INSTITUTO DE EDUCAÇÃO SUPERIOR DE BRASÍLIA - IESB

Pós-Graduação em Inteligencia Artificial

Introdução

O `scippy.stats` fornece ao usuário do Pythom ferramentas para trabalhar com variáveis aleatórias (v.a) discretas ou contínuas. Os métodos mais comuns são:

a) `rvs`: gerar valores aleatórios

b) `pdf`: função de densidade de probabilidade

c) `cdf`: função de probabilidade acumulada $P(X \leq x)$

d) `sf`: função de sobrevivência $P(X > x)$ (1-cdf)

e) `ppf`: Inverso da função `cdf` (dada uma probabilidade α, retorna um valor x, tal que $P(X \leq x) = \alpha$).

f) `isf`: Inverso da função pdf (dada uma probabilidade α, retorna um valor x, tal que $P(X \leq x) = \alpha$).

g) `stats`: retorna as estatísticas: media, variâncias, assimétria (Fisher) e kurtose (Fisher).

h) `moment`: momentos não centrais.

As variáveis contínuas têm os argumentos `loc` e `scale` para ajustar a locação e a escala da distribuição. Em uma distribuição Normal, parâmetro de locação é a media μ e o de escala é o desvio padrão σ.

Texto de base I

É a variável aleatória contínua X com parâmetros μ e σ, e função de densidade de probabilidade

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ para } x \in \mathbb{R}$$

com μ, $\sigma \in \mathbb{R}$, $\sigma > 0$ e denotada $X \sim N(\mu, \sigma^2)$, cujo valor esperado é $E(X) = \mu$ e a variância é $V(X) = \sigma^2$.

Seja $X \sim N(\mu, \sigma)$ uma variável normal comcom μ, $\sigma \in \mathbb{R}$, $\sigma > 0$. No Python, o módulo `from scipy.stats import norm` traz o método **nomr.pdf(x,μ,σ)** para obter valores da função de densidade e **norm.cdf(x,loc=μ,scale=σ)** para calcular $P(X \leq x)$.

O código abaixo calcula para $x = 5$ o seu valor da densidade e da probabilidade $P(X \leq x)$ par $X \sim N(\mu = 5, \sigma = 1)$.

```
import scipy.stats as spy
x = 5
fx = spy.norm.pdf(x,loc=5, scale=2)
print("Valor da densidade:",fx)
F  = spy.norm.cdf(x,loc=5, scale=2)
print("Probabilidade P(X<=x):",F)
```

Exercícios

1 - A função `linspace(a,b,n)` do numpy gera um array `n` valores igualmente espaçados entre `a` e `b`, incluíndo esses valores limites.

   a) Crie o vetor x com 10 valores que iniciem em 1 terminem 2.

   b) Crie o vetor y com 10 valores que iniciem em 2 terminem 1.

2 - A função `cumsum(x)` do numpy gera um array com os valores acumulados.

   a) Seja `x=[1, 2, 3]`, execute o comando `cumsum(x)`.

   b) Crie y com 20 valores de 0 a 1 e gere z com as somas acumuladas de y.

3 - Com a função `linspace` crie o vetor x com 100 valores de -3 a 3. Seja $X \sim N(0, 1)$.

   a) Obtenha o valor da função de densidade de x e armazene o valor em y.

   b) Crie um gráfico de dispersão com a variável `x`, nas abcissas, e `y` nas ordenadas.

   c) Obtenha a probabilidade de $P(X \leq x)$ e armazene o valor em y.

   d) Crie um gráfico de dispersão com a variável `x`, nas abcissas, e `y` nas ordenadas.

4 - Para $X \sim N(\mu = 0, \sigma = 1)$, calcule:

   a) $P(X < 0)$;

   b) $P(X > 1.96)$;

   c) $P(-1 < X < 1)$;

   d) c tal que $P(X \leq c) = 0.5$;

   e) c tal que $P(X \leq c) = 0.025$.

5 - Obtenha os percentis 0.01, 0.025, 0.05 , 0.50, 0.95, 0.975 e 0.99 de $Y \sim N(\mu = 10, \sigma = 2)$.

6 - Obtenha os valores para **a** de uma distribuição $X \sim N(\mu = 15, \sigma = 3)$ tais que:

a) $P(X \leq a) = 0.50$;

b) $P(-a \leq X \leq a) = 0.95$

c) $P(X \geq a) = 0.10$

7 - (Motta 2006) Em uma distribuição de valores de glicose plasmática em jejum em homens normais entre 30 a 39 anos de idade, a média observada foi $\mu = 100$ mg/dL e o desvio padrão $\sigma = 15$ mg/dL.

a) Qual a proporção de homens com glicose plasmática entre 100 e 120 mg/dL?

b) Qual a proporção de homens com glicose plasmática acima de 120 mg/dL?

c) Qual o percentil 90, 95 e 97,5 da distribuição dos valores de glicose plasmática?

d) Faça o gráfico da função de densidade com x variando do percentil 1 ao percentil 99.

8 - (Cybalista 2005) O número de acidentados que chega a um hospital por dia tem distribuição aproximadamente normal com $\mu = 75$ e desvio padrão $\sigma = 8$. Qual a probabilidade de que em certo dia, cheguem

a) mais de 75 acidentados?

b) entre 60 e 80 acidentados?

9 - Gere duas amostras de tamanho 1000, uma de $X \sim N(\mu = 0, \sigma = 1)$ e outra de $Y \sim N(\mu = 3, \sigma = 0.5)$. Seja $Z = X^2 + Y^2$,

a) faça um histograma com as três distribuições;

b) calcule a média e os desvio padrão de Z e compare com as de X e Y;

c) há algum indicativio de Z ter distribuição normal? Se sim, qual?

10 - (SPIEGEL 2004) Se as alturas de 300 estudantes são normalmente distribuídos com média 68 polegadas e desvio padrão 3 polegadas, qual a probabilidade de encontrar estudantes com altura

a) maior que 72 polegadas;

b) menor ou igual a 64 polegadas;

c) entre 65 e 71 polegadas.

11 - (SPIEGEL 2004) A nota média em um exame final foi 72 e o desvio padrão foi 9. Os 10% melhores entre entre os estudantes recebem A. Qual a nota mínima que um estudante deve obter para receber um A?

12 - Uma forma de simular os dados $X \sim N(0, 1)$ é:

a) tire um valor aleatório x de uma distrbuição aleatória uniforme entre 0 e 1;

b) obtenha o percentil x de uma $N(0, 1)$.

Segue um exemplo de código:

```
 from scipy.stats import uniform
from scipy.stats import norm
import numpy as pd
amostra = np.zeros(1000)
for i in np.arange(0,1000):
    x = uniform.rvs(size=1)
    amostra[i]=norm.ppf(x)
plt.hist(amostra,normed=True)

print('media:', amostra.mean(), '\n', 'desvio padrão:', amostra.std(),"\n")
plt.show()
```

Assim, simule de uma distribuição normal de média 5 e variância 2, uma amostra de tamanho 5000. Verifique se a amostra tem pelo menos 3 características da população simulada.

13 - Retire uma amostra aleatória de tamanho 200 de $X \sim N(0, 1)$ e armazene em x. Qual o gráfico resultante de `plt.plot(np.cumsum(x))`.