Tema 4. MODELOS DE DISTRIBUCIONES DISCRETOS.

Objetivos

Conceptos:

Conocer los siguientes modelos discretos de probabilidad: uniforme, binomial, geométrico y Poisson. De cada uno de ellos:

* Tipo de experimentos que modelizan
* Función de masa
* Esperanza y varianza
* Propiedades gráficas y de asimetría
* Propiedades de reproductividad (suma de v.a.)

Saber hacer:

Dada una variable **aleatoria**:
* Reconocer el modelo de probabilidad que sigue
* Calcular probabilidades, utilizando las tablas o bien a partir de las fórmulas adecuadas
* Hallar esperanza y varianza.

Dada una variable **estadística**:
* Estudiar si se ajusta a alguno de los modelos de probabilidad estudiados
* Determinar los parámetros correspondientes a dicho modelo

Problemas de exámenes (web):

SIN: Febrero 2006: Problema 2 (a) (b) (c)
Septiembre 2006. Problema 1 (c)(d)
Febrero 2005: Problema 2
Junio 2004: Problema 1

CON: Septiembre 2006: 4 (d) (modelos A y B)
Diciembre 2004: (m) (n)
Junio 2005: (d) (e)
Septiembre 2005: (c)(d)

4.1.- Introducción.

Hay situaciones que siguen modelos de distribución de probabilidad muy similares:

- Resultados al lanzar una moneda, un dado, una ruleta, el sorteo de la ONCE, ...
- Número de caras al lanzar 3 monedas, número de "seises" al lanzar tres dados, ...
- Número de tiradas hasta que sale la primera "cara", número de tiradas hasta que sale el primer "seis", ...

Buscaremos "patrones", modelos que se adapten a situaciones frecuentes. En general, los modelos son **simplificaciones** de la realidad, no se ajustan exactamente a ella, pero nos sirven para poder comprenderla mejor.

4.2.- Distribución uniforme discreta.

Modelo	**Uniforme discreta**	
Tipo de experimento	X: resultado de un experimento en el que **todos los valores** posibles tienen la **misma probabilidad**	
Ejemplos	1) Resultados al lanzar un dado . 2) Resultados al jugar a una ruleta	
Función de masa	$\{x_i \in R \ / \ i = 1,\dots,n\}$; $P(X = x_i) = \frac{1}{n}$	
E(X) = $\sum_{i \in I} x_i \frac{1}{n}$	V(X)= $\sum_{i \in I} x_i^2 \frac{1}{n} - E(X)^2$	CAF=0 (**simétrica**)

Ejemplos:

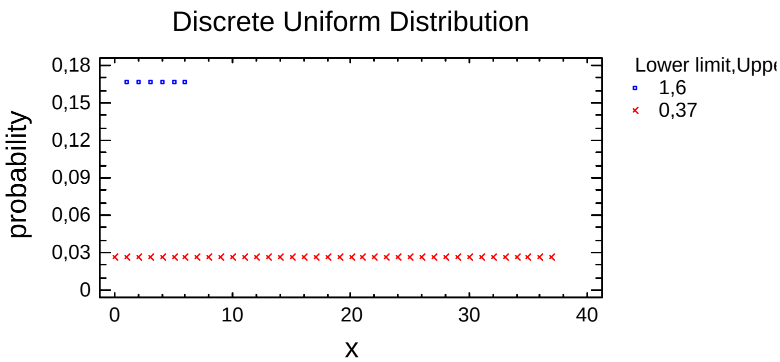
1) Resultados al lanzar un dado

$$P(X = x_i) = \frac{1}{6};\ x_i = 1,2,3,4,5,6\ ;\ \text{E(X)} = \sum_{k=1}^{6} k \cdot \frac{1}{6} = 3.5\ ;\ \text{V(X)} = \sum_{k=1}^{6} k^2 \frac{1}{6} - (3.5)^2 = \frac{35}{12} = 2.91\widehat{6}$$

2) Resultados al jugar a una ruleta

$$P(X = x_i) = \frac{1}{37};\ x_i = 0,1,2,\dots,36\ ;\ \text{E(X)} = \sum_{k=0}^{36} k \cdot \frac{1}{37} = 18\ ;\ \text{V(X)} = \sum_{k=0}^{36} k^2 \frac{1}{37} - (18)^2 = 114$$

Observamos en la representación gráfica, la simetría en los dos casos:



4.3.- Distribución binomial.

