Estatística para Metrologia

Aula 6

Maio de 2008

Mônica Barros, D.Sc.

Aula 6

- Transformações de v.a. discretas
- Transformações de v.a. contínuas
- O método da função de distribuição
- O método do Jacobiano

Objetivos

- Seja X uma v.a. discreta ou contínua com função de probabilidade (ou densidade) conhecida. Queremos encontrar a densidade (ou função de probabilidade de Y=h(X) onde h(.) é uma uma função conhecida.

- Transformações de uma variável aleatória
  - Funções de uma variável discreta
  - Funções de uma variável contínua – o método da função de distribuição

Transformações de uma v.a. discreta

- Exemplo 1
- Seja X o número de "caras" em três jogadas de uma moeda. A função de probabilidade de X é:

x	Pr(X =x) = f(x)
0	1/8
1	3/8
2	3/8
3	1/8

- Qual a função de probabilidade de Y= 2X-1?
- Y é também uma v.a. discreta, e cada valor de X leva a um valor de Y diferente (ou seja, Y = 2X –1 é uma *função injetora*). Especificamente, os valores possíveis de Y são:

Transformações de uma v.a. discreta

- Exemplo 1 (continuação)

y	Pr(Y =y) = f(y)
-1	1/8
1	3/8
3	3/8
5	1/8

- Note que o valor Y = -1 ocorre apenas quando X = 0, Y = 1 apenas quando X = 1 e assim sucessivamente.
- Logo, a tabela anterior nos fornece a função de probabilidade de Y, basta associar cada valor de Y ao valor correspondente(s) de X.

Transformações de uma v.a. discreta

- Exemplo 2
- Seja X uma v.a. discreta com função de probabilidade:
- f(x) = Pr(X = x) = $(1/2)^x$ onde x = 1, 2, 3, ...

- Seja Y = +1 se X é par, e Y = -1 se X é ímpar.
- Obviamente a função h(.), que relaciona X e Y ***não é injetora*** pois, por exemplo, todos os números pares são levados em Y = 1.
- Ache a função de probabilidade de Y.

Transformações de uma v.a. discreta

- Exemplo 2
- Dica: Para resolver o problema precisamos usar a série geométrica infinita:

$$\sum_{k=1}^{\infty} a^k = \frac{a}{1-a} \quad \text{se } |a| < 1$$

- g(1) = Pr (Y = 1) = Pr (X par) = Pr(X = 2, 4, 6, ...) = $(1/2)^2$ + $(1/2)^4$ + $(1/2)^6$ + ... =

$$= \sum_{k=1}^{\infty}\left(\frac{1}{2}\right)^{2k} = \sum_{k=1}^{\infty}\left(\frac{1}{4}\right)^{k} = \sum_{k=0}^{\infty}\left(\frac{1}{4}\right)^{k} - \left(\frac{1}{4}\right)^{0} = \sum_{k=0}^{\infty}\left(\frac{1}{4}\right)^{k} - 1 =$$

$$= \frac{1}{1-1/4} - 1 = \frac{4}{3} - 1 = \frac{1}{3}$$

- Note que Pr(Y = -1) = Pr(X ímpar) = 1 - Pr(Y = 1) = 2/3

Transformação de uma variável aleatória contínua

- Objetivos
- Seja X uma v.a. contínua, e h(.) uma função conhecida. Então Y = h(X) é também uma v.a. e desejamos encontrar sua densidade.

- Dois métodos serão apresentados:
  - O método da função de distribuição
  - O método do jacobiano

- Cada método tem (obviamente) suas vantagens e limitações

Método da Função de Distribuição

- Sejam X e Y = h(X) variáveis aleatórias contínuas. A densidade de Y pode ser encontrada através do seguinte procedimento:
- 1) Encontre o conjunto de todos os valores possíveis de Y.
- 2) Calcule a função de distribuição de Y, ou seja, para cada valor y da variável aleatória Y compute G(y) = Pr (Y ≤ y) escrevendo-a em termos do evento equivalente para X.
- 3) Calcule a derivada de G(y) com relação a y. Isto fornece a densidade de Y, g(y).

Método da Função de Distribuição

- Estes 3 passos são usualmente conhecidos como o "método da função de distribuição".
- Note que o método é bastante geral, e nenhuma condição é imposta à função h(.) que relaciona as variáveis X e Y. Por exemplo, não é necessário que esta função seja injetora.

