Inferência Estatística

Aula 1

Agosto de 2008

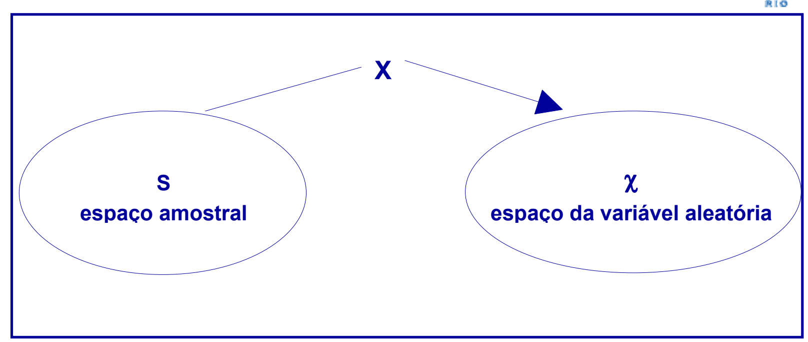
Mônica Barros

Conteúdo – Revisão de Probabilidade

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:

$$\boxed{\begin{array}{l} \Pr(X=x)=f(x)\geq 0 \quad \text{para todo x} \\ \sum_{\text{todo x}}\Pr(X=x)=\sum_{\text{todo x}} f(x)=1 \end{array}}$$

- 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 $\in$ 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:

$$\sum_{n=1}^{\infty} f(n) = \sum_{n=1}^{\infty}(0.9)^{n-1}.(0.1) = 0.1\sum_{n=1}^{\infty}(0.9)^{n-1} = 0.1\sum_{k=0}^{\infty}(0.9)^{k} =$$

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

- 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 Geométrica

- O exemplo anterior é típico da experiência que cria uma variável Geométrica.
- Suponha agora, para generalizar, que p (entre 0 e 1) é a probabilidade de um item defeituoso.
- A experiência é repetida um número *indefinido* de vezes (X), até que o 1o. Sucesso (1a. Peça defeituosa) seja observada.
- Então a função de probabilidade de X é:

Variável Aleatória Geométrica

- A função de probabilidade de X é f(n) = Pr(X = n) dada por:

$$f(n) = \Pr(X = n) = (1-p)^{n-1} p \quad \text{onde } n = 1, 2, 3, ...$$

- X mede o número de tentativas (repetições) necessárias até obter o 1o. “sucesso”.

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

- Exemplo
- Seja X uma variável aleatória contínua com espaço ℵ = {x: 0 < x < 1}. Seja f(x) = cx^2 para todo x ∈ ℵ, 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 ℵ, isto é, para fazer com que a densidade integre a um no intervalo (0,1).

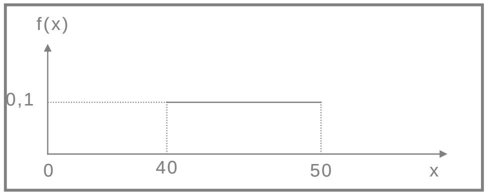
Distribuição Uniforme

- A probabilidade de ocorrência em dois intervalos quaisquer de mesmo tamanho é a mesma – a função de densidade de probabilidade é uma reta paralela ao eixo horizontal.

- Se considerarmos os limites de ocorrência de x como sendo a e b (a < b) devemos ter necessariamente f(x) = 1/(b – a) para que a integral da densidade seja 1.

Distribuição Uniforme

- Exemplo
- Um vôo da ponte aérea RJ-SP leva entre 40 e 50 minutos, com igual probabilidade de ocorrência dentro desse intervalo
  - A distribuição é Uniforme no intervalo (40, 50)
  - f(x) = 1/(50 – 40) para x no intervalo (40,50) e zero fora desse intervalo



Distribuição Uniforme

- Qual a probabilidade de um vôo durar mais de 48 minutos?

$$\Pr(X > 48) = \int_{48}^{50} \frac{1}{50-40}\,dx = \frac{2}{10}$$

- Qual a probabilidade de um vôo durar entre 43 e 45 minutos?

$$\Pr(43 < X < 45) = \int_{43}^{45} \frac{1}{50-40}\,dx = \frac{2}{10}$$

- Uma característica importante da densidade Uniforme é: dois subintervalos de comprimento l que estão totalmente "dentro" de (a, b) têm a mesma probabilidade. Isso não ocorre em geral, no caso de outras densidades.

