

Distribuições Contínuas: Normal

Prof. Alejandra, Cristian, Paula e Rafael

Introdução à Probabilidade e Estatística



Estes slides tem por objetivo apenas complementar e esclarecer os textos principais indicados para a disciplina, e não devem ser usados como único texto de referência.

Junto à essas notas o aluno deve ler os textos além da referência principal do curso.

Vimos na semana passada que se X é uma v. a. contínua então:

- existe uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ tal que

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx.$$

Tal função é chamada de *densidade* de X ;

- para todo $a \in \mathbb{R}$, $P(X = a) = 0$;
- a esperança de X é dada por

$$E[X] = \int_{-\infty}^{\infty} xf(x) \, dx,$$

onde f é a densidade de X ;

- e para $g : \mathbb{R} \rightarrow \mathbb{R}$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) \, dx,$$

A partir de agora vamos nos concentrar no estudo da chamada distribuição Normal, que ocupa um papel central na teoria da probabilidade e na estatística.

Introduzida em 1733 pelo matemático francês Abraham De Moivre, a distribuição normal foi primeiro usada como uma forma de aproximar variáveis Binomiais com parâmetro n muito grande.

Tal ideia foi expandida por outros matemáticos, como Laplace, que identificaram a distribuição normal como a protagonista do chamado Teorema Central do Limite, que introduziremos no próximo módulo deste curso.

Parte da fama da distribuição normal veio com o matemático Karl Friedrich Gauss, que em 1809 usou-a na estimativa da localização de objetos astronômicos. Até hoje ocasionalmente nos referimos à uma variável normal como *Gaussiana*.

Definição 1 (Distribuição Normal)

Diremos que uma variável aleatória contínua X tem distribuição Normal (ou Gaussiana) de parâmetros $\mu \in \mathbb{R}$ e σ^2 , com $\sigma > 0$, se sua função densidade é dada por

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Dizemos também que X é **normalmente distribuída**, ou que X é uma variável aleatória **Normal** ou **Gaussiana**.

Notação: $X \sim \text{Normal}(\mu, \sigma^2)$.

Observação 1

É importante observar que a função densidade $\phi(x)$ não possui primitiva expressa em termos de funções elementares, e portanto não conseguimos usar o Teorema Fundamental do Cálculo para calcular as probabilidades que precisaremos.

Algumas probabilidades específicas, como $P(X > \mu)$, podem ser calculadas usando características próprias da função $\phi(x)$.

Para todas as outras será necessário o uso de aproximações numéricas, que podem ser feitos em diversos sites.

Para esse curso disponibilizamos uma calculadora de probabilidades, que pode ser acessada também em <https://www.geogebra.org/m/hwa3fvs7>.

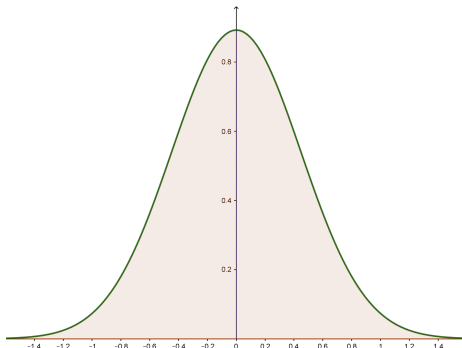
Exemplo

Seja Z uma v.a. normalmente distribuída com parâmetros $\mu = 0$ e $\sigma^2 = 1$. Calcule $P(Z > 0)$ e $P(-2 < Z < 1)$.

Solução: Para encontrar $P(Z > 0)$, observe primeiro que

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

E portanto ϕ é simétrica em relação à $\mu = 0$, como vemos na figura abaixo.



Pela simetria de $\phi(x)$ em torno de $\mu = 0$, temos que

$$P(Z > 0) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{(x-2)^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{(x-2)^2}{2}} dx = P(Z \leq 0).$$

Temos assim que

$$1 = P(Z > 0) + P(Z \leq 0) = 2P(Z > 0),$$

e portanto $P(Z > 0) = 1/2$.

Para calcular $P(-2 < Z < 1)$ temos que usar a [calculadora de probabilidades](#), que nos diz que

$$P(-2 < Z < 1) = 0,8186.$$

Para estudar as propriedades da distribuição normal, vamos começar por uma em particular: a chamada normal padrão.

A distribuição Normal com parâmetros $\mu = 0$ e $\sigma = 1$ é geralmente conhecida como **Normal Padrão**. As razões para isso ficarão claras mais à frente, mas por enquanto vamos estudar algumas de suas propriedades.

