Ejercicios Sobre Estimación Puntual y por Intervalos

1. a) 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.

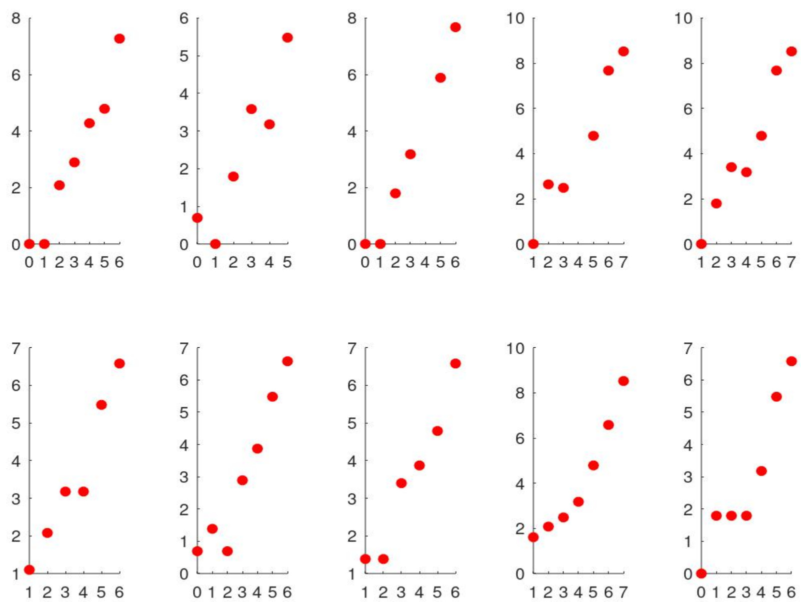
   Una forma de hacerlo es la siguiente:

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

   b) 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,l))];
   (j) endif
   (k) endfor
   (l) Yi=log(fobsi).+log(factorial(Xi));
   (m) subplot (2,5,aux);
   (n) scatter(Xi,Yi,'r','filled')
   (o) aux = aux + 1;
   (p) endfor



..

Fig. 1: Gráficas de la parte b)

Las gráficas pueden verse en la figura (1).

c) 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).

   Una forma de hacerlo es la siguiente:

   (a) lambda=3;
   (b) n=100;
   (c) col=30;
   (d) X=poissrnd(lambda,n,col);
   (e) Xmean=mean(X,2);

d) Construya el histograma de dichas medias. El histograma aparecerá con 10 intervalos.

   Una forma de hacerlo es la siguiente:

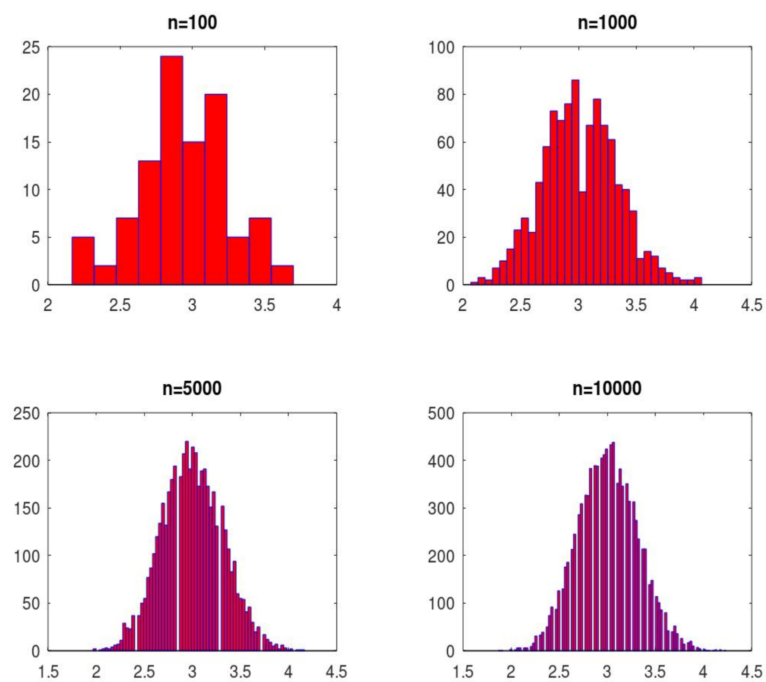
   (a) figure()
   (b) hold on
   (c) subplot (2,2,1);
   (d) hist(Xmean,10,'facecolor','r','edgecolor','b')
   (e) title('n=100')

e) 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;
   (b) subplot (2,2,2);
   (c) hist(mean(poissrnd(lambda,n,col),2),32,'facecolor','r','edgecolor','b')
   (d) title('n=1000')
   (e) n=5000;
   (f) subplot (2,2,3);
   (g) hist(mean(poissrnd(lambda,n,col),2),71,'facecolor','r','edgecolor','b')
   (h) title('n=5000')
   (i) n=10000;
   (j) subplot (2,2,4);
   (k) hist(mean(poissrnd(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 (2).



