Procesamiento y Análisis de Datos Astronómicos

6.- Probabilidades

R. Gil-Hutton

Marzo 2020

Actividades:

Práctica 5:

- Para una variable cualquiera de su archivo de datos calcule todos sus estadísticos.
- Para esa misma variable verificar si el porcentaje de valores dentro de $1\,s$, $2\,s$ y $3\,s$ de $\bar{x}$ representan el $\sim 68{,}3\,\%$, $\sim 95{,}5\,\%$ y $\sim 99{,}7\,\%$ de la muestra, respectivamente.
- Para esa misma variable extraiga al azar 20 muestras con 100 elementos cada una, calcule los estadísticos de cada una de ellas y compare con los valores obtenidos en el primer punto.
- Los valores medios de las muestras usadas en el ejemplo para demostrar el Teorema del Límite central (pág. 24) tienden a un valor de 0,9228. Por qué?.

Actividades:

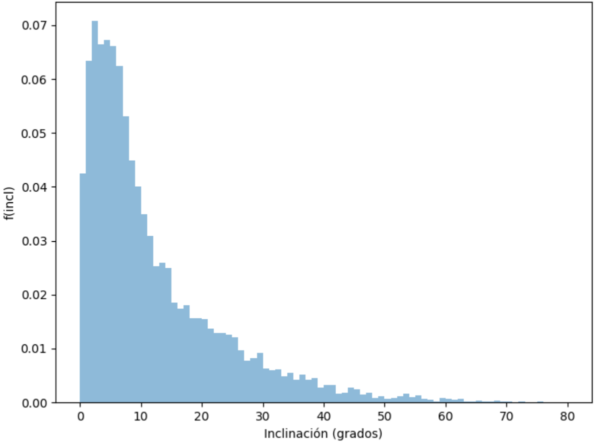
Práctica 5 (cont.):

- Utilizando la función de luminosidad de Schechter, $\phi(L) = \left(\frac{\phi^*}{L^*}\right)\left(\frac{L}{L^*}\right)^{\alpha} \exp(-L/L^*)$, con $L^* = 1$, $\phi^* = 1$ y $\alpha = -0{,}5$, generen una población de galaxias para $0 < L < 6$ y extraigan muestras de 10 galaxias. Cómo se distribuye el valor máximo de cada muestra?. Y en el caso de extraer 50 galaxias?.

Actividades:

```
In [1]: cd cursos/proc-datos/clases/actividades/practica-5
/home/rgh/cursos/proc-datos/clases/actividades/practica-5

In [2]: apo=np.loadtxt('apollo-aeih.dat')

In [3]: import scipy.stats as sts

In [4]: val=sts.describe(apo[:,2],ddof=1)

In [5]: val
Out[5]: DescribeResult(nobs=9239, minmax=(0.023188, 154.372007), mean=12.0879821
1343219, variance=129.8878243300749, skewness=1.842808552339338, kurtosis=5.5546
74671451179)

In [6]: sts.mode(apo[:,2]),np.median(apo[:,2]),np.std(apo[:,2],ddof=1)
Out[6]:
(ModeResult(mode=array([7.756518]), count=array([2])),
 8.160087,
 11.396833960801345)

In [7]:
```

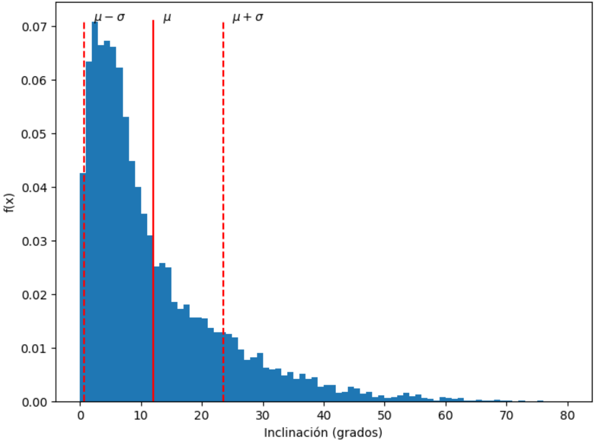
Actividades:



Actividades:

```
In [7]: ind=np.where(np.abs(apo[:,2]-val[2]) < np.sqrt(val[3]))

In [8]: uno=apo[ind]

In [9]: len(uno[:,2])/len(apo[:,2])
Out[9]: 0.8259551899556229

In [10]: ind=np.where(np.abs(apo[:,2]-val[2]) < 2.*np.sqrt(val[3]))

In [11]: dos=apo[ind]

In [12]: len(dos[:,2])/len(apo[:,2])
Out[12]: 0.9460980625608832

In [13]: ind=np.where(np.abs(apo[:,2]-val[2]) < 3.*np.sqrt(val[3]))

In [14]: tres=apo[ind]

In [15]: len(tres[:,2])/len(apo[:,2])
Out[15]: 0.9826821084532958

In [16]:
```

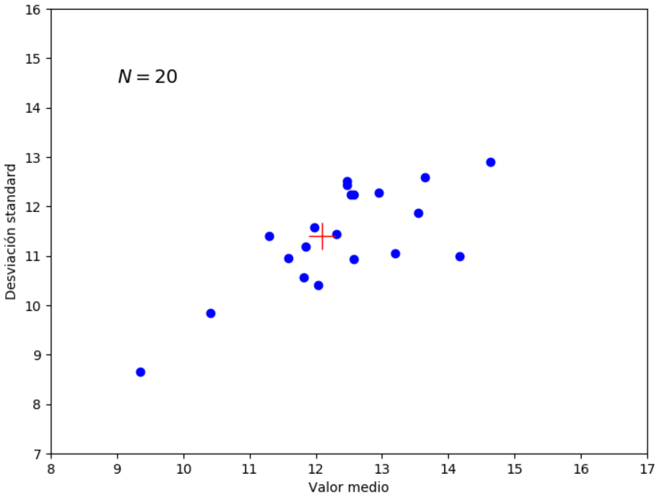
Actividades:



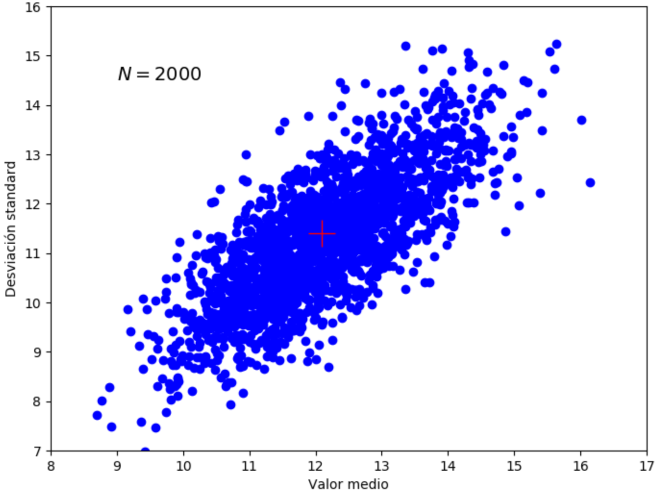
Actividades:

```
In [10]: cd cursos/proc-datos/clases/actividades/practica-5
/home/rgh/cursos/proc-datos/clases/actividades/practica-5

In [11]: apo=np.loadtxt('apollo-aeih.dat')

In [12]: his,lim=np.histogram(apo[:,2],bins=80,range=(0,80))

In [13]: his=his/np.sum(his)

In [14]: ac=np.cumsum(his)

In [15]: mue=np.zeros((20,100))

In [16]: for ii in range(20):
    ...:     kk=np.random.random(100)
    ...:     for jj in range(100):
    ...:         vp=0
    ...:         while(kk[jj]>ac[vp]):
    ...:             vp+=1
    ...:         mue[ii,jj]=lim[vp]+0.5+(np.random.random()-0.5)
    ...:

In [17]: muemed=np.mean(mue,axis=1)

In [18]: muestd=np.std(mue,axis=1,ddof=1)

In [19]:
```

