V.A.s continuas

O paradoxo do contínuo

- Seja X uma v.a. cujos valores possíveis formam um intervalo da reta [a,b]
- Temos uma situação paradoxal:
  - Seja x <u>qualquer</u> valor específico em [a,b]. Por exemplo, x=0.2367123
  - Então P(X=0.2367123) = 0
  - Isto vale para qualquer valor especifico x em [a,b]
  - No entanto, $P(X \in [a,b]) = 1$
  - Assim, todo valor especifico em [a,b] tem probabilidade ZERO de acontecer mas algum numero em [a,b] acontece com certeza
- Isto e’ similar ao paradoxo de uma barra de densidade uniforme ser representada pelo segmento [0,1] ter uma massa 1 kg mas nenhum ponto no segmento [0,1] poder ter massa > 0

Função densidade

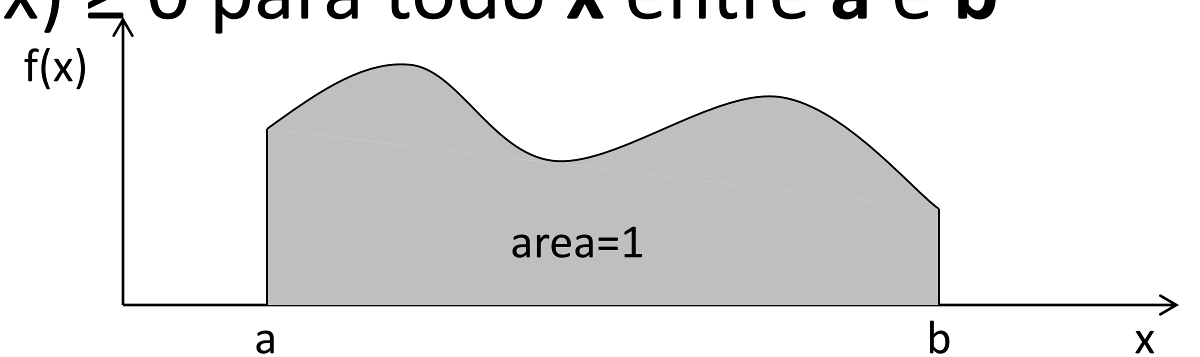
- Como sair dessa situação e trabalhar com v.a.'s continuas?
- A resposta esta na função densidade de probabilidade.
- Uma função f(x) > 0 definida na reta tal que

$$1 = \int_{-\infty}^{+\infty} f(x)dx$$

- QUALQUER função f(x) com esta propriedade representa a densidade de uma v.a.

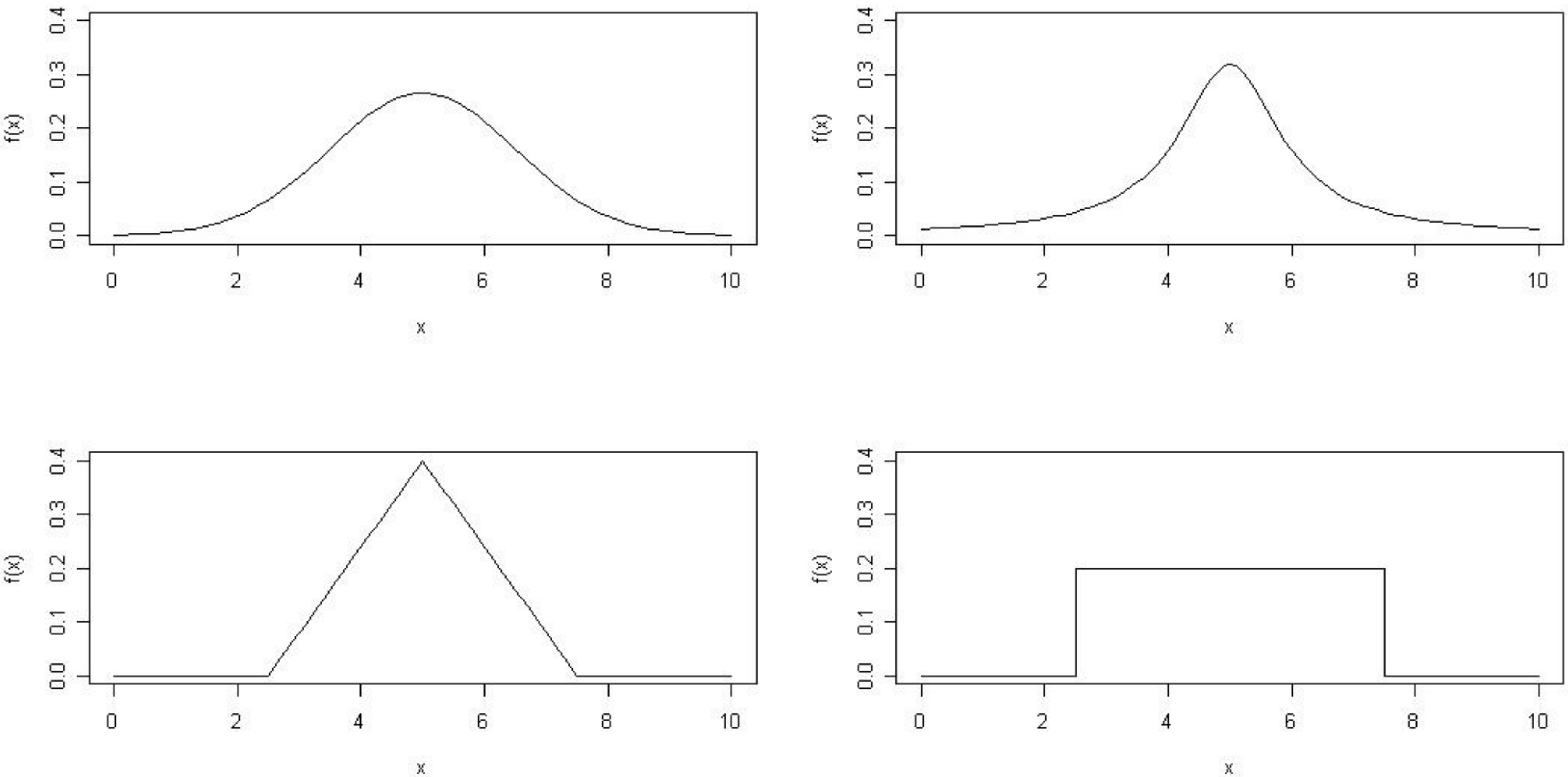
Funcao densidade de probabilidade

- $f(x)$ e' chamada uma ***funcao densidade de probabilidade*** (na regiao $-\infty \leq \mathbf{a} \leq \mathbf{x} \leq \mathbf{b} \leq \infty$) se ela atende aos requisitos:
- $f(x) \geq 0$ para todo **x** entre **a** e **b**

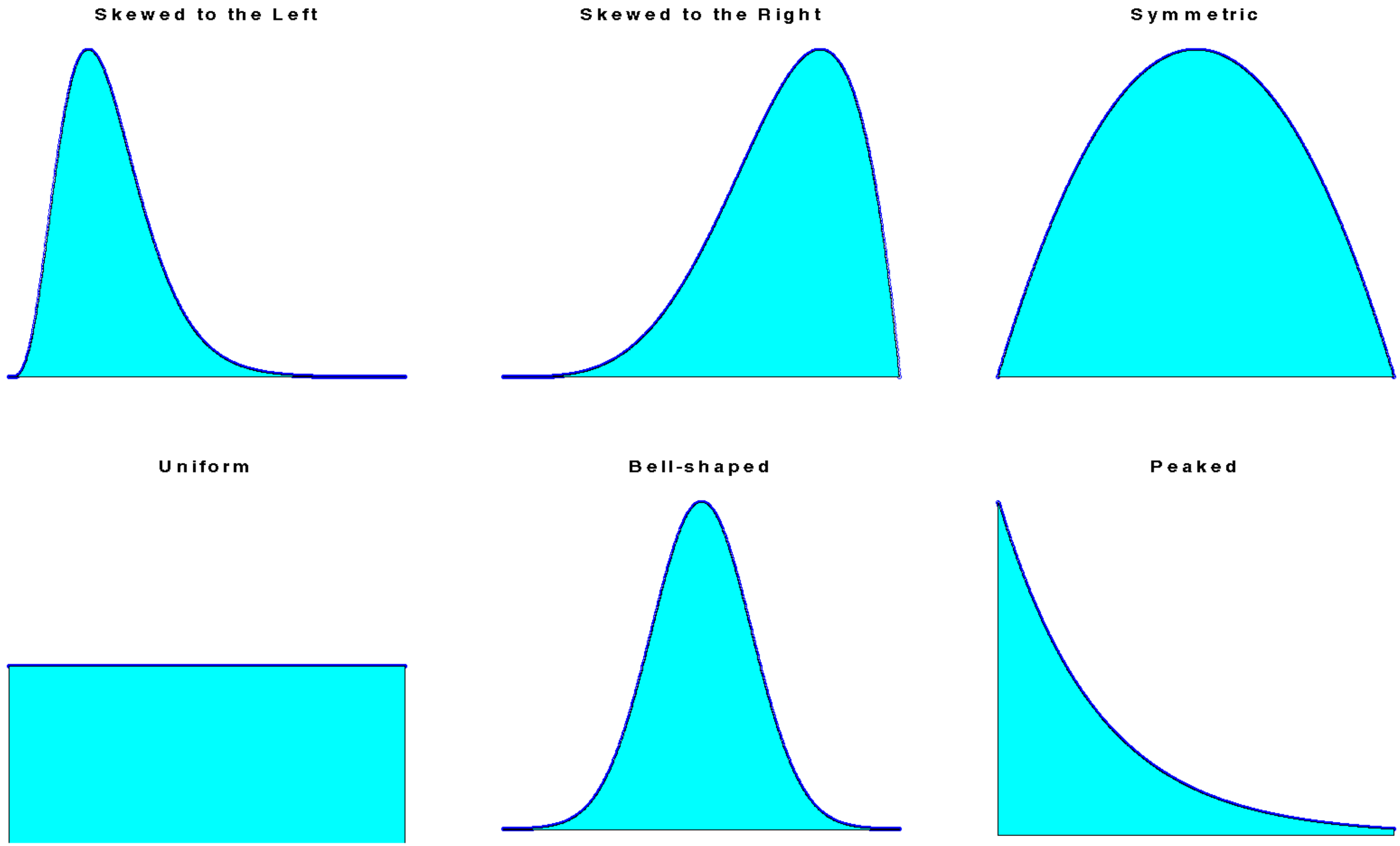


1) A area total sob a curva $f(x)$ entre **a** e **b** e' 1.0

Exemplos de possíveis shapes para f(x)



Mais exemplos de shapes para f(x)

