita pelas leis da mecânica newtoniana.

A **impossibilidade**, prática ou teórica, de conhecer as condições iniciais com precisão arbitrária e a sensibilidade às condições iniciais próprias das leis da física que descrevem sistemas complexos,

fazem inútil qualquer tentativa de **prever** o resultado de **um** lançamento de **um** dado.



Frequências empíricas

No entanto, é uma **observação empírica** que ao lançar um **número grande** n de vezes um dado (ou uma vez n dados), e ao contar o número S_n das vezes em que é observada, por exemplo, uma pinta, a **frequência observada** é próxima de

$$f_n = \frac{S_n}{n} \simeq \frac{1}{6}$$

Isto é o que o homen na rua entende ao dizer que “a probabilidade de observar uma pinta ao lançar um dado é igual a um sexto”.



Giocatori di dadi, Pompei, Osteria della via di Mercurio

Outro arquétipo de experiência aleatória é o lançamento de uma **moeda**.

Os resultados possíveis são apenas dois,

CARA ou COROA

A conjectura natural é que observar CARA ou observar COROA são eventos **equiprováveis**.

Ao lançar **muitas vezes** uma moeda, ou ao lançar **muitas moedas**, esperamos, por razões de simetria, observar CARA mais ou menos a **metade das vezes**.

Ou seja, se S_n denota o número de vezes que observamos CARA ao lançar n moedas, a nossa expectativa é

$$S_n \sim \frac{n}{2}$$

O problema é que $n/2$ é apenas a nossa **melhor aposta** para S_n ,

e de fato ninguém espera obter **exatamente** o mesmo número de CARAs e de COROAs em n lançamentos. Isto seria ter muita sorte!

Probabilidade

A ideia da teoria das probabilidades é **quantificar** a expectativa acerca de um **evento** como

$$(S_n = k) := \text{"observar } k \text{ vezes CARA em } n \text{ moedas"}$$

Os probabilistas associam um **número entre zero e um** a cada um destes eventos, que chamam

$$\mathbf{P}(S_n = k)$$

ou seja, "**probabilidade** de observar k vezes CARA em n moedas".

Uma maneira natural de definir estas probabilidades é contar a cardinalidade dos **casos favoráveis**, todos os que conduzem ao resultado $S_n = k$, e dividir este número pela cardinalidade de todos os **casos possíveis**.

Ou seja, um modelo natural é

$$\mathbf{P}(S_n = k) = \frac{\text{número de casos favoráveis}}{\text{número de casos possíveis}}$$

Interpretação operacional

Um modelo da teoria das probabilidades permite calcular, ou pelo menos estimar, probabilidades de eventos.

Se o modelo dizer que um certo evento tem **probabilidade muito grande**, por exemplo superior a

0.99 ou 0.999

então o evento **é observado** praticamente em todas as vezes que repetimos a experiência.

(se não for observado em uma experiência, podemos pensar que tivemos muito azar, se não for observado em duas, três, quatro experiências seguidas, podemos considerar não credível o nosso modelo)

Se conseguimos encontrar um tal evento, o que estamos a fazer é uma **previsão**.

Muitas moedas

Os probabilistas sabem calcular cada uma das probabilidades $\mathbf{P}(S_n = k)$.

O número dos **casos possíveis** é

$$2^n$$

pois este é o número de todas as palavras de comprimento n formadas pelas duas letras CARA e COROA.

O número dos **casos favoráveis**, e isto obriga a uma pequena reflexão, é

$$\frac{n!}{k! \cdot (n - k)!}$$

pois esta é a cardinalidade de todas as palavras de comprimento n nas letras CARA e COROA que contêm k vezes a letra CARA.

O resultado é que

$$\mathbf{P}(S_n = k) = \frac{\frac{n!}{k! \cdot (n - k)!}}{2^n}$$

(não se assustem!)

Se n é pequeno ...

Quando n é pequeno, as probabilidades **não dizem coisas interessantes**.