..

Fig. 2: Histogramas partes d) y e)

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

2. La resolución de este ejercicio es similar a la del anterior.

a) Genere una matriz de n=10 filas y 25 columnas de datos con distribución Exponencial de parámetro $\lambda = 2$. Para cada fila $X = \{x_1, \ldots, x_{25}\}$ 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.

b) A partir de aquí proceda como en la parte b) del ejercicio anterior, solo que ahora $F(x) = \log(f_{obs}) - \log(25)$.

c) Para n=100,1000,5000,10000 genere una matriz de n filas y 40 columnas de datos con la misma distribución de la parte a). Halle la media de cada fila (Para la distribución Exponencial el estimador por momentos y el de máxima verosimilitud son el mismo, el promedio).

d) Construya el histograma de dichas medias. El histograma aparecerá con 10 intervalos. Pruebe con distinta cantidad de intervalos hasta obtener un histograma que a su juicio sea satisfactorio.

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

3. a) Genere una matriz de n=10 filas y 20 columnas de datos con distribución Normal estándar. Para cada fila $X = \{x_1, \ldots, x_{20}\}$ 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.

Una forma de hacerlo es la siguiente:

(a) pkg load statistics
(b) n=10;
(c) col=20;
(d) X=randn(n,col);
(e) x=round(sort(X,2));

b) 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.

Una forma de hacerlo es la siguiente:

(a) clf;
(b) aux=1;
(c) for i=1 : n
(d) xi=unique(x(i,:));
(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,l))];
(j) endif
(k) endfor
(l) Yi=log(fobsi).-log(n).+log(sqrt(pi));
(m) subplot (2,5,aux);
(n) scatter(xi,Yi,'r','filled')
(o) aux = aux + 1;
(p) endfor

Para n=100,1000,5000,10000 genere una matriz de n filas y 40 columnas de datos con la misma distribución de la parte a).

c) Para cada valor de n, estime μ y σ usando cada fila mediante el método de los momentos y el método de máxima verosimilitud. Esto le proporcionará n estimaciones de μ obtenida mediante momentos y máxima verosimilitud (para el caso de μ no hay diferencia) y n estimaciones para σ (pare este caso sí hay diferencia entre ambos estimadores).

Una forma de hacerlo (ejemplificamos con n=100) es la siguiente (repasar los estimadores de μ y σ^2 por los métodos de los momentos y máxima verosimilitud):

(a) n=100;
(b) X=randn(n,col);
(c) hold on
(d) figure()
(e) subplot (1,3,1);
(f) hist(mean(X,2),10,'facecolor','r')
(g) title('MU - n=100')
(h) subplot (1,3,2);
(i) $hist(mean(X.^2, 2) - mean(X, 2).^2, 10,'facecolor','b')$
(j) title('SIGMA - Método Momentos - n=100')
(k) subplot (1,3,3);
(l) hist(sqrt(var(X,1,2)),10,'facecolor','b')
(m) title('SIGMA - EMV - n=100')
(n) hold off

d) Para cada valor de n construya los histogramas de estos estimadores y estudie su evolución a medida que n crece. Con base en dichos histogramas, qué método de estimación luce mejor para σ, momentos o máxima verosimilitud?

Una forma de hacerlo es la siguiente (ejemplificamos para n=1000, el resto es igual):

(a) n=1000;
(b) X=randn(n,col);
(c) hold on
(d) figure()
(e) subplot (1,3,1);
(f) hist(mean(X,2),32,'facecolor','r')
(g) title('MU - n=1000')
(h) subplot (1,3,2);
(i) $hist(mean(X.^2,2) - mean(X,2).^2, 32,' facecolor',' b')$
(j) title('SIGMA - Método Momentos - n=1000')
(k) subplot (1,3,3);
(l) hist(sqrt(var(X,1,2)),32,'facecolor','b')
(m) title('SIGMA - EMV - n=1000')
(n) hold off

Se obtienen histogramas como los de la figura (3):

4. La resolución de este ejercicio es similar a la del anterior.

   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 μ, suponiendo σ conocido.