Primeiro vamos lembrar que se $Z \sim \text{Normal}(0,1)$ então sua função densidade de probabilidade é dada por

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Como vimos no exemplo anterior, a função densidade de probabilidade de uma normal padrão é simétrica em relação ao eixo- y . Ou seja

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

e isso tem consequências interessantes para algumas probabilidades.

Em particular, notamos que se $a \in \mathbb{R}$ e $Z \sim \text{Normal}(0,1)$ então

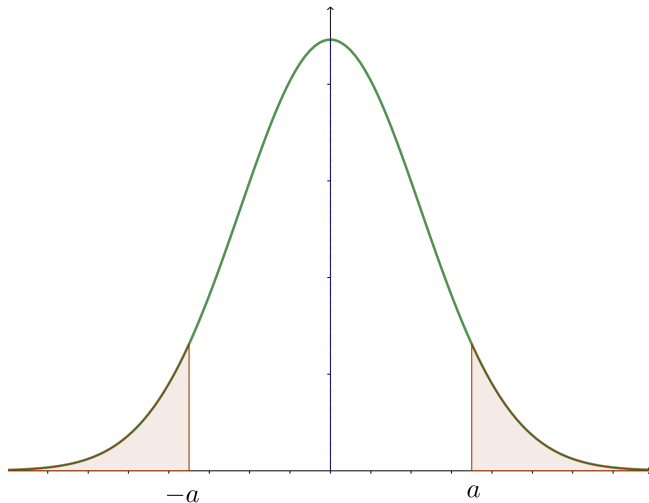
$$P(Z > a) = P(Z < -a).$$

Analiticamente, podemos ver isso fazendo uma simples substituição $u = -x$ na integral de $P(Z < -a)$, que nos daria

$$P(Z < -a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-\frac{u^2}{2}} du = P(Z > a).$$

A Normal Padrão - Simetria

Geometricamente, isso se replete no gráfico abaixo. Aqui a área da região à esquerda representa $P(Z < -a)$, enquanto a área da região à direita representa $P(Z > a)$.



Seguimos agora para o cálculo de $E[Z]$ e $Var[Z]$ quando Z tem distribuição normal padrão.

Lembrando, a f.d.p. de Z é dada por

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

e portanto

$$E[Z] = \int_{-\infty}^{\infty} x \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx.$$

Primeiro encontremos a primitiva de $x e^{-x^2/2}$. E para isso podemos fazer a substituição $u = -x^2/2$. Isso nos dá $du = -x dx$ e

$$\int x e^{-x^2/2} dx = - \int e^u du = -e^u + K = -e^{-x^2/2} + K.$$

Com isso temos

$$\int_{-\infty}^{\infty} x e^{-x^2/2} dx = \left[-e^{-x^2/2} \right]_{-\infty}^{+\infty} = 0 - 0 = 0,$$

e portanto

$$E[Z] = 0.$$

Para calcular $Var[Z]$ vamos primeiro calcular

$$E[Z^2] = \int_{-\infty}^{\infty} x^2 \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx.$$

Dando sequência, vamos usar integração por partes. Fazendo $u = x$ e $dv = x e^{-x^2/2} dx$, encontramos

$$\left. \begin{array}{l} u = x \\ du = dx \end{array} \right| \begin{array}{l} dv = x e^{-x^2/2} dx \\ v = -e^{-x^2/2} \end{array}$$

Segue assim que

$$\begin{aligned}\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx &= -\frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx \\ &= 0 - 0 + 1 \\ &= 1.\end{aligned}$$

Concluimos assim que

$$\text{Var}[Z] = E[Z^2] - (E[Z])^2 = 1 - 0 = 1.$$

E resumindo, se $Z \sim \text{Normal}(0,1)$ então

$$E[Z] = 0 \quad \text{e} \quad \text{Var}[Z] = 1.$$

Voltando nossa atenção para uma variável $X \sim Normal(\mu, \sigma^2)$ qualquer, comecemos com uma propriedade importante da função densidade

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

que é sua simetria em relação ao parâmetro $\mu \in \mathbb{R}$.

