# Distributions de probabilités avec R

Dhafer Malouche

Septembre-Octobre 2011, ESPRIT, 4ième $\infty$

# Introduction

- Les concepts d'expérience aléatoire et de probabilités sont centrales en statistique
- La considération des données comme des réalisations aléatoires et qu'elles proviennent de distributions de probabilités est vital pour comprendre les méthodes statistiques.
- Contenu de ce chapitre :
  1. Tirage aléatoire,
  2. Calcul des probabilités
  3. Distributions de probabilités :
     - Discrètes
     - Continues
  4. L'utilisation de R
  5. Convergence des variables aléatoires.

# Tirage aléatoire.

- Beaucoup d'expériences aléatoires sont décrites avec des jeux de hasard : jeter un dé, une pièce de monnaie, tirer une carte...
- Dans `R`, pour simuler des situations aléatoires on utilise `sample`
  - pour tirer 5 nombres d'une façon aléatoires entre 1 et 40
  - `code R : > sample(1:40,5)`
  - On peut juste sépcifier la longueur de la liste aussi.
  - `code R : > sample(40,5)`

# Tirage aléatoire.

- Par défaut dans `sample` le tirage se fait sans remplacement, si on veut faire un tirage avec remplacement il suffit d'introduire l'argument `replace=T`.
- Ainsi, lancer 10 fois une pièce de monnaie se simule de la façon suivante
  - code R : `> sample(c("H","T"), 10, replace=T)`
  - Dans l'exemple précédent on n'a pas spécifier la probabilité de tirage qui est par défaut égale à 1/2. Si on veut préciser la probabilité de tirage, par exemple avoir un "succés" avec la probabilité 0.9 :
  - code R :`> sample(c("succ", "fail"), 10, replace=T, prob=c(0.9, 0.1))`

# Calcul de probabilités et combinatoire

- La probabilité d'obtenir cinq nombres dans l'ordre compris entre 1 et 40 est $1/(40 \times 39 \times 38 \times 37 \times 36)$, et ce nombre se calcule
  - code R : `> 1/prod(40:36)`
- Si l'ordre de tirage n'intervient pas, la probabilité devient
  - code R : `> prod(5:1)/prod(40:36)`
- Ce nombre est l'inverse du nombre de combinaisons 5 éléments dans un ensemble de 40 éléments :
  - code R : `> 1/choose(40,5)`

# Exercice

On considère un jeu de 52 cartes dont on tire au hasard et simultanément une main de 8 cartes. Calculer

1. Le nombre de mains possibles
2. Les probabilité d'avoir 0,1,2 ou 4 As
3. La probabilité d'avoir 2 As et 3 Trèfles.
4. La probabilité d'obtenir un carré (4 cartes identiques)
5. La probabilité d'obtenir un carré et un berlan (3 cartes identiques).
6. La probabilité d'obtenir 2 As sachant qu'on a obtenu 3 Trèfles.
7. La probabilité d'obtenir 3 Trèfles sachant qu'on a obtenu 2 As.

# Variables aléatoires.

- En probabilité, il y a toujours le besoin d'associer à une expérience aléatoire une application numérique.
- L'application définie de l'ensemble de tous les résultats de l'expérience aléatoire à valeur dans un ensemble numérique est appelée variable aléatoire. Elle est souvent notée par des lettres en majiscule : $X$, $Y$...
- Deux types de variables peuvent être définies :
  - discrètes : l'ensemble des valeurs de $X$ est un sous-ensemble *discret* de $\mathbb{R}$ : $\{0,1\}$, $\{0,1,\dots,n\}$, $\mathbb{N}$
  - continues : l'ensemble des valeurs de $X$ est un sous-ensemble *continue* de $\mathbb{R}$ : $[0,1]$, $\mathbb{R}_+^*$
- Une fois la variable est définie, il faut être capable de calculer les probabilités des valeurs possibles de $X$ : c'est la *loi* ou la *distribution* de probabilité de $X$.

# Variables aléatoires discrètes

Soit $X$ la v.a. représentant le nombre de cartes de même couleur quand on tire d'une façon simultané 8 cartes dans un jeu de 52 cartes.