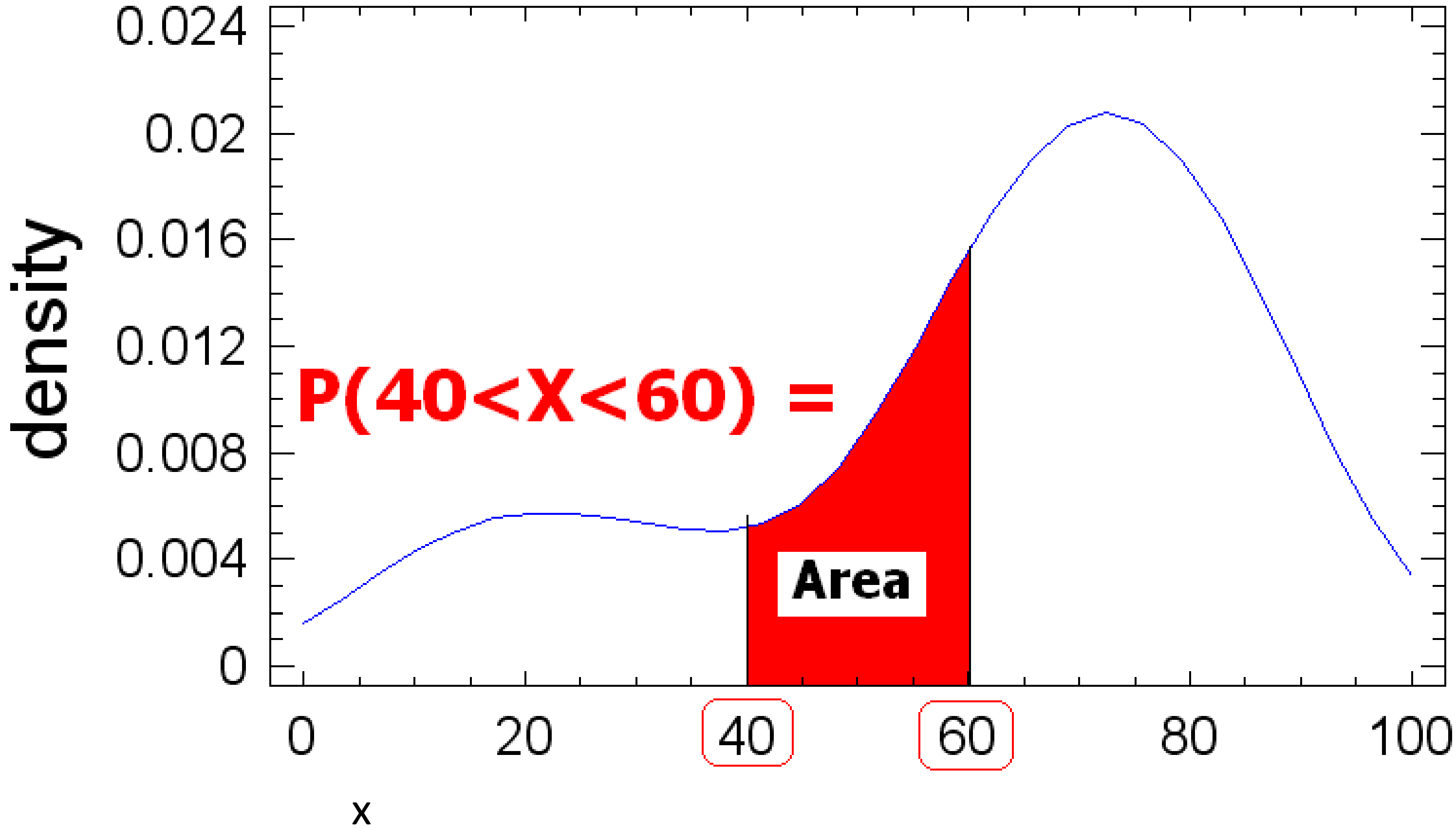


Probab = area sob a curva

- Densidade fornece uma maneira visual de calcular probabilidades

$$P(X \in [c,d]) = \int_c^d f(x)dx = \text{area sob a curva}$$

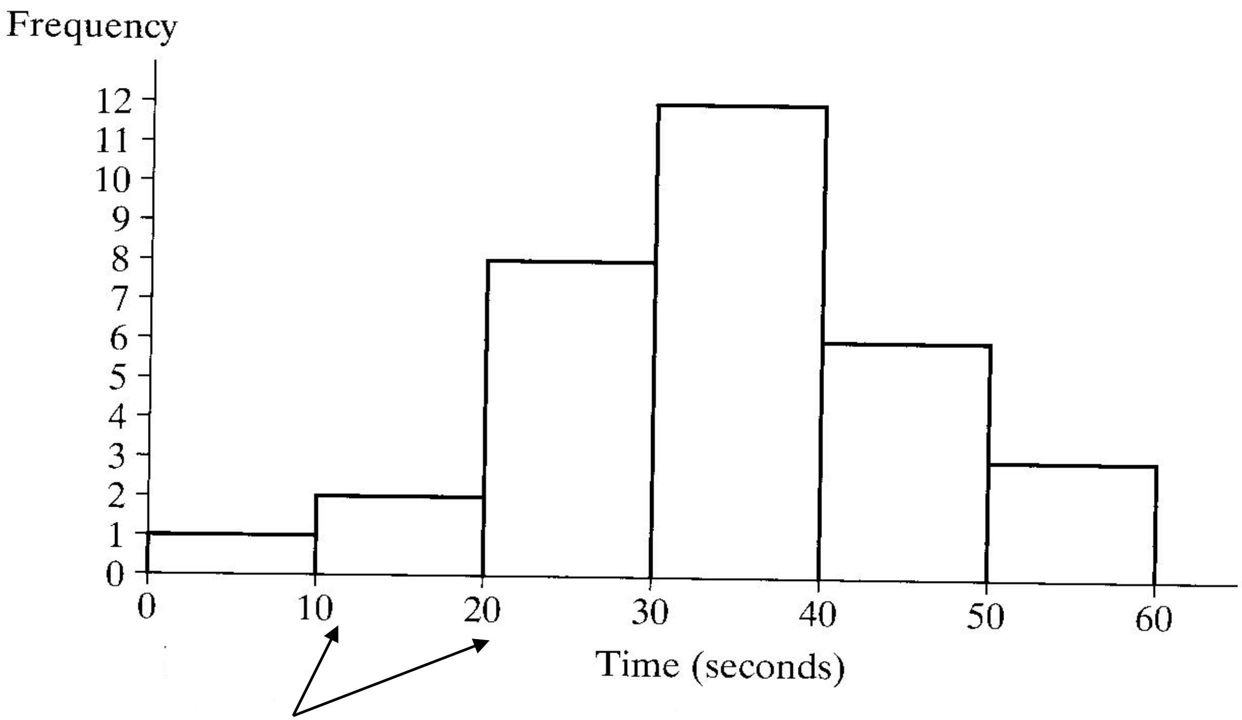
Continuous Distribution



Identificando uma densidade

- Uma amostra de n instancias
- Medimos v.a.'s continuas nas n instancias
- $X_1, X_2, ..., X_n$: copias independentes de uma v.a. X
- Qual a distribuição da v.a. X?
- Podemos ter uma ideia da densidade de X olhando para o histograma dos dados.
- O que e' o histograma?

Histogram



A histogram is made up of a series of bars or rectangles

For continuous data, the class boundaries are written as part of a continuous scale

The area of each rectangle represents the frequency of a class interval.