Por exemplo, a fórmula complicada

$$\mathbf{P}(S_n = k) = \frac{n!}{k! \cdot (n - k)!} 2^{-n}$$

diz que

$$\mathbf{P}(S_2 = 1) = \frac{1}{2}$$

ou seja, que a probabilidade de obter uma vez CARA ao lançar duas moedas é 0.5,

ou diz que

$$\mathbf{P}(S_2 = 0) = \frac{1}{4}$$

ou seja, que a probabilidade de não obter nunca CARA ao lançar duas moedas é 0.25.

...

Média

Uma possibilidade para o “best guess” é o valor de k que maximiza a probabilidade $\mathbf{P}(S_n = k)$. Este valor, se existir, é chamado **moda**.

Uma alternativa é a **média**, ou **valor esperado**, da variável S_n , que é, naturalmente,

$$\mathbf{E}(S_n) = \frac{n}{2}$$

Se querem saber como é calculada, a fórmula que define a média é a soma pesada

$$\mathbf{E}(S_n) := \sum_{k=0}^n k \cdot \mathbf{P}(S_n = k)$$

onde cada resultado possível k , entre 0 e n , é **pesado** com a sua probabilidade.

A média ainda não diz muito sobre o que de fato é observado!

Apenas exprime uma **simetria**.

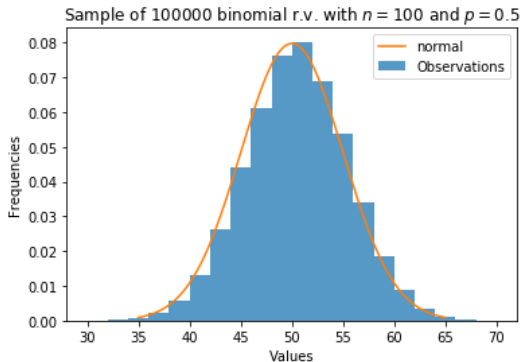
Se n é grande ...

É ao observar um gráfico da função

$$k \mapsto \mathbf{P}(S_n = k)$$

quando n é **grande** que descobrimos um fenómeno interessante.

Os histogramas das simulações parecem desenhar uma função que tem a forma de um “sino” centrado no ponto $n/2$, e rapidamente decresce para valores praticamente nulos quando k se afasta de $n/2$ por algo da ordem de $\sqrt{n}$.



Regularidades probabilísticas

O máximo é no ponto $n/2$ (a melhor aposta), mas é da ordem

$$\mathbf{P}(S_n = n/2) \sim 1/\sqrt{n}$$

um número muito pequeno se n é grande.

Por outro lado, ao somar todos os valores da função $\mathbf{P}(S_n = k)$ em um intervalo de comprimento da ordem $\sqrt{n}$ em torno de $n/2$ (os valores de k para os quais a função é significativamente superior a zero) obtemos algo da ordem de

$$\sqrt{n} \cdot \frac{1}{\sqrt{n}} \sim 1$$

ou seja,

$$\boxed{\mathbf{P}(S_n \simeq n/2 \pm \sqrt{n}) \sim 1}$$

Encontramos um **evento quase certo** !

Esta é uma **previsão**: ao lançar um número grande n de moedas, esperamos observar um número de CARAS no **intervalo**

$$S_n \simeq \frac{n}{2} \pm \sqrt{n}$$

Passeio aleatório

É conveniente interpretar os lançamentos sucessivos de uma moeda como um **passeio aleatório**: deslocamo-nos ao longo de uma reta lançando repetidamente uma moeda, e dando um passo para a frente se sair CARA ou um passo para trás se sair COROA.

Chamamos

$$P_n = S_n - (n - S_n) = 2S_n - n$$

a **posição** do passeio depois do n -ésimo lançamento, i.e. no **tempo** n .

(se cada lançamento corresponde a uma aposta em um jogo honesto, então S_n é o lucro do jogador passadas n apostas)

Qual será o aspeto das **trajetórias típicas** ?

Por razões de simetria, a posição média passado um tempo n é

$$\mathbf{E}(P_n) = 0$$

No entanto, qualquer bêbado sabe que, se começar a andar pela rua do Souto dando um passo para frente ou para trás ao acaso, no fim da noite ele estará bastante longe de casa!

Quanto longe?

