Traitement de données et modélisation avec

https://elearn.univ-tlemcen.dz/

tewfik.mahdjoub@univ-tlemcen.dz

SCAN ME

SCAN ME

Chapitre 1: Principales lois de probabilités

Section 1: Lois discrètes

1. Loi binomiale.

Une expérience aléatoire **Λ** est une expérience dont il est impossible de prévoir le résultat et si elle est répétée, dans des conditions identiques, elle donnera des résultats différents.

Soit **Λ** une expérience aléatoire. L'ensemble des résultats possibles **Ω** est appelé ensemble fondamental.

On appelle expérience (essai) aléatoire de Bernoulli, une expérience aléatoire telle que :

$$\Omega = \{\text{succès}\} \cup \{\text{échec}\}$$

On appelle schéma de Bernoulli , une répétition de n essais de Bernoulli identiques et indépendants de probabilité ***p***

Exemple.
Dans le jeu de pile ou face (coin-tossing), une pièce de monnaie est jetée 3 fois.

Cette expérience peut être simulée sous R par les instructions suivantes :

Commande 1 :

```
> (jeu = sample (c ("P","F"), 3, replace = T ))
[1] "P" "F" "F"
```

✓ La liste des résultats de l'expérience aléatoire s'appelle « jeu ».

✓ L'instruction «sample() » permet de simuler une expérience aléatoire.

✓ «c( )» est la liste des résultats possibles.

✓ 3 est le nombre des essais réalisés.

✓ L'instruction «replace = T» permet d'obtenir un résultat déjà obtenu.

✓ Les parenthèses extérieures permettent l'affichage direct des résultats stockés dans «jeu».

Remarque :
Cette même expérience simulée plusieurs fois ne donne pas le même résultat :

Expérience #1 :

```
> (jeu = sample (c ("P","F"), 3, replace = T ))
[1] "F" "P" "P"
```

Expérience #2 :

```
> (jeu = sample (c ("P","F"), 3, replace = T
))
[1] "P" "F" "P"

> (jeu = sample (c ("P","F"), 3, replace = T
))
[1] "F" "F" "F"
```

C'est un schéma de Bernoulli à 3 essais.
L'ensemble fondamental est :

Ω = {PPP, PPF, PFF, PFP, FFF, FFP, FPF, FPP}

Le nombre d'éléments de Ω (appelé cardinal de Ω, noté card(Ω)) est défini par :

$$\text{Card }(\Omega)=2^n$$

où n est le nombre d' essais.

Exercice :
Ecrire l'expérience aléatoire qui représente le tirage de 2 boules d'une urne contenant 3 boules : une boule rouge, une boule noire et une boule blanche.

Dans ce cas, « replace =T » correspond à un tirage avec remise, « replace =F » est un tirage sans remise.

Par la suite, dans le cas :
Λ = lancer une pièce de monnaie
"P" représente le succès et "F" l'échec.
On se pose la question de savoir quelle est la probabilité ***p*** d'avoir "P" lors du jet d'une pièce de monnaie.

1.2. Probabilité d'obtenir "P" lors d'un lancer.

A cet effet, l'expérience est répétée 100 fois. Le nombre de "P" est noté sous R "nombre_piles", il est calculé par la commande :

Commande 2 :

```
> (jeu = sample (c ("P","F"), 100,
replace=T))
  [1] "F" "F" "P" "F" "F" "P" "P" "P" "P"
"F" "P" "F" "P" "F" "F" "P" "F" "P" "P"
"P" "F" "F"
 [23] "P" "P" "P" "P" "F" "P" "F" "F"
"P" "P" "P" "P" "P" "F" "P" "P" "P" "F"
"F" "P" "P" "P"
 [45] "P" "P" "P" "P" "F" "F" "F" "P"
"P" "F" "P" "F" "F" "P" "P" "F" "F" "P"
"P" "F" "P" "F"
 [67] "F" "F" "F" "P" "F" "P" "P" "P"
"P" "P" "F" "P" "P" "P" "F" "P" "F" "P"
"F" "P" "P" "P"
 [89] "F" "P" "F" "F" "F" "F" "P" "F"
"P" "F" "F" "F"
> (nombre_piles = sum(jeu == "P"))
[1] 56
```

✓ Les nombres entre [ ] indiquent le rang de la valeur juste à côté.

