Estatística para Metrologia

Aula 3

Mônica Barros, D.Sc.

Abril de 2008

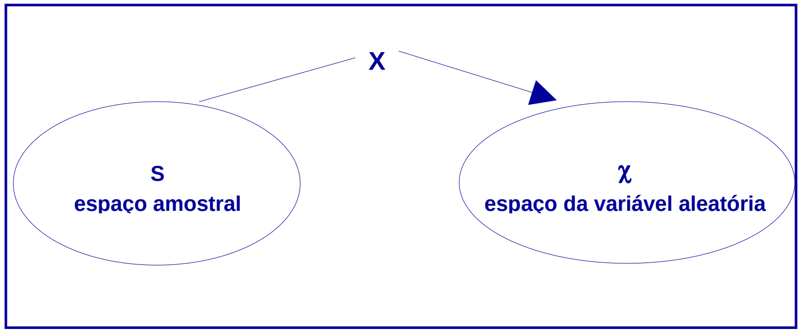
Aula 3

- Variáveis Aleatórias Contínuas e Discretas
- Função de Probabilidade
- Função Densidade
- Função de Distribuição
- Momentos de uma variável aleatória
  Média, Variância e Desvio Padrão

Variáveis Aleatórias

- Muitas vezes o espaço amostral não é um conjunto de valores numéricos. Por exemplo, se jogamos uma moeda 3 vezes, o espaço amostral é S = { HHH, HHT, HTH, THH, HTT, THT, TTH, TTT } , onde cada resultado tem a mesma probabilidade, e T indica "coroa", H indica "cara".
- Seja S o espaço amostral e X uma função que "pega" elementos deste espaço (resultados da experiência) e os leva num subconjunto dos números reais. Esta função X é chamada de ***variável aleatória***.
- **Atenção:** usaremos aqui **X (maiúscula)** para denotar a **variável aleatória** e **x (minúscula)** para indicar um **valor específico da variável**, isto é, um número.

Variáveis Aleatórias



Seja X uma variável aleatória definida num espaço amostral *S* e seja ℵ o espaço de X. Seja A um subconjunto de ℵ e S um subconjunto de *S* (espaço amostral).

Variáveis Aleatórias

- Já definimos a probabilidade de um evento S $\subseteq$ *S (espaço amostral)*, e agora gostaríamos de estender esta definição e falar da probabilidade de um evento A $\subseteq \aleph$.
- O nosso objetivo agora é definir probabilidades a partir de valores possíveis da variável aleatória, sem referência explícita aos pontos do espaço amostral que deram origem aqueles valores da variável aleatória.
- Como definir Pr (X $\in$ A)?
- A maneira mais natural de fazer isso é associar a probabilidade do evento X $\in$ A à probabilidade do evento S no espaço amostral *S*.

Variável Aleatória Discreta

- Nota: freqüentemente iremos abreviar "variável aleatória" por v.a.
- Variável aleatória discreta – pode assumir apenas valores num conjunto finito ou contável, por exemplo, número inteiros ou inteiros positivos.
- Exemplos
  - Número de expectadores em uma sessão de cinema,
  - Resultado do lançamento de um dado,
  - Número de ligações recebidas por uma central de telemarketing num intervalo de tempo especificado,
  - número de assaltos numa esquina.

Função de Probabilidade

- É uma função que associa a cada possível valor de uma variável aleatória discreta a sua probabilidade de ocorrência.
- A função de probabilidade deve satisfazer:

$$\Pr(X = x) = f(x) \geq 0 \quad \text{para todo x}$$

$$\sum_{\text{todo x}} \Pr(X = x) = \sum_{\text{todo x}} f(x) = 1$$

- Também, a probabilidade de qualquer subconjunto A de valores da v.a. é apenas o somatório de f(x) para os valores da v.a. contidos em A.

