Operaciones lineales con una variable aleatoria: ejemplo Matlab (distr. uniforme)

Este código ejecutó sin errores en Matlab **`R2021b`**

Presentación en vídeo: http://personales.upv.es/asala/YT/V/opvau.html

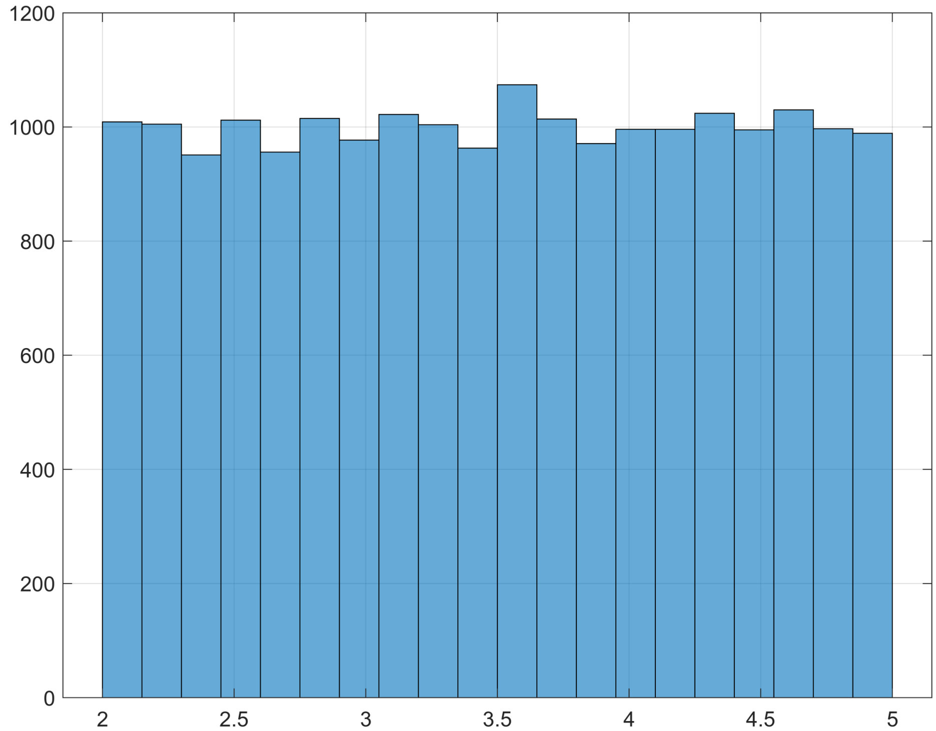
Objetivos: En muchas aplicaciones, una variable aleatoria es una "entrada" a un cierto modelo $y = h(x)$, de modo que se desea obtener información sobre la aleatoriedad de la "salida" y del modelo, si conocemos datos sobre la distribución de probabilidad de la entrada x. Aquí, buscamos comprender el significado de las operaciones con variables aleatorias, y las transformaciones que sufren media y varianza con operaciones lineales en concreto. Usaremos una distribución uniforme, por simplicidad.

Tabla de Contenidos

Variables aleatorias: media, varianza, desv. típica teóricas versus muestrales

Distr. Uniforme

```
N=2e4;
MuestrasX=3*rand(N,1)+2; %uniforme, desplazada
histogram(MuestrasX,20), grid on
```



Cálculos muestrales:

```
MediaXMuestral=mean(MuestrasX)
```

```
MediaXMuestral = 3.5041
```

```
var(MuestrasX) %varianza muestral
```

```
ans = 0.7487
```

```
std(MuestrasX) %desv. típica muestral
```

```
ans = 0.8653
```

Cálculos teóricos (esperanza matemática):

Con función de densidad $f(x)$, con $\int_2^5 f(x)\,dx = 1$, tenemos:

```
syms x real %variable aleatoria simbólica
dens_x=1/3; %es constante en el intervalo [2 5]
int(sym(dens_x),2,5) %esto da 1
```

ans = 1

$$E[x] = \int_2^5 x\,f(x)\,dx$$

```
E_x=int(x*dens_x,2,5) %esperanza matemática (media teórica, valor esperado)
```

