PRÀCTICA 3. DISTRIBUCIONS DISCRETES

Objectiu

L'objecte de la present sessió de pràctica informàtica és complementar i refermar els conceptes relatius a les distribucions discretes vistos a classe (apartats 1, 2.1 i 2.2 de la UD4).

Per a això, en primer lloc (apartat 1), es proposen diversos exercicis per a completar o reforçar qüestions que no s'hagen pogut tractar en les classes de teoria i seminari.

El els apartats 2 i 3, que l'alumne ha de fer a casa, es repassen els aspectes bàsics de la distribució binomial i de Poisson. Així mateix, s'hi expliquen les possibilitats que ofereix l'Statgraphics respecte a aquest tipus de variables, per mitjà de la resolució de dos exemples senzills.

NOTA: Es recomana que, durant el treball no presencial de l'alumne, els resultats obtinguts a partir de l'Stagraphics en aquesta pràctica es calculen *a mà* i es confronten amb els primers

1. Exercicis

Exercici 1 Una indústria que utilitza massivament en els seus productes cert component electrònic vol garantir que el percentatge de components defectuosos en cada partida que compra és inferior al 10%. Per a fer-ho prova en cada partida N unitats seleccionades a l'atzar, i accepta la partida només si totes les unitats resulten correctes.

Quant ha de valdre com a mínim N perquè la probabilitat d'admetre una partida amd un 10% o més d'unitats defectuoses no supere el 5%?

RESPOSTA:

X="nº de components defectuoses en la mostra de N unitats"

X~B(N, p≥0,1) situació més difícil de detectar p=0,1.

$P(\text{acceptar}) \leq 0{,}05 \Rightarrow P(X=0) \leq 0{,}05 \Rightarrow \binom{N}{0} 0{,}1^0 (1-0{,}1)^N \leq 0{,}05$

$\Rightarrow \; 0{,}9^N \leq 0{,}05 \Rightarrow N \log 0{,}9 \leq \log 0{,}05 \Rightarrow$

$\Rightarrow \; N \geq (\log 0{,}05)/(\log 0{,}9) \Rightarrow N \geq 28{,}43 \Rightarrow N_{mínim}=29$

Exercici 2 Una empresa dedicada a la comercialització de material informàtic produeix certes unitats utilitzades en el muntatge dels repetidors de xarxa. L'empresa vol garantir que la proporció d'unitats defectuoses en els lots que ven no supera el 7 per mil. Per a fer-ho selecciona de cada lot N unitats a l'atzar i accepta el lot si totes resulten correctes. Determineu el valor mínim de N si es pretén que la probabilitat d'acceptar com a bo un lot amb un 7 per mil o més d'unitats defectuoses no supere el 10%.

RESPOSTA:

X= " Nº d'unitats defectuoses en les N seleccionades"

X $\sim$ *Poisson(λ=Nx0,007)*

P(acceptar)=P(X=0)≤0,1 $\Rightarrow \lambda \geq 2,2 \Rightarrow$

N ≥2,2/0,007=314,28 $\Rightarrow$ ***N mínim=315***

2. Distribucions de probabilitat

En primer lloc hem de triar el tipus de variable aleatòria amb què treballarem i per fer-ho seleccionem **Plot....Probability Distributions....**

Així arribem a una finestra en què ens apareixen les diferents distribucions de probabilitat que es poden manejar amb l'Statgraphics. Si seleccionem una d'aquestes, la distribució binomial, per exemple obtenim per defecte quatre panells.