De fato, para todo $a \in \mathbb{R}$, vale que

$$\phi(\mu + a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(a)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(-a)^2}{2\sigma^2}} = \phi(\mu - a).$$

Isso é ilustrado na figura a seguir

Gráfico da Função Densidade

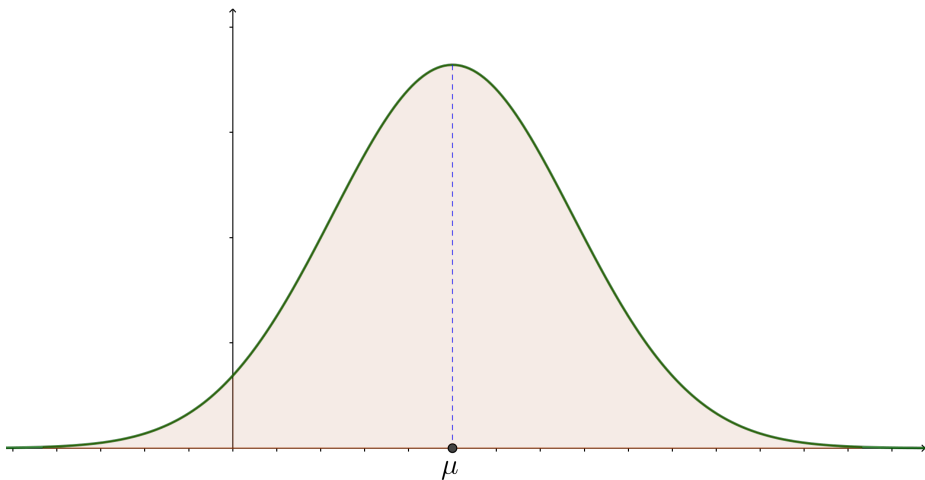


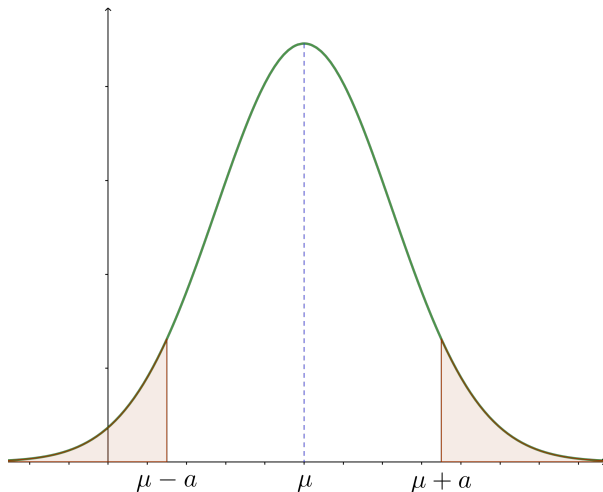
Figura: Gráfico da f.d.p. de uma $Normal(\mu, \sigma^2)$. Observe a simetria do gráfico em torno de μ .

Simetria da Função Densidade

Podemos mostrar que se $X \sim \text{Normal}(\mu, \sigma^2)$, então

$$P(X < \mu - a) = P(X > \mu + a),$$

para todo $a \in \mathbb{R}$, como vemos na imagem abaixo.



Exemplo

Seja Z uma v.a. normalmente distribuída com parâmetros $\mu = 2$ e $\sigma^2 = 1$. Calcule $P(Z > 2)$ e $P(-2 < Z < 4)$.

Solução: Note pela simetria de $\phi(x)$ em torno de $\mu = 2$, temos que

$$P(Z > 2) = \frac{1}{\sqrt{2\pi}} \int_2^{\infty} e^{-\frac{(x-2)^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^2 e^{-\frac{(x-2)^2}{2}} dx = P(Z \leq 2).$$

Temos assim que

$$1 = P(Z > 2) + P(Z \leq 2) = 2P(Z > 2),$$

e portanto $P(Z > 2) = 1/2$.

Para calcular $P(-2 < Z < 4)$ temos que usar a [calculadora de probabilidades](#), que nos diz que

$$P(-2 < Z < 4) = 0,9772.$$

Antes de prosseguir para o cálculo da esperança e da variância da normal para os demais parâmetros, vamos ver mais uma propriedade essencial da distribuição normal.

Para isso, tome $X \sim \text{Normal}(\mu, \sigma^2)$ e $a, b \in \mathbb{R}$, e faça $Y = aX + b$. Note que se F_X e F_Y são as funções distribuição de X e Y , respectivamente, e $a > 0$ então

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(aX + b \leq x) \\ &= P\left(X \leq \frac{x-b}{a}\right) \\ &= F_X\left(\frac{x-b}{a}\right). \end{aligned}$$

E derivando a equação, encontramos que a densidade de Y é dada por

$$f_Y(x) = F'_Y(x) = \frac{1}{a} F'_X\left(\frac{x-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{x-b}{a} - \mu\right)^2}{2\sigma^2}}$$

Arrumando a equação, encontramos

$$f_Y(x) = \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{(x-(b+a\mu))^2}{2(a\sigma)^2}},$$

e portanto $Y \sim \text{Normal}(a\mu + b, (a\sigma)^2)$.

Escolhendo a e b de maneira adequada, mostramos então que se $X \sim \text{Normal}(\mu, \sigma^2)$, então

$$Z = \frac{X - \mu}{\sigma}$$

é tal que $Z \sim \text{Normal}(0, 1)$.

