1 2 9 0 FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE DE COIMBRA

DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES

Processamento Digital de Sinal

Trabalho Prático Nº 7

Variáveis aleatórias, histogramas e normalização de amostras.

Descrição

Neste trabalho pretende-se elaborar os conceitos de variável aleatória (VA), amostra, função densidade de probabilidade e normalização de variáveis aleatórias. Para isso, iremos usar os dois geradores de números pseudoaleatórios do Matlab: `rand` (distribuição uniforme entre 0 e 1) e `randn` (distribuição normal ou Gaussiana, de média zero e variância 1). Os números gerados são *pseudoaleatórios* pois podemos repetir a amostra inicializando o gerador no mesmo estado (`rng(seed)`).

Trabalho Prático

1. Amostra 1D com distribuição triangular.

Gere uma amostra (pseudoaleatória) de uma VA com distribuição triangular com 10^5 realizações usando o seguinte código Matlab (soma de duas VAs iid):

```
rng(0) %para o caso de querermos repetir a experiência.
x1 = rand(1,1e5); %uma amostra com 10^5 realizações
x2 = rand(1,1e5); %outra: independente e identicamente distribuída.
x  = x1+x2;
```

a) Analise empiricamente a distribuição da amostra fazendo um histograma: `histogram(x)`. Verifique que a moda da distribuição (o valor mais frequente) é 1 e que a VA **x** se distribui de 0 a 2. <u>Estime</u> a média e a variância da distribuição, usando as funções `mean` e `var`.

b) A amostra `x` foi gerada como a soma de duas VAs iid (independentes e identicamente distribuídas) uniformes entre 0 e 1, logo a distribuição da soma é a <u>convolução</u> das distribuições originais. Assim, a VA **x** é a convolução da distribuição uniforme com ela própria. Defina a equação desta PDF (*probability distribution function*). Determine a média e variância da distribuição.

2. Amostra 1D com distribuição Gaussiana.

Gere uma amostra com o seguinte código Matlab:

```
rng(0) %para o caso de querermos repetir a experiência.
x = 7*randn(1,100000)+3; %amostra com 100k pontos
```

a) Determine a média e variância da variável $\mathbf{x} = 7\mathbf{z} + 3$, onde $\mathbf{z} \sim \mathcal{N}(0,1)$ (significando que **z** é normal com média zero e variância 1). Estime a média e a variância (empíricas, ou da amostra) a partir da amostra `x`.

b) Verifique empiricamente que função densidade de distribuição da VA **x** é também gaussiana, fazendo um histograma. Para isso, faça um *plot* conjunto da PDF de **x**, com o histograma normalizado, isto é, o histograma com área 1:

```
h=histogram(x,100,'Normalization','pdf'); %histograma com 100 bins, modo PDF
```

Nota: pode ver os valores possíveis das propriedades do objeto `h` com `set(h)`.

Para sobrepor a PDF $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ no gráfico, considere um vetor `xpdf` linearmente espaçado entre -30 e 40, com 201 pontos, e o correspondente vetor `fpdf` da função Gaussiana indicada.

Processamento Digital de Sinal

Trabalho Prático Nº 7

Variáveis aleatórias, histogramas e normalização de amostras.

3. Vetor de duas variáveis aleatórias

Vamos tomar uma amostra de uma distribuição normal em 2D com média $\boldsymbol{\mu}_y = \begin{bmatrix} 27 \\ 7 \end{bmatrix}$ e matriz de covariância $\mathbf{C}_y = \begin{bmatrix} 4 & 5 \\ 5 & 9 \end{bmatrix}$. Para tal vamos proceder de forma idêntica ao caso 1D, fazendo uma transformação a partir de uma amostra $\mathbf{x}$ de média zero e matriz de covariância identidade, isto é, $\boldsymbol{\mu}_x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ e $\mathbf{C}_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Para gerar a amostra `x` basta simplesmente fazer:

```
rng(1);
x=randn(2,100000); %cada coluna é um dos 100000 pontos 2D
```

A transformação a fazer (desnormalizar a variável $\mathbf{x}$) corresponde a : $\mathbf{y} = \mathbf{C}_y^{\frac{1}{2}} \cdot \mathbf{x} + \boldsymbol{\mu}_y$ (se fosse 1D seria $\mathbf{y} = \sigma_y \mathbf{x} + \mu_y$). Use: `S=sqrtm(Cy) %matriz raiz quadrada: S*S=Cy (real pois |Cy|>=0)`

a) Calcule as médias e variâncias (empíricas) da amostra `y` e compare-as com os valores teóricos relativos à distribuição da VA $\mathbf{y}$ (que é $\boldsymbol{\mu}_y$).

b) Calcule o coeficiente de correlação, ρ_y, entre as duas componentes do vetor aleatório $\mathbf{y}$, as VAs y_1 e y_2 onde $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Estime-o através da amostra e compare-o com $\rho_y = \frac{5}{6}$. Nota: compare o resultado com `corrcoef(y')`. (`y'` pois esta função assume uma observação em cada linha da matriz de entrada).

c) Faça um "scatter plot" (plot com pontos, `'.'`) da amostra e coloque no gráfico a elipse de concentração de nível r=4. Para tal defina parametricamente uma circunferência de raio 4 com 101 pontos relativos aos ângulos entre 0 e 2π. Esta circunferência simula o corte, numa uma dada cota, de uma gaussiana 2D com matriz de covariância identidade e média nula. Se transformarmos os pontos desta circunferência da mesma maneira que desnormalizámos os pontos iniciais, obtemos a elipse pretendida.

```
r = 4;
teta=(0:100)/100*2*pi; cx=r*cos(teta);cy=r*sin(teta);
plot(cx,cy), axis equal
```

Transforme os pontos da circunferência com a mesma transformação que fez para a amostra `y`. Faça o "scatter plot" da amostra conjuntamente com a elipse obtida. Deve observar que os pontos da amostra estão concentrados dentro deste "elipsoide de concentração", neste caso elipse de concentração. Se fizer r=5, praticamente todos os pontos da amostra ficam dentro da elipse.

d) Normalize agora a amostra `y` e chame-lhe `z`. Use a média e matriz de covariância estimadas a partir da amostra (com `cov(y')` e `mean(y,2)`) em vez dos valores exatos. Verifique o resultado.

Nota: no caso 1D faria $z = \dfrac{y - \hat{\mu}_y}{\hat{\sigma}_y}$; neste caso será $\mathbf{z} = \hat{\mathbf{C}}_y^{-1/2}\left(\mathbf{y} - \hat{\boldsymbol{\mu}}_y\right)$.

No final da aula entregue o script que produziu, na "Submissão de Trabalhos" do Nónio.