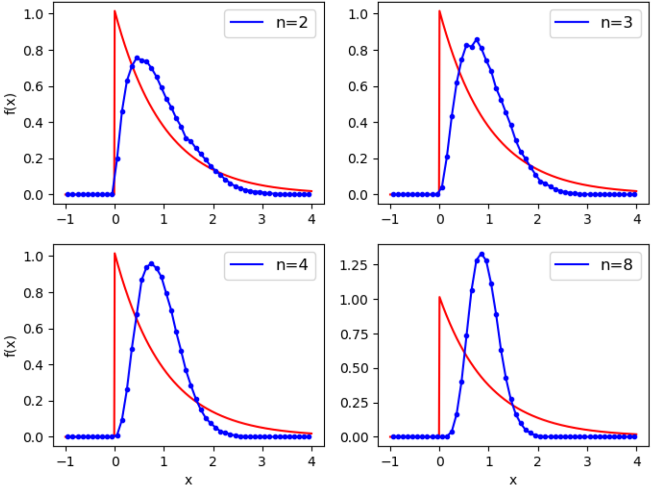
Actividades:



Actividades:



Actividades:



Actividades:

```
In [25]: xx=np.linspace(-1,4,500)

In [26]: yy=np.zeros(len(xx))

In [27]: for ii in range(len(yy)):
    ...:     if(xx[ii]>0.):
    ...:         yy[ii]=np.exp(-xx[ii])
    ...:

In [28]: yy=yy/np.sum(yy)

In [29]: np.mean(xx)
Out[29]: 1.5

In [30]: np.average(xx,weights=yy)
Out[30]: 0.9228314385225089

In [31]:
```

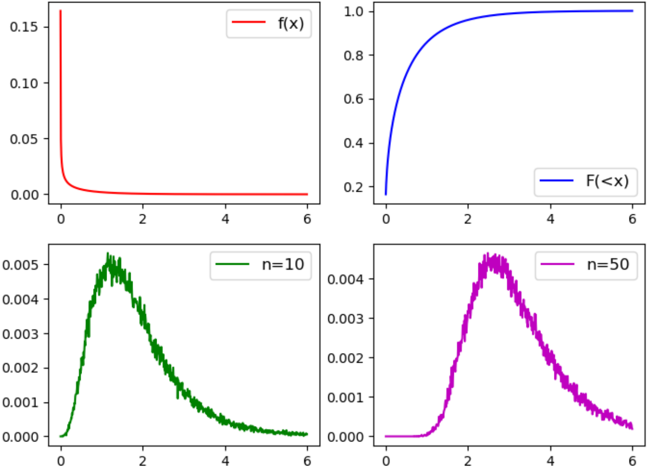
Actividades:

```
import numpy as np

def prom(x,ac,n,p):
  """
  Extrae p elementos de n valores cada uno del array x tomados al azar,
obtiene el máximo y  el histograma correspondiente. El ancho del bin es de
0.05 para 600 bins entre 0 y 6.
  """
  vx=[]
  for ii in range(p):
      kk=np.random.random(n)
      sx=[]
      for jj in range(n):
          vp=0
          while(kk[jj] > ac[vp]):
            vp+=1
          sx.append(x[vp])
      vx.append(np.max(sx))
  vx=np.array(vx)
  pes=np.ones(len(vx))/float(p)
  his=np.histogram(vx,bins=600,range=(0,6),weights=pes)
  return his
```

Actividades:

```
xx=np.linspace(0.001,6,600)
yy=xx**(-0.5)*np.exp(-xx)

yy=yy/np.sum(yy)

ac=np.cumsum(yy,dtype=float)

plt.figure()
plt.subplot(221)
plt.plot(xx,yy,'-r',label='f(x)')
plt.legend(fontsize=12)
plt.subplot(222)
plt.plot(xx,ac,'-b',label='F(<x)')
plt.legend(fontsize=12)

his0=prom(xx,ac,10,100000)
his1=prom(xx,ac,50,100000)

plt.subplot(223)
plt.plot(his0[1][1:]-0.005,his0[0],'-g',label='n=10')
plt.legend(fontsize=12)
plt.subplot(224)
plt.plot(his1[1][1:]-0.005,his1[0],'-m',label='n=50')
plt.legend(fontsize=12)
```

