Métodos Estadísticos

Tarea 1

Métodos Estadísticos. Ciencia de datos

González Paz Edgar

1. Considere una muestra aleatoria $X_1, ..., X_n$ con función de densidad

$$f(x|\theta) = 1_{\{\theta < x < \theta+1\}}, \theta \in \mathrm{R}$$

1.1 Calcule un estimador para θ por el método de momentos y por el método de máxima verosimilitud, diga si estos son insesgados y suficientes para theta.

Primero calculamos la esperanza de nuestra función de densidad

$$\mathrm{E}(x) = \int_{\mathrm{R}} x f(x|\theta)\, dx = \int_{\theta}^{\theta+1} x\, dx = \frac{x^2}{2}\Big|_{\theta}^{\theta+1} = \frac{(\theta+1)^2}{2} - \frac{\theta^2}{2} = \frac{\theta^2 + 2\theta + 1 - \theta^2}{2} = \frac{2\theta+1}{2}$$

Entonces, igualamos el primer momento muestral a la esperanza para obtener el estimador de momentos:

$$\Rightarrow \frac{2\theta+1}{2} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\Rightarrow 2\theta + 1 = \frac{2}{n}\sum_{i=1}^{n} x_i$$

$$\Rightarrow 2\theta = \frac{2}{n}\sum_{i=1}^{n} x_i - 1$$

$$\Rightarrow \hat{\theta}_{MM} = \frac{2\bar{X} - 1}{2}$$

Para el $\hat{\theta}_{MLE}$ obtenemos la función de verosimilitud del parámetro θ:

$$L(\theta|X) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} 1_{\{\theta < x_i < \theta+1\}} = \prod_{i=1}^{n} 1_{\{\theta < x_i\}} \prod_{i=1}^{n} 1_{\{x_i < \theta+1\}} = 1_{\{X_1 > \theta\}} 1_{\{X_n < \theta+1\}}$$

$$= 1_{\{X_1 > \theta\}} 1_{\{X_n - 1 < \theta\}} = 1_{\{X_{(n)} - 1 < \theta < X_{(1)}\}}$$

$$1_{\{X_{(n)} - 1 < \theta < X_{(1)}\}} = \begin{cases} 1, & si\ X_{(n)} - 1 \leq \theta \leq X_{(1)} \\ 0, & en\ cualquier\ otro\ caso \end{cases}$$

No hay un estimador único de Máxima Verosimiltud, puede existir más de un estimador de θ tal que cumpla:

$$X_{(n)} - 1 \leq \theta \leq X_{(1)}$$

Es decir, aquellos que sean de la forma

$$\hat{\theta}_{MLE} = w_1(X_{(n)} - 1) + w_2 X_{(1)}$$

donde $w_1, w_2 \in \mathrm{R}^+$, es decir, $w_1, w_2 \geq 0$ t.q. $w_1 + w_2 = 1$.

¿Son insesgados y suficientes para θ?

Primero, probamos para el $\hat{\theta}_{MLE}$. Necesitamos obtener la esperanza de este estadístico.

$$\mathrm{E}(\hat{\theta}_{MLE}) = \mathrm{E}(w_1(X_{(n)} - 1) + w_2 X_{(n)}) = w1\mathrm{E}(X_{(n)}) - \mathrm{E}(w_1) + w_2\mathrm{E}(X_{(n)})$$

Debemos obtener la esperanza del estadístico $X_{(n)}$. Sea $Y = X_{(n)}$

$$\mathrm{E}(X_{(n)}) = \mathrm{E}(Y) = \int_{\mathrm{R}} y \frac{n(y-\theta)^{n-1}}{(\theta+1-\theta)^n} 1_{\{\theta < y < \theta+1\}}\, dy = n\left(\int_{\theta}^{\theta+1} y(y-\theta)^{n-1}\, dy\right)$$

Hacemos un cambio de variable.