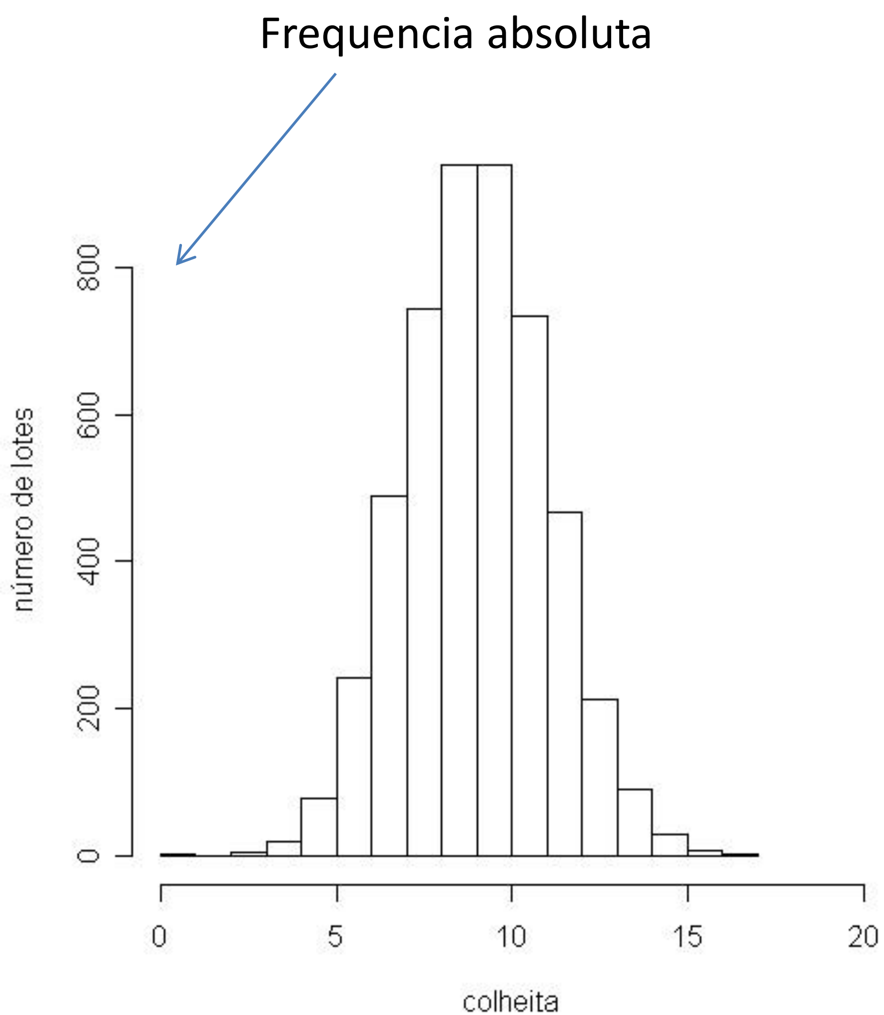
Histograma normalizado

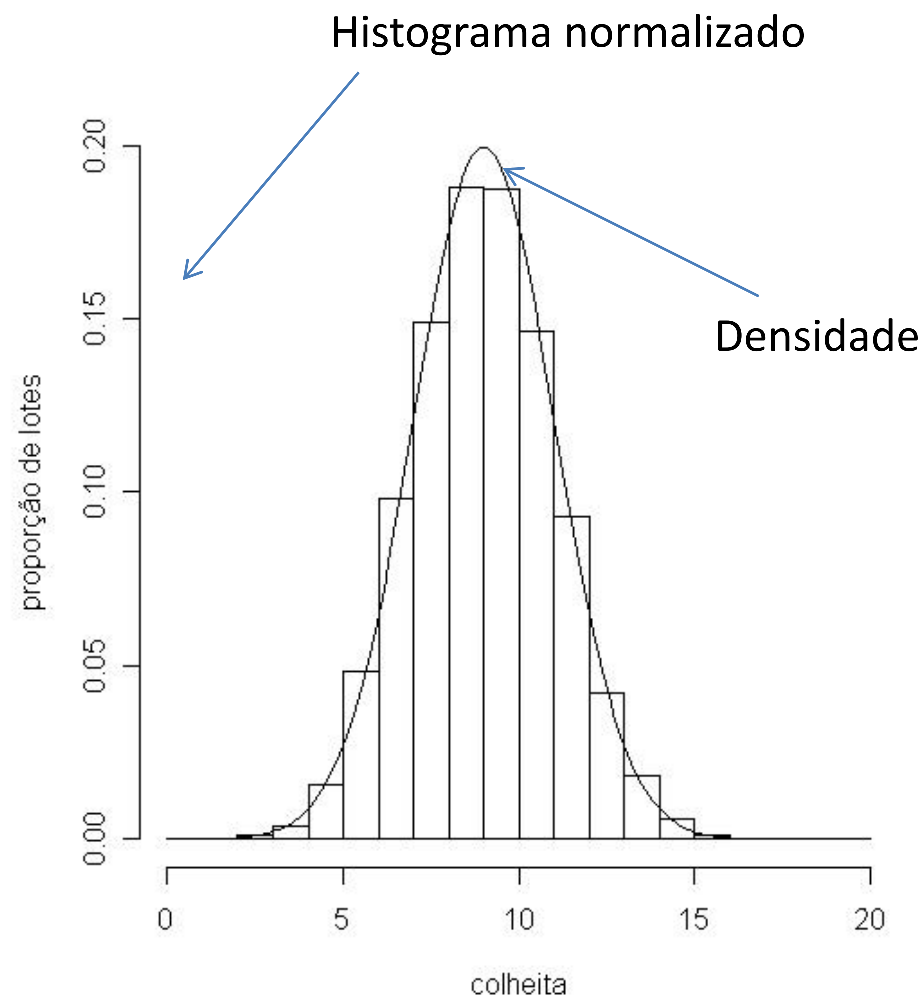
- Normalize a histogram: make the total area of all rectangles equal to 1.
- Se altura do retangulo = frequencia na classe
- Solucao: divida a altura dos retangulos por n*Δ onde Δ e' o comprimento da classe.
- Verifique que a area total agora e' 1:

$$\text{area total} = \sum_i \frac{\#\{X_i \in \text{classe i}\}}{n\Delta} * \Delta = \frac{1}{n} \sum_i \#\{X_i \in \text{classe i}\} = 1$$

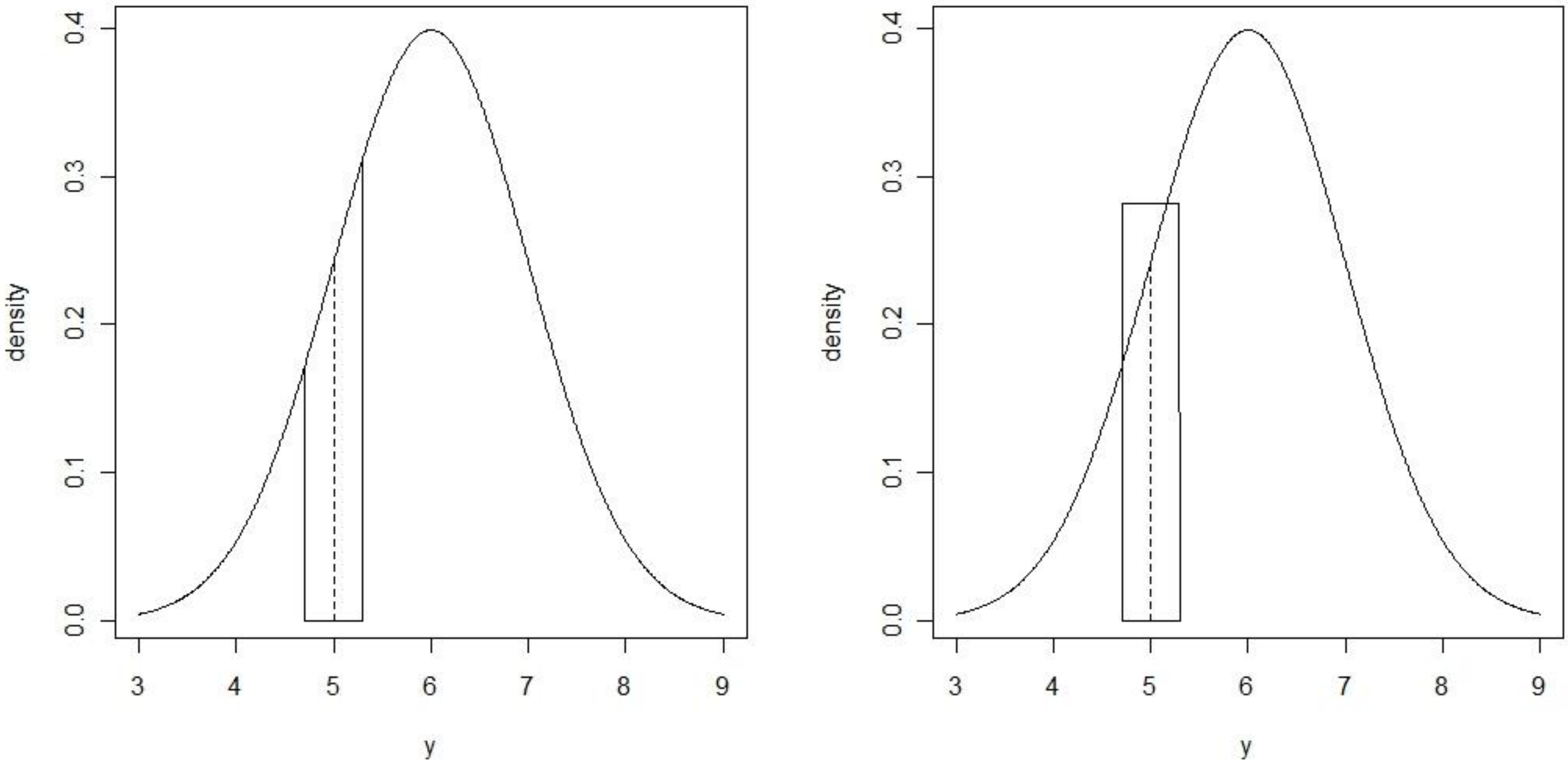
- O histograma vai ser parecido com a densidade subjacente. Por que?

Histograma ≈ densidade





Similaridade de histograma e densidade



Histograma ≈ densidade

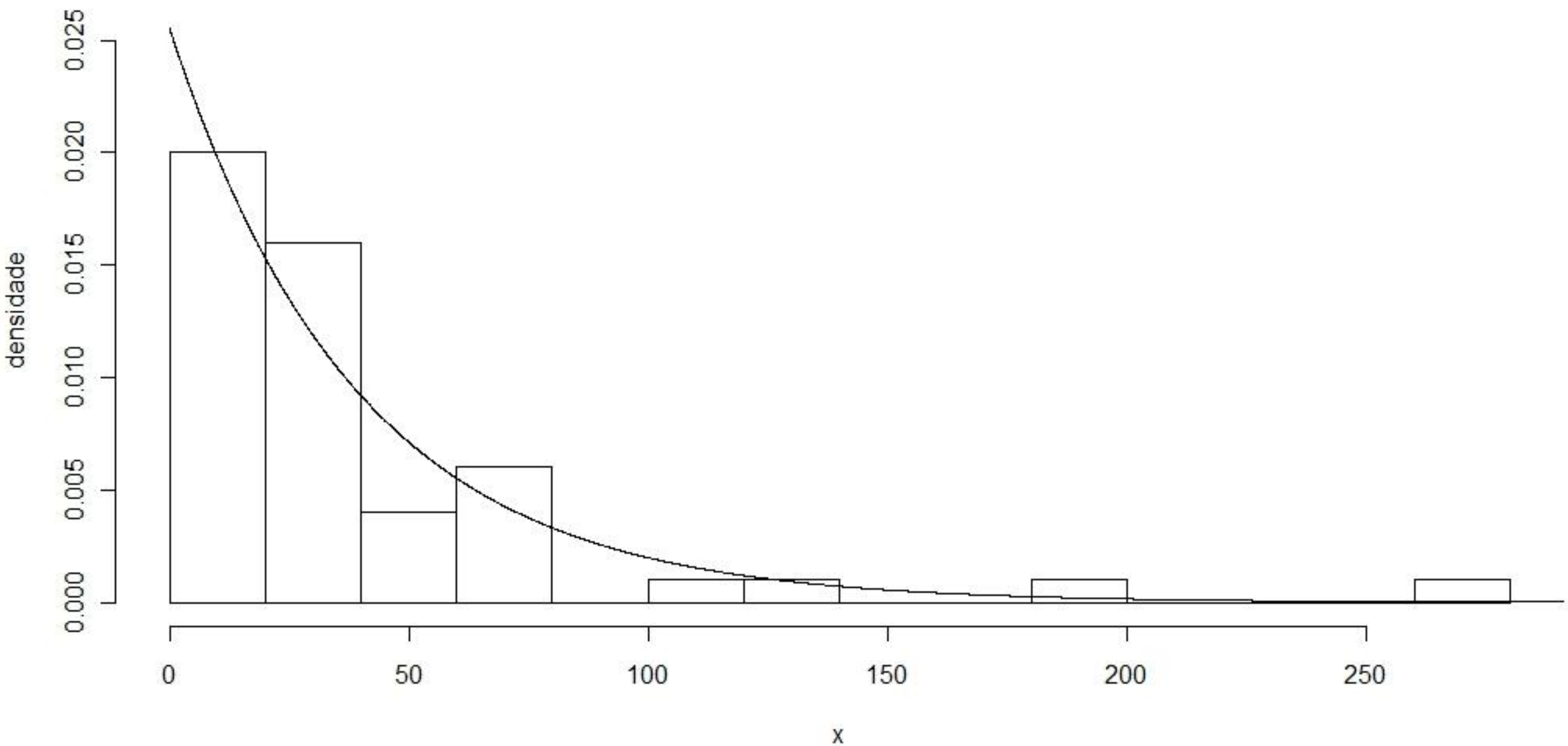
- Histograma normalizado da a dica de qual e' a densidade dos dados.
- Probab de cair num intervalo pequeno centrado em y_0 e de comprimento δ e' aproximadamente