Para compreender quanto longe anda uma trajetória típica do passeio aleatório, é útil calcular a **média do quadrado** da posição, ou seja, $\langle P_n^2 \rangle$.

É claro que

$$\mathbf{E} (P_0^2) = 0$$

se o passeio começa na origem.

Por outro lado,

$$P_{n+1} = P_n \pm 1,$$

as duas possibilidades sendo equiprováveis. Logo, a média do quadrado de P_{n+1} é

$$\begin{aligned}\mathbf{E} (P_{n+1}^2) &= \mathbf{E} (P_n \pm 1)^2 \\ &= \mathbf{E} (P_n^2 \pm 2P_n + 1) \\ &= \mathbf{E} (P_n^2) + 1\end{aligned}$$

porque a média é uma operação linear, e as duas possibilidades são equiprováveis.

Esta é uma equação recursiva que diz que a média do quadrado cresce de uma unidade em cada passo. Consequentemente,

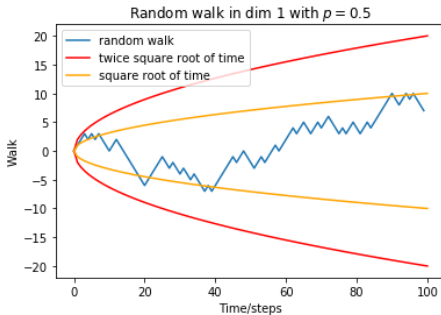
$$\mathbf{E} (P_n^2) = n$$

Princípio da raiz quadrada

O **desvio médio** das trajetórias é portanto

$$|P_n| \sim \sqrt{n}$$

Ou seja, as **trajetórias típicas** se afastam da origem como a **raiz quadrada** do tempo !



Lei dos grandes números

Em termos da frequência de CARAs, sendo $P_n = 2S_n - n$, redescobrimos a conjectura

$$\frac{S_n}{n} \simeq \frac{1}{2} \pm \frac{1}{\sqrt{n}}$$

Se n é grande, e os “instrumentos” não observam um erro da ordem de $1/\sqrt{n}$, podemos estimar

$$\frac{S_n}{n} \sim \frac{1}{2}$$

Esta afirmação é chamada **lei dos grandes números**.

É o que um probabilista entende ao dizer que “ $1/2$ é a probabilidade de obter CARA ao lançar uma moeda honesta”.

Intervalos de confiança

Outra maneira de ler a nossa estimação é

$$\frac{1}{2} \simeq \frac{S_n}{n} \pm \frac{1}{\sqrt{n}}$$

Ou seja, podemos **estimar** a “probabilidade de obter CARA” com a frequência observada S_n/n ,

sabendo que a **precisão** da estimação é da ordem de $1/\sqrt{n}$.

Em geral, se lançamos n vezes um dado com um número desconhecido de faces, e queremos estimar a probabilidade p de sair uma certa face,

podemos contar o número S_n de vezes em que é observada esta face, e estimar que p estará no **intervalo de confiança**

$$p \simeq \frac{S_n}{n} \pm \frac{1}{\sqrt{n}}$$

Provas de Bernoulli

As considerações anteriores generalizam ao caso das **provas de Bernoulli**.

São provas repetidas e **independentes** de uma experiência com dois resultados possíveis, **sucesso** ou **insucesso**.

(por exemplo, escolher ao acaso uma bolinha de uma população de r bolinhas vermelhas e b bolinhas azuis)

É conveniente chamar X_k a variável que assume o valor $X_k = 1$ se foi observado um sucesso na k -ésima prova, e o valor $X_k = 0$ se foi observado insucesso.

O único parâmetro do modelo é a probabilidade

$$p = \mathbf{P}(X_k = 1)$$

de obter **sucesso** em uma prova, sendo $q = 1 - p$ a probabilidade de obter insucesso.