$$Sea\ u = y - \theta,\ entonces\ y = u + \theta. Para\ los\ límites,\ Si\ y = \theta \rightarrow u = 0,\ Si\ y = \theta + 1 \rightarrow u = 1$$

$$\Rightarrow n\left(\int_0^1 (u+\theta)u^{n-1}\, du\right) = n\left(\int_0^1 u^n + \theta u^{n-1}\, du\right) = n\left(\int_0^1 u^n\, du + \theta\int_0^1 u^{n-1}\, du\right) = n\left(\frac{u^{n+1}}{n+1}\Big|_0^1 + \theta\frac{u^n}{n}\Big|_0^1\right) = n\left(\frac{1}{n+1} + \frac{\theta}{n}\right)$$

$$= \frac{n}{n+1} + \theta^n = \frac{n + \theta(n+1)}{n+1}$$

También debemos obtener la esperanza del estadístico $X_{(1)}$. Sea $Y = X_{(1)}$

$$\Rightarrow \mathrm{E}(X_{(1)}) = \mathrm{E}(y) = \int_{\mathrm{R}} y n(\theta+1-y)^{n-1} 1_{\{\theta < y < \theta+1\}}\, dy = n\left(\int_{\theta}^{\theta+1} y(\theta+1-y)\, dy\right)$$

Hacemos un cambio de variable.

$$Sea\ u = \theta + 1 - y,\ entonces\ y = \theta + 1 - u. Para\ los\ límites,\ Si\ y = \theta \rightarrow u = 1,\ Si\ y = \theta + 1 \rightarrow u = 0 dx = -du$$

$$= n\left(\int_1^0 (\theta + 1 - u)u^{n-1}\ - du\right) = n\left(\int_0^1 \theta u^{n-1}\, du + \int_0^1 u^{n-1}\, du - \int_0^1 u^n\, du\right) = n\left(\theta\frac{u^n}{n}\Big|_1^0 + \frac{u^n}{n}\Big|_1^0 - \frac{u^{n+1}}{n+1}\Big|_1^0\right) = n\left(\frac{\theta}{n} + \frac{1}{n} - \frac{1}{n+1}\right)$$

$$= \theta + 1 - \frac{n}{n+1} = \theta + \frac{1}{n+1}$$

Retomando

$$\Rightarrow \mathrm{E}(\hat{\theta}_{MLE}) = w_1\left(\frac{n + \theta(n+1)}{n+1}\right) - w_1 + w_2\left(\theta + \frac{1}{n+1}\right) = w_1\left(\frac{n + \theta(n+1)}{n+1} - 1\right) + w_2\left(\theta + \frac{1}{n+1}\right)$$

$$= w_1\left(\frac{n + \theta n + \theta - n - 1}{n+1}\right) + w_2\left(\frac{\theta n + \theta + 1}{n+1}\right) = \frac{w_1 n\theta + w_1\theta - w_1 + w_2 n\theta + w_2 n\theta + w_2\theta + w_2}{n+1}$$

Entonces $\mathrm{E}(\hat{\theta}_{MLE}) = \theta \Leftrightarrow w_1 = \frac{1}{2} w_2 = \frac{1}{2}$. Por lo tanto, un MLE insesgado sería

$$\hat{\theta}_{MLE} = \frac{X_{(n)} - 1}{2} + \frac{X_{(1)}}{2}$$

Ahora bien, para $\hat{\theta}_{MM}$, tenemos que:

$$\mathrm{E}(\hat{\theta}_{MM}) = \mathrm{E}\left(\frac{2\bar{X} - 1}{2}\right) = \mathrm{E}(\bar{X}) - \mathrm{E}(\frac{1}{2}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(x_i) - \frac{1}{2} = \frac{2\theta+1}{2} - \frac{1}{2} = \theta + \frac{1}{2} - \frac{1}{2} = \theta$$

Por lo tanto, $\hat{\theta}_{MM}$ sí es insesgado.

Suficiencia

Utilizamos el teorema del cociente del estadístico suficiente minimal para θ. Dado una función $T(\mathrm{X})$ y para cualesquiera muestras observardas X, Y, realizamos el cociente $\frac{f(\mathrm{X}|\theta)}{f(\mathrm{Y}|\theta)}$, si no depende de θ sii $T(\mathrm{X}) = T(\mathrm{Y})$ entonces $T(\mathrm{X})$ es suficiente minimal para θ.

$$\frac{f(\mathrm{X}|\theta)}{f(\mathrm{Y}|\theta)} = \frac{1_{\{X_{(n)} - 1 < \theta < X_{(1)}\}}}{1_{\{Y_{(n)} - 1 < \theta < Y_{(1)}\}}}$$

¿Qué valores puede tomar el numerador y el denominador?

$$1_{\{X_{(n)} - 1 < \theta < X_{(1)}\}} = \begin{cases} 1, & si\ \ \theta \in (X_{(n)} - 1, X_{(1)}) \\ 0, & en\ \ \theta \notin (X_{(n)} - 1, X_{(1)}) \end{cases}$$

$$1_{\{Y_{(n)} - 1 < \theta < Y_{(1)}\}} = \begin{cases} 1, & si\ \ \theta \in (Y_{(n)} - 1, Y_{(1)}) \\ 0, & en\ \ \theta \notin (Y_{(n)} - 1, Y_{(1)}) \end{cases}$$