Distribuição Uniforme

- Exemplo (para casa)
- O retorno de uma aplicação financeira de risco num intervalo de uma semana é uma variável com distribuição Uniforme no intervalo –2% a 1.8%. Calcule:

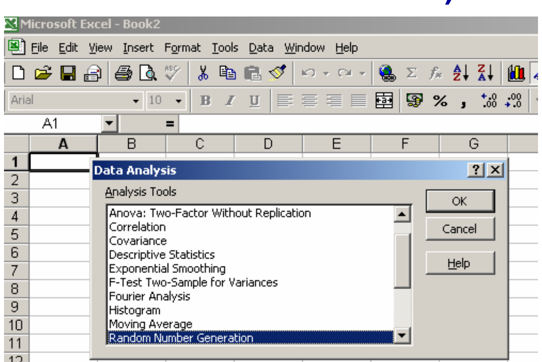
  A probabilidade do retorno do investimento nesta semana ser positivo.

  A probabilidade do retorno estar entre –1% e +1%.

  A probabilidade do retorno exceder 0.5%.

Distribuição Uniforme

- Geração de v.a. Uniformes no Excel
- (É necessária a instalação prévia do suplemento "Análise de Dados")

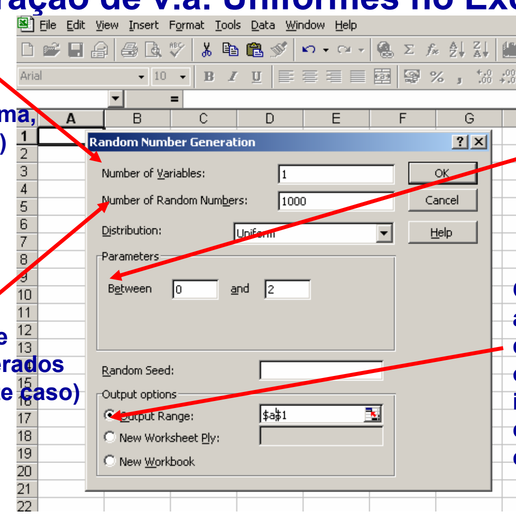


Distribuição Uniforme

- Geração de v.a. Uniformes no Excel

Número de variáveis geradas (uma, neste caso)

Número de valores gerados (1000 neste caso)



Intervalo de definição, neste caso, densidade Unif(0,2)

Célula inicial de armazenamento dos dados – neste caso os números gerados irão preencher a coluna A, a partir da célula A1

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.

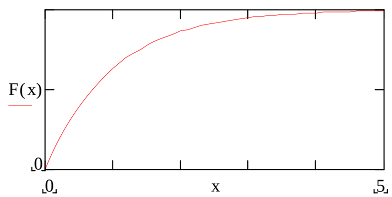
  Se X é discreta, F(x) é uma função contínua à direita, isto é , a função de distribuição apresenta "pulos" (descontinuidades) que só são "sentidos" quando nos aproximamos do ponto onde existe o "pulo" pela esquerda.

Função de Distribuição - Exemplo

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

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

- 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:

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

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

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.**

Distribuição Uniforme

- **Exemplo (para casa)**
- **Se X tem densidade Unif(a,b), mostre que sua função de distribuição acumulada é:**

$$F(x) = \begin{cases} 0 \text{ se } x < a \\ \dfrac{x-a}{b-a} \text{ se } x \in [a,b] \\ 1 \text{ se } x > b \end{cases}$$

Densidade Exponencial

- **Serve para:**
  - **Modelar tempos de duração de equipamentos;**
  - **Modelar tempos entre chegadas de carros num pedágio, entre a chegada de pessoas num caixa de banco, ou seja, tempo entre ocorrências.**
- **Outros exemplos de variáveis Exponenciais**
  - **Tempo entre "beeps" num contador Geiger**
  - **Distância entre mutações num pedaço de DNA**
  - **Tempo que um caixa demora para atender um cliente num banco, loja ou call center**

Densidade Exponencial

- **Densidade**

$$f(x) = \lambda.\exp(-\lambda.x) \quad \text{onde } \lambda > 0 \quad \text{e} \quad x \ge 0$$

- **Função de Distribuição**

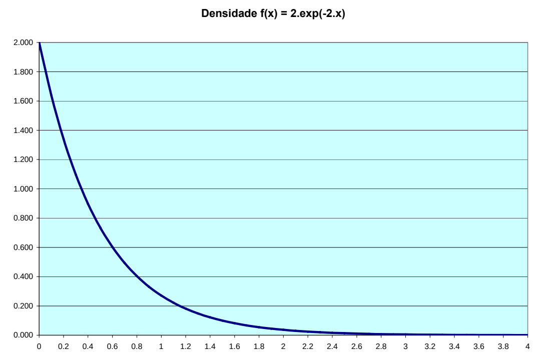
$$F(x) = \Pr(X \le x) = \int_0^x \lambda.\exp(-\lambda.u)du = -e^{-\lambda.u}\Big|_0^x = 1 - e^{-\lambda.x}$$