Actividades:



Probabilidad:

- Definición clásica de Laplace: la probabilidad $p(A)$ de un evento A en un experimento aleatorio es:

$$p(A) = \frac{w}{m}$$

  donde w es el número de casos en que A ocurre y m es el número de casos igualmente elegibles en el experimento.
- Concepto estadístico: decimos que el evento E tiene una probabilidad $p(E)$ si al repetir muchas veces un experimento se obtiene que $p(E) \simeq f(E)$.
- Ley de los grandes números de Bernoulli:
  $\lim_{m\to\infty} p\left(\left|\frac{w}{m} - p(E)\right| \leq \epsilon\right) \to 1$ para un infinitésimo ϵ.

Reglas de operación:

Axiomas de Kolmogorov:

1. Cualquier evento A tiene una $p(A)$ entre 0 y 1.
2. El evento seguro tiene $p(A) = 1$.
3. Si A y B son eventos excluyentes, tenemos que $p(A \text{ or } B) = p(A) + p(B)$.

Además:

- Si A y B son eventos independientes, tenemos que $p(A \text{ and } B) = p(A) \times p(B)$.
- Si A y B son eventos no independientes, tenemos que $p(A \text{ and } B) = p(A|B) \times p(B)$, donde $p(A|B)$ es la probabilidad condicional que suceda A siendo conocido B.
- Si existen varias posibilidades para el evento B, tenemos que $p(A \text{ and } B_i) = \sum_i p(A|B_i) \times p(B_i)$ (marginalización).

Reglas de operación:

- Si A y B no son independientes, entonces:

$$p(A) = p(A \text{ and } B) + p(A \text{ and } \sim B)$$

$$p(A) = p(A|B) \times p(B) + p(A| \sim B) \times p(\sim B)$$

$$p(A \text{ and } B) = p(A|B) \times p(B)$$

$$p(B \text{ and } A) = p(B|A) \times p(A)$$

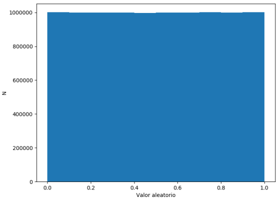
- Finalmente, como $p(A \text{ and } B) = p(B \text{ and } A)$:

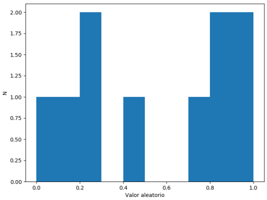
$$p(B|A) = \frac{p(A|B) \times p(B)}{p(A)}$$

que se denomina Teorema de Bayes.

Asignando probabilidades:

- Formalmente, el concepto estadístico asocia la probabilidad con la frecuencia relativa en la población.
- Además, el concepto estadístico tiene una fuerte dependencia con el tamaño de la muestra.
- Si la frecuencia relativa de un evento es baja y la muestra es pequeña posiblemente ese evento no se observe pero, si de todos modos se observa, su probabilidad estará distorsionada respecto de la real.





Asignando probabilidades:

- El principio de indiferencia nos dice que, a menos que podamos distinguir entre ellos, debemos asignar igual probabilidad a los eventos.
- Si es posible identificar eventos igualmente probables el problema se reduce simplemente a contar.
- Por ejemplo:
  - Al arrojar un dado asumimos que las 6 caras tienen igual posibilidad y asignamos una probabilidad a cada una de $1/6$. Es fácil asignar probabilidades.
  - Si el dado está cargado una de las caras tendrá mayor probabilidad de aparecer. Cómo asignamos probabilidad en estos casos?.

Asignando probabilidades:

- Muchas veces se estiman probabilidades directamente de los datos, pero la cantidad de datos usualmente es limitada y no siempre se pueden identificar casos igualmente probables.
- Si se definen probabilidades en base a la frecuencia de los datos la definición usualmente se vuelve circular porque los datos se obtienen en base a su frecuencia en la población.
- Definir probabilidades de este modo es parte de la **estadística clásica** o **estadística frecuentista**.
- La probabilidad refleja solo lo que conocemos, pero a veces se formula en base a valoraciones intuitivas o subjetivas.