Variável Aleatória Discreta - Exemplo

- Seja X uma variável aleatória discreta com espaço $\aleph$ = {X: x = 0,1,2,3,4}.
- Seja $f(x) = \Pr(X = x) = \binom{4}{x} . \left(\frac{1}{2}\right)^4 = \frac{4!}{x!(4-x)!}\left(\frac{1}{16}\right) \quad x = 0,1,2,3,4$
- Note que f(x) é uma função de probabilidade, pois:

i) f (x) $\geq$ 0 para todo x $\in \aleph$, isto é, x = 0, 1, 2, 3, 4

Também:

ii) $\sum_{\aleph} f(x) = \sum_{x=0}^{4} \frac{4!}{x!(4-x)!}\left(\frac{1}{2}\right)^4 = \sum_{x=0}^{4} \frac{3}{2} \frac{1}{x!(4-x)!} = \frac{3}{2}\left[\frac{1}{4!} + \frac{1}{3!} + \frac{1}{2!(2)!} + \frac{1}{3!} + \frac{1}{4!}\right]$

$$= \frac{3}{2}\left[\frac{1}{12} + \frac{1}{3} + \frac{1}{4}\right] = \frac{3}{2}\left[\frac{8}{12}\right] = \frac{24}{24} = 1$$

Variável Aleatória Discreta - Exemplo

- Seja A = {0,1}. Então:
- Pr (X ∈ A) = f (0) + f (1) = Pr(X=0) + Pr(X=1)=

$$= \frac{4!}{0!4!}\left(\frac{1}{2}\right)^4 + \frac{4!}{1!3!}\left(\frac{1}{2}\right)^4 = \frac{5}{16}$$

- Veremos depois que este é um caso particular da função de probabilidade Binomial, com parâmetros n = 4 e p = 1/2.

Variável Aleatória Discreta - Exemplo

- **Uma fábrica produz fusíveis. A probabilidade de um fusível produzido ser defeituoso é 10%. Testa-se fusíveis encerrando o teste assim que o primeiro fusível defeituoso é encontrado.**
- **Seja X o número de testes realizados até encontrar o primeiro fusível defeituoso.**
- **Ache a função de probabilidade de X.**

Variável Aleatória Discreta - Exemplo

- **Solução**

  O espaço amostral é constituído por seqüências como:

  D
  BD
  BBD
  BBBD
  BBBBD
  ....

- **Onde B indica que o fusível está perfeito, e D indica que o fusível tem defeito.**

Variável Aleatória Discreta - Exemplo

- **Logo, os valores possíveis de X são: 1, 2, ..., n, ... (não há um valor máximo).**
- **Mas, X = n se, e somente se, os (n-1) primeiros fusíveis testados estão OK e o n-ésimo tem defeito. Isto é, X = n corresponde à seqüência: BBBBB.........BD, que tem n-1 fusíveis OK e 1 com defeito.**
- **Se o estado de um fusível não afetar a condição do próximo podemos supor que:**

  f(n) = Pr(X = n) = $(0.9)^{n-1}$.(0.1) para n = 1, 2, ...

Variável Aleatória Discreta - Exemplo

- Note que f(n) > 0 para todo n e também:

$$\sum_{n=1}^{\infty} f(n) = \sum_{n=1}^{\infty}(0.9)^{n-1}.(0.1) = 0.1\sum_{n=1}^{\infty}(0.9)^{n-1} = 0.1\left\{1+0.9+(0.9)^2+....\right\} = 0.1\left\{\frac{1}{1-0.9}\right\} = 1$$

- Logo, f(n) = Pr(X = n) assim definida é uma função de probabilidade válida.
- Veremos mais tarde que a variável X que surge neste exemplo é chamada de v.a. Geométrica.

Variável Aleatória Discreta - Exemplo

- Nota:
- Neste exemplo empregamos a série geométrica para demonstrar que o somatório das probabilidades para todos os valores de X era um.
- A série geométrica é:

$$\sum_{k=0}^{\infty} a^k = 1 + a + a^2 + a^3 + ... = \frac{1}{1-a} \text{ desde que } |a| < 1$$

- Alternativamente, se começarmos a série em k=1:

$$\sum_{k=1}^{\infty} a^k = a + a^2 + a^3 + ... = \sum_{k=0}^{\infty} a^k - 1 =$$

$$= \frac{1}{1-a} - 1 = \frac{a}{1-a} \text{ desde que } |a| < 1$$

Variável Aleatória Contínua

- Se uma variável puder assumir qualquer valor num intervalo real, é uma variável aleatória contínua.
- Exemplos
  - Tempo de atendimento em um caixa de banco,
  - Peso real de um pacote de 1 Kg de açúcar,
  - Custo de construção de uma fábrica,
  - Custo de lançamento de uma campanha publicitária,
  - Altura dos homens brasileiros com idades entre 18 e 30 anos,
  - Retorno diário de uma ação,
  - Proporção de eleitores a favor da reeleição do prefeito.