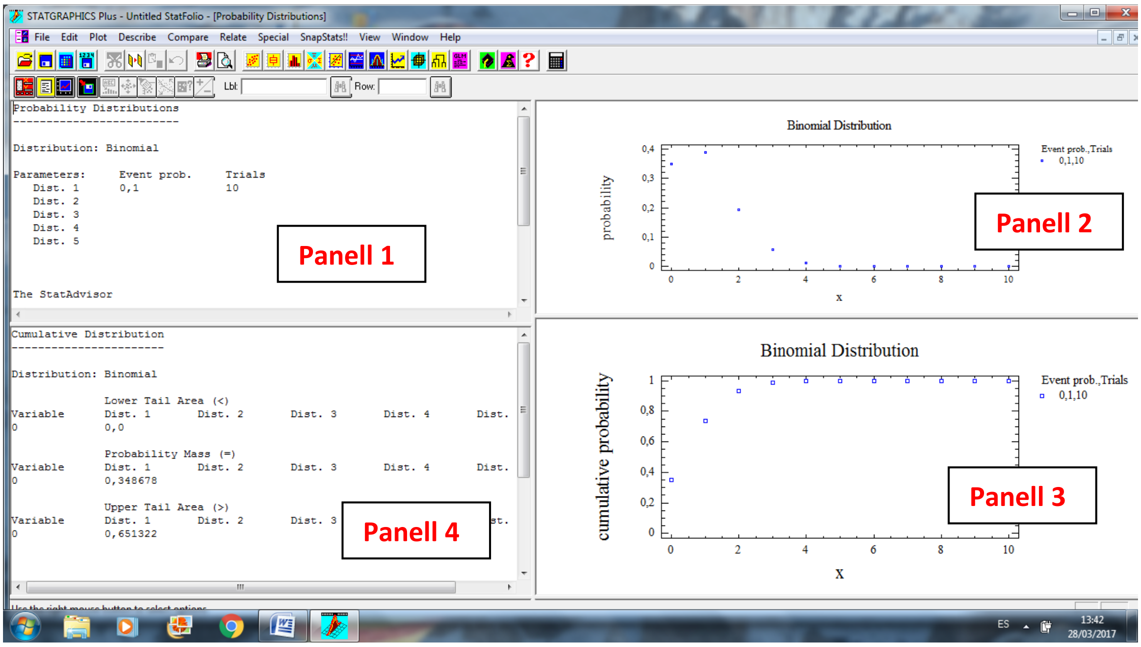

Figura 1. Finestres (panells) amb els resultats de les diferents anàlisis sobre una variable aleatòria amb una determinada distribució de probabilitat

Per mitjà del botó dret del ratolí i fent clic en **Analysis Options** sobre qualsevol dels panells s'especifiquen els **paràmetres** que caracterizen la distribució de probabilitat triada (binomial, Poisson,...).

Es pot especificar fins a 5 distribucions del mateix tipus amd diferente paràmetres simultàniament. Aquesta informació s'arreplega en el **panell 1**.

El **panell 2** mostra **la funció de probabilitat P(X=x) (Mass function)** si la variable aleatòria és discreta, o la **funció de densitat f(x) (Density function)** si la variable aleatòria és contínua.

Anàlogament, el **panell 3** mostra la **funció de probabilitat acumulada F(x)=P(X≤x) (CDF Cumulative Distribution Function)** de la distribució.

En el **panell 4** es pot calcular la probabilitat que una variable aleatòria prenga un valor major, igual o menor que un introduït per nosaltres. Inversament, visualitzant el panell *Inverse CDF* amb la icona **Tabular Options,** es pot obtenir el valor de una variable aleatòria (*critical value*) per al qual hi ha una probabilitat predeterminada que la variable aleatòria prenga com a màxim aquest valor.

3. Càlcul de probabilitats i representacions gràfiques

Una vegada triada la distribució de probabilitat que s'adequa a la pauta de variabilitat de la variable aleatòria en estudi, es poden calcular les probabilitats associades als valors d'aquesta variable aleatòria d'acord amb el model triat (binomial, Poisson, normal,...). També es poden obtindre les representacions gràfiques de la **funció de densitat** o **probabilitat**.

3.1 Exemple 1

Una factoria fabrica disquets de baixa qualitat a baix preu. Siga la variable aleatòria X={nombre de disquets defectuosos en una mostra de grandària n}. Se sap que aquesta factoria produeix un 20% de disquets defectuosos. Es pren una mostra d'11 disquets. Quina és la probabilitat que 3 disquets en la mostra siguen defectuosos? I la probabilitat que en troben en la mostra més de 3 defectuosos?

