

## Probabilités, Statistiques, Combinatoire : TD7

Convergences, loi des grands nombres

### Petits compléments sur les variables gaussiennes

On a vu en cours la définition de la loi normale; on définit de manière un peu plus complète, une famille plus large de lois dites *gaussiennes* :

- si la variable aléatoire  $N$  est normale, alors pour  $a$  et  $b$  deux constantes, la variable aléatoire  $G = a.N + b$  (d'espérance  $b$  et de variance  $a^2$ , d'après les propriétés générales) est dite gaussienne d'espérance  $b$  et de variance  $a^2$ ;
- une loi gaussienne est caractérisée par son espérance et sa variance : deux variables gaussiennes de même espérance et même variance, ont la même loi; et la loi normale n'est rien d'autre que la loi gaussienne d'espérance 0 et de variance 1;
- si  $G$  est gaussienne d'espérance  $m$  et de variance  $\sigma^2$ , alors  $G' = a.G + b$  est également gaussienne, d'espérance  $a.m + b$  et de variance  $a^2\sigma^2$  (comme on peut le voir en posant  $N = (G - m)/\sigma$ , qui est alors normale, et en exprimant  $G'$  en fonction de  $N$ );
- on a la propriété suivante, qui se démontre par le calcul : si  $G$  et  $G'$  sont deux variables gaussiennes, indépendantes, d'espérances respectives  $m$  et  $m'$  et de variances respectives  $\sigma^2$  et  $\sigma'^2$ , alors  $G + G'$  est également gaussienne, d'espérance  $m + m'$  et de variance  $\sigma^2 + \sigma'^2$ ;
- enfin, la loi normale est “symétrique” : si  $N$  est une variable normale, alors  $N$  et  $-N$  ont la même loi (pour une gaussienne  $G$  d'espérance  $m$  et de variance  $\sigma^2$ , c'est plutôt  $G$  et  $2m - G$  qui ont même loi).

### Exercice 7.1

#### Un problème de sondages

On se place ici dans la situation de l'institut de sondages qui cherche à suivre l'évolution, au cours du temps, du taux de popularité d'une proposition ou d'une idée.

On suppose qu'à un moment donné  $t$ , la proportion des électeurs qui sont favorables est  $p$ ; et qu'à un autre moment  $t'$  (une ou deux semaines plus tard par exemple), cette proportion est de  $p'$  ( $p$  et  $p'$  sont inconnus, bien sûr). On aimerait savoir si  $p'$  est ou non supérieur à  $p$ . La méthode est la suivante : au temps  $t$ , on interroge  $n$  personnes prises au hasard pour savoir leur opinion; et, au temps  $t'$ , on recommence la même expérience. On suppose que les réponses données reflètent la réalité (ce qui peut être irréaliste pour les sondages réels). On suppose aussi que les personnes interrogées sont choisies uniformément, et indépendamment (ce qui signifie qu'une même personne peut être interrogée plusieurs fois; mais la probabilité que cela se produise est faible si la taille de la population est très importante par rapport à  $n$ ). On suppose également que les choix de personnes interrogées aux temps  $t$  et  $t'$  sont indépendants.

On peut penser que, si  $p' > p$ , la proportion de réponses “oui” au temps  $t'$  a de bonnes chances d'être supérieures à celle au temps  $t$ ; mais cela ne peut pas être une certitude.

L'objectif de cet exercice est d'arriver à une méthode pour déterminer une valeur de  $n$  qui assure une certaine fiabilité à ce critère.

1. On note  $X$  le nombre de personnes interrogées qui répondent "oui" à la question au temps  $t$ , et  $X'$  le nombre de personnes interrogées qui répondent "oui" au temps  $t'$ .  
Quelles lois peut-on supposer pour les variables aléatoires  $X$  et  $X'$ ? Peut-on considérer que  $X$  et  $X'$  sont indépendantes?
2. Soient  $A$  et  $B$  deux variables aléatoires indépendantes; on suppose que  $A$  suit la loi binomiale de paramètres  $n$  et  $p$ , et que  $B$  suit la loi binomiale de paramètres  $n$  et  $p'$ , avec  $p' > p$ .  
Donner l'espérance et la variance de  $A$ ,  $B$ , et  $B - A$ .
3. Donner un argument simple permettant d'affirmer que, si  $n$  tend vers  $+\infty$ , la probabilité que  $B - A$  soit négatif (on considère que  $A$  et  $B$  dépendent de  $n$ ; on peut les noter  $A_n$  et  $B_n$ ) tend vers 0.
4. En utilisant l'inégalité de Tchebycheff sur  $B - A$ , donner, en fonction de  $n$ ,  $p$  et  $p'$ , une majoration de la probabilité de l'événement  $\{A \geq B\}$  (l'objectif est d'obtenir une majoration par une valeur la plus petite possible).
5. Supposons que l'on ait  $p = 1/5 = 0.2$  et  $p' = 0.25 = 1/4$ . D'après vos résultats de la question précédente, à partir de quelle valeur de  $n$  est-on assuré que la probabilité de l'événement  $\{A \geq B\}$  est inférieure à  $0.05 = 1/20$ ?
6. Si on applique l'approximation gaussienne suggérée par le théorème central limite à  $A$  et  $B$ , par quelles lois gaussiennes devrait-on approximer  $A$  et  $B$ ?  $B - A$ ? (pour  $B - A$  on utilisera le fait que  $A$  et  $B$  sont indépendantes, et les propriétés données en préambule)
7. Supposons de nouveau que l'on ait  $p = 0.2$  et  $p' = 0.25$ . En utilisant l'approximation gaussienne, estimer à partir de quelle valeur de  $n$  on a  $\mathbb{P}(A \geq B) \leq 0.05$  (On pourra utiliser la valeur numérique suivante : la probabilité qu'une variable normale prenne une valeur inférieure à  $-1.96$  est très proche de 0.05).

## Exercice 7.2

### Convergence en loi, en probabilités

On considère une suite de variables aléatoires  $(X_n)_{n \geq 1}$ , indépendantes, et de même loi (supposée être une loi discrète).

1. Soit  $x$  une valeur possible pour la variable aléatoire  $X_1$ , *i.e.*  $\mathbb{P}(X_1 = x) > 0$ . Montrer qu'avec probabilité 1, l'ensemble des  $n$  tels que  $X_n = x$  est infini.
2. On suppose que les  $X_n$  ne sont pas constantes (il existe au moins deux valeurs  $a$  et  $b$  telles qu'on ait  $\mathbb{P}(X_1 = a) > 0$  et  $\mathbb{P}(X_1 = b) > 0$ ). Montrer que la suite  $(X_n)_{n \geq 1}$  ne converge pas en probabilités.
3. Qu'en est-il de la convergence en loi de la séquence  $(X_n)_{n \geq 1}$ ?