(no exemplo, $p = r/(r + b)$)

Binomial

A variável que **conta** o número de sucessos em n provas de Bernoulli é

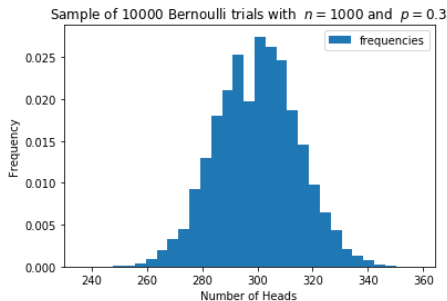
$$S_n = X_1 + X_2 + \cdots + X_n$$

e assume valores entre 0 e n .

Então a probabilidade de obter k sucessos em n provas é

$$\mathbf{P}(S_n = k) = \frac{n!}{k! (n - k)!} p^k q^{n-k}$$

Esta é chamada **distribuição binomial**.



Média e variância da binomial

A média de uma variável binomial $B(n, p)$ é

$$\mathbf{E}(S_n) = np$$

e representa o **valor esperado** (ou **best guess**) do número de sucessos obtidos em n provas de Bernoulli.

A **variância**, ou seja, o valor esperado do quadrado do desvio $S_n - \mathbf{E}(S_n)$, é

$$\begin{aligned}\mathbf{V}(S_n) &:= \mathbf{E}(S_n - \mathbf{E}(S_n))^2 \\ &= npq\end{aligned}$$

A sua raiz quadrada, chamada **desvio padrão** (em inglês, **standard deviation**), é

$$\sqrt{\mathbf{V}(S_n)} = \sqrt{npq}$$

Assim, o número típico de sucessos em n provas de Bernoulli tem oscilações da ordem de $\sqrt{npq}$ em torno do valor médio np , ou seja,

$$S_n = np \pm \sqrt{npq}$$

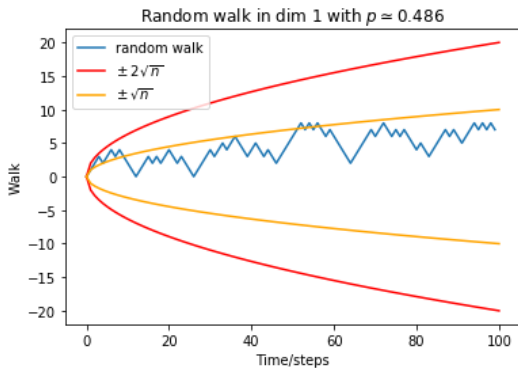
Porque os Casinos ganham?

Se $p \neq 1/2$, as trajetórias de um passeio aleatório escolhem a direção mais provável, desde que o tempo seja suficientemente grande.

Por exemplo, se

$$p = \frac{18}{37}$$

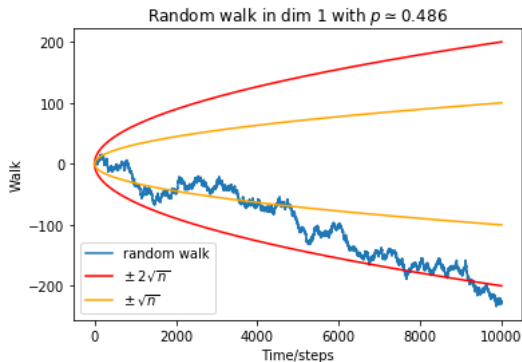
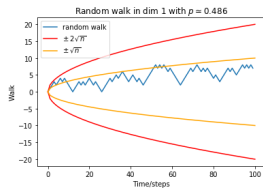
e o tempo é pequeno, como $n \simeq 100$ (as apostas de um jogador durante uma noite), as trajetórias parecem oscilar como as trajetórias de um passeio aleatório simétrico, assim que o jogo **parece honesto** ao jogador.



Ruina do jogador

... mas depois de um tempo muito maior, por exemplo da ordem de $n \simeq 10000$, (as apostas de 100 jogadores durante uma noite, ou as apostas de um jogador durante 100 noites),

o que se observa é a **ruína** certa dos jogadores, ou, melhor, o lucro certo da casa de jogos !



Tempo de espera

A variável **tempo de espera** T calcula o número de provas de Bernoulli que temos de esperar até obter sucesso pela primeira vez. A sua lei é chamada **geométrica**.