En primer lloc cal determinar quina és la distribució de probabilitat de la variable aleatòria, així com els corresponents paràmetres. En aquest cas la variable X té distribució binomial de paràmetres n=11 y p=0,2:

$$X \sim B(n=11,\ p=0{,}2)$$

RECORDEU: n i **p** són els **paràmetres de la distribució**. **n** és la grandària de la mostra i **p** la probabilitat de que un disquet siga defectuós, o el que és el mateix la proporció de disquets defectuosos (*tant per u*) en la població.

Per a calcular les probabilitats esmentades seleccionem la distribució pertinent, binomial en aquest exemple, i a continuació proporcionem els paràmetres d'aquesta distribució. Per a fer-ho, seleccionem **Analysis Options** amb el botó dret del ratolí sobre qualsevol dels panells, i llavors apareix la finestra de diàleg corresponent. Em aquesta finestra introduïm els paràmetres **n** (*Trials*) i **p** (*Event Probability*) de la distribució, en l'exemple 11 i 0,2 respectivament.

Automàticament apareixen en els panells de la dreta la funció de probabilitat i la de probabilitat acumulada. Abans de respondre a les preguntes plantejades, vegem quina forma tenen les funcions de probabilitat i distribució per a la variable aleatòria X.

Observant el gràfic (**figura 2**) corresponent a la funció de probabilitat del **panell 2** (gràfica de dalt) es té que la probabilitat que 3 disquets de la mostra siguen defectuosos és aproximadament del 22%, es a dir, P(X=3)=0,22.

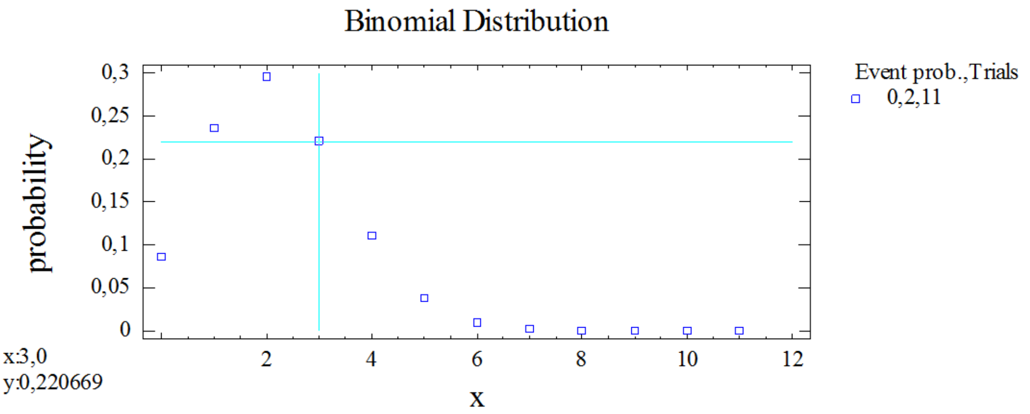

Figura 2. Representació gràfica de la **funció de probabilitat P(X=x)** de la variable aleatòria X~B(n=11, p=0,2)

Igualment, s'obté que la probabilitat que no es trobe cap disquet defectuós (P(X=0)) és, amb caràcter aproximat, 0,086.

A fi de poder calcular aquestes probabilitats amb precisió, hem d'utilitzar el **panell 4**. Per a indicar el valor de la variable aleatòria per al qual volem conèixer la probabilitat (*Random Variable*), seleccionem **Pane Options** amb el botó dret del ratolí sobre el **panell 4**, i llavors apareix una finestra de diàleg, en la que introduïm els valors 0 i 3 del nostre exemple (el valor 0 sempre apareix per defecte en qualsevol distribució).

Els valors obtinguts en el quadre **Cumulative Distribution** de la pàgina següent indiquen:

P(X<0)=0

P(X<3)=P(X≤2)=0,617402

P(X=0)=0,0858993

P(X=3)=0,221459

P(X>0)=P(X≥1)=0,914101

P(X>3)=P(X≥2)= 0,161139

A partir dels valors anteriors es pot calcular, per exemple:

P(X≤3)= P(X<3) + P(X=3) = 0,617402 + 0,221459 = 0,838861

RECORDEU: Per definició P(X≤3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)