Tenemos dos casos, (omitimos $f(\mathrm{Y}|\theta) = 0$ en el denominador porque es una indeterminación, sin embargo, no dependería de θ)

$$\frac{f(\mathrm{X}|\theta)}{f(\mathrm{Y}|\theta)} = 1 \Leftrightarrow \theta \in (X_{(n)} - 1, X_{(1)}), \theta \notin (Y_{(n)} - 1, Y_{(1)})$$

$$\frac{f(\mathrm{X}|\theta)}{f(\mathrm{Y}|\theta)} = 0 \Leftrightarrow \theta \notin (X_{(n)} - 1, X_{(1)}), \theta \in (Y_{(n)} - 1, Y_{(1)})$$

El cociente no depende de θ sii

$$(X_{(n)} - 1, X_{(1)}) = (Y_{(n)} - 1, Y_{(1)})$$

$$\Rightarrow X_{(n)} - 1 = Y_{(n)} - 1 \Rightarrow X_{(n)} = Y_{(n)} \wedge X_{(1)} = Y_{(1)}$$

$\Rightarrow$ El vector de estadísticos $(X_{(n)}, X_{(1)})$ son los únicos estimadores suficientes minimales.

In [1]:
```
#Código para cargar el conjutno de datos, calcular los estimadores y realizar las gráficas.
```

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import pylab
```

In [3]:
```
datos1 = pd.read_csv('dataEj1.txt')
```

In [4]:
```
datos1.head()
```

Out[4]:

	x
0	0.257329
1	-0.111362
2	0.376463
3	-0.002884
4	-0.229582

In [5]:
```
MuestraAleatoria1 = np.asarray(datos1['x'].values.tolist())
```

In [6]:
```
#Función de densidad Uniforme
def Uniforme(x,a,b):
    if a<x<b:
        r= 1/(b-a)
    else:
        r = 0
    return r
#Estimador por el método de momentos
def ThetaMM(X):
    n= len(X)
    Promedio = sum(X[i] for i in range(n))
    Xbarra = (1/n)*Promedio
    return (2*Xbarra - 1)/2
#Estimador de Máxima Verosimilitud
def ThetaMLE(X,w1,w2):
    Xn = max(X)
    X1 = min(X)
    if w1>=0 and w2 >=0 and (w1+w2==1):
        ThetaMLE = w1*(Xn-1)+w2*(X1)
    return ThetaMLE
```

1.2 Con ayuda de la muestra aleatoria observada, dataEj1.txt , calcule la estimaciones en ambos casos y quédese con alguna de las dos, justificando su respuesta.

Posibles estimadores a elegir:

In [7]:
```
ThetaMM(MuestraAleatoria1)
```

Out[7]: `-0.498538219173369`

In [8]:
```
ThetaMLE(MuestraAleatoria1,0.5,0.5)
```

Out[8]: `-0.499965088092722`

In [9]:
```
min(MuestraAleatoria1)
```

Out[9]: `-0.499822691315785`

In [10]:
```
max(MuestraAleatoria1)-1
```

Out[10]: `-0.500107484869659`

En este caso, tomaremos la segunda estimación $\hat{\theta}_{MLE}$ porque es insesgado y (no estoy seguro aquí) es función de los estadísticos suficientes minimales.

1.3 Finalmente haga un histograma (recuerdo normalizarlo) con los datos y encime la función de densidad $f(x|\hat{\theta})$, con $\hat{\theta}$ la estimación elegida del inciso anterior.

In [11]:
```
Dominio=np.arange(min(MuestraAleatoria1)-0.1,max(MuestraAleatoria1)+0.1,0.01)
Rango=[]
for i in range(len(Dominio)):
        Rango.append(Uniforme(a=ThetaMLE(MuestraAleatoria1,0.5,0.5), b=ThetaMLE(MuestraAleat
oria1,0.5,0.5) +1 ,x=Dominio[i]))
fig, (ax1, ax2) = plt.subplots(1,2, figsize=[12,5])
ax1.hist(MuestraAleatoria1,density=True)
ax1.title.set_text('Histograma de datos')
ax2.hist(MuestraAleatoria1,density=True)
ax2.plot(Dominio,Rango)
ax2.title.set_text('Histograma de datos con '+ r'$f(x|\hat{\theta})$')
plt.show()
```