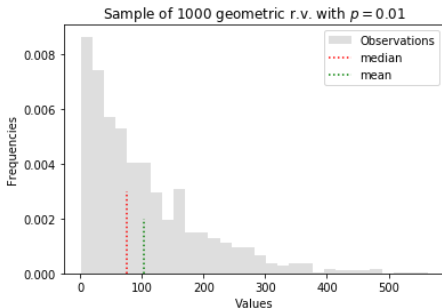
Se p é a probabilidade de obter sucesso em uma prova de Bernoulli, e $q = 1 - p$, então a probabilidade de ter sucesso pela primeira vez na n -ésima prova é

$$\mathbf{P}(T = n) = q^{n-1} p$$

A média (que é sensível aos eventos grandes mas raros) é

$$\mathbf{E}(T) = \frac{1}{p}$$

A mediana, que é menor, é provavelmente mais representativa dos valores típicos.



Eventos raros

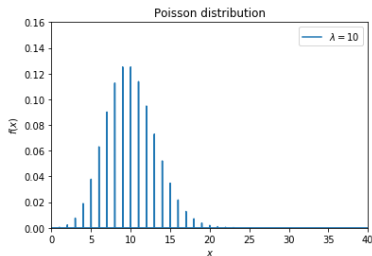
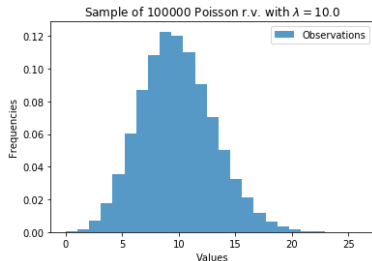
Quando $p \ll 1$, o evento “ter sucesso” é um **evento raro**.

É natural esperar que apenas são observados valores pequenos do número de sucessos.

Se n é muito grande, e $pn = \lambda$, a distribuição binomial $\mathbf{P}(S_n = k)$ é bem aproximada pela **lei de Poisson**

$$\mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, 3, \dots$$

O único parâmetro λ é a média e também a variância.



Exponencial

Se λ é o valor esperado do número de eventos raros na unidade de tempo, então a probabilidade de não observar nenhum destes eventos duante um tempo t , ou seja, a probabilidade que o **tempo de espera** T seja superior a t , é

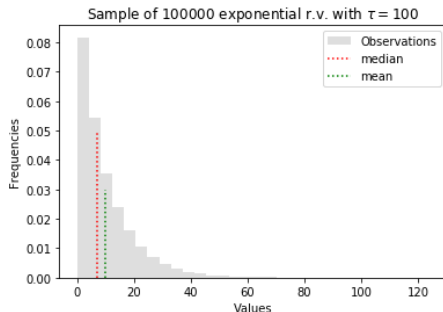
$$\mathbf{P}(T > t) = e^{-\lambda t}$$

Assim, a variável tempo de espera T tem **lei exponencial** (uma versão contínua da geométrica), cuja densidade de probabilidades é

$$p(t) = \frac{1}{\tau} e^{-t/\tau}$$

onde $\tau = 1/\lambda = \mathbf{E}(T)$ é a média.

A mediana, que é inferior a τ , é mais representativa dos valores típicos.



Falta de memória

Uma variável exponencial, que modela um **tempo de vida**, ou de espera,

(como o tempo de decaimento de um núcleo de uma substância radioativa, o tempo de vida de uma lâmpada ou de um aparelho eletrônico, o tempo que devo esperar até ver sair o número 13 na Roulette, ...)

não tem memória, ou seja, a probabilidade de esperar um tempo $t + s$ **sabendo** que já esperamos um tempo s é igual a probabilidade de esperar um tempo t sem ter mais informações, que os probabilistas escrevem assim:

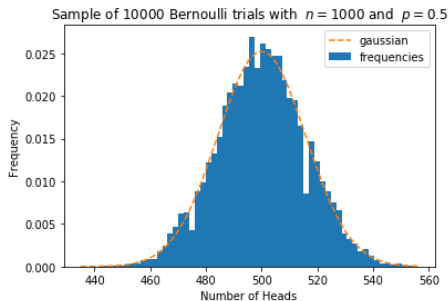
$$\mathbf{P}(T > t + s | T > s) = \mathbf{P}(T > t)$$

Em outras palavras, um sistema físico cujo tempo de vida tem lei exponencial **não tem idade**: se viveu até hoje, o seu futuro é igual ao futuro de um sistema recém nascido

(o que explica porque apostar em um número que não sai desde muito tempo, por exemplo o número 13, não faz crescer a probabilidade de ganhar!)

