

## Ejercicios Sobre Estimación Puntual y por Intervalos

1. a) Genere una matriz de  $n=100$  filas y 30 columnas de datos con distribución Poisson de parámetro  $\lambda = 3$ . Halle la media de cada fila (Para la distribución Poisson el estimador por momentos y el de máxima verosimilitud son el mismo, el promedio, o sea  $\hat{\lambda} = \bar{X}_n$ ).

Una forma de hacerlo es la siguiente:

- (a) `lambda=3;`
  - (b) `n=100;`
  - (c) `col=30;`
  - (d) `X=poissrnd(lambda,n,col);` esto genera una matriz de  $n$  filas y 30 columnas, cada fila tiene 30 realizaciones de una variable aleatoria con distribución Poisson de parámetro  $\lambda = 3$ . Cada fila podemos verla como una muestra i.i.d.  $\{x_1, x_2, \dots, x_{30}\}$ .
  - (e) La otra forma es `X=poissrnd(3,100,30);` el punto y coma que va al final de cada línea es para que lo que uno ha cargado no se muestre en la ventana de comandos. Esto no importa si se trata de mostrar un valor, pero si creamos una matriz  $100 \times 30$  y no terminamos la sentencia con el punto y coma, al hacer enter toda la ventana de comandos se llena con la matriz.
  - (f) `Xmean=mean(X,2);` esta sentencia calcula el promedio de cada fila, el resultado es una matriz  $100 \times 1$ , o sea una columna con 100 filas.
- b) Construya el histograma de dichas medias (el histograma aparecerá con 10 intervalos).  
Una forma de hacerlo es la siguiente:
    - (a) `figure()` (esta sentencia crea una ventana para gráficos).
    - (b) `hold on` (esta sentencia mantiene la ventana).
    - (c) `subplot (2,2,1);` con esto indicamos que la ventana creada va a tener 4 gráficos, organizados en 2 filas y 2 columnas, el número 1 indica que el gráfico que se cree a continuación irá en el primer lugar, el superior izquierdo.
    - (d) `hist(Xmean,10,'facecolor','r','edgecolor','b')` (hacemos un histograma con los 100 promedios, con las barras en rojo y sus bordes en azul).

- (e) `title('n=100')` (título del histograma)
- c) Repita los pasos anteriores para  $n=1000, 5000, 10000$ . Puede ser conveniente para estos valores de  $n$  probar cambiando la cantidad de intervalos del histograma (recordar la sugerencia de usar  $\sqrt{n}$ ) hasta obtener un resultado ni demasiado grueso ni demasiado fino. Una forma de hacerlo es la siguiente (los histogramas aparecerán en la misma ventana usada para  $n=100$ ):
- (a)  $n=1000$ ; (cambiamos la cantidad de filas de 100 a 1000).
- (b) `subplot (2,2,2)`; (el gráfico que creamos a continuación irá arriba a la derecha).
- (c) `hist(mean(poisrnd(lambda,n,col),2),32,'facecolor','r','edgecolor','b')` (notemos que si vamos de adentro hacia afuera, hemos creado una matriz  $1000 \times 30$ , hemos calculado el promedio de cada fila y luego hacemos el histograma de estos 1000 promedios usando 32 intervalos. Asimismo si a la función `hist` no le ponemos la cantidad de intervalos, la asigna por default).
- (d) `title('n=1000')`
- (e)  $n=5000$ ;
- (f) `subplot (2,2,3)`;
- (g) `hist(mean(poisrnd(lambda,n,col),2),71,'facecolor','r','edgecolor','b')`
- (h) `title('n=5000')`
- (i)  $n=10000$ ;
- (j) `subplot (2,2,4)`;
- (k) `hist(mean(poisrnd(lambda,n,col),2),100,'facecolor','r','edgecolor','b')`
- (l) `title('n=10000')`
- (m) `hold off`

Obtenemos los histogramas que se muestran en la figura (1).