Variável Aleatória Contínua

- Como já foi dito, variáveis aleatórias contínuas são aquelas que podem assumir quaisquer valores dentro de um intervalo.
- Para variáveis aleatórias discretas, nós podíamos atribuir uma probabilidade a um determinado valor da variável.
- Para variáveis aleatórias contínuas a situação é bem diferente. Como uma variável contínua pode assumir qualquer valor em um intervalo, na realidade ela pode assumir infinitos valores.

Variável Aleatória Contínua

- Portanto, **não podemos falar da probabilidade de ocorrência de um valor em particular**. Ao invés disso, devemos pensar na probabilidade de ocorrência associada a um intervalo.

- Na discussão anterior sobre distribuições discretas de probabilidades introduzimos o conceito de função de probabilidade (f(x)).

- **No caso contínuo, utilizaremos a função densidade de probabilidade, também representada por f(x).**

Variável Aleatória Contínua

- Nesse caso, a função densidade de probabilidade fornece um valor para cada possível valor (infinitos) da variável X.

- No entanto, os valores de f(x) não representam as probabilidades associadas a x.

- **Ao invés disso, a área (isto é, a integral!) sob a função de densidade de probabilidade em um determinado intervalo fornece a probabilidade de ocorrência de um valor dentro desse intervalo.**

Função Densidade de Probabilidade

- É uma função que satisfaz:

$$f(x) \geq 0 \quad \text{para todo x}$$

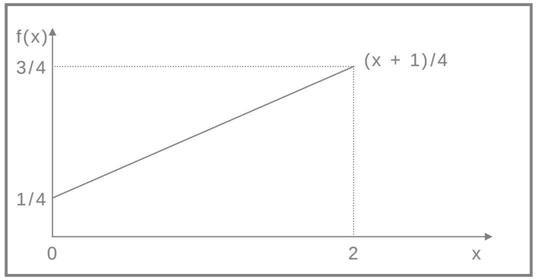
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

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

- Da definição de densidade, segue que, **para uma v.a. contínua, a probabilidade de um único ponto é zero, isto é: P(X = a) = 0 para qualquer número a.**

Distribuições contínuas de probabilidade - exemplo

- Considere a seguinte função de densidade de probabilidade: f(x) = (x + 1)/4 para $0 \leq x \leq 2$.
- Verifique se esta é uma função de densidade de probabilidade válida para o intervalo considerado.
- Calcule a probabilidade de $X \geq 1$



Distribuições contínuas de probabilidade - exemplo

- Solução

  a) Para que f(x) seja uma função de densidade de probabilidade válida, devemos ter a sua área = 1 no domínio da função.

- Neste caso, devemos calcular a área sob a função no intervalo de 0 a 2.
- A área dessa região é dada por:

$$\text{Área} = \frac{f(2)+f(0)}{2}(2-0) = \frac{3/4+1/4}{2}2 = 1$$

- **Logo, f(x) é uma função de densidade de probabilidade válida, pois sua integral é 1 no seu domínio de definição e f(x) é sempre maior ou igual a zero.**

Distribuições contínuas de probabilidade - exemplo

- Solução

  (b) A probabilidade para um determinado intervalo de x é dada pela área sob a função de densidade de probabilidade nesse intervalo.

  P($X \geq 1$) corresponde à área sob a função para $1 \leq x \leq 2$

$$\text{Área} = \frac{f(2)+f(1)}{2}(2-1) = \left(\frac{3/4+2/4}{2}\right)(1) = 0.625$$

Distribuições contínuas de probabilidade - exemplo

- Exemplo
- Seja X uma variável aleatória contínua com espaço $\aleph$ = {x: 0 < x < 1}. Seja f(x) = cx^2 para todo $x \in \aleph$, onde c é uma constante a determinar. Qual o valor de c?
- Solução

$$\int_0^1 cx^2 dx = \left.\frac{cx^3}{3}\right|_0^1 = \frac{c}{3} = 1 \Leftrightarrow c = 3$$

Logo c = 3 é a constante necessária para fazer de f(x) uma densidade em $\aleph$, isto é, para fazer com que a densidade integre a um no intervalo (0,1).

Função de Distribuição

- Para cada valor x_0 da variável aleatória, a Função de Distribuição (ou Função de Distribuição Acumulada, ou Função de Distribuição Cumulativa) é a probabilidade de estar naquele valor, ou abaixo dele, isto é:
- $F(x_0) = Pr(X \leq x_0)$ para todo x_0

> Note que, como $F(x_0)$ é uma probabilidade, ela está limitada ao intervalo (0,1).