     Una forma de hacerlo ($\alpha = 0.10$, para el otro valor de α es similar) es la siguiente:

     (a) pkg load statistics
     (b) n=100;
     (c) matriz=normrnd(40,5,n,35);
     (d) $z_\alpha = norminv(.95, 0, 1)$;(es el $z_{\alpha/2}$que deja a su derecha el 0.05 de masa)
     (e) b=a;
     (f) captura=b;(pondremos un 1 en el lugar i si el IC i-simo captura a mu)
     (g) promedios=mean(matriz,2);(aquí guardaremos los promedios de cada fila)

(h) a=zeros(n,1); (aquí guardaremos los extremos izquierdos de los intervalos de confianza)

(i) b=a;

(j) captura=a; (aquí pondremos un 1 cada vez que el intervalo [a,b] capture a μ)

(k) for i=1:n

(l) $a(i,1) = promedios(i,1) - z_\alpha * 5/10;$

(m) $b(i,1) = promedios(i,1) + z_\alpha * 5/10;$

(n) Si a(i,1)<=40 y 40<=b(1,1)

(o) captura(i,1)=1;

(p) endif

(q) endfor

- Repetir lo anterior pero ahora suponiendo que no se conoce σ.

  Igual que lo anterior, solamente que en lugar de z se halla t, mediante

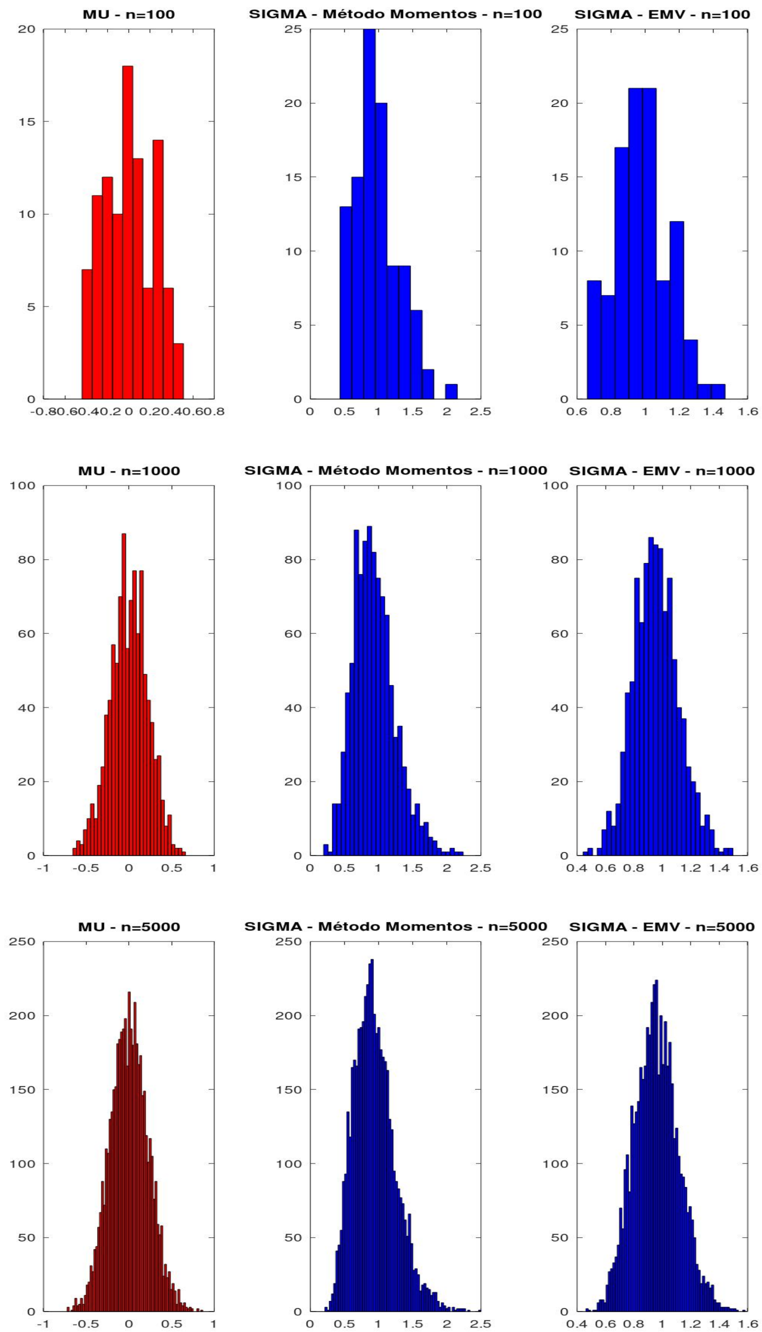
  (a) $t_\alpha = tinv(.95, 34);$ (usamos la t de Student con 34 grados de libertad)

  (b) Al no conocer σ usamos el estimador calculado a partir de los datos, en cada fila i (cada iteración del for), debemos calcular $s = std(matriz(i,:));$

  (c) El radio del intervalo ahora es $\pm t_\alpha * s/10$, el resto sigue igual.

- Que porcentaje de los 100 IC captura a μ en su interior?

  Para cada caso contamos cuantos lugares de la matriz columna captura tiene unos y lo dividimos entre n.



..

..

..

Fig. 3: Ejercicio 3) d)