Asignando probabilidades:

Ejemplo 1: Estamos estudiando un pequeño cúmulo de 100 estrellas de las cuales 80 son de secuencia principal mientras que 20 ya salieron de ella. Si selecciono al azar 6 estrellas, cuál es la probabilidad de obtener 2 estrellas que estén fuera de la secuencia?.

- El problema corresponde a un número total de T estrellas, con E estrellas en secuencia y F fuera, y seleccionando con reposición.
- Entonces, la probabilidad de obtener 2 estrellas fuera de secuencia de una muestra de 6 será:

$$p = \binom{6}{2} \left(\frac{F}{T}\right)^2 \left(\frac{E}{T}\right)^{(6-2)} = 0{,}24576$$

Asignando probabilidades:

Ejemplo 2: Tenemos asignada una noche de observación y queremos estimar cuál es la probabilidad de que esté nublada.

- En este caso podemos estimar la probabilidad considerando el número de noches nubladas durante el último año: $p =$ noches nubladas / 365.
- Si los datos del último año no nos parece suficiente podemos tomar los datos de cualquier otro período.
- La pregunta es si tienen la misma probabilidad de estar nublada una noche de invierno que de verano. Cómo definimos invierno?.

Otra definición: Probabilidad es una formalización numérica de nuestro grado de creencia sobre un hecho particular.

Asignando probabilidades:

Ejemplo 3: Lanzamos 4 monedas y registramos cuántas caras se obtienen.

- Con 4 monedas la probabilidad de no obtener caras es $\begin{pmatrix} 4 \\ 0 \end{pmatrix} (1/2)^4$.
- Con 4 monedas la probabilidad de obtener 1 cara es $\begin{pmatrix} 4 \\ 1 \end{pmatrix} \times (1/2)^4$.
- Con 4 monedas la probabilidad de obtener 2 caras es $\begin{pmatrix} 4 \\ 2 \end{pmatrix} \times (1/2)^4$.
- ...etc.

Asignando probabilidades:

- Si x es el número de caras, tenemos que $x = (0, 1, 2, 3, 4)$, y su probabilidad es $p(x) = (1/16,\ 1/4,\ 3/8,\ 1/4,\ 1/16)$.
- La suma es $\sum_{i=0}^{4} p(x) = 1$, por lo cual la distribución de probabilidades corresponde a la distribución de las frecuencias obtenidas para los posibles eventos.
- En este caso la variable aleatoria adquiere valores enteros, pero en otros casos puede tener valores reales. En esos casos podemos agrupar valores próximos de x y definir una densidad de probabilidades (área $\propto$ 1).

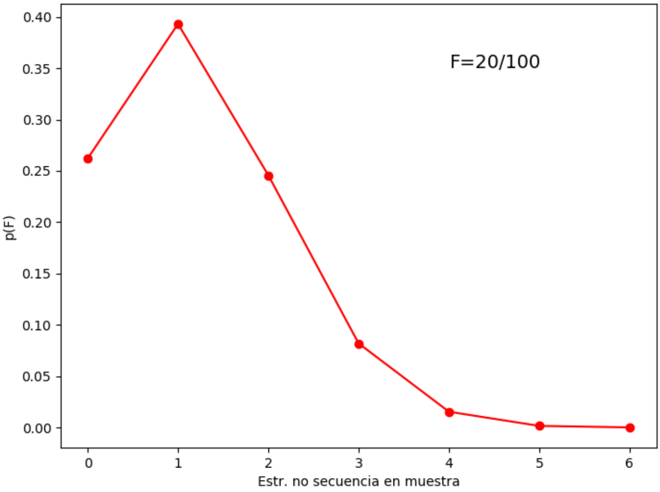
Asignando probabilidades:

Ejemplo 4: En el ejemplo 1, cuál es la probabilidad de obtener en una muestra al azar de 6 estrellas entre 0 y 6 estrellas fuera de la secuencia principal?.