- d) ¿ Los histogramas sugieren algún patrón, evolucionan de alguna manera ?

A medida que aumenta  $n$ , los histogramas tienden a quedar más simétricos y “suaves”, centrados en el valor  $\lambda = 3$ , ya que grosso modo la mitad de las filas producen estimadores  $\hat{\lambda} \geq \lambda$  y la otra mitad produce estimadores  $\hat{\lambda} \leq \lambda$ .

Asimismo son más frecuentes las estimaciones que no se apartan mucho del valor real  $\lambda = 3$  que las que se alejan del mismo.

Si miramos el último histograma, el generado con  $n=10000$ , a ojo se ve que la gran mayoría de dichas estimaciones se encuentran entre 2.5 y 3.5.

Podemos calcular dicha proporción contando la cantidad de  $\hat{\lambda}$  que se encuentran entre 2.5 y 3.5.

Si guardamos la matriz  $n \times 1$  que contiene los promedios de cada fila como `Xmean` (queda  $10000 \times 1$ ), y hacemos

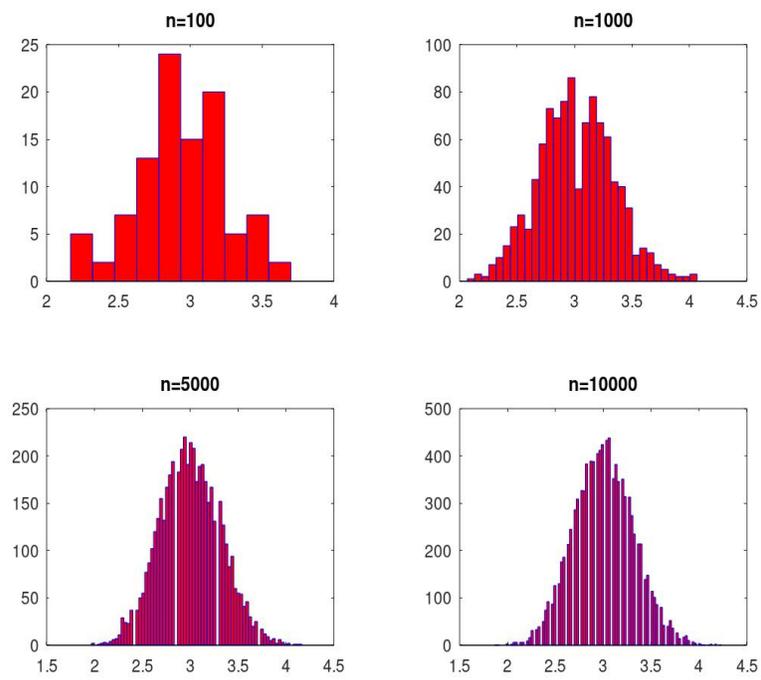


Fig. 1: Histogramas

`size(find(Xmean>=2.5 Xmean<=3.5))`, el comando `find` busca en la matriz `Xmean` aquellas filas con valores entre 2.5 y 3.5, dando como salida una matriz 10000 x 1 con ceros en aquellos lugares donde el promedio no estaba entre 2.5 y 3.5, y unos donde sí se cumple.

Luego el comando `size` cuenta la cantidad de unos, el resultado para esa matriz concreta es 8990.

Esto es, el 89.9% de las 10000 veces obtuvimos estimadores  $\hat{\lambda}$  entre 2.5 y 3.5, lo que confirma la impresión visual del histograma.

**Estas 2 últimas partes pueden omitirse, refieren al apéndice de la clase de estimación puntual, métodos que se usaban antes cuando se tenían pocos datos.**

- e) Genere una matriz de  $n=10$  filas y 15 columnas de datos con distribución Poisson de parámetro  $\lambda = 3$ . Para cada fila  $X = \{x_1, \dots, x_{15}\}$  llamemos  $x$  al vector con los valores distintos que toma  $X$  y  $f_{obs}(x)$  a la frecuencia (cantidad de veces) de cada uno de dichos valores distintos.