$$\frac{\#\{Y_i's \in (y_0 - \delta/2,\ y_0 + \delta/2)\}}{n}$$

- Mas isto deve ser aproximadamente igual a

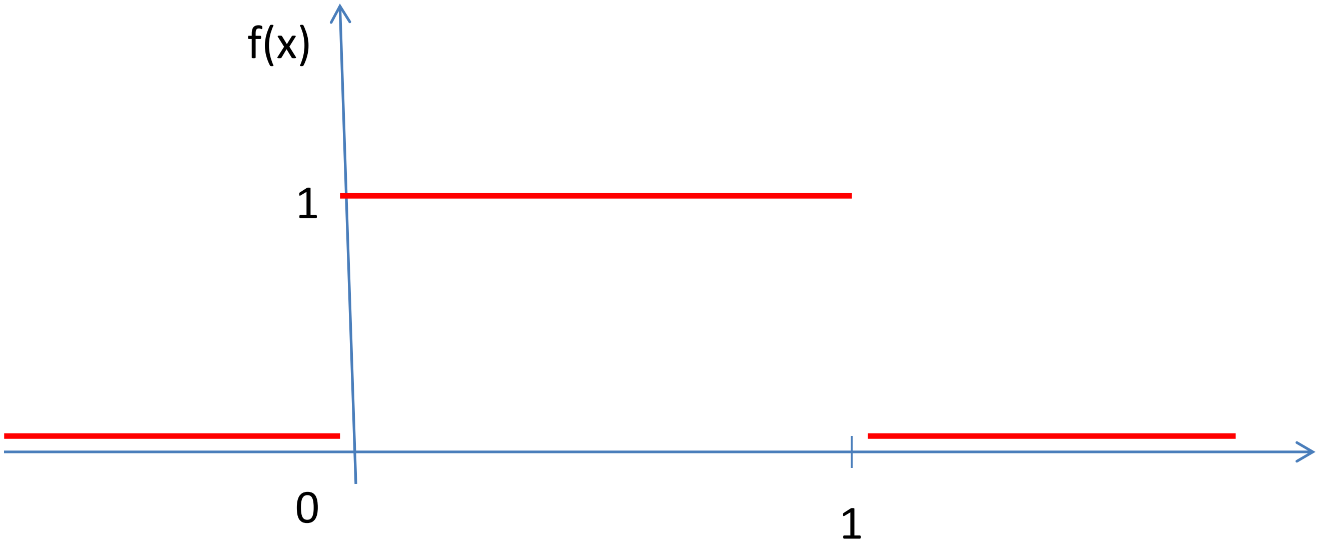
$$\mathbb{P}\left(Y \in (y_0 - \delta/2, y_0 + \delta/2)\right) = \int_{y_0-\delta/2}^{y_0+\delta/2} f^*(y)\, dy \approx f^*(y_0)\delta$$

Exemplo



Distribuição uniforme

- X ~ Uniforme em (0,1): U(0,1)

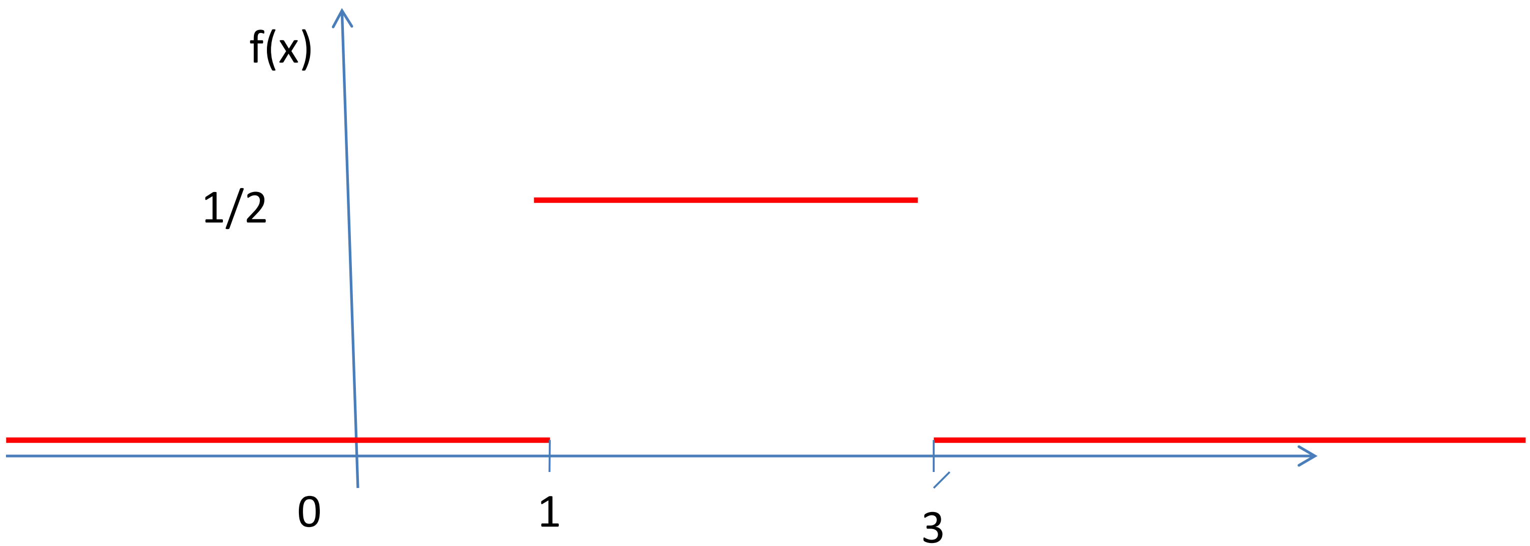


$$P(X \in [c,d]) = \int_c^d f(x)dx = \int_{[c,d]\cap[0,1]} f(x)dx = \int_{[c,d]\cap[0,1]} 1dx =$$

$$= \text{comprimento de } [c,d] \cap [0,1]$$

Distribuição uniforme

- Uniforme em (1,3) → notação: U(1,3)



$$P(X \in [c,d]) = \int_c^d f(x)dx = \int_{[c,d]\cap[1,3]} f(x)dx = \int_{[c,d]\cap[1,3]} 0.5dx =$$

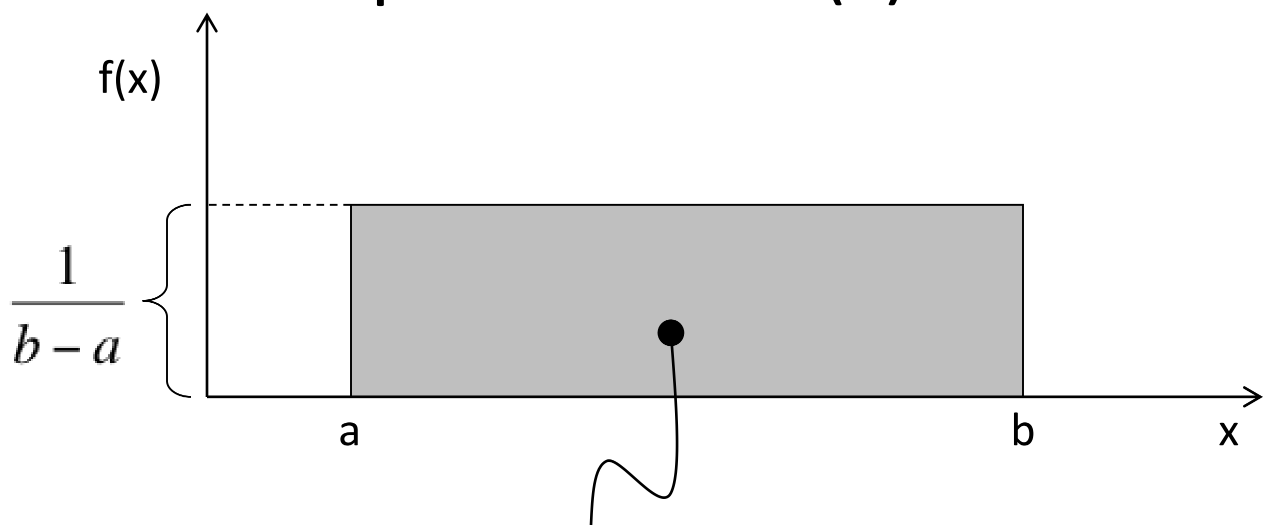
$$= 0.5 * \text{comprimento de} [c,d] \cap [1,3]$$

Distribuicao Uniforme em geral

- X ~ Unif(a,b) se a sua densidade e' igual a

$$\boxed{f(x) = \frac{1}{b-a}, \ where \ \ a \le x \le b}$$

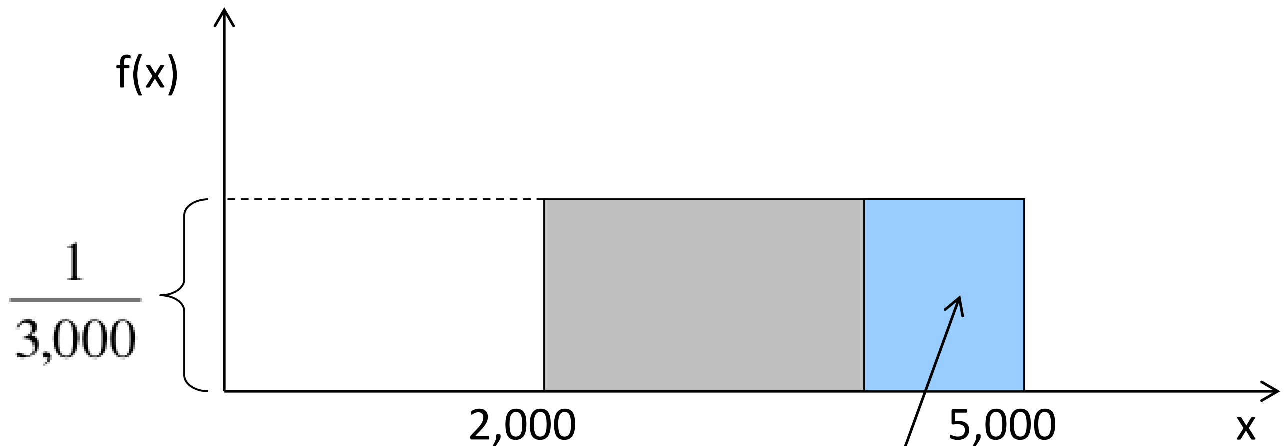
- E' descrita pela funcao f(x) abaixo:



area = width x height = (b – a) x $\frac{1}{b-a}$ = 1

Exemplo

- The amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons.