Així mateix, l'última probabilitat calculada (P(X≤3)) també es pot obtenir de manera aproximada a partir del gràfic de la funció de distribució (figura 3).

```
Cumulative Distribution
-----------------------

Distribution: Binomial

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,0
3             0,617402

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,0858993
3             0,221459

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,914101
3             0,161139
```

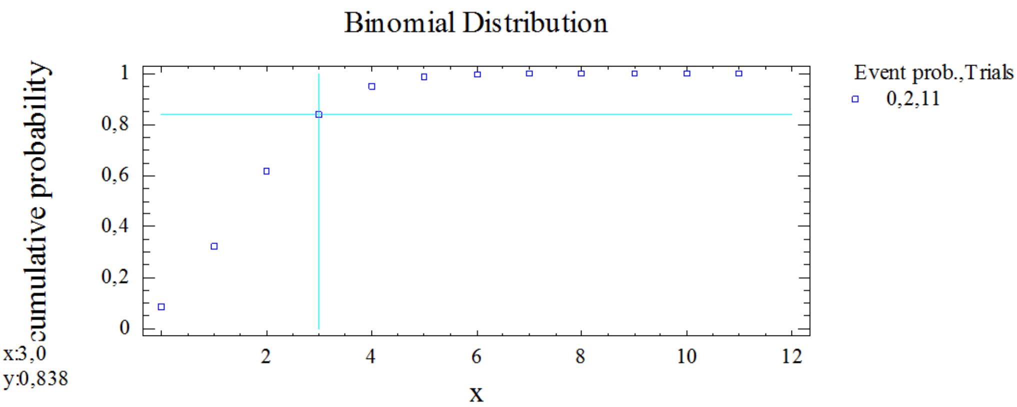

Figura 3. Representació gràfica de P(X≤x) de la variable aleatòria X~B(n=11, p=0,2)

Pregunta 1. Calculeu la probabilitat de trobar 4 disquets defectuosos.

Pregunta 2. Calculeu la probabilitat de trobar almenys 2 disquets defectuosos.

Pregunta 3. Calculeu la probabilitat de trobar 5 disquets defectuosos en una mostra de 20 disquets. Quina seria ara la distribució de la variable aleatòria?

L'Stagraphics permet calcular els valors que pren la variable aleatòria donades les probabilitats.

Per a fer-ho obrim el panell **Inverse CDF** amb la icona **Tabular Options** (com ja s'ha indicat abans) i apareix el quadre que es mostra a continuació:

```
Inverse CDF
-----------

Distribution: Binomial

CDF           Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0,01          0
0,1           1
0,5           2
0,9           4
0,99          6
```

El quadre indica:

Si $P(X \leq x)=0,01 \Rightarrow x=0$

Si $P(X \leq x)=0,1 \Rightarrow x=1$

Si $P(X \leq x)=0,5 \Rightarrow x=2$

Si $P(X \leq x)=0,9 \Rightarrow x=4$

Si $P(X \leq x)=0,99 \Rightarrow x=6$

Per a indicar una probabilitat diferent de la que apareix per defecte en el quadre, seleccionem **Pane Options** amb el botó dret del ratolí perquè ens mostre la finestra de diàleg corresponent.

Per exemple, per a obtenir el valor de x associat a una probabilitat de 0,3, introduïm aquest valor en el camp **CDF**.

Pregunta 4. Seguint amb el primer exemple, si la probabilitat de trobar una quantitat de disquets defectuosos menor o igual que x és 0,95, quin és el valor de x?

IMPORTANT: Tot el que hem vist ara s'utilitza de la mateixa forma per a qualsevol tipus de distribució, siga quin siga el tipus triat. Només canvien, com és obvi, el tipus i valor dels paràmetres que la determinen en cada cas (**Analysis Options**).