Una forma de hacerlo es la siguiente:

- (a) `pkg load statistics`
- (b) `lambda=3;`
- (c) `n=10;`
- (d) `col=15;`
- (e) `X=poissrnd(lambda,n,col);`

- f) Para cada una de las 10 filas grafique  $F(x) = \log(f_{obs}) + \log(x!)$  (con el comando `hold-on` puede tener las 10 gráficas juntas). ¿Las gráficas de estas 10  $F(x)$  lucen como rectas?.

Una forma de hacerlo es la siguiente:

- (a) `clf;`
- (b) `aux=1;`
- (c) `for i=1 : n`
- (d) `Xi=unique(sort(X(i,:),2));`
- (e) `for l=1 : length(Xi)`
- (f) `if (l==1)`
- (g) `fobsi=[sum(X(i,:)==Xi(1,1))];`
- (h) `else`
- (i) `fobsi=[fobsi sum(X(i,:)==Xi(1,1))];`
- (j) `endif`
- (k) `endfor`
- (l) `Yi=log(fobsi).+log(factorial(Xi));`

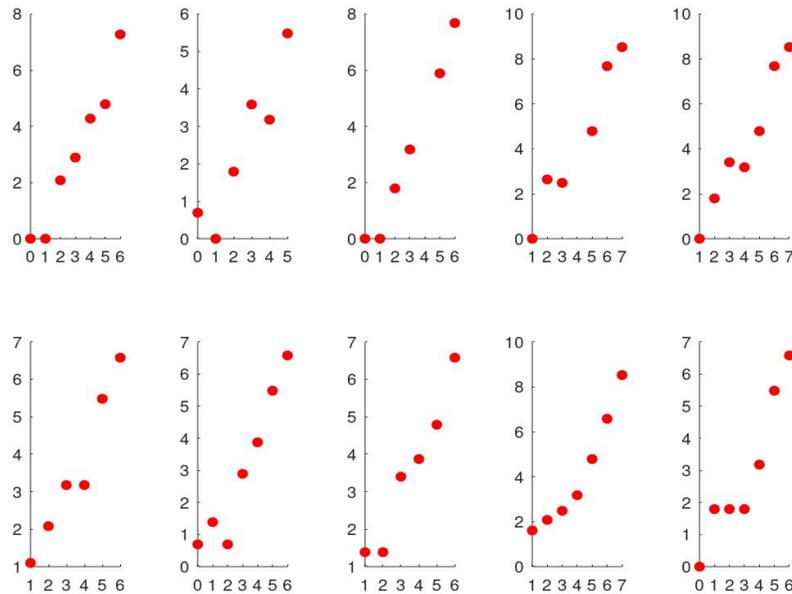


Fig. 2: Gráficas de la parte f)

- (m) subplot (2,5,aux);
- (n) scatter(Xi,Yi,'r','filled')
- (o) aux = aux + 1;
- (p) endfor

Las gráficas pueden verse en la figura (2).

2. a) Para  $n=100,1000,5000,10000$  genere una matriz de  $n$  filas y 40 columnas de datos con distribución exponencial de parámetro  $\lambda = 2$ ). Halle la media de cada fila (Para la distribución Exponencial el estimador por momentos y el de máxima verosimilitud son el mismo, el promedio).  
 Generamos la matriz (usamos  $n=1000$ ) con el comando:  
`mexp=exprnd(2,1000,40);`  
 generamos la matriz-columna con los promedios de cada fila:  
`mediasexp=mean(mexp,2);`
- b) Construya el histograma de dichas medias. El histograma aparecerá con 10 intervalos. Lo hacemos con el comando:  
`hist(mediasexp).`

---

Pruebe con distinta cantidad de intervalos hasta obtener un histograma que a su juicio sea satisfactorio.