- ***What is the probability that the service station will sell <u>at least</u> 4,000 gallons?***
- Algebraically: what is **P(X ≥ 4,000)** ?
- **P(X ≥ 4,000)** = (5,000 – 4,000) x (1/3000) = .3333

Exponential Distribution

- Another important continuous distribution is the ***exponential distribution*** which has this probability density function:

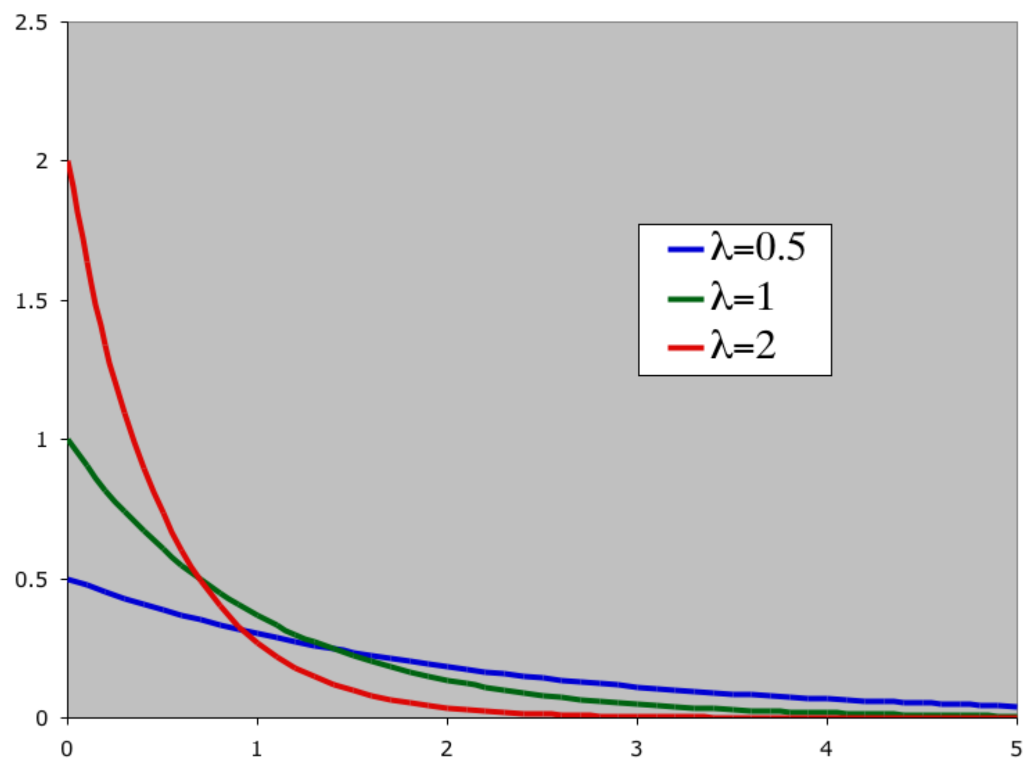
$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

- Note that x ≥ 0.
- λ is a constant > 0
- Time (for example) is a non-negative quantity; the exponential distribution is often used for time related phenomena such as the length of time between phone calls or between parts arriving at an assembly station.

Exponential Distribution

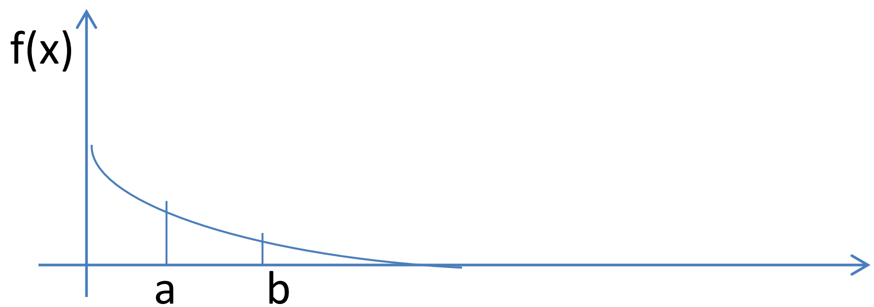
- The exponential distribution depends upon the value of λ
- Smaller values of λ "flatten" the curve → increase the chance of larger values

- Exponential
- distributions for
- λ = .5, 1, 2)



Calculando probab com exp(λ)

- Algumas probabilidades (com a > 0 e x > 0):



$$P(X \in (a,b)) = \int_a^b \lambda e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}$$

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t} dt = e^{-\lambda x} - e^{-\lambda \infty} = e^{-\lambda x}$$

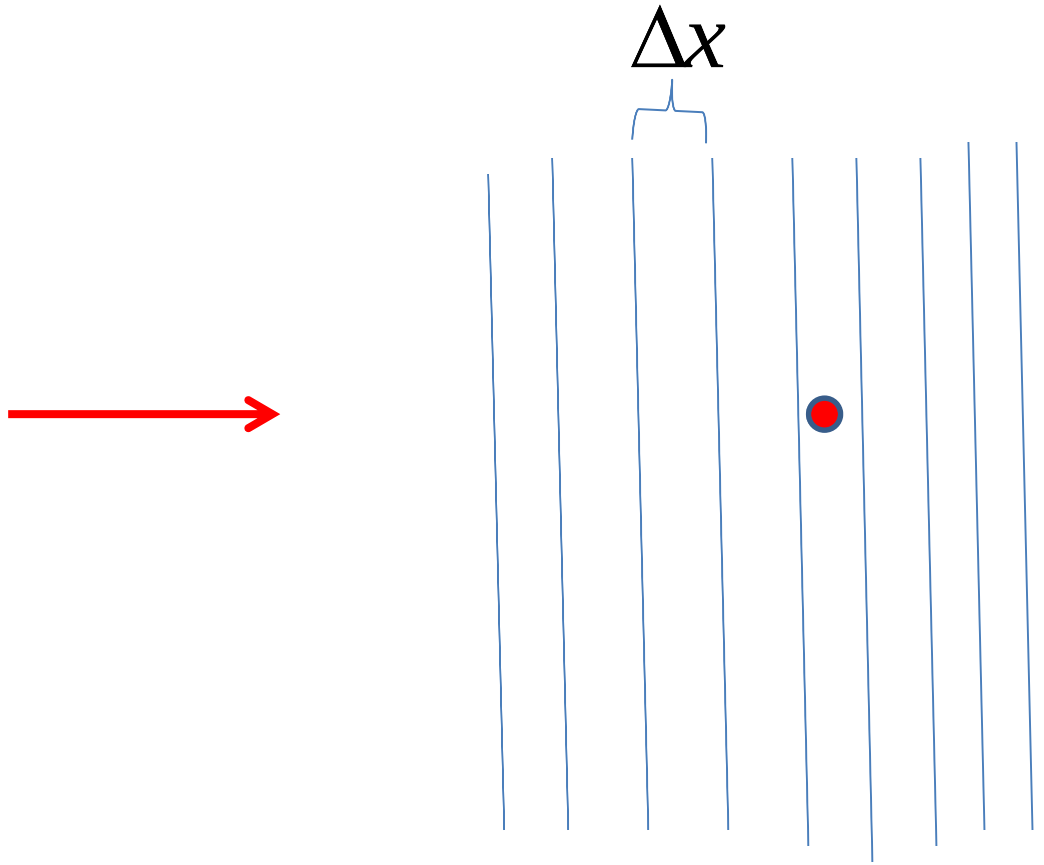
Quando uma v.a. e’ exp(λ)?

- Se X e’ uma v.a. com valores possiveis na semi-reta $(0, \infty)$
- Se $f(x) = \lambda e^{-\lambda x}$
- OU
- Se $P(X > x) = e^{-\lambda x}$ para $x > 0$

- ENTAO $X \sim \exp(\lambda)$

Como aparece a distribuição exponencial?

- Um modelo para a penetração de uma partícula em material metálico



Como aparece a distribuição exponencial?

- Um modelo para a penetração de uma partícula em material metálico
- O metal constituído de placas finas, cada uma com a mesma espessura Δx.
- Se a partícula passa através da *n*-ésima placa:
  - ela possui probabilidade $\rho \Delta x$ de ser absorvida na placa *n+1*
  - e probabilidade $1- \rho \Delta x$ de passar através da placa *n+1*
- $\rho > 0$ é uma constante que depende do material.
- E’ como jogar uma moeda com P(sucesso)= $1- \rho \Delta x$ ate’ obter o primeiro fracasso ...

Probab(penetrar k placas)

- $P(A_k)$ = P(penetrar pelo menos k placas) = ??
- E’ o mesmo que jogar a moeda para cima k vezes e obter k sucessos sucessivos
- $P(A_k)$ = P(k sucessos sucessivos) = $(1-\rho\Delta x)^k$

Probab(penetrar mais que x)

- $P(A_k)$ = P(penetrar k placas ou mais) = $(1-\rho\Delta x)^k$
- Suponha que x = certa profundidade FIXA no material (um numero real > 0)
- P(penetrar mais que x) = ??
- Divida o intervalo (0, x) em muitos pequenos intervalos de comprimento Δx
- Isto e', se x = k * Δx → k = x/Δx
- Portanto P(penetrar mais que x) ≈ P(penetrar k placas ou mais) =

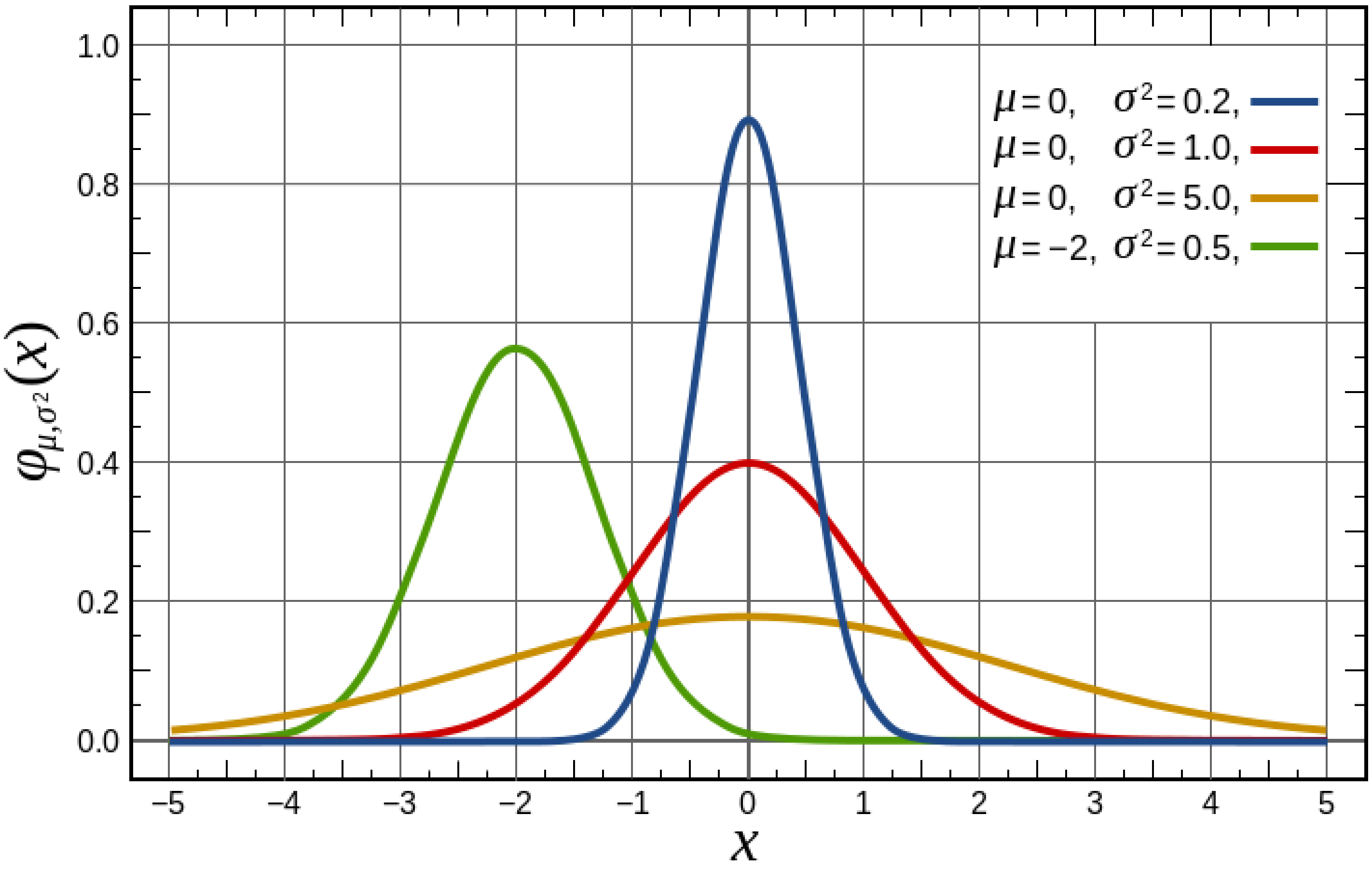
$$(1-\rho\Delta x)^{x/\Delta x} = \left[(1-\rho\Delta x)^{1/\Delta x}\right]^x \rightarrow \left[e^{-\rho}\right]^x = e^{-\rho x}$$