Método da Função de Distribuição - exemplo

- Seja X uma v.a. Uniforme(0,1). Seja Y = - log (X) onde log indica o logaritmo na base e.
- Encontre a função de distribuição e a densidade de Y.
- Solução
- Os valores possíveis de Y estão no intervalo [0, + ∞), pois quando X tende a zero, log(X) tende a -∞, e Y tende a +∞. Também, quando X tende a 1, log(X) = 0.
- A função de distribuição de Y é: G(y) = Pr (Y ≤ y)

  = Pr (-log X ≤ y) = Pr (log X ≥ -y) = Pr (X ≥ e^{-y}) =

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

Método da Função de Distribuição - exemplo

- A densidade de Y é obtida por diferenciação de G(y) com respeito a y.

$$g(y) = \frac{dG(y)}{dy} = \frac{d(1-e^{-y})}{dy} = -(-1)e^{-y} = e^{-y}, y \geq 0$$

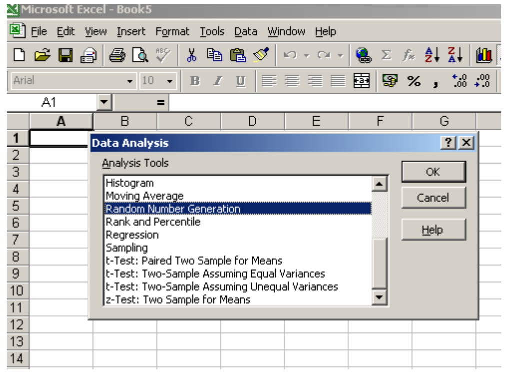
- Note que Y assim gerado é uma v.a. com densidade Exponencial e média 1.
- Este exemplo é uma aplicação importante do método da função de distribuição que pode ser utilizado na geração de variáveis aleatórias com densidade Exponencial, como mostrado a seguir.

Método da Função de Distribuição - exemplo

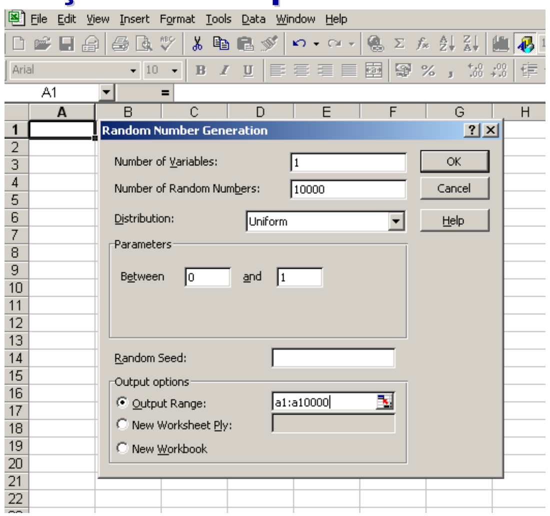
- Logo, se X é Unif(0,1) então Y = - log(X) é Exponencial(1).
- Qual a importância disso? Variáveis exponenciais servem para modelar tempos de duração de equipamentos, ou tempos entre ocorrências (quando o número de ocorrências é Poisson).
- O Excel possui um simulador para diversas distribuições de probabilidade, mas **não** para a Exponencial. Por que? Porque o algoritmo padrão é exatamente este que acabamos de mostrar, ou seja, é muito FÁCIL gerar variáveis exponenciais.

Método da Função de Distribuição - exemplo

- Considere o exemplo anterior e suponha que geramos uma amostra aleatória de 10000 observações da densidade Unif(0,1) no Excel, como mostrado nas próximas figuras.

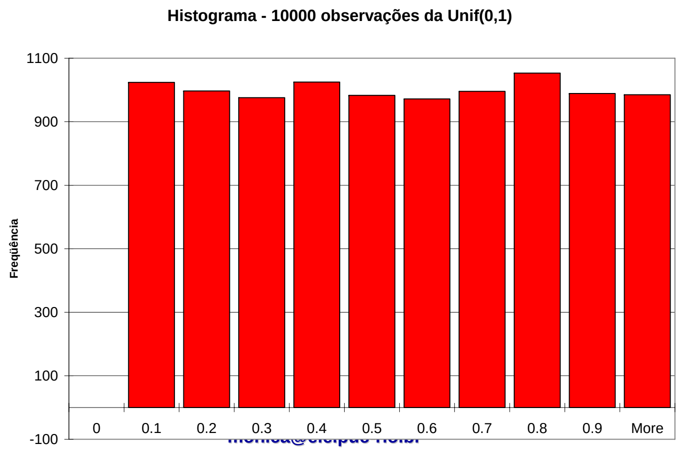


Método da Função de Distribuição - exemplo



Método da Função de Distribuição - exemplo

- O histograma das 10000 observações geradas é:

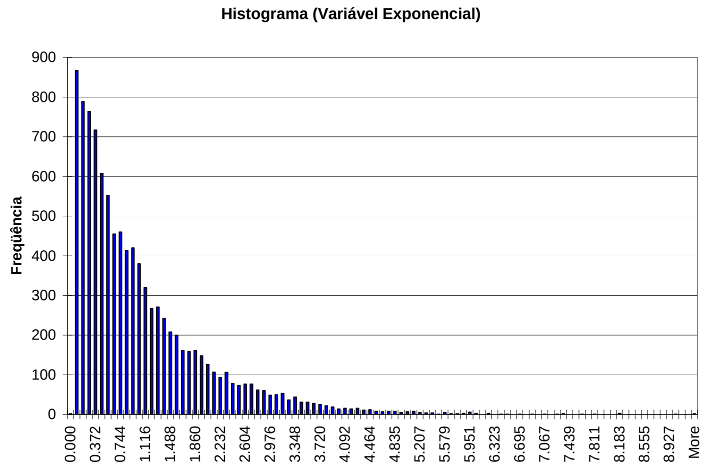


Método da Função de Distribuição - exemplo

- Agora criamos uma nova coluna de 10000 observações usando a transformação Y = - log (X) onde X 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.

Método da Função de Distribuição - exemplo



Método da Função de Distribuição - exemplo

- Suponha que desejamos generalizar este exemplo de maneira a gerar uma v.a. Exponencial com parâmetro λ qualquer.
- Novamente, X é Unif(0,1).
- Note que a função de distribuição de X é:

$$F(x) = \Pr(X \le x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \le x \le 1 \\ 1 & \text{se } x > 1 \end{cases}$$

- Neste caso só nos interessam os valores de x entre 0 e 1.

Método da Função de Distribuição - exemplo

- Seja $Y = \frac{-1}{\lambda}\log(X)$ onde λ é > 0 e log indica o logaritmo natural.

- Qual a função de distribuição de Y? Note que Y ≥ 0 sempre.

$$G(y) = \Pr(Y \le y) = \Pr\left(\frac{-1}{\lambda}\log X \le y\right) = \Pr(\log X \ge -\lambda y) =$$
$$= \Pr\left(X \ge e^{-\lambda y}\right) = 1 - \Pr\left(X \le e^{-\lambda y}\right) = 1 - e^{-\lambda y}$$

- Mas, das propriedades da densidade Exponencial, notamos que Y assim definido é Exponencial com parâmetro λ, isto é:

Método da Função de Distribuição - exemplo

$g(y) = \lambda.\exp(-\lambda y)$ onde $y \geq 0$ e $\lambda > 0$

- Logo, a transformação:

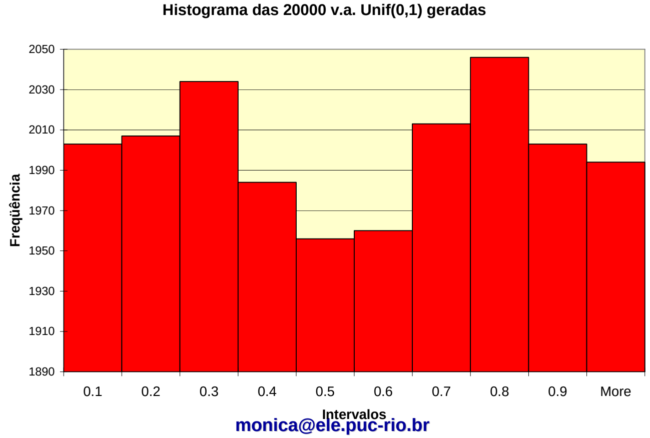
$$Y = \frac{-1}{\lambda}\log(X)$$

- Gera, a partir de uma variável Unif(0,1), uma variável Exponencial com média $1/\lambda$.

- No próximo "slide" exibimos uma aplicação deste resultado.

Método da Função de Distribuição - exemplo

- O histograma de 20000 observações geradas da Unif(0,1) está a seguir.

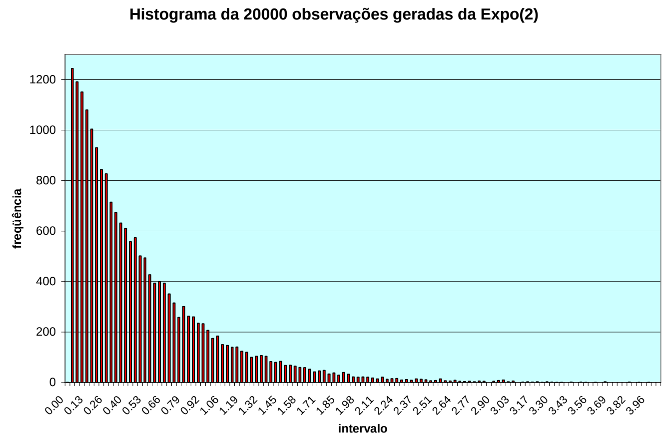


Método da Função de Distribuição - exemplo

- A seguir está o histograma de 20000 observações geradas a partir da transformação $Y = \frac{-1}{2}\log(X)$.