- Um ponto importante aqui é: a definição de Função de Distribuição é a mesma para variáveis contínuas ou discretas.

Função de Distribuição

- Algumas funções de distribuição são tabeladas, por exemplo, a da distribuição Normal (0,1).

- O Excel normalmente fornece a opção de calcular a função de probabilidade (ou a densidade) ou a função de distribuição acumulada, através de um argumento lógico nas suas diversas funções estatísticas – por exemplo, vide o “help” da função dist.binom.

Função de Distribuição

- **Propriedades da Função de Distribuição**

  i) $0 \leq F(x) \leq 1$ pois $0 \leq Pr(X \leq x) \leq 1$

  ii) F(x) é uma função não decrescente

  iii) $\lim_{x \to +\infty} F(x) = 1$

  iv) $\lim_{x \to -\infty} F(x) = 0$

Função de Distribuição

- **Propriedades da Função de Distribuição**

  v) Se X é uma variável aleatória contínua, sua função de distribuição é contínua.

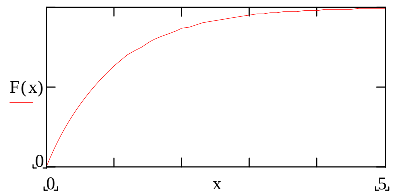
  Se X é discreta, F(x) é uma função contínua à direita, isto é , a função de distribuição apresenta "pulos" (descontinuidades) que só são "sentidos" quando nos aproximamos do ponto onde existe o "pulo" pela esquerda.

Função de Distribuição - Exemplo

- Seja X uma variável aleatória com função de distribuição definida por:

$$\mathbf{F}(\mathbf{x}) = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 - \mathbf{e}^{-x} & \text{se } x > 0 \end{cases}$$

- O gráfico desta função de distribuição é mostrado a seguir.



Função de Distribuição – Exemplo 2

- Considere uma variável discreta com a seguinte função de probabilidade:

$$f(x) = \Pr(X = x) = \frac{4!}{x!(4-x)!}\left(\frac{1}{2}\right)^4 \quad \text{para } x = 0,1,2,3,4$$

- A função de distribuição é:

$$F(x) = \Pr(X \le x) = \sum_{k=0}^{x} \frac{4!}{x!(4-x)!}\left(\frac{1}{2}\right)^4 = \sum_{k=0}^{x}\left(\frac{3}{2}\right)\frac{1}{x!(4-x)!}$$

$$= \left(\frac{3}{2}\right)\sum_{k=0}^{x}\frac{1}{x!(4-x)!} \quad \text{para } x = 0,1,2,3,4$$

Função de Distribuição – Exemplo 2

- Assim:
- F (0) = 1/16 = 0.0625 = Pr (X ≤ 0) = Pr (X = 0)
- F (1) = 5/16 = 0.3125 = Pr (X ≤ 1) = Pr (X = 0) + Pr (X = 1)
- F (2) = 11/16 = 0.6875 = Pr (X ≤ 2) = Pr (X=0) + Pr(X=1) + Pr(X=2)
- F (3) = 15/16 = 0.9375
- F (4) = 1
- Também F(x) = 0 se x < 0 e F(x) = 1 se x > 4

Relação entre a densidade e a função de distribuição

- Seja X uma v.a. contínua com densidade f(x) e função de distribuição acumulada F(x). Então:

$$\Pr(a < X < b) = \int_a^b f(x)dx$$

- Mas:

$$F(a) = \Pr(X \le a) = \int_{-\infty}^{a} f(x)dx \quad \text{e}$$

$$F(b) = \Pr(X \le b) = \int_{-\infty}^{b} f(x)dx$$

Relação entre a densidade e a função de distribuição

- Então:

$$\Pr(a \le X \le b) = \int_a^b f(x)dx = F(b) - F(a)$$

- Pelo teorema fundamental do cálculo:

$$f(x) = \frac{dF(x)}{dx}$$

- Logo, **a densidade é a derivada da função de distribuição**.