Modelo	**Bernouilli B(1,p)**	
Tipo de experimento	X: **número de éxitos** en un experimento tal que: • Sólo hay dos resultados posibles {0,1} • La probabilidad de "éxito" (1) es constante: **p**	
Ejemplos	1) Acertar la respuesta de una pregunta de test contestando al azar Si es de Verdadero o Falso, p=1/2 Si hay 3 alternativas p=1/3 En general, con n alternativas p=1/n. 2) Obtener una pieza correcta o defectuosa p=P(obtener pieza defectuosa)	
Función de masa	$\{0,1\}$; $\boldsymbol{P}(\boldsymbol{X}=0)=1-\boldsymbol{p}$; $\boldsymbol{P}(\boldsymbol{X}=1)=\boldsymbol{p}$	
E(X) = $\boldsymbol{p}$	**V(X)=** $\boldsymbol{p}(1-\boldsymbol{p})$	$CAF=\dfrac{1-2\boldsymbol{p}}{\sqrt{\boldsymbol{p}(1-\boldsymbol{p})}}$; la simetría depende de p: es simétrica si p=0.5.

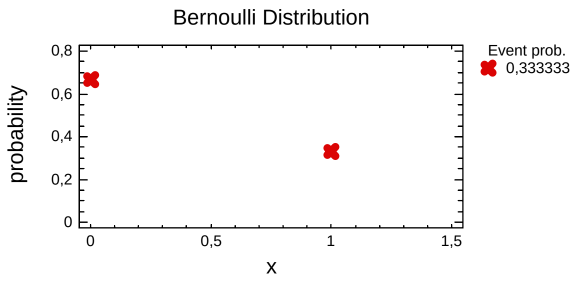
Ejemplo:

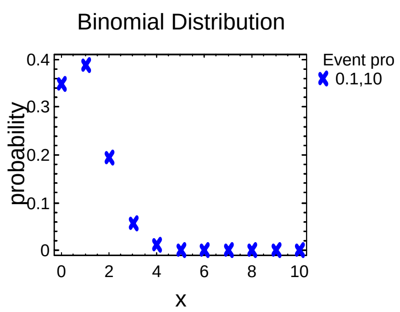
1) Acertar la respuesta de una pregunta de test con 3 alternativas contestando al azar:

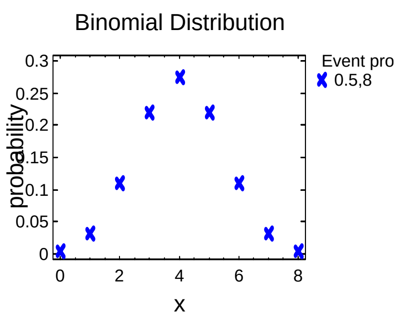
Como p=1/3, E(X)=1/3. V(X)= $\frac{1}{3}(1-\frac{1}{3})=\frac{2}{9}$; $dt=\sqrt{\frac{2}{9}}=0.47$.

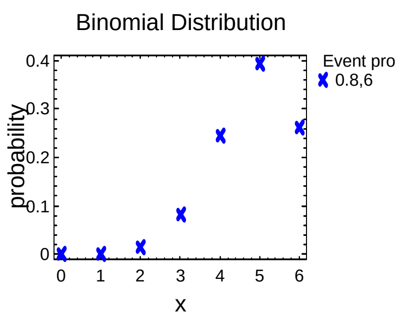
$$CAF=\frac{1-2\frac{1}{3}}{\sqrt{\frac{1}{3}(1-\frac{1}{3})}}=0.7>0$$

(asimetría a la derecha)



<table>
<tr><td>Modelo</td><td colspan="3">Binomial B(n,p)</td></tr>
<tr><td>Tipo de experimento</td><td colspan="3">X: número de “éxitos” en n experimentos de Bernouilli independientes
(p: probabilidad de “éxito”)</td></tr>
<tr><td>Ejemplos</td><td colspan="3">X: Número de respuestas acertadas en un examen de 10 preguntas de test, con 3 alternativas, contestando al azar.
X ~ B(10,1/3) (n=10, p=1/3)
Y: Número de piezas defectuosas en una partida de 20 piezas, sabiendo que la probabilidad de obtener una pieza defectuosa es del 1%.
Y ~ B(20,0.01) (n=20, p=0.01)</td></tr>
<tr><td>Función de masa</td><td colspan="3">$r = 0,1,2,\ldots,n \quad ; \quad P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$</td></tr>
<tr><td colspan="2">E(X) = np</td><td colspan="2">V(X) = $np(1-p)$</td></tr>
<tr><td>Simetría</td><td colspan="3">$CAF = \dfrac{1-2p}{\sqrt{np(1-p)}}$; la simetría depende de p:</td></tr>
<tr><td colspan="2">