Universalidade assintótica

Os histogramas das simulações desenham curvas muito regulares ...



Em particular, a lei da variável reduzida

$$s_n = \frac{S_n - np}{\sqrt{npq}}$$

(que tem média nula e variância unitária) se estabiliza perto de uma lei universal dita **normal (reduzida)** ou **gaussiana** ao crescer o número de provas n .

Teorema limite central

O teorema limite central de De Moivre e Laplace diz que para estimar as probabilidades

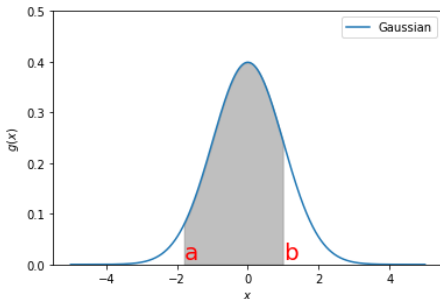
$$\mathbf{P}(a < s_n < b)$$

quando n é grande, é suficiente calcular as áreas (que os matemático chamam **integrals**) das regiões limitadas entre o eixo dos x , o gráfico da gaussiana

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

e as retas verticais $x = a$ e $x = b$, ou seja,

$$\mathbf{P}(a < s_n < b) \simeq \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$



Gaussianas

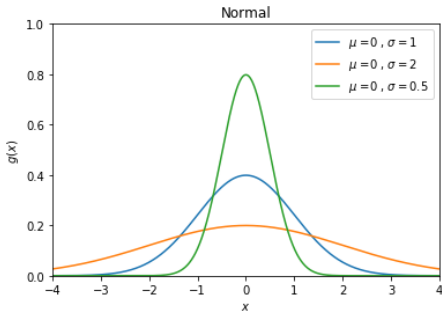
Em geral, se a variável X tem lei normal reduzida $N(0, 1)$, então a variável

$$Y = \sigma \cdot (X + \mu)$$

tem lei **normal** $N(\mu, \sigma)$, ou seja, tem densidade de probabilidades

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Os parâmetros são a média $\mu = \mathbf{E}(Y)$, o parâmetro de localização, e o desvio padrão $\sigma = \sqrt{\mathbf{V}(Y)}$, o parâmetro de escala.

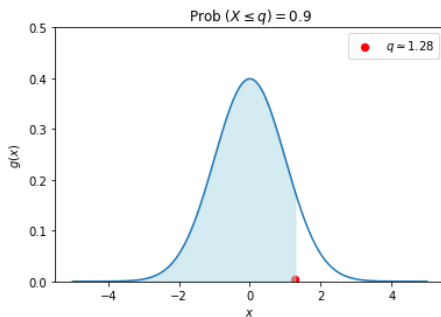


Quantis da lei normal

Seja X uma variável normal reduzida, i.e. com lei $N(0, 1)$.

Dada uma probabilidade α , o **quantil** q_α é o valor q_α tal que

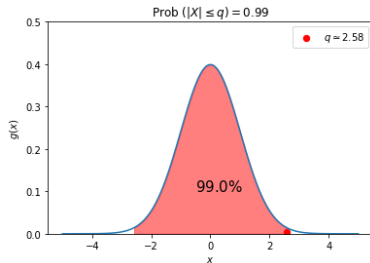
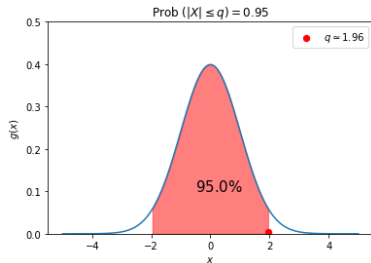
$$\mathbf{P}(X \leq q_\alpha) = \alpha$$