- Estas observações devem ter densidade Exponencial com parâmetro 2.

Método da Função de Distribuição - exemplo



Método da Função de Distribuição - exemplo

- Podemos calcular empiricamente algumas probabilidades e compará-las com os valores reais, obtidos da densidade Exponencial.

- O resultado teórico é: $\Pr(Y \le y) = 1 - \exp(-2y)$
- A mesma probabilidade pode ser estimada através de:

$$\Pr(Y \le y) \cong \frac{\text{número de observações geradas} \le y}{20000}$$

Método da Função de Distribuição - exemplo

y	Pr(Y<=y) empírica	Pr(Y<=y) teórica	y	Pr(Y<=y) empírica	Pr(Y<=y) teórica
0.10	18.08%	18.13%	1.20	90.87%	90.93%
0.20	33.27%	32.97%	1.30	92.53%	92.57%
0.30	45.30%	45.12%	1.40	93.94%	93.92%
0.40	54.95%	55.07%	1.50	95.08%	95.02%
0.50	63.05%	63.21%	1.75	97.12%	96.98%
0.60	69.65%	69.88%	2.00	98.28%	98.17%
0.70	75.36%	75.34%	2.25	98.91%	98.89%
0.80	79.72%	79.81%	2.50	99.32%	99.33%
0.90	83.45%	83.47%	2.75	99.58%	99.59%
1.00	86.50%	86.47%	3.00	99.77%	99.75%
1.10	88.91%	88.92%			

Método da Função de Distribuição

- A Transformação $Y = X^2$
- Seja X uma variável aleatória contínua com densidade f(x) e função de distribuição F(x).
- Seja $Y = X^2$. Então a densidade de Y é:

$$\boxed{g(y) = \frac{1}{2.\sqrt{y}}.\left[f(\sqrt{y}) + f(-\sqrt{y})\right]}$$

- Nota: o único cuidado que dever ser tomado ao usar esta fórmula é fazer as adaptações necessárias quando X (a variável original) for definida apenas na região $x \ge 0$ (ou $x \le 0$), pois neste caso um dos termos $\sqrt{y}$ ou $-\sqrt{y}$ acima será nulo.

Método da Função de Distribuição

- Demonstração
- A função de distribuição de Y é:

$$G(y) = \Pr(Y \le y) = \Pr(X^2 \le y) = \Pr(-\sqrt{y} \le X \le \sqrt{y}) =$$
$$= F(\sqrt{y}) - F(-\sqrt{y})$$

- A densidade de Y é encontrada por diferenciação:

$$g(y) = \frac{dG(y)}{dy} = \left(\frac{y^{-1/2}}{2}\right).f(\sqrt{y}) - \left(-\frac{y^{-1/2}}{2}\right).f(-\sqrt{y}) =$$
$$= \frac{1}{2.\sqrt{y}}.f(\sqrt{y}) + \frac{1}{2.\sqrt{y}}.f(-\sqrt{y}) = \frac{1}{2.\sqrt{y}}.\left[f(\sqrt{y}) + f(-\sqrt{y})\right]$$

Método da Função de Distribuição

- Exemplo 3
- Seja X uma v.a. contínua com densidade:

$$f(x) = \frac{1}{\sqrt{2\pi}}.e^{-x^2/2}, \text{onde x e um numero real}$$

- Veremos depois que esta é a densidade Normal (ou Gaussiana) com média zero e variância 1. Seja $Y = X^2$. Encontre a densidade de Y utilizando o teorema anterior.

Método da Função de Distribuição

- Exemplo 3 (continuação)

$$g(y) = \frac{1}{2.\sqrt{y}}.\left[f(\sqrt{y}) + f(-\sqrt{y})\right]$$

$$g(y) = \frac{1}{2.\sqrt{y}}.\left[\frac{1}{\sqrt{2\pi}}e^{-(\sqrt{y})^2/2} + \frac{1}{\sqrt{2\pi}}e^{-(-\sqrt{y})^2/2}\right] =$$

$$= \frac{1}{2.\sqrt{y}}.\left[\frac{2}{\sqrt{2\pi}}e^{-y/2}\right] = \frac{1}{\sqrt{2\pi}}.y^{-1/2}e^{-y/2} \quad \text{para } y \geq 0$$

- Veremos depois que esta é a densidade Qui-Quadrado com 1 grau de liberdade.