Conclusão: X ~ exp(λ)

- Seja X = profundidade penetrada
- Então $P(X > x) = e^{-\rho x}$
- Mas se $X \sim \exp(\rho)$ então $P(X > x) = e^{-\rho x}$
- Isto implica que a profundidade penetrada X e’ uma v.a. com distribuição exponencial.

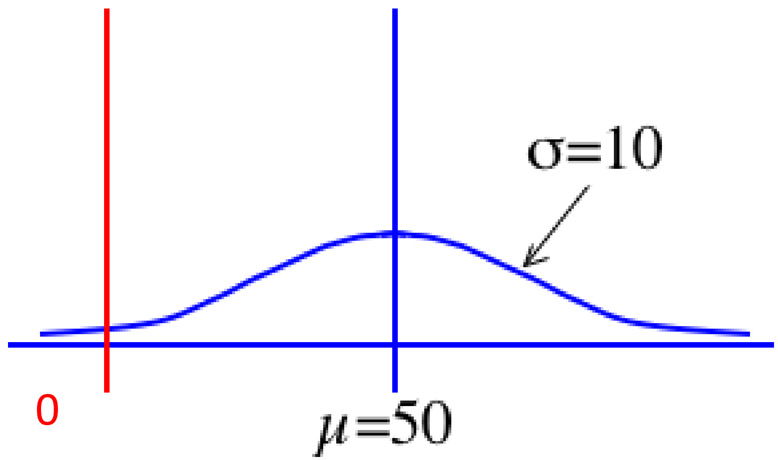
Distribuição normal (ou gaussiana)

- Densidade: $f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$



Calculating Normal Probabilities…

- Sempre caimos na normal com mu=0 e sigma=1
- Example: The time required to build a computer is ***normally distributed*** with a mean of 50 minutes a of 10 minutes:

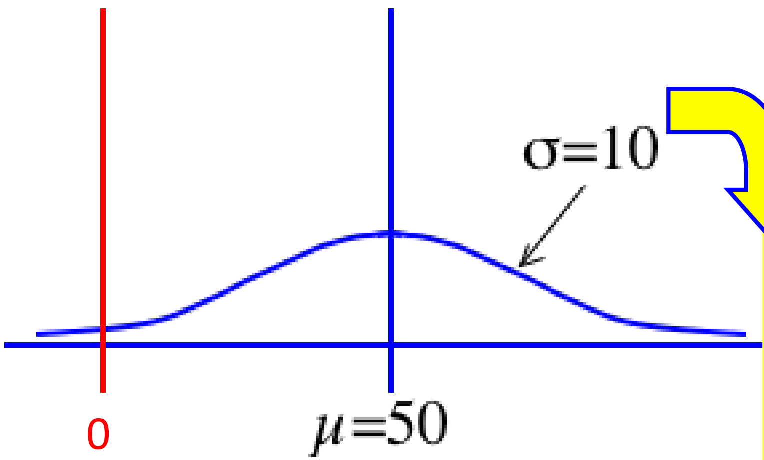


- What is the probability that a computer is assembled in a time between 45 and 60 minutes?

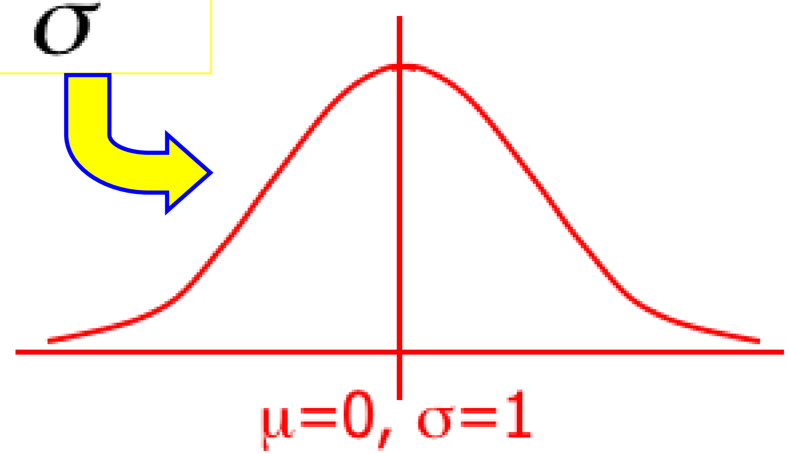
- Algebraically speaking, what is **P(45 < X < 60)** ?

Calculating Normal Probabilities…

…mean of 50 minutes and a standard deviation of 10 minutes…



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



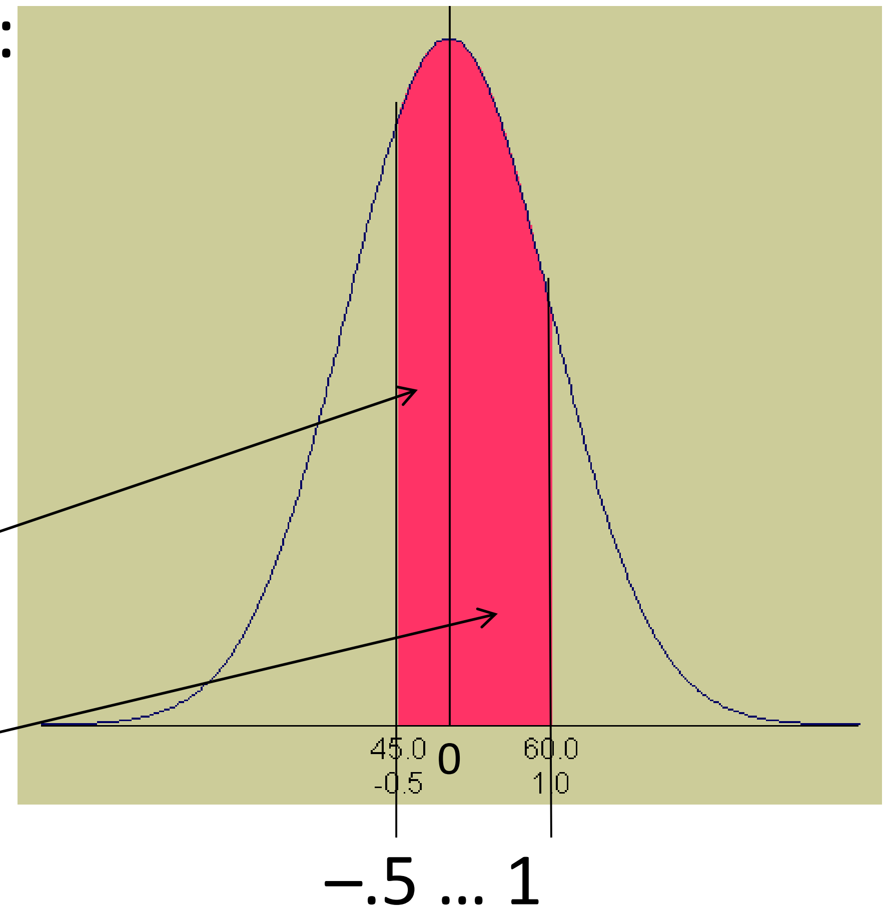
$$P(45 < X < 60) =$$

$$P\left(\frac{45-50}{10} < \frac{X-\mu}{\sigma} < \frac{60-50}{10}\right) =$$

$$P(-.5 < Z < 1)$$

Calculating Normal Probabilities…

- **P(–.5 < Z < 1)** looks like this:

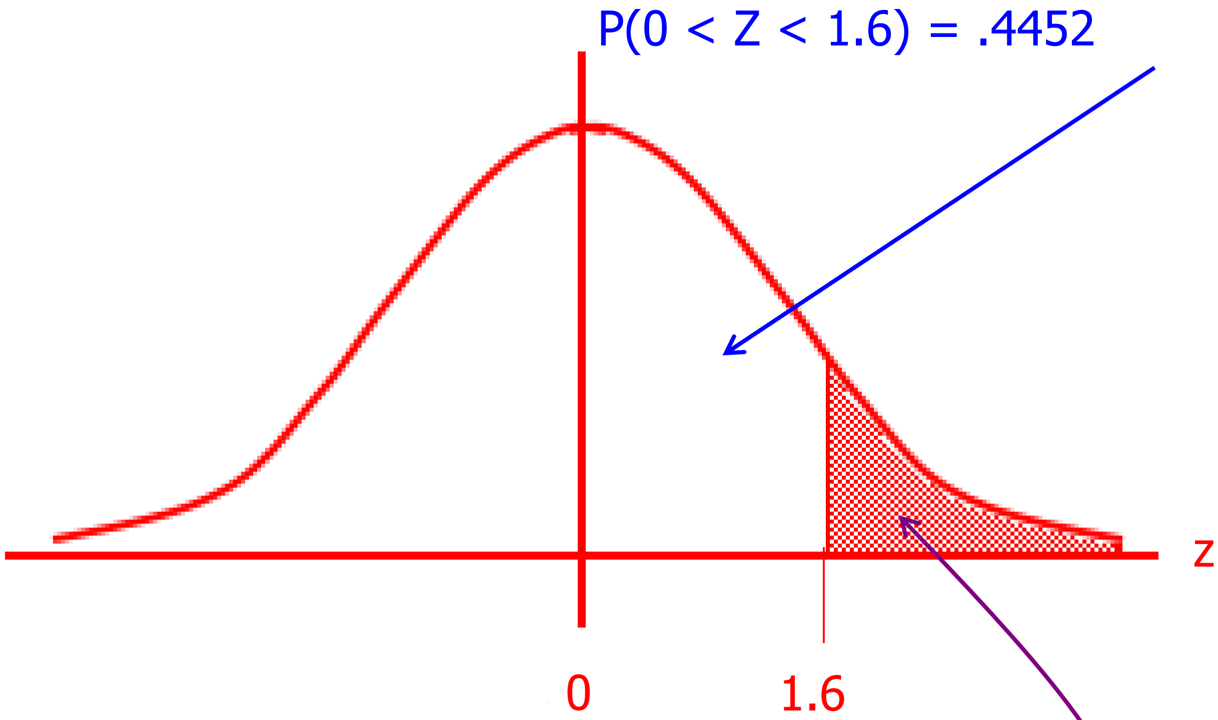


- The probability is the area
- under the curve…

- We will add up the
- two sections:
- P(–.5 < Z < 0) and
- P(0 < Z < 1)