p<0.5, CAF>0,
asimetría a la derecha</td><td>

p=0.5, CAF=0,
simétrica</td><td>

p>0.5, CAF<0,
asimetría a la izquierda</td></tr>
<tr><td>Otras propiedades</td><td colspan="3">Reproductividad de la binomial:
Si X e Y son v.a. independientes tales que X ~ B(n,p) , Y ~ B(m,p), entonces X+Y ~ B(n+m,p)</td></tr>
</table>

Ejemplo 1:

X: Número de respuestas acertadas en un examen de 10 preguntas de test, con 3 alternativas, contestando al azar.

Hallar la probabilidad de acertar 7 respuestas y de acertar 5 o más respuestas:

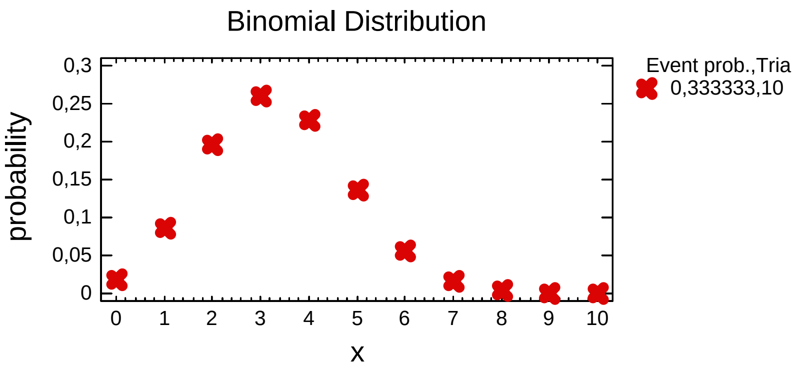
$$P(X = 7) = \binom{10}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^3 = 0.016$$

$$P(X \geq 5) = \binom{10}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^5 + \binom{10}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^4 + \ldots + \binom{10}{9}\left(\frac{1}{3}\right)^9\left(\frac{2}{3}\right)^1 + \left(\frac{1}{3}\right)^{10} = 0.21$$

El número esperado de respuestas acertadas contestando al azar es **E(X)**=10·(1/3)=3.33.

Y la desviación típica: $dt = \sqrt{V(X)} = \sqrt{10 \cdot \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)} = 1.49$.

Como p<0.5, es asimétrica a la derecha:



Ejemplo 2:

Sea *Y:* Número de piezas defectuosas en una partida de 20 piezas, sabiendo que la probabilidad de obtener una pieza defectuosa es del 1%. $Y \sim B(20,0.01)$ (n=20, p=0.01).

Se quiere conocer el número esperado de piezas defectuosas si se reciben 5 partidas de 20 piezas.

Sea Z: Número de piezas defectuosas en 5 partidas de 20 piezas, sabiendo que la probabilidad de obtener una pieza defectuosa es del 1%.

Suponiendo que el número de piezas defectuosas en cada partida son independientes, tenemos que $Z=Y_1+Y_2+\ldots+Y_5$, siendo $Y_i \sim B(20,0.01)$.

Por la propiedad de reproductividad de la binomial:

$Z=Y_1+Y_2+\ldots+Y_5 \sim B(20+20+20+20+20,0.01) \sim B(100,0.01)$.

Por tanto, el valor esperado es E(Z)=100·0.01=1 pieza.

Si se quiere hallar la probabilidad de obtener más de una pieza defectuosa en las 5 partidas:

$$P(Z>1) = 1 - P(Z \leq 1) = 1 - \left(P(Z=0) + P(Z=1)\right) = 1 - \left((0.99)^{100} + \binom{100}{1}(0.01)^1(0.99)^{99}\right) = 1 - 0.74$$

$P(Z>1) = 0.26$.

4.5.- Distribución geométrica.

