Probabilités et Statistiques

Chapitre 3 : Variables aléatoires discrètes

Une distribution statistique associe à une valeur (ou un intervalle de valeurs) d'une variable statistique sa fréquence d'apparition à partir de résultats observés. On veut introduire une notion analogue mais probabiliste : la fréquence d'apparition du résultat observé est remplacée par la "probabilité" d'apparition du résultat attendu. On va donc introduire des "variables probabilistes" (on dit plutôt variables aléatoires) avec leur "distribution de probabilités" (on dit plutôt loi de probabilité).

1 Variables aléatoires et lois de probabilité.

Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé. Une variable aléatoire réelle sur cet espace est une application qui à chaque élément de Ω (c'est-à-dire à chaque résultat d'une expérience aléatoire) associe une donnée numérique réelle $X : \Omega \longrightarrow \mathbb{R}$. On demande seulement que l'image inverse d'un intervalle par X soit un évènement, afin de pouvoir en évaluer la probabilité par p. La variable aléatoire est dite *discrète* si elle ne peut prendre ses valeurs que dans un ensemble discret (une suite de valeurs isolées $\{x_1, x_2, ...x_i, ...\}$). Elle est dite *continue* lorsqu'elle peut prendre ses valeurs dans tout un intervalle de $\mathbb{R}$.

Exemple 1. (discret) On reprend le jet de deux dés. On pose : $X(\{i, j\}) = i + j$. Sur la même expérience on aurait pu considérer d'autres variables aléatoires (produit, différence des valeurs des dés etc...)

Exemple 2. (continu) Le test des trois ampoules électriques. Si on s'intéresse à la durée de vie moyenne, la variable aléatoire naturelle est ici : $X(x, y, z) = \frac{1}{3}(x + y + z)$.

Loi de probabilité d'une variable aléatoire réelle X. Dans le cas discret, il s'agit de la donnée des $p(X = i)$, expression qui désigne la probabilité de l'évènement formé de tous les résultats ω sur lesquels X prend la valeur i :

$$p(X = i) := p(\{\omega \in \Omega, X(\omega) = i\}).$$

De même, pour le cas continu, $p(i \leq X \leq j)$ est la probabilité de l'évènement dont les éléments prennent par X une valeur entre i et j :

$$p(i \leq X \leq j) = p(\{\omega \in \Omega, X(\omega) \in [i, j]\}).$$

Exercice 1. *Sur l'exemple des deux dés muni de la variable aléatoire X : "somme des deux dés", calculer (en précisant l'évènement correspondant) :*
$p(X = 4)$ *;*
$p(X < 5)$ *;*

Fonction de répartition de X. C'est la fonction définie par :

$$F(a) = p(X < a).$$

Elle est croissante de 0 à 1, et permet de trouver la probabilité de n'importe quel évènement par :

$$p(X \geq a) = 1 - F(a), \qquad p(a \leq X < b) = F(b) - F(a).$$

Pour une variable aléatoire discrète, c'est une fonction en escalier qui prend la valeur $p_1 + \cdots + p_k$ sur l'intervalle $]x_k, x_{k+1}]$.

> ***Exercice 2.***
> *- Démontrer les deux formules ci-dessus.*
> *- Sur l'exemple des dés, que vaut $p(X \geq 5)$?*

2 Les principales lois de probabilités discrètes.

Loi de Bernouilli. On l'utilise dès qu'une situation est analogue au modèle suivant : on considère une urne avec des boules blanches en proportion p et noires en proportion $q = 1-p$. On tire une boule. On pose : $X = 1$ si la boule est blanche, $X = 0$ sinon. La loi de probabilité de cette variable aléatoire est :

$$p(X = 1) = p, \; p(X = 0) = q.$$

Loi binomiale. Le modèle correspondant est le suivant : dans la même urne, on tire n boules avec remise. X est le nombre de boules blanches obtenues. On va montrer que pour tout k entre 0 et n,

$$p(X = k) = C_n^k p^k q^{n-k}.$$

En effet, on peut voir X comme une somme $X_1 + \cdots + X_n$ de n variables de Bernouilli. On a alors

$$\begin{array}{rcl} p(X = k) & = & p(X_1 + \cdots + X_n = k) \\ & = & p(X_1 = 1 \cdots \text{ et } X_k = 1 \text{ et } X_{k+1} = 0, \cdots \text{ et } X_n = 0) \\ & + & \cdots \\ & + & p(X_1 = 0 \cdots \text{ et } X_{n-k} = 0 \text{ et } X_{n-k+1} = 1, \cdots \text{ et } X_n = 1) \end{array}$$

les lignes intermédiaires correspondant à tous les changements d'ordre possibles pour la répartition des valeurs 1 et 0. Il y en a C_n^k. En considérant que ces évènements sont indépendants, on a donc

$$p(X = k) = p^k q^{n-k} + \cdots p^k q^{n-k} = C_n^k p^k q^{n-k}.$$

Remarque importante. Si le nombre de boules dans l'urne est très grand et le nombre de tirages petit, on peut considérer que le fait de remettre ou non les boules change peu le résultat. **On pourra alors utiliser la loi binomiale même s'il n'y a pas remise.**