- Número total de T estrellas, con E estrellas en secuencia y F fuera, y seleccionando con reposición.
- Entonces, para A estrellas fuera de secuencia en la muestra:

$$p = \binom{6}{\mathrm{A}} \left(\frac{F}{T}\right)^{A} \left(\frac{E}{T}\right)^{(6-A)}$$

Asignando probabilidades:



Distribuciones contínuas:

En el ejemplo la distribución es discreta, pero puede también ser contínua.

Si x es una variable aleatoria contínua, $f(x)$ es su función de densidad de probabilidad, usualmente designada como distribución de probabilidad, cuando:

1. $p(a < x < b) = \int_a^b f(x)dx$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. $f(x)$ es siempre un número real positivo para todo x.

En este caso, la correspondiente función acumulativa de densidad es $F(< y) = \int_{-\infty}^{y} f(x)dx$.

Distribuciones de probabilidad:

Distr.	Func. de densidad	μ	σ^2
Binomial	$f(x\|p,q) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$	np	npq
Poisson	$f(x\|\mu) = \frac{e^{-\mu}\mu^x}{x!}$	μ	μ
Normal	$f(x\|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	μ	σ^2
χ^2	$f(x\|\nu) = \frac{x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \exp\left(-\frac{x}{2}\right)$	ν	2ν

Distribuciones de probabilidad:

Distr.	Func. de densidad	μ	σ^2
t	$f(x\|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}}$	0	$\frac{\nu}{\nu-2}$
Beta	$f(x\|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$
Gamma	$f(x\|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\exp(-\beta x)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Maxwell	$f(x\|a) = \sqrt{\frac{2}{\pi}}\frac{x^2}{a^3}\exp\left(-\frac{x^2}{2a^2}\right)$	$2a\sqrt{\frac{2}{\pi}}$	$\frac{a^2(3\pi-8)}{\pi}$

Estadísticos y sus distribuciones:

Si tenemos una muestra de N elementos x_i con media $\bar{x}$ y varianza s^2 extraídos de una población con media μ y varianza σ^2, entonces:

- $\bar{x}$ tiene una distribución normal con media μ y varianza σ^2/N.
- s^2 se distribuye como $\sigma^2\chi^2/(N-1)$ con $(N-1)$ grados de libertad.
- el cociente:

  $$\frac{\sqrt{N}(\bar{x}-\mu)}{s^2}$$

  tiene una distribución t con $(N-1)$ grados de libertad.

Estadísticos y sus distribuciones:

- si tenemos dos muestra de N y M elementos obtenidas de una distribución normal, el cociente entre s_1^2 y s_2^2 siguen una distribución F.
- los estadísticos de orden k de muestras con N elementos obtenidas de una población con distribución de probabilidad $f(x)$ y distribución acumulativa $F(x)$ tienen una distribución:

$$g_{(k)}(x) = \frac{N!}{(k-1)!(N-k)!} f(x)[F(x)]^{k-1}[1-F(x)]^{N-k}$$

Estadísticos de orden:

Si tenemos una muestra $x_1, x_2, \cdots, x_k$ extraída de una población con distribución de probabilidad $f(x)$ y distribución acumulativa $F(x)$ y se la ordena de menor a mayor, se obtiene la muestra ordenada:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)}$$

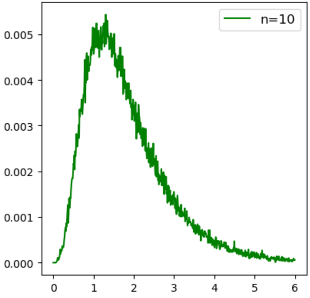
donde:

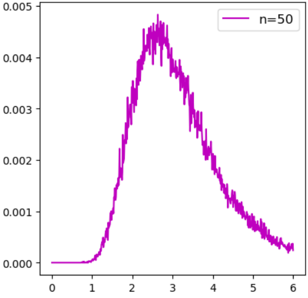
$$x_{(1)} = min(x_1, x_2, \cdots, x_k)$$

$$x_{(k)} = max(x_1, x_2, \cdots, x_k)$$

$$x_{(k/2)} = mediana(x_1, x_2, \cdots, x_k)$$

Ejemplo 5: Utilizando la función de luminosidad de Schechter, $\phi(L) = \left(\frac{\phi^*}{L^*}\right)\left(\frac{L}{L^*}\right)^{\alpha}\exp(-L/L^*)$, con $L^* = 1$, $\phi^* = 1$ y $\alpha = -0{,}5$, generen una población de galaxias para $0 < L < 6$ y extraigan muestras de 10 galaxias. Cómo se distribuye el valor máximo de cada muestra?. Y en el caso de extraer 50 galaxias?.





Estadísticos de orden:

A partir de la distribución:

$$g_{(k)}(x) = \frac{N!}{(k-1)!(N-k)!} f(x)[F(x)]^{k-1}[1-F(x)]^{N-k}$$

haciendo $k = 1$ y $k = N$ se obtienen las distribuciones para el mínimo y el máximo, respectivamente:

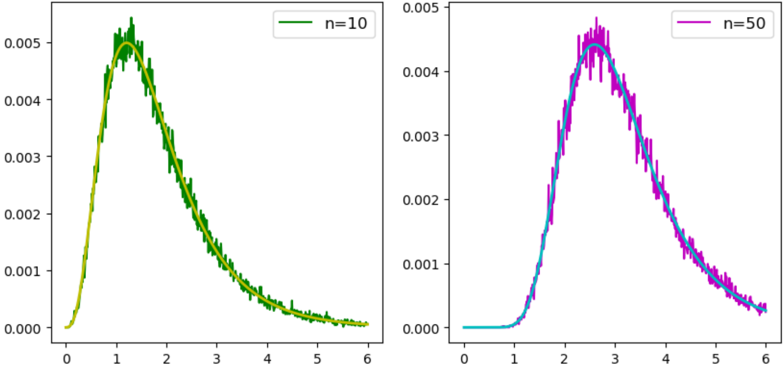
$$g_{(1)}(x) = N\ f(x)\ [1-F(x)]^{N-1}$$

$$g_{(N)}(x) = N\ f(x)\ [F(x)]^{N-1}$$

y haciendo $k = N/2$ se obtiene la distribución para la mediana, siempre que k sea entero.

Estadísticos de orden:

Distribución para el máximo de muestras con $N = 10$ y $N = 50$:



Probabilidad bayesiana:

Otra definición: Probabilidad es una formalización numérica de nuestro grado de creencia sobre un hecho particular.

A partir del Teorema de Bayes se define la **estadística bayesiana**:

$$p(B|A) = \frac{p(A|B) \times p(B)}{p(A)}$$

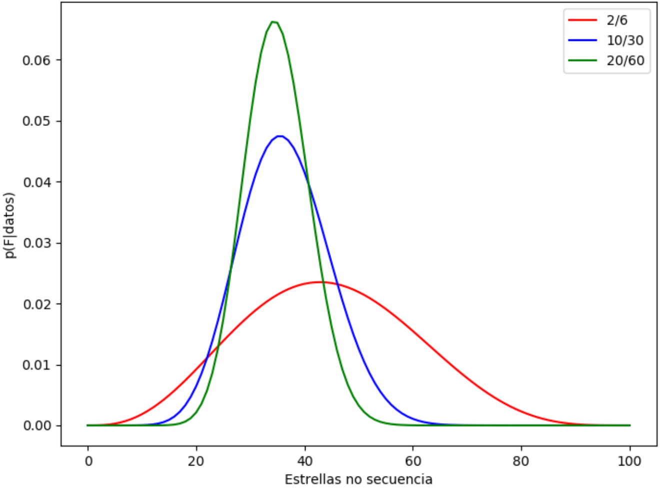
- $p(B)$ se denomina probabilidad a priori.
- $p(A|B)$ se denomina verosimilitud o probabilidad condicional.
- $p(B|A)$ se denomina probabilidad a posteriori.
- $p(A)$ se denomina probabilidad marginal.