Para  $n=1000$  iremos generando distintos histogramas con distinta cantidad de intervalos.

Hcemos 6 histogramas con 10, 30, 50, 70, 90 y 110 intervalos (fig. 3)

```
figure()
hold
subplot(2,3,1)
hist(mediasexp,10)
title('n=10 intervalos')
subplot(2,3,2)
hist(mediasexp,30)
title('30 intervalos')
subplot(2,3,3)
hist(mediasexp,50)
title('50 intervalos')
subplot(2,3,4)
hist(mediasexp,70)
title('70 intervalos')
subplot(2,3,5)
hist(mediasexp,90)
title('90 intervalos')
subplot(2,3,6)
hist(mediasexp,110)
title('110 intervalos')
print('6histogramas.png')
esta sentencia guarda la ventana con los 6 histogramas.
```

- c) ¿ Los histogramas sugieren algún patrón, evolucionan de alguna manera ?

Vemos que hay una cantidad óptima de intervalos, (del orden de  $\sqrt{n}$ ), si son demasiados o demasiado pocos el histograma queda o bien excesivamente irregular y con pinchos, o excesivamente grueso.

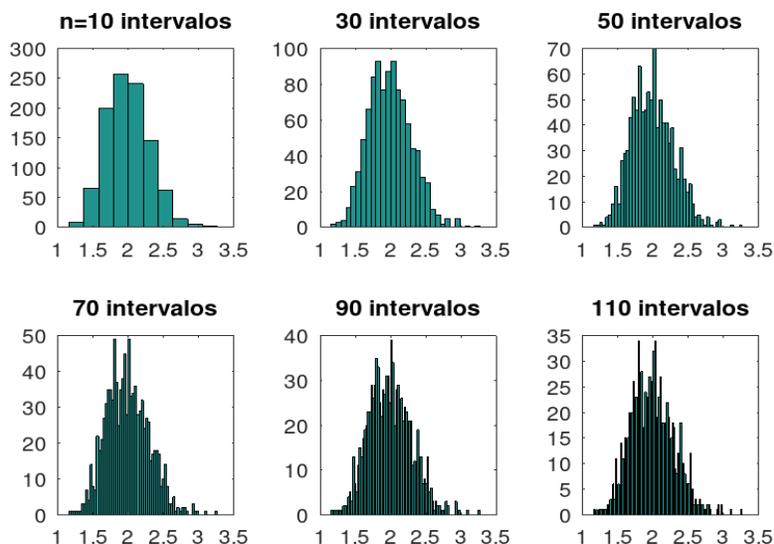


Fig. 3: Histogramas del ejercicio 2

**Estas 2 últimas partes pueden omitirse, refieren al apéndice de la clase de estimación puntual, métodos que se usaban antes cuando se tenían pocos datos.**

- d) Genere una matriz de  $n=10$  filas y 25 columnas de datos con distribución Exponencial de parámetro  $\lambda = 2$ . Para cada fila  $X = \{x_1, \dots, x_{25}\}$  ordene los valores de menor a mayor y divida dicho rango en 5 o 6 subintervalos de igual longitud. Tomaremos el punto medio de cada intervalo para definir  $x$ , y la frecuencia que asignaremos a cada uno de esos valores será la cantidad de datos que caen en el subintervalo en cuestión.
- e) A partir de aquí proceda como en la parte f) del ejercicio anterior, solo que ahora  $F(x) = \log(f_{obs}) - \log(25)$ .
3. Para  $n=100, 1000, 5000, 10000$  genere una matriz de  $n$  filas y 40 columnas de datos con distribución normal estándar.
- Para cada valor de  $n$ , estime  $\mu$  y  $\sigma$  usando cada fila mediante el método máxima verosimilitud. Esto le proporcionará  $n$  estimaciones de  $\mu$  y  $n$  estimaciones para  $\sigma$ .
- b) Para cada valor de  $n$  construya los histogramas de estos estimadores y estudie su evolución a medida que  $n$  crece.
- for  $v=[100\ 1000\ 5000\ 10000]$ (la variable  $v$  tomara cada valor del vector)

```

matriz=normrnd(0,1,v,40);(generamos la matriz)
mu=mean(matriz,2);(guardamos el promedio de cada fila de la ma-
triz)
mediacuadrados=mean(matriz.^2,2);(elpromediodeloscuadrados)
sigma = mediacuadrados - mu.^2;(estimador de sigma cuadrado)
figure()
subplot(2,1,1)
hist(mu,sqrt(v))
title(['estimaciones de mu, n=',num2str(v)])
subplot(2,1,2)
hist(sqrt(sigmaest),sqrt(v))
title(['estimaciones de sigma, n=',num2str(v)])
end.