E_x =

$\frac{7}{2}$

$$E[(\Delta x)^2] = \int_2^5 (x - E[x])^2 f(x)\, dx$$

```
vza_teo_x=int((x-E_x)^2*dens_x,2,5)  %varianza teórica
```

vza_teo_x =

$\frac{3}{4}$

```
std_teo_x=sqrt(vza_teo_x) %desv. típica teórica
```

std_teo_x =

$\frac{\sqrt{3}}{2}$

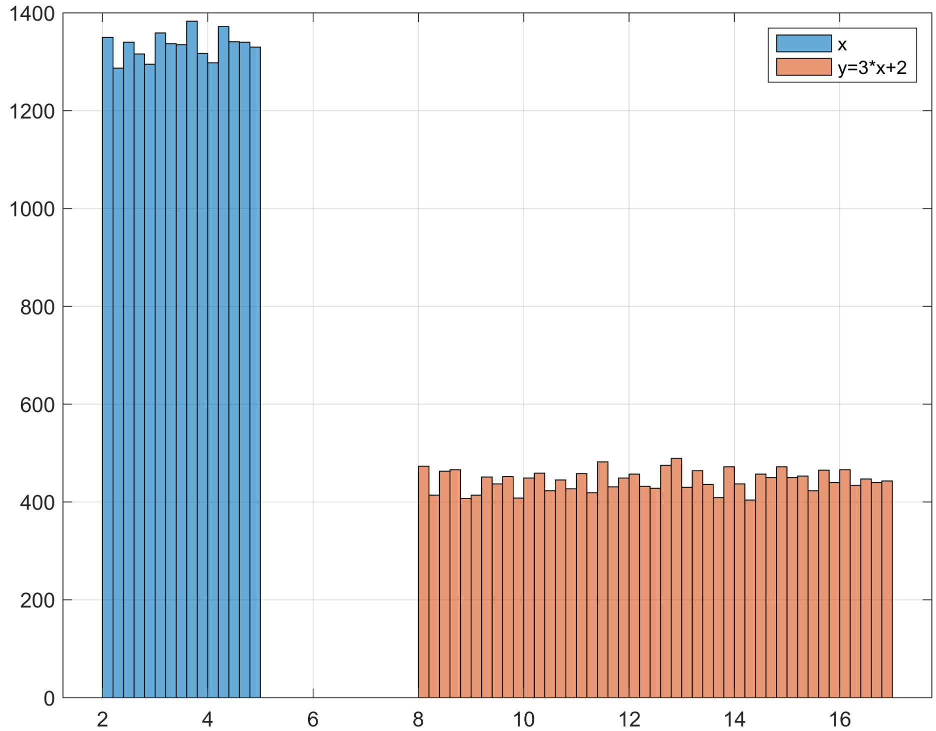
```
eval(std_teo_x)
```

ans = 0.8660

Funciones lineales: escalado/desplazamiento de una variable aleatoria uniforme

Generaremos una variable aleatoria "nueva" $y = h(x) = 3x + 2$.

```
MuestrasY=3*MuestrasX+2;
histogram(MuestrasX,15), hold on,
histogram(MuestrasY,45), hold off, legend('x','y=3*x+2'), grid on
```



Calculos teóricos (esperanza matemática):

Para calcular media y varianza de $y = h(x)$ podemos hacerlo sobre la distribución de "x":

```
h(x)=3*x+2;
```

$$E[h(x)] = \int_2^5 h(x)\, f(x)\, dx$$

```
E2=int(h(x)*dens_x,2,5)
```

E2 =

$$\frac{25}{2}$$

Como la integral es un operador lineal $\int (b+ax)\cdot f(x)\cdot dx = b + a\int x \cdot f(x)\cdot dx$, podemos calcular E2 a partir de la media de x:

```
3*E_x+2 %coincide... en funciones lineales la "media de la función" es la "función de l
```

ans =

$$\frac{25}{2}$$

En cuanto a la varianza, formalmente

```
Var2=int((h(x)-E2)^2*dens_x,2,5)
```

Var2 =

$\frac{27}{4}$

```
sqrt(Var2) %desviación típica
```

ans =

$\frac{3\sqrt{3}}{2}$

```
eval(ans)
```

ans = 2.5981

También, en incrementales (restando la media)

$\int y^2 \cdot f(x) \cdot dx = \int a^2 x^2 \cdot f(x) \cdot dx = a^2 \int x^2 \cdot f(x) \cdot dx = a^2 \sigma_x^2$, la varianza se multiplica por la "ganancia al cuadrado":

```
3^2*vza_teo_x
```

ans =

$\frac{27}{4}$

Obviamente, la desviación típica se multiplica sólo por la "ganancia":

```
3*std_teo_x
```

ans =

$\frac{3\sqrt{3}}{2}$

Alternativamente, tambien podemos hacerlo pensando en la nueva distribución que tendrá y, "desconectándola" conceptualmente de x.

```
dens2=1/9; %función de densidad uniforme, entre 8 y 17.
int(sym(dens2),2*3+2,5*3+2) %esto da 1
```

ans = 1

```
syms y real %ahora "y" es una variable aleatoria "por sí misma"...
E_VbleProp=int(y*dens2,2*3+2,5*3+2); %esperanza matemática (media esperada)
eval(E_VbleProp)
```

ans = 12.5000

```
vza_again=int((y-E_VbleProp)^2*dens2,2*3+2,5*3+2)
```

```
vza_again =
```

$$\frac{27}{4}$$

```
eval(sqrt(vza_again))
```

```
ans = 2.5981
```

Cálculos "muestrales":