Loi hypergéométrique. C'est le même modèle que pour la loi binomiale, mais on fait les n tirages sans remise. X est encore le nombre de boules blanches obtenues. Soit a le nombre de boules blanches dans l'urne, b le nombre de boules noires, $a+b$ le nombre total. On a (en raisonnant sur le nombre de cas favorables / nombres de cas possibles) :

$$p(X=k) = \frac{C_a^k C_b^{n-k}}{C_{a+b}^n}.$$

Remarque. Contrairement à la loi binomiale, l'utilisation de cette loi nécessite la connaissance du nombre total $a+b$ d'éléments dans l'urne.

Loi de Pascal, (ou loi géométrique quand $r=1$.) Dans la même urne, on tire maintenant des boules avec remise jusqu'à obtenir r boules blanches (r fixé). X = le nombre de tirages nécessaires pour obtenir r boules blanches. On a, pour tout $k \geq r$:

$$p(X=k) = C_{k-1}^{r-1} p^r q^{k-r}.$$

Loi de Poisson. On l'utilise dans une situation consistant à compter le nombre d'évènements (appels téléphoniques, passage d'une voiture, accidents,...) pendant une durée donnée. La variable X est le nombre d'évènements apparus pendant cette durée. On a pour tout $k \geq 0$:

$$p(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

où λ est un paramètre positif fixé.

Exercice 3.
1- Une machine fabrique 2% de pièces défectueuses. On prélève un échantillon de 50 pièces. Repérer la loi de probabilité qui modélise cette situation, puis déterminer la probabilité qu'il contienne :
- deux pièces défectueuses ;
- une ;
- aucune.

2- On tire 10 cartes dans un jeu de 52 cartes. Repérer la loi de probabilité qui modélise cette situation, puis déterminer la probabilité d'obtenir 3 cœurs.

3- Le nombre d'accidents suit une loi de Poisson de paramètre 2. Quelle est la probabilité que dans une journée donnée il y ait :
- aucun accident ?
- 2 accidents ?
- 4 accidents ?

3 Paramètres d'une variable aléatoire discrète.

Ils sont analogues aux paramètres de position ou dispersion d'une distribution statistique, en remplaçant les fréquences f_i par les probabilités $p_i = p(X = x_i)$:

Espérance mathématique de X (ou "moyenne probabiliste"). On fait la moyenne des valeurs x_i, $i \in I$ en les pondérant par leur probabilité d'apparition $p_i = p(X = x_i)$:

$$E(X) = \sum_{i \in I} p_i x_i.$$

Ecart-type de X. Posons $\overline{x} = E(X)$. On mesure la dispersion des valeurs de X autour de $\overline{x}$ par le paramètre :

$$\sigma = \sqrt{\sum_{i \in I} p_i (x_i - \overline{x})^2} = \sqrt{(\sum_{i \in I} p_i x_i^2) - \overline{x}^2}.$$

Son carré est aussi appelé variance de X.
La variable centrée est $X' = X - \overline{x}$. Sa moyenne est nulle.
La variable centrée réduite est $Y = \frac{X - \overline{x}}{\sigma}$. Sa moyenne est nulle et son écart-type vaut 1.

Exercice 4. *Calculer l'espérance et la variance de la loi de Bernouilli.*

Plus généralement, les paramètres des lois usuelles sont :

	Espérance	Variance
Loi de Bernouilli	p	pq
Loi binomiale	np	npq
Loi hypergéométrique	np	$npq\frac{a+b-n}{a+b-1}$
Loi de Pascal	$\frac{r}{p}$	$\frac{rq}{p^2}$
Loi de Poisson	λ	λ

Remarque. Dans le cas d'une loi de Poisson, la connaissance de $E(X)$ détermine le paramètre λ. Par exemple dans l'exercice 3, question (3-), si on sait que le nombre moyen d'accidents par jour est de 2, le paramètre à utiliser est $\lambda = 2$.

4 Couple de variables aléatoires discrètes.

On a étudié en statistique descriptive des distributions à deux dimensions et leurs relations ou indépendance. On va reprendre ces notions pour des distributions de probabilités.