Esperança matemática

- **Definição (média ou valor esperado)**
- **A média (ou valor esperado ou primeiro momento) de uma variável aleatória é definida como:**

$$\mu = E(X) = \begin{cases} \int_{-\infty}^{\infty} x.f(x)dx & \text{se X é v.a. contínua} \\ \sum_{\text{todo x}} x.f(x) = \sum_{\text{todo x}} x.\Pr(X = x) & \text{se X é v.a. discreta} \end{cases}$$

- **A média de uma variável aleatória é uma medida de tendência central da distribuição de probabilidade desta variável aleatória.**

Esperança matemática

- **Exemplo**
- **Seja X uma variável contínua com densidade: f(x) = cx^2 para 0 < x < 1**
- **1) Ache a constante c que faz de f(x) uma densidade.**
- **2) Encontre a média desta densidade.**

Esperança Matemática

- **Solução**
- **1) Para que f(x) seja uma densidade:**

$$\int_0^1 f(x)dx = 1 \Leftrightarrow \int_0^1 cx^2 dx = 1 \Leftrightarrow c\frac{x^3}{3}\Big|_0^1 = \frac{c}{3} = 1 \Leftrightarrow c = 3$$

- **2) A média desta densidade é:**

$$\int_0^1 x(3x^2)dx = 3\int_0^1 x^3 dx = 3\frac{x^4}{4}\Big|_0^1 = \frac{3}{4}$$

Esperança matemática

- **Definição (Variância)**
- **A variância de uma variável aleatória mede a dispersão da distribuição de probabilidade, e é definida como:**

$$\sigma^2 = VAR(X) = E\left((X-\mu)^2\right) = \begin{cases} \int_{-\infty}^{\infty} (x-\mu)^2 .f(x)dx & \text{se X contínua} \\ \sum_{\text{todo x}} (x-\mu)^2 .f(x) = \sum_{\text{todo x}} (x-\mu)^2 .\Pr(X = x) & \text{se X discreta} \end{cases}$$

Esperança matemática

- Onde novamente f(x) representa a densidade de probabilidade (se X contínua) ou a função de probabilidade (se X é discreta) e μ é a média da variável aleatória.

- A variância é o segundo momento em torno da média, e corresponde ao momento de inércia em Mecânica.

Da própria definição segue que **a variância é uma quantidade sempre maior ou igual a zero**.

Esperança matemática

- **Definição (desvio padrão)**
- O **desvio padrão** de uma variável aleatória é a **raiz quadrada positiva da sua variância**, e denotado por σ, isto é:

$$\sigma = \sqrt{\sigma^2} = \sqrt{VAR(X)}$$

- O **desvio padrão** é expresso nas **mesmas unidades que a variável aleatória**, e a variância é dada nas unidades da variável aleatória ao quadrado.

Esperança matemática

- Se o **desvio padrão** é **pequeno** existe **pouca dispersão em torno da média**. Se ele é **grande**, os **valores da variável aleatória** estão **muito dispersos** em torno da média.

- A média e a variância são casos particulares dos momentos de uma distribuição de probabilidade.

Esperança matemática

- Os momentos de uma distribuição servem para caracterizar esta distribuição, não apenas no que se refere à sua centralidade e dispersão, mas também com relação a outras características, como a simetria ou assimetria da densidade de probabilidade.

- **A notação E(...) indica o valor esperado (ou “esperança”, ou “expectância”), e pode ser estendida para funções mais gerais que X^k ou $(X - \mu)^k$.**

Esperança matemática

- **Definição (valor esperado de uma função de uma variável aleatória)**
- **Seja X uma variável aleatória com densidade f(x) e seja u(X) uma função qualquer tal que as integrais ou somatórios mostrados a seguir existem.**
- **O valor esperado (ou esperança matemática) de u(X) é:**

$$E(u(X)) = \begin{cases} \int_{-\infty}^{\infty} u(x).f(x)dx & \text{se X é v.a. contínua} \\ \sum_{\text{todo x}} u(x).f(x) = \sum_{\text{todo x}} u(x).\Pr(X = x) & \text{se X é v.a. discreta} \end{cases}$$

Esperança matemática

- **Note que u(X) é também uma v.a.!**
- **A definição anterior inclui, como casos particulares, as definições de média e variância.**
- **O próximo teorema é útil na manipulação de combinações lineares de v.a. (ou suas funções).**

Esperança matemática

- **Teorema (Linearidade do valor esperado)**
- **Sejam <u>a</u> e <u>b</u> constantes e <u>u</u>, <u>v</u> funções quaisquer de X com valores esperados finitos. Então:**

$$\boxed{E[a.u(X) + b.v(X)] = a\,E[u(X)] + b\,E[v(X)]}$$

A expressão acima segue diretamente das propriedades de integrais e somatórios e tem impacto significativo em todos os cálculos envolvendo valores esperados.