```

Los histogramas obtenidos se muestran en la figura (4).

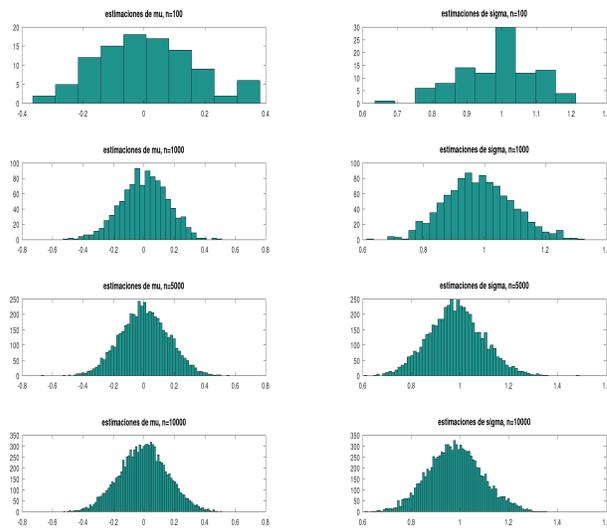


Fig. 4: Histogramas de los estimadores de  $\mu$  y  $\sigma$

**Estas 2 últimas partes pueden omitirse, refieren al apéndice de la clase de estimación puntual, métodos que se usaban antes cuando se tenían pocos datos.**

- c) Genere una matriz de  $n=10$  filas y 20 columnas de datos con distribución Normal estándar. Para cada fila  $X = \{x_1, \dots, x_{20}\}$  ordene los valores de menor a mayor y redondéelos al entero más cercano. Llamemos  $x$  al vector con estos valores redondeados y  $f_{obs}$  al vector con la cantidad de veces que observamos cada valor de  $x$ .

- d) Al igual que en los otros ejercicios plotee las  $F(x) = \log(f_{obs}) - \log(20) + \log(\sqrt{2\pi})$  y evalúe si lucen como parábolas.
4. a) Para  $n=100,1000,5000,10000$  genere una matriz de  $n$  filas y 35 columnas de datos con distribución uniforme  $(0,b)$  con  $b=1$ . Para cada fila estime el parámetro  $b$  mediante momentos y máxima verosimilitud.
- b) Para cada  $n$ , plotee el histograma de ambos estimadores de  $b$ , el de momentos y el de máxima verosimilitud. ¿ Cual de los dos métodos parece funcionar mejor?
5. Genere una matriz de 100 filas y 35 columnas con datos de distribución Normal de parámetros  $\mu = 40$  y  $\sigma = 5$ .
- Para  $\alpha = 0.10$  y  $\alpha = 0.05$  y para  $c/u$  de las 100 muestras encuentre los respectivos IC al nivel  $1 - \alpha$  para  $\mu$ , suponiendo  $\sigma$  conocido.
  - Repetir lo anterior pero ahora suponiendo que no se conoce  $\sigma$ .
  - Que porcentaje de los 100 IC captura a  $\mu$  en su interior?