Using the Normal Distribution

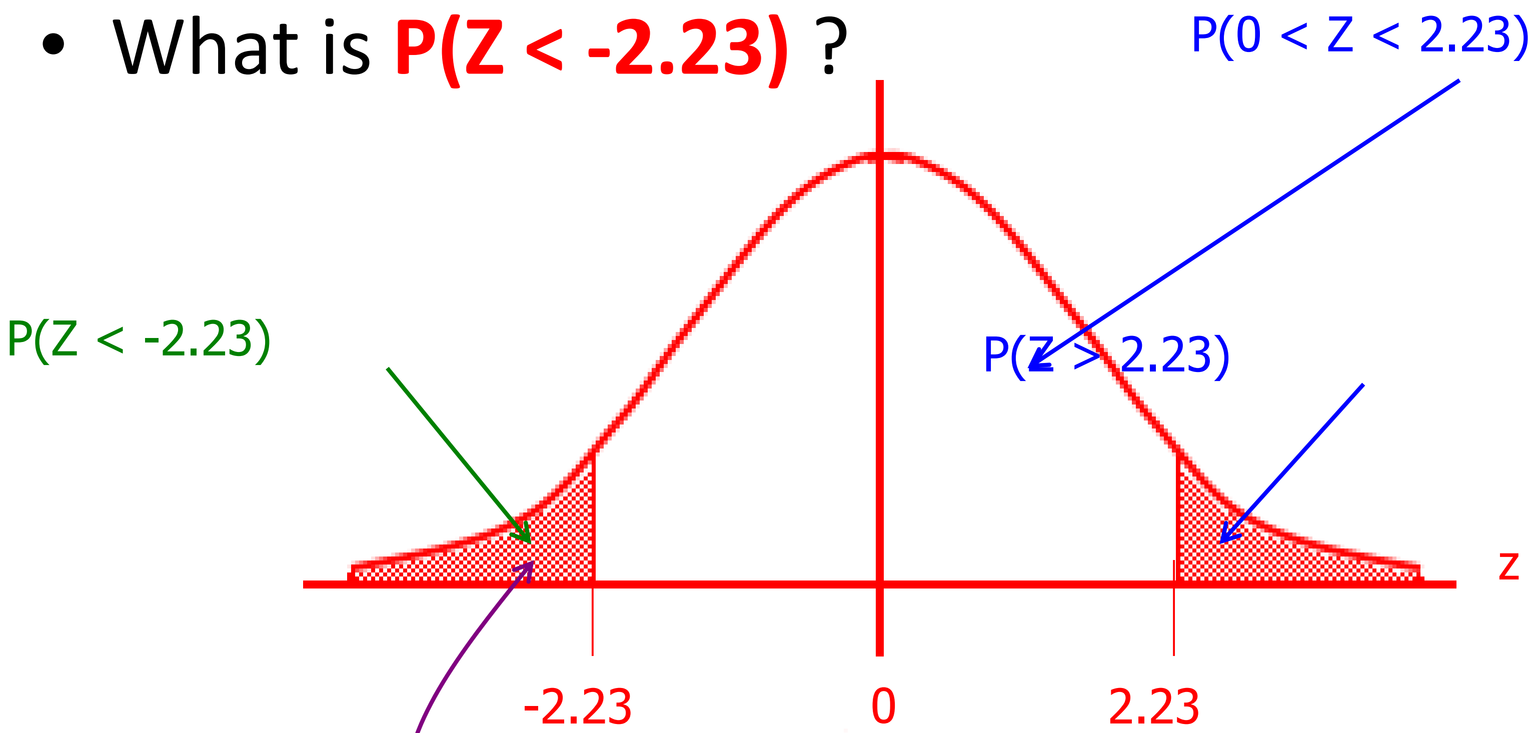
- What is **P(Z > 1.6)** ?



P(Z > 1.6) = .5 – P(0 < Z < 1.6)
= .5 – .4452
= **.0548**

Using the Normal Distribution

- What is **P(Z < -2.23)** ?



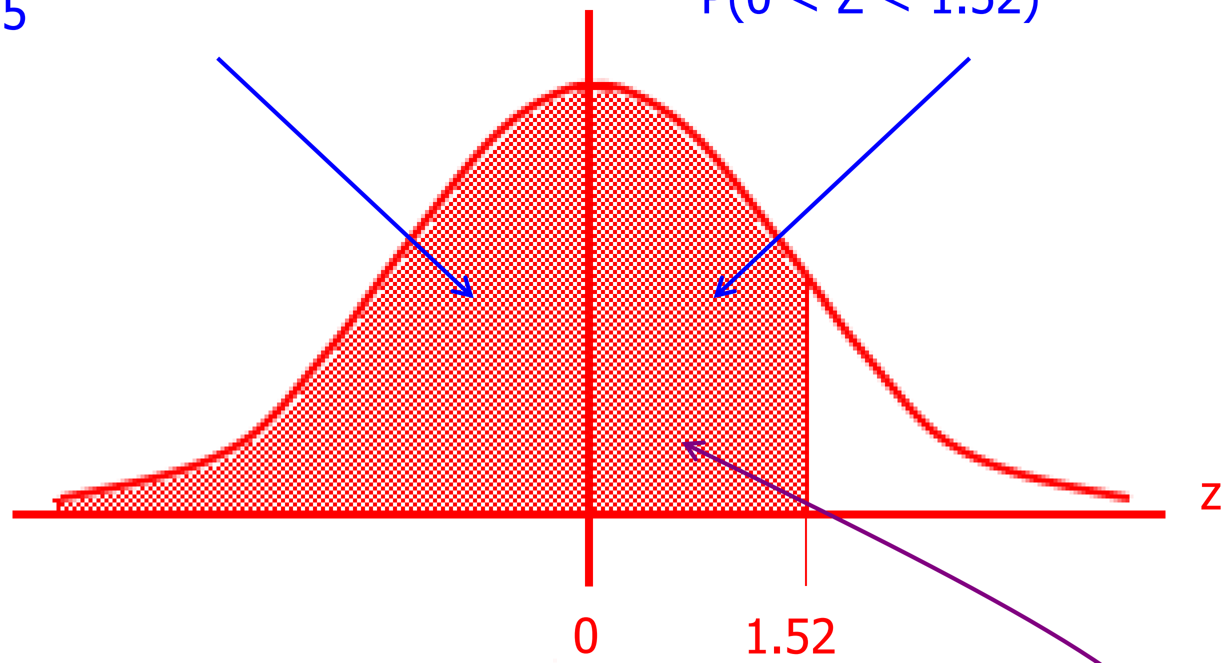
P(Z < -2.23) = P(Z > 2.23)
= .5 – P(0 < Z < 2.23)
= **.0129**

Using the Normal Distribution

- What is **P(Z < 1.52)** ?

P(Z < 0) = .5

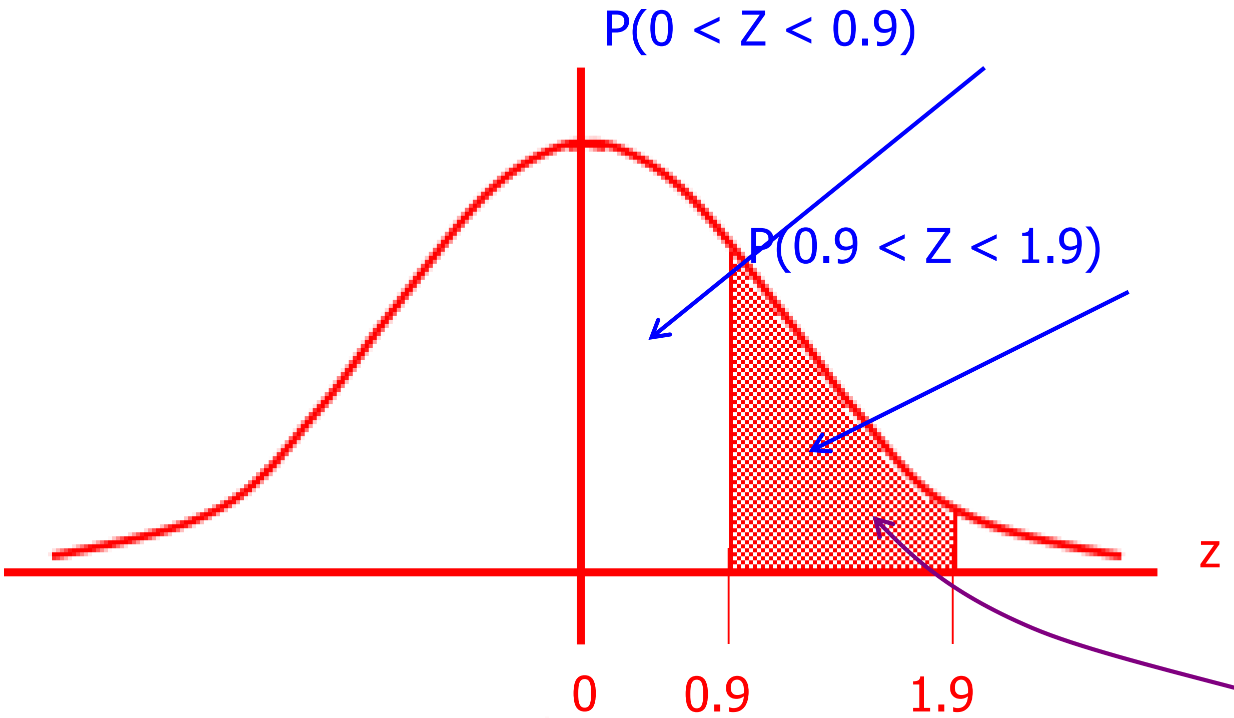
P(0 < Z < 1.52)



P(Z < 1.52) = .5 + P(0 < Z < 1.52)
= .5 + .4357
= **.9357**

Using the Normal Distribution

- What is **P(0.9 < Z < 1.9)** ?