- **Daí segue que Pr(X > x) = 1-F(x) = exp(-λ.x)**

Densidade Exponencial

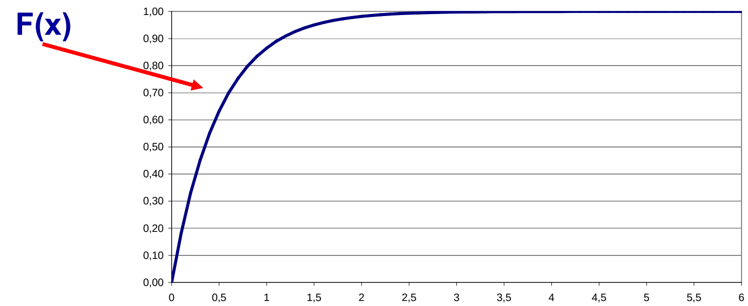
- Gráfico – densidade Exponencial com $\lambda = 2$

$$f(x) = 2.\exp(-2.x) \quad \text{onde } x \geq 0$$



Densidade Exponencial

- O gráfico a seguir apresenta a **função de distribuição** de uma v.a. Exponencial com parâmetro $\lambda = 2$, isto é, a função de distribuição associada à densidade da página anterior.



Densidade Exponencial

- Exemplo
- O tempo entre as chegadas de táxi num cruzamento é uma variável Exponencial com λ = 1/10 chegadas por minutos. Calcule:

  a) A probabilidade de alguém ter que esperar mais de 60 minutos por um táxi.

  b) A probabilidade de um táxi demorar menos de 10 minutos para passar.

Densidade Exponencial

- Solução

  Seja T o tempo entre chegadas de um táxi, isto é, o tempo que você terá que esperar por um táxi nesta esquina.

  T é uma variável Exponencial com λ = 1/10.
  Para uma variável Exponencial, a função de distribuição é $F(t) = \Pr(T \leq t) = 1 - \exp(-\lambda.t)$ e também $\Pr(T > t) = 1 - F(t) = \exp(-\lambda.t)$. Logo:

a) $\Pr(T > 60) = \exp(-60/10) = \exp(-6) = 0.0025$

b) $\Pr(T < 10) = 1 - \exp(-10/10) = 1 - \exp(-1) = 0.6321$

Densidade Exponencial

- Exemplo (para casa)
- O tempo até a ocorrência de um defeito (isto é, o tempo de duração) numa TV é uma variável Exponencial com parâmetro λ = 1/3 anos.
- Calcule a probabilidade de uma TV "pifar" nos primeiros 2 anos de uso.
- Calcule a probabilidade de uma TV durar mais de 5 anos.
- Calcule a probabilidade de uma TV durar entre 3 e 5 anos.

Variável Exponencial - simulação

- O próximo exemplo apresenta a geração de 10000 variáveis Exponenciais com parâmetro 1 a partir de uma amostra do mesmo tamanho da Uniforme(0,1).

- Neste exemplo usamos o suplemento "Análise de dados" do Excel, que permite a geração de v.a. e a construção dos histogramas indicados.

Variável Exponencial - simulação

- Exemplo - Simulação
- A maioria das linguagens de programação tem um gerador de variáveis Uniforme (0,1) "embutido".

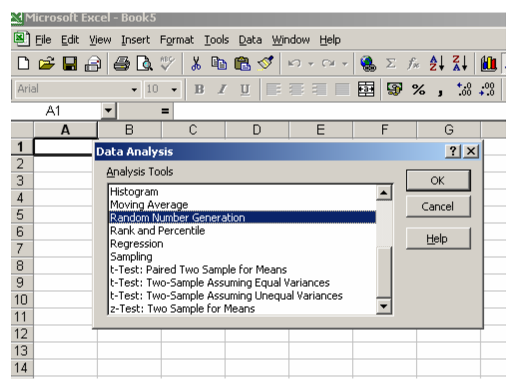
- Mas, é conveniente ser capaz de gerar variáveis com outras densidades.
- Pode-se mostrar (e faremos isso eventualmente) que, se U ~ Unif(0,1) então:

$$Y = \frac{-1}{\lambda}.\log(U)$$

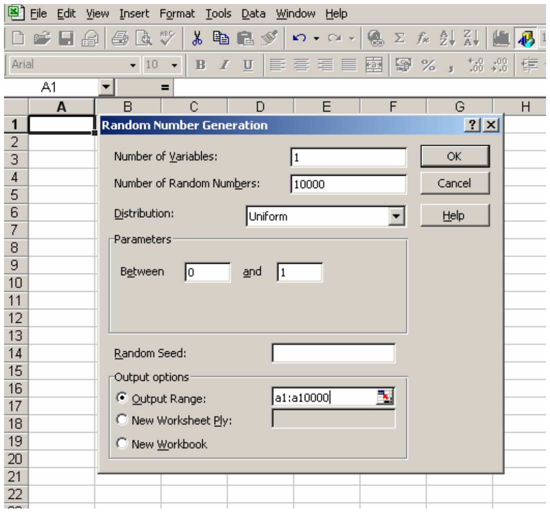
- tem densidade Exponencial com parâmetro λ.

Variável Exponencial - simulação

- Suponha que geramos uma amostra aleatória de 10000 observações da densidade Unif(0,1) no Excel, como mostrado nas próximas figuras.

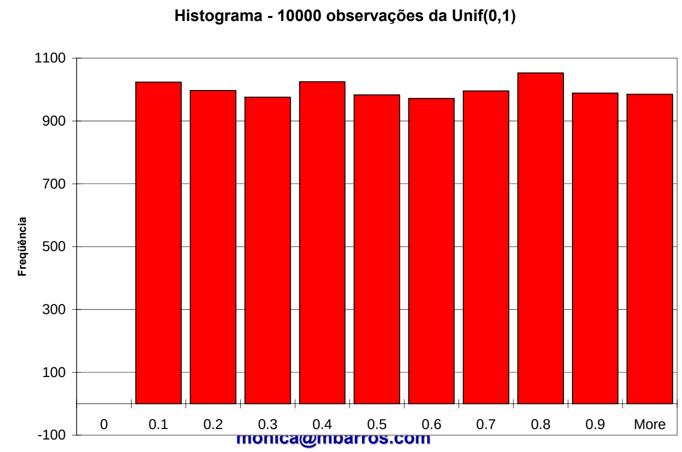


Variável Exponencial - simulação



Variável Exponencial - simulação

- O histograma das 10000 observações geradas é:

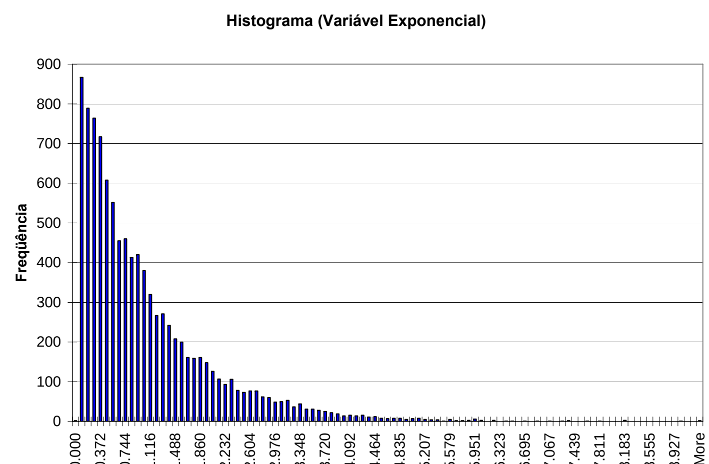


Variável Exponencial - simulação

- **Agora criamos uma nova coluna de 10000 observações usando a transformação Y = - log (U) onde U é um valor gerado da distribuição Unif(0,1).**

- **O histograma da nova amostra deve ter um comportamento decrescente, que se "pareça" com uma densidade Exponencial com média 1. Este histograma é mostrado na próxima figura.**

Variável Exponencial - simulação



Variável Exponencial - simulação

- **Nota:**
- **O Excel não tem um gerador de variáveis Exponenciais. O procedimento que você deve usar para simulá-las é apenas uma extensão do método mostrado neste exemplo.**
- **Para gerar uma variável X com densidade:**

$$f(x) = \lambda.\exp(-\lambda.x) \quad \text{onde } \lambda > 0 \quad \text{e} \quad x \geq 0$$

- **Faça X = (-1/λ). Log(U) onde U é uma v.a. Uniforme(0,1).**