✓ Pour 100 lancers, on a obtenu 56 "P".

Pour approcher la probabilité d'obtenir "P", il suffit d'augmenter le nombre des essais : 494 "P" sont obtenues pour 1000 essais ; 4938 "P" sont obtenues pour 10^4 essais.

Les probabilités d'avoir "P" sont donc : 56/100, 494/1000, 4938/10000 qui convergent vers la probabilité théorique

$$\boldsymbol{p}=\frac{1}{2}$$

Nous avons donc dans ce cas :

Probabilité (succès) = P (succès)= 0.5
P (échec) = 1- P (succès) = 0.5

Théoriquement, la probabilité du succès P(succès) est définie par la relation :

$$\text{P(succès)} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

1.3. Probabilité d'obtenir *k succès* lors de *n* essais de Bernoulli.

Soit le schéma de Bernoulli qui consiste à lancer une pièce de monnaie 3 fois.

On se pose alors les questions :
Quelle sont les probabilités d'obtenir :
a) 0 "P" b) 1 "P" c)2 "P" d) 3 "P".

On sait déjà que pour l'essai de Bernoulli :

$$\text{P("P")} = \text{P(succès)} = \boldsymbol{p} = \frac{1}{2}$$

Ω = {PPP, PPF, PFF, PFP, FFF, FFP, FPF, FPP}
Card(**Ω**)=8

Si on s'intéresse à la probabilité d'avoir 0 succès, alors le nombre de ces cas est :

0P = **{**FFF**}** Card(**0P**)=1

soit donc : $\text{P(0 succès)} = \frac{1}{8}$
L'ensemble des cas possibles pour avoir un succès est :

1P= **{**PFF, FPF, FFP**}** Card(**1P**)=3

soit donc : $\text{P(1 succès)} = \frac{3}{8}$

Les autres probabilités se calculent de la même façon et on a : $\text{P(2 succès)} = \frac{3}{8}$ et $\text{P(3 succès)} = \frac{1}{8}$.

Remarque :

$$\text{P(0 succès)} + \text{P(1 succès)} + \text{P(2 succès)} + \text{P(3 succès)} = 1$$

De façon générale, la probabilité d'avoir exactement *k* succès lors de *n* essais de Bernoulli (il y aura (*n*-*k*) échecs) est donnée par la formule :

$$\boxed{\mathbf{P}(\boldsymbol{k}\ \textbf{succès}) = \boldsymbol{C_n^k\, p^k\, (1-p)^{n-k}}}$$

où : $C_n^k = \frac{n!}{k!(n-k)!}$

p : probabilité du succès ;

et *n*!=*n*.(*n*-1).(*n*-2)...2.1 (lire factoriel de *n*)

Le calcul du terme C_n^k se fait par l'instruction **choose(*n*,*k*)**.
Ainsi la probabilité d'avoir zéro succès définie par :

$$\text{P(0 succès)} = C_3^0 \left(\frac{1}{2}\right)^0 \left(1-\frac{1}{2}\right)^{3-0}$$

se calcule via la commande :

Commande 3 :

```
> ( proba0success = choose(3,0)  *
((0.5)^0) * (0.5)^3)
[1] 0.125
```

✓ A noter que : 0.125=$\frac{1}{8}$

L'instruction **prod(*n*,*k*)**, avec $k \leq n$, calcule le produit des termes de *n* jusqu'à *k*. Par exemple :

Commande 4 :

```
> (prod(10:9))
[1] 90
```

✓ Pour calculer le factoriel de n, il suffit de calculer 'prod(n:1)'