3.2. Exemple 2

Els codis CRC utilitzats en l'enviament de paquets a través de la xarxa, són capaços de corregir com a màxim 5 errors per paquet. Se sap que el nombre mitjà d'errors per paquet en un determinat enviament és de 2. Si volem saber quina és la probabilitat que tan sols hi haja un error en un paquet pres a l'atzar, abans que res n'hem de determinar la variable aleatòria, la distribució i els paràmetres:

Variable: X= nombre d'errors en un paquet

Distribució: Poisson

Paràmetre: mitjana λ=2 és el nombre mitjà d'errors en un paquet

$$X \sim \text{Poisson}\ (\lambda=2)$$

Per a calcular la probabilitat que ens demanen, procedim com s'ha explicat en els apartats 1 i 2. Primer triem la distribució corresponent com es mostra en la Figura 4, seleccionant Poisson en la finestra **Probability Distributions**. A continuació introduïm el paràmetre λ amb clic en **Analysis Options** (botó dret del ratolí sobre qualsevol dels panells).

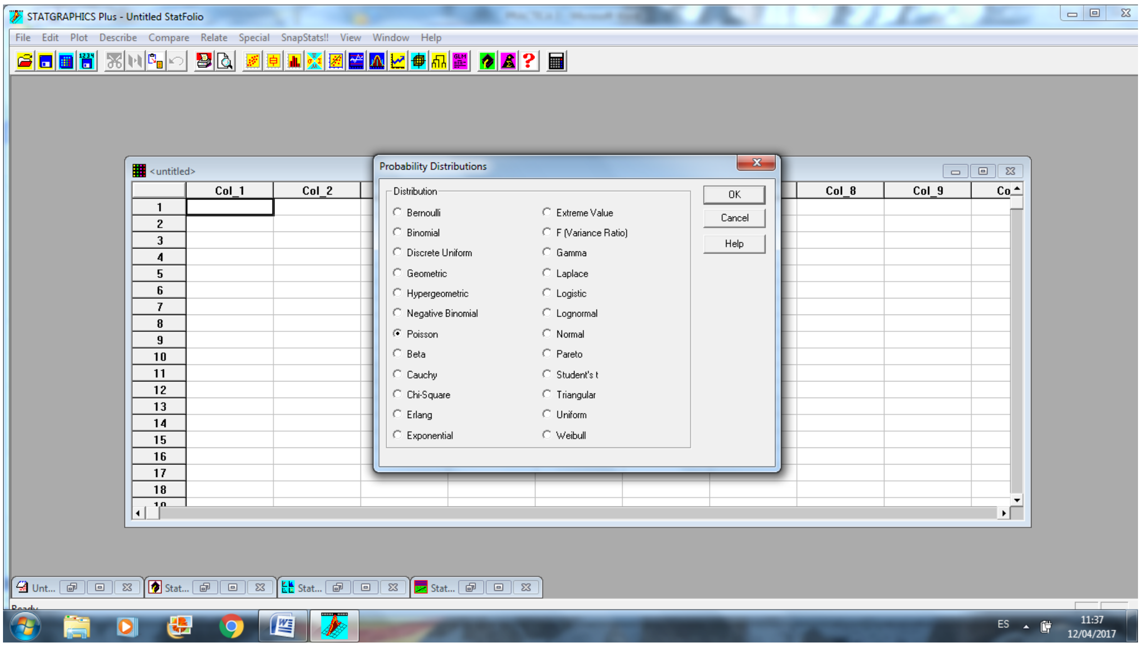

Figura 4. Quadre de selecció del model de distribució de probabilitat

Si a més ens demanen calcular la probabilitat que un paquet haja de ser rebutjat per no poder ser corregit, o siga, P(X>5), llavors introduïm el valor 5 (valor de la variable aleatòria per al què volen conèixer la probabilitat o *Random Variable*) seleccionant **Pane Options** (botó dret del ratolí sobre el **panell 4**, tal com s'ha descrit en l'apartat 2).

Les probabilitats obtingudes són:

La probabilitat que un paquet triat a l'atzar no continga cap error és

P(X=0)= **0,135335**

La probabilitat que un paquet haja de ser rebutjat per no poder ser corregit és

P(X>5)= **0,0165636**

Pregunta 5. Si prenem a l'atzar 10 paquets consecutius, quina és la probabilitat que el nombre total d'errors siga major que 20?

```
Cumulative Distribution
-----------------------

Distribution: Poisson

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0             0,0
5             0,947347

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0             0,135335
5             0,0360894

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0             0,864665
5             0,0165636
```