Por essa razão, a distribuição $\text{Normal}(0, 1)$ é conhecida como **normal padrão**.

Seja X uma v.a. com distribuição normal de parâmetros $\mu = 1$ e $\sigma^2 = 2$. Encontre $E[X]$ e $Var[X]$.

Solução: Seguindo o que acabamos de ver, como $X \sim Normal(1,2)$, faça

$$Z = \frac{X-1}{\sqrt{2}}.$$

Sabemos então que $Z \sim Normal(0,1)$, e portanto $E[Z] = 0$ e $Var[Z] = 1$.

Mas vale que $X = \sqrt{2}Z + 1$, e portanto

$$E[X] = E[\sqrt{2}Z + 1] = \sqrt{2}E[Z] + 1 = 1,$$

e

$$Var(X) = E[(X-1)^2] = E[2Z^2] = 2E[Z^2] = 2Var[Z] = 2.$$

Propriedades

Se $X \sim \text{Normal}(\mu, \sigma^2)$, então

$$E[X] = \mu \quad \text{e} \quad \text{Var}[X] = \sigma^2,$$

e portanto o desvio padrão de X é σ .

Além disso, se $a, b \in \mathbb{R}$, então

$$aX + b \sim \text{Normal}(a\mu + b, (a\sigma)^2)$$

e, em particular,

$$\frac{X - \mu}{\sigma} \sim \text{Normal}(0, 1).$$

Outra propriedade importante da distribuição normal, além da simetria em torno de da média μ , diz respeito a como as probabilidades se distribuem em torno da média μ .

Observe que para todo $a > 0$,

$$P(\mu - a\sigma < X < \mu + a\sigma) = P(-a\sigma < X - \mu < a\sigma) = P\left(-a < \frac{X - \mu}{\sigma} < a\right),$$

onde, como já vimos, $(X - \mu)/\sigma = Z \sim \text{Normal}(0,1)$.

Em outras palavras, a probabilidade de uma v.a. Normal estar à uma distância de “ a desvios padrão” da média, **não depende** da média μ ou do desvio padrão σ .

Alguns valores interessantes de se ter na cabeça são:

$$P(\mu - \sigma < X < \mu + \sigma) \approx 0,6827$$

e

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0,9545.$$

Em resumo,

$$P(\mu - a\sigma < X < \mu + a\sigma),$$

crece apenas com a , mas não o desvio padrão σ .

Para observar essa propriedade em ação, fizemos um pequeno applet, que está disponível no site no curso no Moodle.

É também possível acessar o applet em

<https://www.geogebra.org/m/m5hg4b2r>

Uma companhia de equipamentos eletrônicos pretende distribuir um dos seus produtos para uma certa cidade. Seus aparelhos não suportariam uma corrente acima de 15 miliamperes, mas por segurança eles não pretendem vender se a probabilidade da corrente passar de 13 miliamperes for maior que 5%.

Baseados nos dados enviados pela companhia elétrica, eles estimam que a corrente elétrica I que passaria pelo seu produto quando ligado nesta cidade tem distribuição normal, com média de 10 miliamperes e desvio padrão de 2 miliamperes.

- a Qual será a decisão da companhia?
- b Qual o valor máximo de σ para que a companhia decidisse vender o produto nesta região?

- a Sabemos que $I \sim \text{Normal}(10,4)$, e que a companhia só venderá o produto se $P(I > 13) < 0,05$.

Usando a [calculadora de probabilidades](#), encontramos que

$$P(I > 13) = 0,0668 > 0,05.$$

E a empresa não vai vender o produto!

- b Dado $X \sim \text{Normal}(10,\sigma)$, queremos saber o valor de σ para o qual

$$P(X > 13) = 0,05.$$

Mas

$$P(X > 13) = P(X - 10 > 3) = P\left(\frac{X - 10}{\sigma} > \frac{3}{\sigma}\right) = P(Z > 3/\sigma),$$

onde $Z \sim \text{Normal}(0,1)$.

Queremos então encontrar $a \in \mathbb{R}$ tal que $P(Z > a) = 0,05$.

E, felizmente, nossa [calculadora de probabilidades](#) é capaz de fazer essa conta!

Fazendo $\mu = 0$ e $\sigma = 1$ na calculadora, selecionando a opção de cálculo para intervalos $[a, +\infty)$, e entrando “0.05” no resultado da probabilidade, obtemos:

$$P(Z > 1,6449) = 0,05.$$

Temos então que $\frac{3}{\sigma} = 1,6449$ e portanto

$$\sigma \approx 1,823819.$$