La probabilité d'avoir un succès est :

Commande 5 :

```
>( proba1success = ( prod(3:1) /
prod(2:1)) *(0.5)*((0.5)^2))
[1] 0.125
```

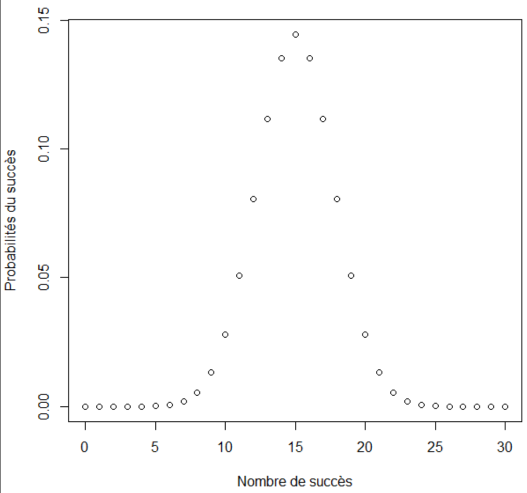
✓ Il est plus simple d'utiliser **choose**.

La distribution du nombre de succès *k* lors de *n* essais est dite une distribution

binomiale. C'est l'exemple d'une distribution **discrète** ($k \in \mathbb{N}$), **finie** ($k \leq n$).
Une représentation graphique de cette distribution est possible :

Commande 6 :

```
> k=0:30
> plot(k, dbinom (k, size=30, prob=0.5), xlab="Nombre de succès", ylab= "Probabilités du succès")
```



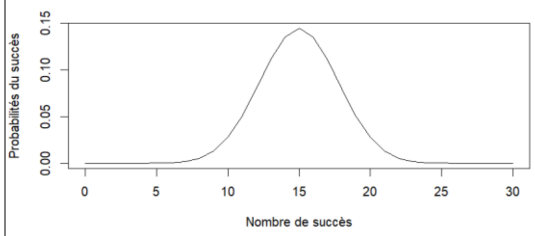
✓ L'instruction **k=0 :30** signifie que k prend les valeurs entières de 0 à 30. k représente le nombre de succès.
✓ dbinom (k, size=30, prob=0.5) donne la probabilité, dans une distribution binomiale, de k succès, lorsque le nombre d'essais de Bernoulli est 30 et la probabilité du succès est 0.5. Par exemple, le lancer d'une pièce de monnaie 30 fois.
✓ Cette fonction est appelée fonction densité de probabilité.

La fonction plot peut être exécutée avec plusieurs types :

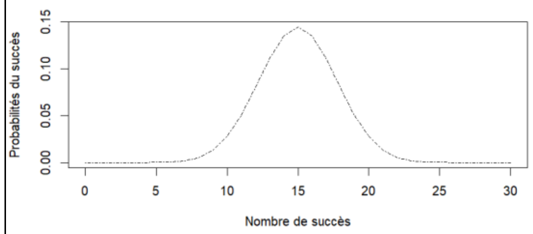
type = "l" (lines)°

Commande 7 :

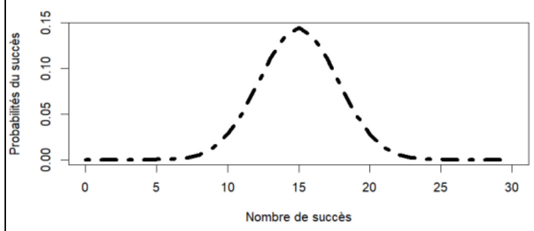
```
> k=0:30
> plot (k, dbinom (k, 30, 0.5), type = "l", xlab="Nombre de succès", ylab= "Probabilités du succès")
```




```
> plot(k, dbinom (k,30,0.5), type = "l", lty = 4, xlab="Nombre de succès", ylab= "Probabilités du succès")
```




```
> plot(k, dbinom (k,30,0.5), type = "l", lty = 4, lwd = 5, xlab="Nombre de succès", ylab= "Probabilités du succès")
```