Media:

```
mean(MuestrasY)
```

```
ans = 12.5124
```

```
3*MediaXMuestral+2 %la multiplicación por constante y suma replica la media
```

```
ans = 12.5124
```

Varianza y desviación típica:

```
std(MuestrasY)
```

```
ans = 2.5959
```

La multiplicación por constante también multiplica la desv. típica, el desplazamiento no cambia la desv. típica (se mide alrededor de la media, que también se mueve).

```
3*std(MuestrasX)
```

```
ans = 2.5959
```

```
var(MuestrasY)
```

```
ans = 6.7387
```

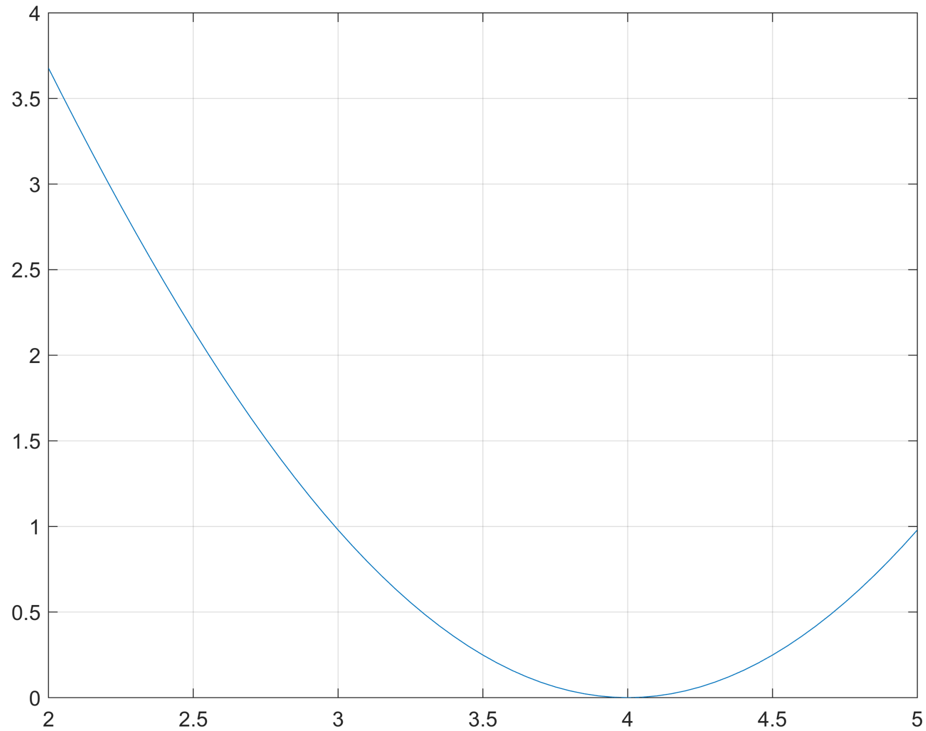
```
3^2*var(MuestrasX) %la varianza tiene unidades del "cuadrado" de la variable...
```

```
ans = 6.7387
```