Esperança matemática

- **A demonstração deste fato segue diretamente da linearidade das integrais ou somatórios. Em particular, se <u>a</u> é uma constante, E (a) = a.**
- **Nota: fórmula alternativa para o cálculo da variância**
- **O cálculo da variância através da definição é, às vezes, bastante trabalhoso. Por exemplo, no caso de uma v.a. discreta, é necessário computar todas as diferenças $x_i - \mu$, elevá-las ao quadrado e multiplicá-las pela probabilidade de ocorrência de cada x_i.**
- **Logo, seria interessante encontrar uma fórmula alternativa (e mais fácil) para o cálculo da variância, e isso pode ser feito empregando-se a linearidade do valor esperado.**

Esperança matemática

- **Fórmula Alternativa para o Cálculo da Variância**

$$\sigma^2 = VAR(X) = E[(X - \mu)^2] = E[X^2 - 2.\mu.X + \mu^2]$$

- **Pela linearidade do valor esperado e notando que μ é uma constante:**

$$\sigma^2 = E(X^2) - 2\mu.E(X) + E(\mu^2)$$

- **Mas, por definição: $\mu = E(X)$ e μ é uma constante, daí $E(\mu^2) = \mu^2$.**

Esperança matemática

- **Logo:**

$$\sigma^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

$$\boxed{\sigma^2 = E(X^2) - \{E(X)\}^2}$$

- **Esta fórmula é válida para qualquer variável aleatória X (contínua ou discreta), desde que a média de X seja finita.**

Esperança matemática

- **Propriedades do valor esperado e da variância de funções lineares**

- **Sejam a e b constantes, e X uma variável aleatória qualquer. Então:**

1) $E(a.X + b) = a.E(X) + b$
2) $E(a) = a$
3) $VAR(a.X + b) = a^2.VAR(X)$
4) $VAR(a) = 0$

Esperança matemática

- **Exemplo**
- **O retorno mensal de certo investimento de risco pode ser modelado pela variável aleatória R com função de probabilidade dada a seguir:**

r	-5 %	0 %	5 %	10 %	15 %
Pr(R = r)	0.40	0.15	0.25	0.15	0.05

- **Calcule o retorno esperado (em %) do investimento e sua variância e desvio padrão.**

- **Solução**
- **A variável R é discreta, e sua média é (pela definição):**

Esperança matemática

- μ = (-5)(0.40) + (0).(0.15) + (5).(0.25) + (10).(0.15) + (15).(0.05) = 1.5
- A variância de R é:
- $\sigma^2 = (-5-1.5)^2.(0.40) + (0-1.5)^2.(0.15) + (5- 1.5)^2.(0.25) + (10- 1.5)^2.(0.15) + (15 - 1.5)^2.(0.05) = 40.25$
- O desvio padrão de R é:
- $\sigma = \sqrt{\sigma^2} =$ 6.344 (em porcentagem, que é a unidade em que estão expressos os retornos)

Esperança matemática – exemplo (para casa)

Seja X uma variável aleatória contínua com densidade f(x) = c.x onde 0 < x < 3.

- Ache a constante c que faz de f(x) uma densidade.
- Encontre a função de distribuição de X.
- Ache a média, a variância e o desvio padrão de X.
- Encontre um ponto m no intervalo (0,3) tal que $Pr(X \geq m) = Pr(X \leq m) = 50\%$. Este ponto é a mediana da distribuição.

Esperança matemática – exemplo (para casa)

- A renda de uma pessoa numa população é uma variável aleatória contínua X com densidade $f(x) = k/x^3$ onde x >1.

  a) Ache a constante k que faz desta expressão uma densidade.

  b) Encontre a renda média nesta população.

  c) Encontre a renda mediana nesta população, onde m, a mediana, é tal que $Pr(X > m) = Pr(X \leq m) = 0.50$.

Esperança matemática – exemplo (para casa)

- O salário (em milhares de reais) dos funcionários numa empresa pode ser modelado por uma variável contínua X com a seguinte densidade:

  $f(x) = c/x^2$ se $2 \leq x \leq 8$
- Ache a constante c que faz de f(x) uma densidade.
- Encontre a função de distribuição de X para qualquer número real x.
- Ache o ponto m entre 2 e 8 tal que $Pr(X \leq m) = 0.50$. Este ponto é a mediana de X, ou seja, o salário mediano dos funcionários desta empresa.