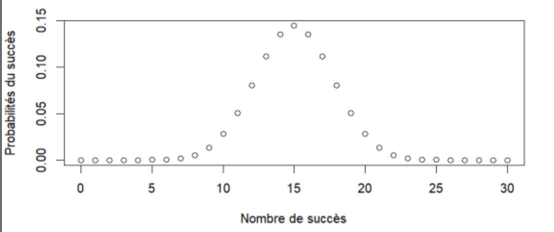


✓ **lty**=type de lignes : valeurs entre 1 et 6
✓ **lwd**= largeur de la ligne

type = "p" (points)°

Commande 8 :

```
> k=0:30
> plot (k, dbinom (k, 30, 0.5), type = "p", xlab="Nombre de succès", ylab= "Probabilités du succès")
```

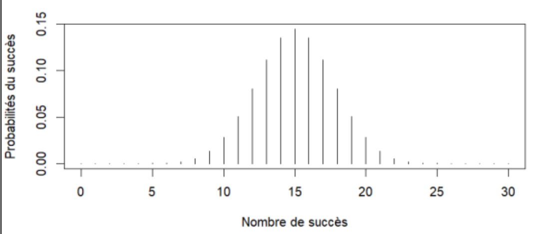


✓ Pour avoir les points et les lignes le type utilisé est type = "b" (both)

type = "h" (histogram)°

Commande 9 :

```
> k=0:30
> plot (k, dbinom (k, 30, 0.5), type = "h", xlab="Nombre de succès", ylab= "Probabilités du succès")
```

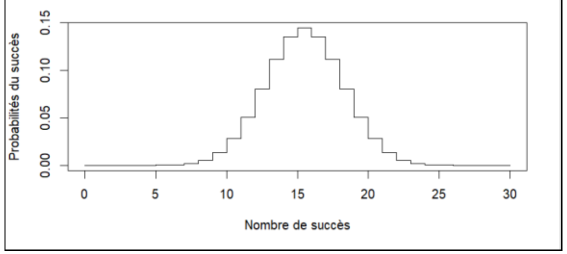


✓ Genre diagramme en bâtons

type = "s" (stairs)°

Commande 10 :

```
> k=0:30
> plot (k, dbinom (k, 30, 0.5), type = "s", xlab="Nombre de succès", ylab= "Probabilités du succès")
```



On demande à un groupe de 30 étudiants de lancer 5 fois une pièce de monnaie. La distribution du nombre théorique exact de "P" obtenu pour 10 lancers parmi les 30 étudiants est :

Commande 11 :

```
> k=0:5
> (eff=dbinom (k ,5 ,0.5) *30)
[1] 0.9375 4.6875 9.3750 9.3750 4.6875 0.9375
> (tab = data.frame (nbre_P=k, effec_the=eff))
  nbre_P  effec_the
1    0     0.9375
2    1     4.6875
3    2     9.3750
4    3     9.3750
5    4     4.6875
6    5     0.9375
```

✓ Théoriquement, il y a 0.9375 étudiants qui ont 0 "P".
✓ C'est le tableau statistique de la distribution théorique du nombre de "P".

La moyenne et la variance de cette série théorique sont :

Commande 12 :

```
> ( moyenne = weighted.mean (tab$nbre_P, tab$effec_the) )
[1] 2.5

> ( variance = weightedVar (tab$nbre_P, tab$effec_the)*29/30 )
[1] 1.25
```

✓ L'ordre dans **weighted.mean** est important : on commence par la colonne des valeurs de la variable puis la colonne des fréquences.
✓ **weightedVar** est une fonction du package **matrixStats.**
✓ Multiplier la variance par le terme correctif $\frac{n-1}{n}$.

Exercice :
Un QCU (questions à choix unique) se compose de 10 questions. Dans chaque

question, il est proposé 4 réponses dont une est correcte.