Modelo	**Geométrica G(p)**	
Tipo de experimento	*X:* **número de experimentos de Bernouilli** independientes realizados **antes del primer "éxito"** (**p**: probabilidad de "éxito")	
Ejemplos	*X:* Número de tiradas de una moneda hasta que obtenemos una cara. *X* ~ G(0.5) *Y:* Número de piezas revisadas hasta que aparece la primera defectuosa (probabilidad de pieza defectuosa, 1%) *Y* ~ G(0.01)	
Función de masa	$r = 0,1,2,\ldots \quad ; \quad P(X = r) = (1-p)^r \, p$	
E(X) = $\dfrac{1-p}{p}$	V(X) = $\dfrac{1-p}{p^2}$	$CAF = \dfrac{2-p}{\sqrt{1-p}} > 0$ asimetría a la derecha.

Ejemplo:

Y: Número de piezas revisadas antes de que aparezca la primera defectuosa. *Y* ~ G(0.01)

Hallar la probabilidad de revisar 100 piezas antes de que aparezca la primera pieza defectuosa y la probabilidad de revisar más de 100 piezas antes de que aparezca la primera pieza defectuosa:

$$P(Y = 100) = (0.99)^{100}(0.01) = 0.0036$$

$$P(Y > 100) = (0.99)^{101}(0.01) + (0.99)^{102}(0.01) + \ldots = (0.01)\sum_{k=101}^{\infty}(0.99)^k = (0.1)\frac{(0.99)^{101}}{1-0.99}$$

$$P(Y > 100) = 0.36$$

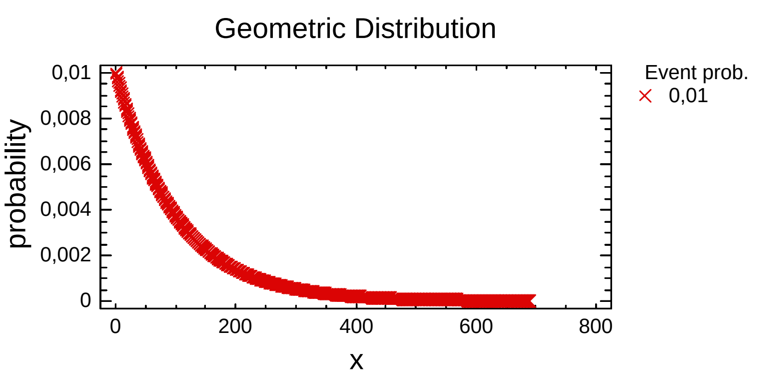
El número medio de piezas revisadas hasta que aparece la primera defectuosa es:

E(Y)= $\dfrac{1-0.01}{0.01} = 99$ piezas.

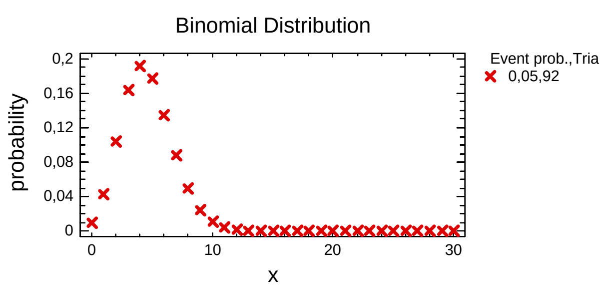
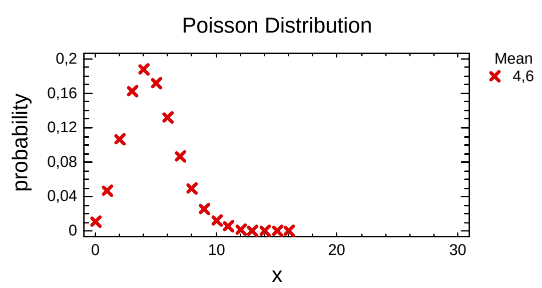
Y su desviación típica:

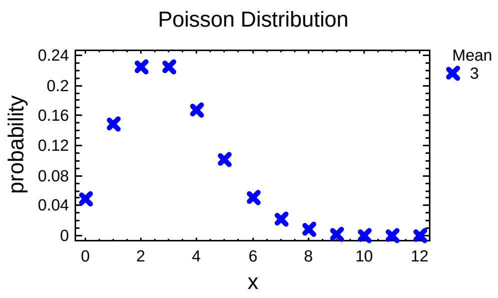
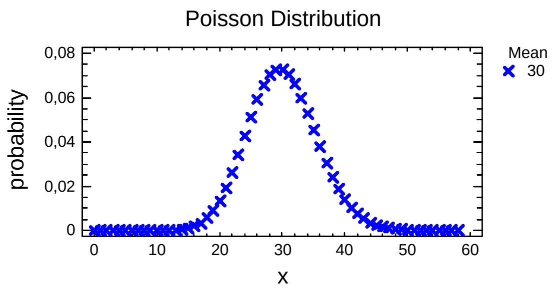
$$dt = \sqrt{V(X)} = \sqrt{\frac{1-0.01}{(0.01)^2}} = 99.5$$

En la gráfica se ve su asimetría a la derecha.



4.4.- Distribución de Poisson.

Modelo	**Poisson P(λ)**
Tipo de experimento	X: **número de "éxitos"** en un **intervalo** [a,b] (tiempo, espacio, ...) en alguna de las siguientes condiciones: • **Proceso de Poisson:** o Los sucesos ocurren de forma independiente (sin "memoria": el número de "exitos" en un intervalo no influye en el número de "éxitos" en el intervalo siguiente) o El **número medio de "éxitos"** (λ) permanece estable • ***n* experimentos de Bernouilli** tales que: o $n \to +\infty$, $p \to 0$ ($n \geq 50$, $p \leq 0.1$) o ***np*** permanece constante ($\lambda = np \leq 10$)
Ejemplos	***X:*** Número de erratas por página de un libro (número medio de erratas: 2) *X ~ P(2)* ***Y:*** Número de visitas a un sitio web en una hora (número medio de visitas: 8) *X* ~ P(8) ***Z:*** Número de días de lluvia en verano Podemos definir: Z_i: llueve o no llueve en el día *i*-ésimo, Z_i ~ B(1,0.05). Entonces, podríamos decir que $Z=Z_1+Z_2+\ldots+Z_{92}$ ~ B(92,0.05) (por la reproductividad de la binomial) Tenemos $n=92 \geq 50$, $p=0.05 \leq 0.1$, E(Z)=np=4.6. Podemos considerar Z ≈ P(4.6)
	
Función de masa	$r = 0,1,2,\ldots$; $P(X=r) = e^{-\lambda}\dfrac{\lambda^r}{r!}$
E(X) = λ	**V(X) =** λ ; $CAF = \dfrac{1}{\sqrt{\lambda}} > 0$: asimetría a la derecha Según λ crece, se hace cada vez más simétrica

	
Otras propiedades	***Reproductividad de la Poisson***: Si X e Y son v.a. independientes tales que: X ~ P(λ) , Y ~ P(μ), entonces X+Y ~ P($\lambda+\mu$).

Ejemplos:

1) ***Y:*** Número de visitas a un sitio web en una hora. *X* ~ P(8)

Hallar la probabilidad de que el sitio sea visitado por 10 personas, y la probabilidad de que lo visiten menos de 8 personas en una hora:

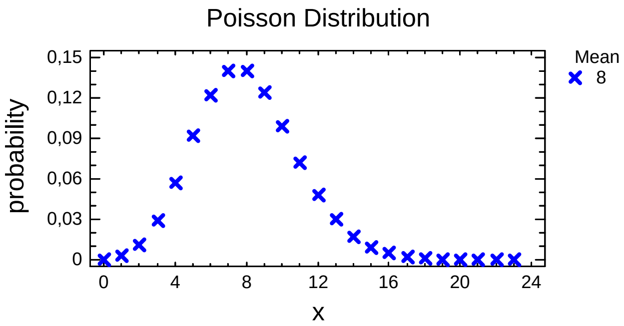
$$P(Y=10)=e^{-8}\frac{8^{10}}{10!}=0.099\approx 0.1 \qquad (\lambda=8)$$

$$P(Y<8)=P(Y=0)+P(Y=1)+\ldots+P(Y=7)=\sum_{k=0}^{7}e^{-8}\frac{8^{k}}{k!}=0.45$$

La esperanza es $\lambda=8$, y la desviación típica $dt=\sqrt{V(X)}=\sqrt{\lambda}=\sqrt{8}=2.8$ visitas.