$P(0.9 < Z < 1.9) = P(0 < Z < 1.9) - P(0 < Z < 0.9)$

$= .4713 - .3159$

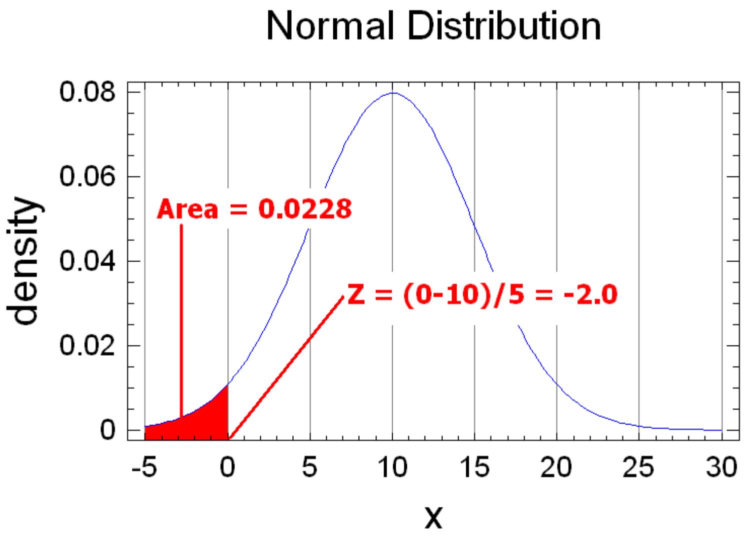
$=$ **.1554**

Example

- The return on investment is normally distributed with a mean of 10% and a standard deviation of 5%. What is the probability of losing money?

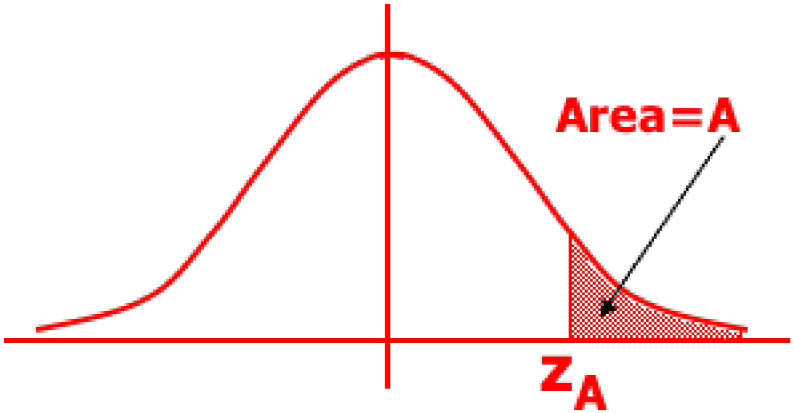
- We want to determine P(X < 0). Thus,

$$
\begin{aligned}
P(X < 0) &= P\left(\frac{X-\mu}{\sigma} < \frac{0-10}{5}\right) \\
&= P(Z < -2) \\
&= .5 - P(0 < Z < 2) \\
&= .5 - .4772 \\
&= .0228
\end{aligned}
$$



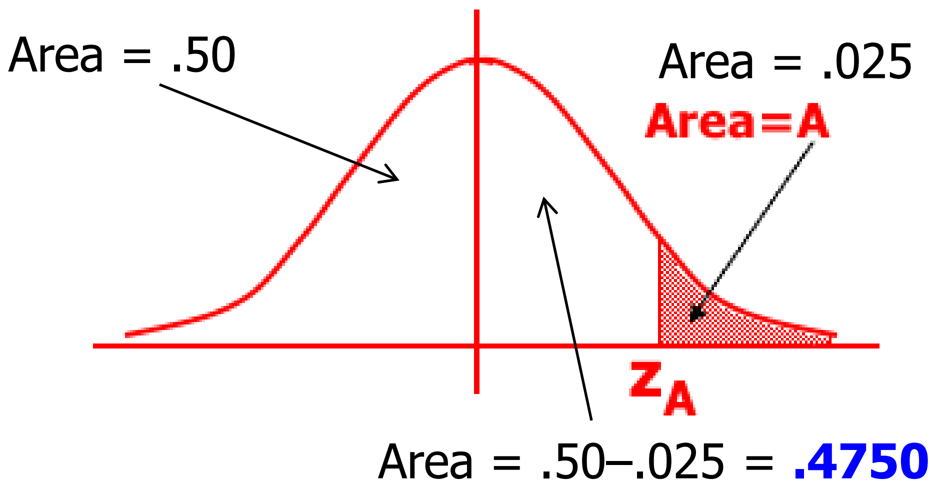
Finding Values of Z_A...

- Often we're asked to find some value of Z for a given probability, i.e. given an area (A) under the curve, what is the corresponding value of z (z_A) on the horizontal axis that gives us this area? That is:



Finding Values of Z…

- What value of z corresponds to an area under the curve of 2.5%? That is, what is $z_{.025}$?



If you do a "reverse look-up" for .4750,
you will get the corresponding $z_A = 1.96$
Since $P(z > 1.96) = .025$, we say: $\mathbf{z_{.025} = 1.96}$

Beta

Existem vários fenômenos cujas variáveis de interesse tem seus valores limitados acima e abaixo por números conhecidos a e b. Um exemplo típico é constituído por dados que aparecam em forma de proporção:

i) em cada empresa industrial brasileira(com produção acima de certo valor fixo) é medido o vetor (X_1,X_2) onde

X_1= proporção de gastos em salários na produção total durante 1985

X_2= proporção de gastos em energia na produção total durante 1985.

ii) a razão entre o comprimento do fêmur e o comprimento total da perna de um indivíduo.

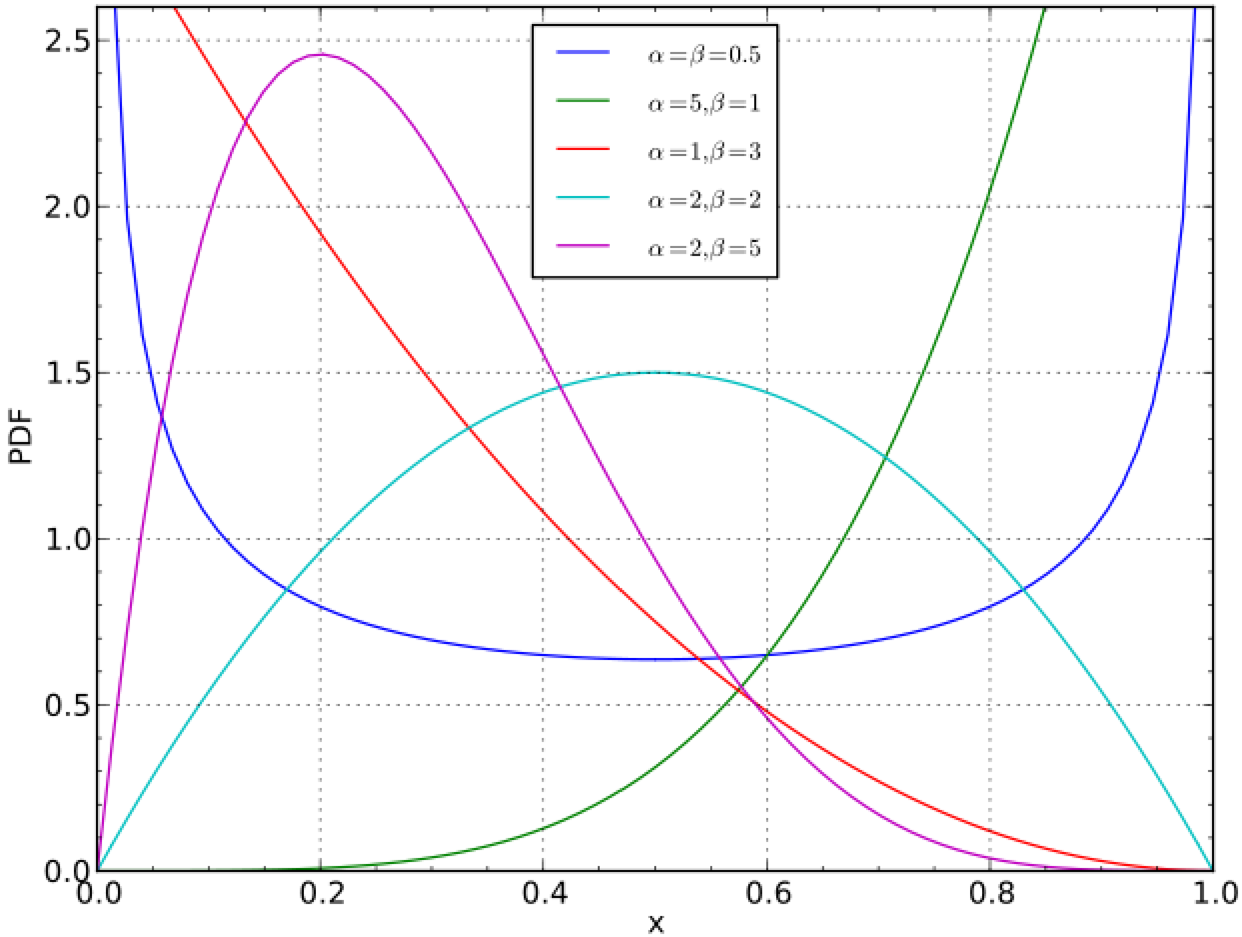
Uma classe de distribuições que inclue a distribuição uniforme e é rica o suficiente para fornecer modelos para a maioria das variáveis aleatórias que têm valores limitados é a classe de distribuições beta.

A variável aleatória Y tem distribuição beta com parâmetros r e s ($Y \sim \beta(r, s)$) onde r > 0 e s > 0 se Y tem densidade dada por

$$f(y) = \begin{cases} \frac{y^{r-1}(1-y)^{s-1}}{B(r,s)} & se \quad 0 \leq y \leq 1 \\ 0 & c.c. \end{cases} \tag{3.3}$$

onde $B(r, s) = \int_0^1 y^{r-1}(1-y)^{s-1}dy$ é a constante que torna f(y) uma densidade. B(r,s) é chamada

Flexibilidade da beta



Variação do preço de ações

Exemplo: Vamos representar por Y a variável que mede a proporção da variação de preços(médios) diários de ações quando estes preços caem. Isto é, se o preço(médio) de uma ação num dia é d_1e o preço desta ação no dia seguinte é $d_2 \leq d_1$ então $y = \dfrac{d_2 - d_1}{d_1}$. Se o preço sobe $(d_1 < d_2)$, Y não é registrada para aquela ação. Um total de 2314 ações com preços que caíram foram observadas. Os dados estão resumidos na forma de distribuição de freqüência na tabela 10 e a figura 16 apresenta um histograma junto com a densidade de uma distribuição beta com parâmetros r=1,038 e s=10,63 . A maneira de escolher estes valores será assunto de um próximo capítulo. Nesta figura parece qeu a distribuição beta fornece um modelo razoável para a variável Y.

Histograma e densidade