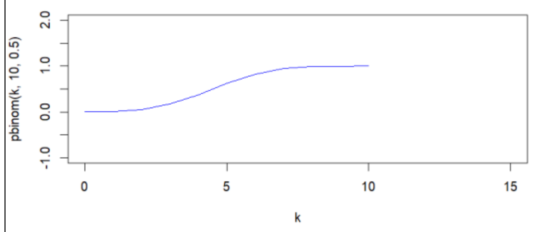
a) Donner la distribution théorique pour 100 étudiants.
b) Montrer que la note moyenne obtenue est : **2.5**.

De façon générale, la moyenne et la variance d'une distribution binomiale sont :

E[x]=n*p
Var[x]=n*p*q

E[x] s'appelle aussi l'espérance mathématique.
La probabilité d'obtenir au moins k "P" pour 10 lancers est :

Commande 13 :
```
> k=0:10
> (pbinom(k, 10, 0.5))
 [1] 0.0009765625  0.0107421875
0.0546875000  0.1718750000
0.3769531250  0.6230468750
0.8281250000
 [8] 0.9453125000  0.9892578125
0.9990234375  1.0000000000
> plot (k , pbinom(k ,10 ,0.5 ), type="l",
col="blue", xlim=c(0,15), ylim=c(-1,2))
```



✓ La probabilité d'obtenir au moins 3 "P" est 0.1718750000. Ceci veut dire qu'on peut obtenir 0 "P", 1 "P", 2 "P" ou 3 "P".
✓ **pbinom(k, n, p)** donne la probabilité cumulée.

Pour générer K schémas de Bernoulli où le nombre de répétitions est n et la probabilité de succès est p, on a la fonction :

Commande 14 :
```
> rbinom ( 5, 10, 0.5)
[1] 5 4 3 6 3
```

✓ L'expérience du lancer de la pièce de monnaie a été répétée 5 fois. On a obtenu 5 "P" la 1ère fois, 4 "P" la 2ème fois etc….

Le nombre de schémas K peut être supérieur an nombre de répétitions n.

Commande 15 :
```
> rbinom ( 15, 10, 0.5)
 [1] 5 7 3 3 4 1 4 4 7 6 8 6 7 5 4
```

2. Loi de Poisson.

Λ: nombre de fois qu'un évènement se produit de manière indépendante pendant un intervalle de temps ou un espace donné.

<u>Exemples :</u>
1. Nombre k de personnes qui arrivent dans un bureau.
2. Nombre k de communications dans un intervalle donné.
3. Nombre k de mutations dans les séquences d'ADN.

La probabilité qu'un évènement, qui suit une loi de Poisson, arrive est :

$$\mathbf{f(k) = \frac{\lambda^k}{k!} e^{-\lambda}}$$

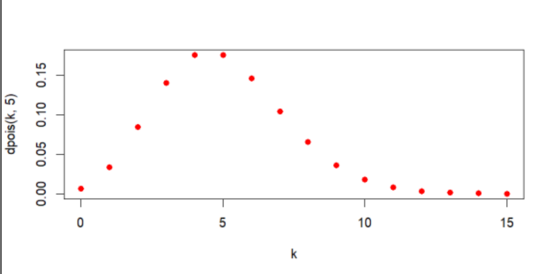
où : k=0, 1,…,+∞
et λ= moyenne que l'évènement arrive pendant une période T ou dans une unité d'espace=> λ>0.
En Algérie, les nouveaux cas de cancer par jour sont de 178 cas. Quelle est la probabilité d'observer 200 cas/jour.

Sous R, cette probabilité se calcule par la commande :

Commande 16 :
> dpois (200, 178) [1] 0.007632979
✓ La probabilité d'avoir 200 cas /jour est 0.007632979.

La loi de Poisson de moyenne λ=5 par exemple peut-être représentée par la commande :

Commande 17 :
> k=0:15 > plot(k, dpois(k,5), col="red", pch=19)



✓ λ > k est possible. ✓ Il existe 25 types de points :