Probabilidad bayesiana:

Ejemplo 1a: En un pequeño cúmulo de 100 estrellas hay estrellas de secuencia principal y otras que ya salieron de ella. Si selecciono al azar 6 estrellas y obtengo 2 estrellas que están fuera de la secuencia, cuántas estrellas fuera de la secuencia tengo en el cúmulo?.

- El planteo sería: $p(F|datos) \propto p(datos|F) \times p(F)$, donde los datos son obtener 2 de 6.
- $p(datos|F)$ es la distribución binomial para F y E variables.
- Para $p(F)$ consideramos una distribución uniforme entre 0 y T.

Probabilidad bayesiana:



Probabilidad bayesiana:

Ejemplo 6: Se observa un cierto tipo de objeto con magnitud m y se asume que la medición se distribuye en forma normal alrededor de la magnitud correcta M, con varianza σ^2. La probabilidad de observar un objeto de ese tipo es $p(M) = KM^{2,5}$. Entonces:

- El planteo bayesiano sería:

$$p(M|m) = \frac{p(m|M) \times p(M)}{p(m)}$$

- Como m se distribuye en forma normal tenemos que:

$$p(m|M) = K' \exp\left[-\frac{(m-M)^2}{2\sigma^2}\right]$$

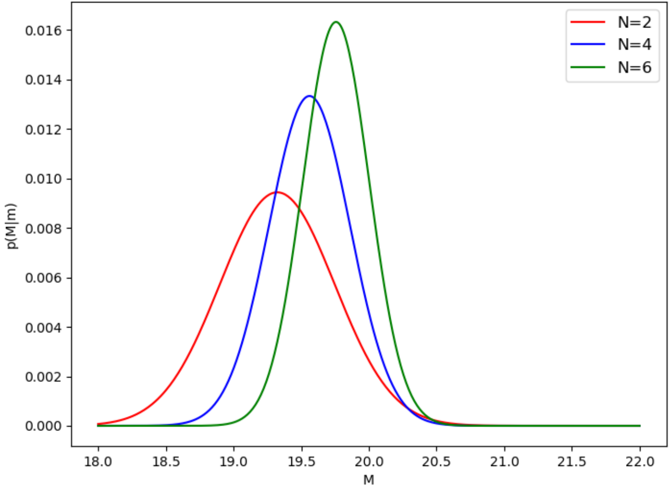
Probabilidad bayesiana:

- La probabilidad a posteriori utilizando varias mediciones será:

$$p(M|m) \propto K'' \exp\left[-\frac{\sum_{i=1}^{N}(m_i - M)^2}{2\sigma^2}\right] M^{2,5}$$

- Por simplicidad, usamos $\sigma^2 = 0{,}6$.
- Los valores observados son $m_i = 20{,}0;\ 18{,}6;\ 19{,}4;\ 20{,}2;$ $19{,}9$ y $20{,}4$.
- Se grafican los dos primeros, luego cuatro y finalmente seis.

Probabilidad bayesiana:



Actividades:

Práctica 6:

- La función de distribución uniforme entre dos valores reales a y b es $f(x; a, b) = (b - a)^{-1}$ para $a \leq x \leq b$. Para esta distribución de probabilidades, cuánto valen μ y σ^2?.
- Un conjunto de 300000 vectores tienen sus tres componentes distribuidos de manera normal con $\sigma^2 = 1$ y $\mu = 0$. Simular la distribución de los módulos de los vectores, calcular cuánto valen μ y σ^2 en este caso y verificar que tienen una distribución Maxwelliana.

Actividades:

Práctica 6 (cont.):

- Repita el último ejercicio de la práctica anterior con muestras de 21 galaxias, pero ahora trabajando con la mediana. Ajuste a los resultados la función de distribución correspondiente.

Entrega

Para la próxima clase

Por consultas: ricardo.gil-hutton@conicet.gov.ar
Grupo de Ciencias Planetarias - CUIM 2