El coeficiente de simetría

$CAF=\frac{1}{\sqrt{8}}=0.35>0$, que indica una ligera asimetría a la derecha:



2) Dada una partida de 1000 piezas, tal que la probabilidad de ser defectuosa es del 1%,

- Hallar la probabilidad de encontrar exactamente 10 piezas defectuosas
- Hallar la probabilidad de encontrar menos de 5 piezas defectuosas

Si ***Z***: Número de piezas defectuosas en una partida de 1000 piezas, el modelo que sigue dicha variable aleatoria es B(1000, 0.01). Como n=1000>50 y p=0.01<0.1, podemos considerar que Z≈P(10) (np=10).

Por tanto,

$$P(Z=10)\approx e^{-10}\frac{10^{10}}{10!}=0.1251$$

$$P(Z<5)=P(Z=0)+P(Z=1)+\ldots+P(Z=4)\approx\sum_{k=0}^{4}e^{-10}\frac{10^{k}}{k!}=0.0292$$

(Si se hacen los cálculos, considerando Z~B(1000, 0.01), se obtiene:

$$P(Z=10)=\binom{1000}{10}(0.01)^{10}(0.99)^{990}=0.1257\approx 0.125$$

$$P(Z<5)=\sum_{k=0}^{4}\binom{1000}{k}(0.01)^{k}(0.99)^{1000-k}=0.0286\approx 0.029$$

Es recomendable para trabajar con la Poisson, pues para cantidadesmuy grandes de n y muy pequeñas de p, puede haber más errores de redondeos y las cuentas son más farragosas.)

3) Sabemos que el número medio de erratas por página en un libro es 2. Hallar la probabilidad de que un libro de 60 páginas tenga más de 100 erratas.

Sea ***X:*** Número de erratas por página de un libro, de la que sabemos que *X ~ P(2)*

Si el libro tiene 60 páginas de texto, y definimos *L: número de erratas del libro, tendremos*

$$L=X_1+X_2+\ldots+X_{60},$$

y por la **reproductividad de la Poisson**: *L ~ P(2+2+…+2) ~ P(120).*

La probabilidad de que el libro tenga más de 100 erratas, sería:

$$P(Y>100)=1-P(Y\le 100)=1-\sum_{k=0}^{100}e^{-120}\frac{120^{k}}{k!}=0.965\,.$$

Ejercicio: Proponer un modelo de probabilidad para cada una de las siguientes v.a.:

X1: Número de ordenadores de una partida de 20 que se estropean en el periodo de garantía, sabiendo que la probabilidad de que un ordenador esté estropeado es del 2%.
X2: Número de tiradas de un dado hasta que sale el primer 6
X3: Resultado en un bombo del sorteo de la ONCE
X4: Número de llamadas diarias que se hacen por teléfono móvil

Encontrar un ejemplo de experimento para cada modelo de distribución.

Ajuste de una variable estadística a un modelo teórico

Objetivo:

- elegir un modelo
- encontrar los parámetros del modelo

Medios:

- definición de la variable: ¿qué mide? ¿en qué condiciones?
- medidas descriptivas (media, varianza, simetrías, frecuencias, ...)
- representaciones gráficas

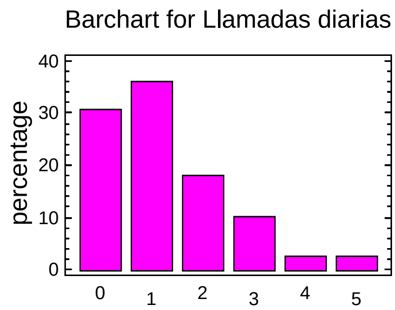
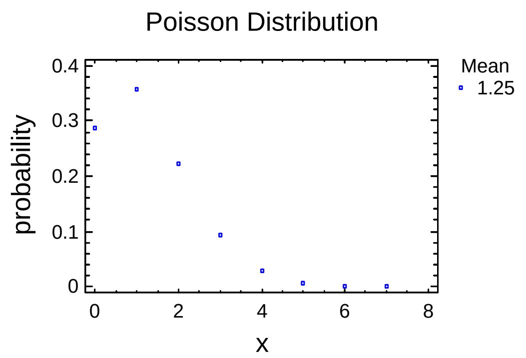
Verificación:

Contrastes no paramétricos con Statgraphics (Distribution Fitting):
p-valor > 0.3 para aceptar la hipótesis
(cuanto mayor sea, con más confianza se acepta el modelo propuesto)

Ejemplo:

X4: Número de llamadas diarias que se hacen por teléfono móvil
(Datos tomados en el curso 2004/05)

- Como la variable mide número de "exitos" (llamadas) en un intervalo de tiempo (un día), eso nos hace pensar en un modelo de Poisson.
- Como E(X)=1.25 y V(X)=1.51, que son valores más o menos similares, sería posible un modelo de Poisson de parámetro λ=1.25 (la media). (Con esos datos, sería menos posible un modelo binomial, pues en dicho modelo la varianza siempre es menor que la media)
- Observamos la representación gráfica del diagrama de barras (asimetría a la derecha) y lo comparamos con la distribución de probabilidad de la P(1.25), y vemos que son similares.

Diagrama de barras **de frecuencias relativas**	**Función de masa** **P(1.25)**
	

Por tanto, la hipótesis que tendríamos que verificar es si X4~P(1.25).

Utilizando la opción de Statgraphics Describe/Distributios/Distribution Fitting, y seleccionando la opción del modelo Poisson, obtenemos:

Goodness-of-Fit Tests for Llamadas diarias :

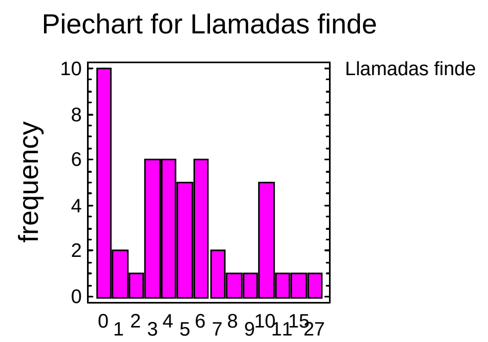
```
Fitted Poisson distribution:
    mean = 1.25641

Chi-Square = 0.555238 with 2 d.f.   P-Value = 0.757586
```

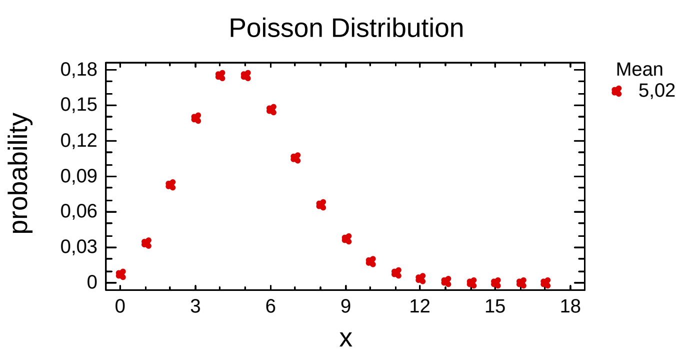
Como p-valor=0.75>0.3, aceptamos que la variable *número de llamadas diarias*, puede tener una distribución de Poisson de parámetro λ=1.25.

Si consideramos la variable Y: Número de llamadas que se hacen por teléfono móvil en un fin de semana, considerando los datos del curso 2006/07, tenemos lo siguiente:

- Como la variable mide número de "exitos" (llamadas) en un intervalo de tiempo (fin de semana), eso nos hace pensar en un modelo de Poisson.
- Como E(X)=5.02 y V(X)=23.5, que son valores muy dispares, esto nos indica que difícilmente sigue un modelo de Poisson.
- La representación gráfica del diagrama de barras y la de la distribución de probabilidad de la P(5), no se parecen.

Por tanto, no es muy probable que siga un modelo de Poisson. Esto se corrobora utilizando la opción de Statgraphics Describe/Distributios/Distribution Fitting, y seleccionando la opción del modelo Poisson, obtenemos P-Value = 0,000567705 < 0.3, por lo que rechazamos la idea de que la variable Y siga un modelo de Poisson.





Ejercicios:

Con los datos recogidos en este curso,estudiar si siguen algún modelo de distribución las siguientes variables (http://www.eui.upm.es/~rafami/Estadistica/Material/material07.html):

- Nº de asignaturas matriculadas
- Nº de asignaturas aprobadas
- Nº de asignaturas aprobadas habiéndose matriculado de 11 asignaturas
- Nº de asignaturas aprobadas habiéndose presentado a 10 asignaturas