RESPOSTES A LES PREGUNTES PROPOSADES EN ELS EXEMPLES 1 I 2

```
Cumulative Distribution
-----------------------

Distribution: Binomial

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
2             0,322123
4             0,838861

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
2             0,295279
4             0,11073

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
2             0,382598
4             0,0504092
```

Quadre 1. Probabilitats associades a una variable aleatòria binomial(n=11,p=0,2)

Pregunta 1

Sent X=nombre de disquets defectuosos en una mostra de grandària 11

X~Binomial(n=11, p=0,2)

Com es veu en el quadre P(X=4)= **0,11073**

La probabilitat de trobar 4 disquets defectuosos és 11,073%

Pregunta 2

Sent X=nombre de disquets defectuosos en una mostra de grandària 11

X~Binomial(n=11, p=0,2) (la mateixa variable aleatòria que abans: la mateixa distribució de probabilitat i els mateixos paràmetres)

P(X≥2)= P(X>2) + P(X=2) com s'observa en el quadre 1:

P(X>2)= **0,382598**

P(X=2)= **0,295279**

P(X≥2)= P(X>2) + P(X=2) = **0,382598** + **0,295279** = **0,677877**

La probabilitat de trobar almenys 2 disquets defectuosos es **67,79%**

Pregunta 3

X= nombre de disquets defectuosos en una mostra de grandària 20

Per tant, X~B(n=20, p=0,2)

Canviant n per 20 mitjançant **Analysis Options** en el panell 4 s'obté:

```
Cumulative Distribution
-----------------------

Distribution: Binomial

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
5             0,629648

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
5             0,17456

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
5             0,195792
```

Quadre 2. Probabilitats associades a una variable aleatòria binomial(n=20,p=0,2)

Observant el quadre 2, P(X=5)= **0,17456**

La probabilitat de trobar 5 disquets defectuosos en una mostra de 20 disquets és 17,46%.

Pregunta 4

X=nombre de disquets defectuosos en una mostra de grandària 11

$X \sim B(n=11, p=0,2)$

Per a contestar aquesta pregunta hem d'obrir el panell **Inverse CDF** amb la icona **Tabular Options** i introduir el valor de la probabilitat 0,95 (si no hi apareix per defecte) en el quadre de diàleg que es mostra amb **Pane Options**. Així apareixen els resultats següents:

```
Inverse CDF
-----------

Distribution: Binomial

CDF            Dist. 1        Dist. 2        Dist. 3        Dist. 4
Dist. 5
0,01           0
0,1            1
0,5            2
0,9            4
0,95           5
```

Quadre 3. Valors de la variable aleatòria associats a les probabilitats donades (per exemple, el valor 5 deixa a l'esquerra el 95% dels valors de la distribució) en una distribució binomial(n=11, p=0,2)

Si $P(X \leq x)=0,95 \Rightarrow x=$5 disquets defectuosos.

Pregunta 5

Ara tenim una nova variable aleatòria que es contrueix a partir d'altres variables aleatòries de Poisson (Ps per a abreujar):

Variable aleatòria Z=nombre d'errors en 10 paquets consecutius

$Z=X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}$, sent:

X_i=nombre d'errors al paquet i ~ $Ps(\lambda_i=2)$

Com la suma de variables aleatòries de Poisson també segueix una distribució de Poisson amb λ igual a la suma dels λ de les variables aleatòries que sumen:

$Z \sim Ps(\lambda_Z)$ on $\lambda_Z=10x\lambda_i=10x2=20 \Rightarrow Z \sim Ps(\lambda_Z=20)$

Els resultats donats per l'Statgraphics per a aquesta variable aleatòria es mostren en el quadre 4. Així, la probabilitat demanada és:

$P(Z>20)=$**0,440907**

Per tant, la probabilitat d'agafar a l'atzar 10 paquets consecutius i que el nombre total d'errors siga major que 20 es 44,09%.

```
Cumulative Distribution
-----------------------

Distribution: Poisson

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0             0,0
20            0,470257

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0             2,06115E-9
20            0,0888353

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4
Dist. 5
0             1,0
20            0,440907
```

Quadre 4. Probabilitats associades a una variable aleatòria amb distribució Ps(λ_Z=20)