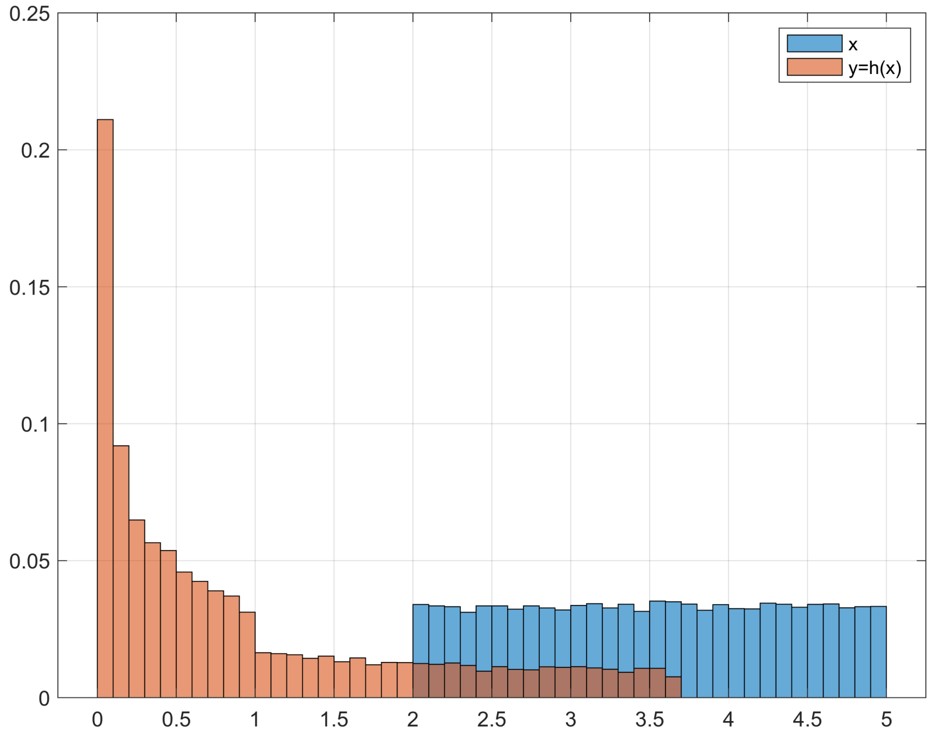
¿Operaciones no lineales?

Si cogemos, p.ej. la uniforme y la pasamos por una no-linealidad:

```
MuestrasY=(4*sin(0.25*MuestrasX-1)).^2;
rangoX=2:0.05:5;
plot(rangoX,(4*sin(0.25*rangoX-1)).^2), grid on
```




```
h1=histogram(MuestrasX,20); hold on,
h2=histogram( MuestrasY,20); hold off, legend('x','y=h(x)'), grid on
h1.Normalization = 'probability';
h1.BinWidth = 0.1;
h2.Normalization = 'probability';
h2.BinWidth = 0.1;
```



El resultado ya no es uniforme: las funciones no lineales deforman la distribución (la función de densidad), a otra cosa más complicada de calcular.

Cálculos teóricos (esperanza matemática)

```
h=(4*sin(0.25*x-1))^2;
Mean_y=int(h*dens_x,2,5);eval(Mean_y)
```

```
ans = 0.9552
```

```
eval(int((h-Mean_y)^2*dens_x,2,5))
```

```
ans = 1.0459
```

Cálculos muestrales

```
mean(MuestrasY)
```

```
ans = 0.9500
```

```
var(MuestrasY)
```

```
ans = 1.0467
```

NOTA 1: Ya no se puede dar, en general, una función de la media y la varianza de la variable transformada a partir de la media o varianza original.

```
WrongMean=(4*sin(0.25*mean(MuestrasX)-1))^2 %esto está FATAL, no lo hagas nunca en No I
```

```
WrongMean = 0.2446
```

NOTA 2: En teoría se puede calcular la nueva función de densidad de y, para analizarla "desconectada de x". Necesitaremos la inversa (si existe) y la derivada...

$$y = h(x), \qquad dy = \dot{h}(x)dx, \qquad dx = \frac{1}{\dot{h}(x)}dy = \frac{1}{\dot{h}(h^{-1}(y))}dy$$

$$E[y] = \int_a^b h(x)f(x)dx = \int_a^b yf(h^{-1}(y))dx = \int_{h(a)}^{h(b)} y \cdot \underbrace{\frac{f(h^{-1}(y))}{\dot{h}(h^{-1}(y))}}_{\text{nueva dens.}} dy$$

En el uniforme $f(h^{-1}(y)) = 1/3$, $\dot{h}(h^{-1}(y)) = 3$ con lo que la nueva densidad es 1/9... Muy complicado en un caso general (y más si f no fuera invertible, como es el caso, donde hay "saltos" en la función de densidad --histograma--).

Conclusión

En una transformación $y = a \cdot x + b$, siendo x una variable aleatoria, a y b constantes (ganancia y offset, respectivamente), entonces y es una nueva variable aleatoria cuya media (ideal $E[y]$ o muestral $\bar{y}$) es $a \cdot E[x] + b$ o bien $a\bar{x} + b$, y cuya desviación típica (ideal o muestral) es $\sigma_y = a \cdot \sigma_x$.

La varianza es, obviamente el cuadrado: $var(y) = a^2 \cdot var(x)$.

En una transformación $y = f(x)$ siendo f no lineal, entonces y es una variable aleatoria, obviamente, pero su media, varianza, función de densidad, ya son más complicadas de calcular.

Usualmente, lo primero que se estudia en estadística son las "transformaciones lineales de variables aleatorias", por razones obvias de simplicidad.