- Les valeurs possibles de $X$ sont $0, 1, \ldots, 8$
- Calcul de la loi de $X$,
  - code R :
    ```
    x=0:8
    loiX=choose(8,x)*choose(44,8-x)/choose(52,8)
    ```
- La représentation graphique de la loi de $X$,
  - code R :
    ```
    plot(x,loiX,type="h",lwd=2,col=2,xlab="X",
    ylab="Loi de X")
    ```

# Paramètres de variables aléatoires

- La moyenne ou l'espérance de $X$ définie par

$$E(X) = \sum_x x\mathbb{P}(X = x)$$

code R : > `m1X=sum(x*loiX)`

- Le moment d'ordre $k$,

$$E(X^k) = \sum_x x^k\mathbb{P}(X = x)$$

code R : > `m2X=sum(x^2*loiX)`

- La variance de $X$

$$\mathrm{Var}(X) = \sum_x (x - E(x))^2\,\mathbb{P}(X = x) = E(X^2) - E(X)^2$$

code R : > `varX=m2X-m1X^1`

# Fonction de répartition

## Definition

Soit $X$ une variable aléatoire discrète. La fonction de répartition de $X$ est définie par

$$F(x) = \mathbb{P}(X \leq x)$$

- code R :

- `> Fx=cumsum(loiX)`
- `> plot(x,Fx,type='s',xlab='X',ylab='F(x)')`

La fonction $F$ vérifie les propriétés suivantes :

1. $F$ est croissante
2. $\lim_{x \to -\infty} F(x) = 0$ et $\lim_{x \to +\infty} F(x) = 1$
3. pour tout $a$ et $b$, on a $\mathbb{P}(a \leq X \leq b) = F(b) - F(a^-)$

# Lois usuelles

- La loi *Binomiale* : Supposons qu'on répète $n$ fois une expérience qui a seulement deux issues possibles "succés", "échec" et on considère la variable $X$ qui compte le nombre de "succès" obtenus. Si $p$ est la probabilité d'un "succès", on dit que $X$ suit la loi Binomiale de paramètres $n$ et $p$ :

  $$X \sim \text{Binomiale}(n, p)$$

  où $X$ est à valeurs dans $\{0, 1, \ldots, n\}$et pour tout$k \in \{0, 1, \ldots, n\}$

  $$\mathbb{P}(X = k) = \mathtt{C}_n^k p^k (1 - p)^{n-k}$$

- `code R` : `dbinom, pbinom, rbinom.`

# Loi Binomiale sous R

Soit $X \sim \text{Binomiale}(10, 0.3)$

- Calcul de $\mathbb{P}(X \leq 4)$ : code R : `pbinom(4,10,0.3)`
- Calcul de $\mathbb{P}(X = 4)$ : code R : `dbinom(4,10,0.3)`
- Simuler 20 expériences aléatoires permettant l'observation de $X$ : code R : `rbinom(20,10,0.3)`
- Représentation graphique de la loi de $X$ : code R

```
plot(0:10,dbinom(0:10,10,0.3),type='h',col='red',
xlab='X',ylab='loi de X')
```

# Exercice

On cherche à constituer une équipe de basket (5 joueurs de taille $\geq 180cm$). On suppose que $p$ est la proportion des individus ayant une taille $\geq 180cm$ dans une population donnée

1. Si on choisit $n$ personnes dans cette population quelle est la loi de la variable $X$ indiquant le nombre de personnes ayant une taille $\geq 180cm$ parmi les $n$ choisis.
2. On suppose que $p = 0.05$, Quelle est le nombre minimal de personnes personnes à consulter pour être sûr à 99% de choisir parmi eux une équipe de basket.
3. Etudier la relation entre $p$ et le $n$ calculé à la question 2.

# Simulation d'une expérience aléatoire.

Bob a dans sa poche deux boîtes d'allumettes indiscernables : l'une contient $n$ allumettes et l'autre $m$. Il choisit au hasard une des deux boîtes, allume sa cigarette avec une seule allumette, puis remet la boîte dans sa poche si elle n'est pas vide, ou la jette s'il est vide. Soit $X$ la variable aléatoire représentant le nombre de cigarettes allumées avant de jeter une des deux boîtes.

1. Ecrire une fonction R permettant la simulation de l'expérience et les valeurs possibles de $X$
2. Utiliser cet algorithme pour simuler un grand nombre de fois l'expérience.
3. Calculer les fréquences de chaque réalisation.
4. Peut-on retrouver ce résultat théoriquement ?

# La loi géométrique

- On repète plusieurs fois une expérience ayant seulement deux issus possibles. On considère $X$ le rang du premier “succès”. Alors
  - $X$ est à valeurs dans $\{1, 2, \dots, n, \dots\}$
  - Pour tout $k \in \{1, 2, \dots, n, \dots\}$, on a

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p.$$

- $E(X) = \dfrac{1}{p}$ et $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$
- `code R` : `pgeom`, `dgeom`, `rgeom`