Método da Função de Distribuição

- Vantagens e Desvantagens do Método
- O método da função de distribuição é bastante geral, pois pode ser empregado para transformações não injetoras.
- Mas muitas vezes é difícil escrever a função de distribuição de Y e derivá-la.
- Por isso apresentamos um método adicional para calcular a densidade de uma função de uma variável aleatória, que é chamado de método do Jacobiano. O método do Jacobiano requer que a função Y = h(X) seja injetora.

Método do Jacobiano

- Seja X uma variável aleatória contínua definida num intervalo (a,b), com densidade f(x) e função de distribuição F(x).
- Seja Y = h(X) onde h(.) é uma função contínua e injetora (ou seja, cada x é levado num y diferente).
- Então a densidade de Y, g(y), pode ser encontrada da seguinte maneira:

$$g(y) = f(x).\left|\frac{dx}{dy}\right|$$

Método do Jacobiano

- Por que o módulo |dx/dy| aparece na fórmula anterior?
  - Para garantir que g(y) seja sempre ≥ 0, pois dx/dy pode ser negativo!

Também, x na expressão anterior está escrito em função de y, ou seja, a variável "velha" está em função da variável "nova".

Método do Jacobiano

- Na expressão anterior, $x = h^{-1}(y)$ é expresso em termos da "nova" variável y.
- Se h(.) for uma função crescente (isto é, $x_1 \le x_2$ implica em $h(x_1) \le h(x_2)$) então o intervalo de valores possíveis para Y é (h(a), h(b)).
- Se h(.) é decrescente, o intervalo de definição de Y é (h(b), h(a)).

Método do Jacobiano - exemplo

- Seja X uma variável aleatória contínua com densidade:

$$f(x) = \begin{cases} \frac{m}{\alpha}.x^{m-1}.e^{-x^m/\alpha} & \text{se } x > 0, m > 0, \alpha > 0 \\ 0 & \text{do contrário} \end{cases}$$

- Esta densidade é chamada de densidade *Weibull*, e é muito usada para modelar o tempo de duração de componentes eletrônicos.
- Seja $Y = X^m$. Encontre a densidade de Y.

Método do Jacobiano - exemplo

- Note que $Y = X^m$ é injetora quando x > 0.

$$y = x^m \Leftrightarrow x = y^{1/m} \Rightarrow \frac{dx}{dy} = \frac{1}{m}.y^{\frac{1}{m}-1}$$

- Pelo método do Jacobiano, a densidade de Y é:

$$g(y) = \frac{m}{\alpha}.\left(y^{1/m}\right)^{m-1}.\exp\left\{-\frac{\left(y^{1/m}\right)^m}{\alpha}\right\}.\left[\frac{y^{\frac{1}{m}-1}}{m}\right]$$

- Após simplificações, encontramos:

$$g(y) = \frac{1}{\alpha}.e^{-y/\alpha} \quad \text{para} \quad y > 0$$

- Note que Y tem densidade Exponencial com média α.

Método do Jacobiano – exemplo (para casa)

- **A velocidade de uma molécula de gás é uma variável aleatória contínua V com densidade dada por:**

$f(v) = a.v^2.e^{-bv^2}$, onde b é uma constante que depende do gás e v > 0

- **E <u>a</u> > 0 é uma constante determinada a partir do fato de f(v) integrar a 1 no intervalo (0, + ∞). Seja Z a energia cinética da molécula de gás, dada por:** $Z = \frac{mV^2}{2}$
- **Encontre a densidade de Z (você pode usar o método do Jacobiano ou o da função de distribuição)**

Método do Jacobiano – exemplo (para casa)

- **A duração (Y) de componentes eletrônicos é às vezes modelada pela densidade Rayleigh, mostrada a seguir.**

$$f(y) = \left(\frac{2y}{\theta}\right)\exp\left\{\frac{-y^2}{\theta}\right\} \quad \text{onde } y > 0$$

- **Encontre a densidade de U = Y^2.**
- **Use o resultado acima para achar a média e variância de U.**

Transformações de Variáveis (para casa)

- **Seja X uma v.a. contínua com densidade:**
- $f(x) = \frac{3}{x^4} \quad x > 1$
- **Encontre a densidade de Y = 1/X**

Transformações de Variáveis (para casa)

- **O preço de um ativo financeiro é uma v.a. contínua com densidade:**

$$f(x) = 2xe^{-x^2} \quad \text{onde } x > 0$$

- **Encontre a densidade de Y = X^2.**