Intervalos de confiança da lei normal

Mais úteis são os intervalos simétricos tais que se X tem lei normal reduzida $N(0, 1)$ então

$$\mathbf{P}(|X| \leq q_{1-\alpha/2}) = \alpha$$



Assim, uma variável Y com lei normal $N(\mu, \sigma)$ está no intervalo de confiança

$$Y = \mu \pm q_{1-\alpha/2} \cdot \sigma$$

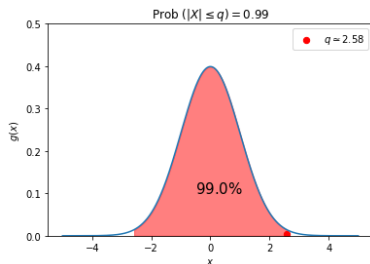
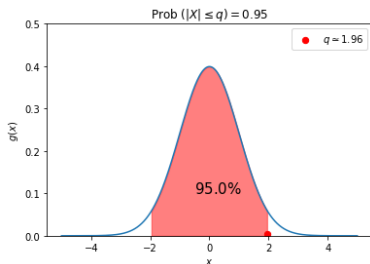
com probabilidade α , chamada **nível** (do intervalo de confiança).

Regra 68 - 95 - 99.7

Mais simples é lembrar os níveis de confiança de intervalos simétricos que correspondem a

1 2 ou 3

desvios padrão.



Os físicos das partículas, que são mais conservativos/prudentes, também gostam do intervalo de 5 desvios padrão, que corresponde a uma probabilidade 99.99994% !

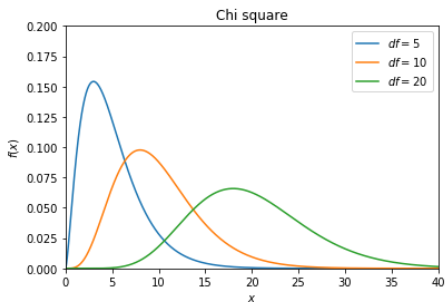
Qui-quadrado

Se $X_1, X_2, \dots, X_n$ são variáveis independentes com lei normal reduzida $N(0, 1)$, então a soma dos quadrados

$$Q_n = X_1^2 + X_2^2 + \dots + X_n^2$$

tem **lei qui-quadrado** com n **graus de liberdade** (em inglês, **degrees of freedom**), denotada por $\chi^2(n)$.

A média e a variância são $\mathbf{E}(Q_n) = n$ e $\mathbf{V}(Q_n) = 2n$, respetivamente.

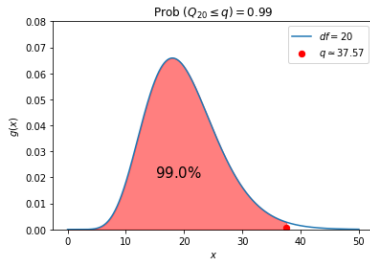
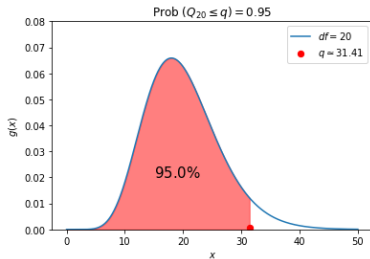


Quantis do Qui-quadrado

Os quantis da lei qui-quadradao mais utilizados nos testes de hipóteses são

$$q_{95\%} \quad \text{e} \quad q_{99\%}$$

e, naturalmente, dependem do número de graus de liberdade.



Qui-quadrado com n grande

Sendo Q_n uma soma de variáveis independentes e identicamente distribuídas, pelo **teorema limite central** a variável normalizada

$$\frac{Q_n - n}{\sqrt{2n}}$$

converge, para n grande, a uma variável normal $N(0, 1)$.

Na prática, se $n \geq 50$, podemos **estimar**

$$q_\alpha \simeq n + \phi_\alpha \sqrt{2n}$$

onde ϕ_α é o quantil da lei normal $N(0, 1)$.

Por exemplo, um intervalo de confiança **generoso** de nível $\geq 95\%$ é

$$Q_n \leq n + 2\sqrt{2n}$$