On considère une expérience aléatoire décrite dans un espace probabilisé $(\Omega, \mathcal{T}, p)$ pour laquelle, à chaque résultat ω de l'expérience, on associe deux nombres $X(\omega)$ et $Y(\omega)$. Si le couple de variables prend des valeurs discrètes, sa loi de probabilité est la donnée de toutes les probabilités (la virgule doit être lue comme un "et") :

$$p_{ij} = p(X = i, Y = j).$$

La fonction de répartition du couple est

$$F(i, j) = p(X < i, Y < j).$$

Les lois marginales sont obtenues en sommant sur toutes les valeurs d'une des deux variables. Elles permettent de calculer $p(X = x_i) = p(X = x_i$ et "Y quelconque") :

$$p_{i\cdot} = \sum_j p_{ij}, \quad p_{\cdot j} = \sum_i p_{ij}.$$

Paramètres d'un couple de variables aléatoires discrètes. Il y a d'abord les paramètres (espérances et écart-types) des lois marginales :

$$\begin{aligned} E(X) &= \sum_i p_{i.} x_i, \quad E(Y) = \sum_j p_{.j} y_j. \\ \sigma(X) &= \sqrt{\sum_i p_{i.}(x_i - \overline{x})^2}, \quad \sigma(Y) = \sqrt{\sum_j p_{.j}(y_j - \overline{y})^2} \end{aligned}$$

où $\overline{x} = E(X)$ et $\overline{y} = E(Y)$. On mesure la liaison ou l'indépendance des deux variables par la covariance :

$$cov(X, Y) = \sum_{i,j} p_{ij}(x_i - \overline{x})(y_j - \overline{y})$$

Le coefficient de corrélation linéaire est

$$r(X, Y) = \frac{cov(X, Y)}{\sigma(X)\sigma(Y)}.$$

On a les propriétés (rappelons que la variance est le carré de l'écart-type : $var(X) = \sigma^2(X)$) :

1. $cov(X, Y) = E(XY) - E(X)E(Y)$ (formule de Koenigs)
2. $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$.
3. $-1 \leq r(X, Y) \leq +1$.

Indépendance de deux variables aléatoires à valeurs discrètes. X et Y sont *indépendantes* lorsque la loi de probabilité du couple vérifie, pour tout (i, j),

$$p_{ij} = p_{i\cdot} \times p_{\cdot j}.$$

En effet, lorq'on a cette condition, la probabilité conditionnelle $p(X = x_i / Y = y_j)$ est indépendante de la condition $Y = y_j$.

Critères d'indépendance. Ils se démontrent de manière analogue aux distributions statistiques (voir ch 1) :

– Si X et Y sont indépendantes alors $cov(X,Y) = r(X,Y) = 0$.
– Si X et Y sont linéairement liées ($Y = aX + b$), alors $r(X,Y) = \pm 1$.

Exercice 5. *On jette un dé, puis un autre autant de fois que nécessaire pour obtenir un nombre supérieur ou égal à celui du premier dé. Soit X la valeur du premier jet, et Y celle du dernier.*
- Quel est l'ensemble des valeurs prises par le couple (X,Y) ?
- Déterminez les lois de probabilité conditionnelle $p(Y = j/X = i)$ pour chaque valeur de i.
- En déduire la loi de probabilité du couple (X,Y).
- Les variables sont-elles indépendantes ?

Test sur le chapitre 3

1. Décrivez la loi de Bernouilli.
   Détaillez le calcul de son espérance et de son écart-type.
2. Que valent l'espérance et l'écart-type d'une loi binomiale ?
3. Quand doit-on utiliser la loi hypergéométrique plutôt que la loi binomiale ?
4. Quelle est la définition d'un couple de variables aléatoires discrètes indépendantes ?

Chapitre 3 : Travaux dirigés

1. (Approximation d'une loi binomiale par une loi de Poisson). Une petite entreprise emploie 20 personnes. Une étude statistique permet d'admettre qu'un jour donné la probabilité qu'un employé donné soit absent est de 0,05. On admet que les absences des employés un jour donné sont des évènements indépendants. On note X la variable aléatoire qui à chaque jour choisi au hasard associe le nombre d'employés absents.
   (a) Expliquer pourquoi X suit une loi binomiale et donner les paramètres n et p de cette loi.
   (b) Calculer la probabilité (à 10^{-3} près) des évènements suivants :
   - E_1 : un jour donné il y a exactement trois absents ;
   - E_2 : un jour donné il y a strictement plus de deux absents ;
   - E_3 : un jour donné le nombre d'absents est compris entre trois et six (bornes comprises).
   (c) Calculer l'espérance mathématique $E(X)$. Que représente-t-elle ?
   (d) On approche cette loi binomiale par une loi de Poisson de paramètre $\lambda = np$, où n et p sont les paramètres de la loi binomiale. Déterminer les probabilités des évènements E_1, E_2 et E_3 avec cette nouvelle loi, et comparer avec les résultats obtenus à la question (b).

2. Combien de pièces de monnaie doit-on jeter pour que la probabilité d'avoir au moins une fois pile dépasse 99% ?
   (Indication : considérer la variable aléatoire X_n= nombre de piles obtenus sur n lancers).

3. Un joueur tire au hasard une carte dans un jeu de 52 cartes. Il gagne 20 euros si c'est un cœur, et perd 10 euros sinon.
   (a) Quelle est son espérance de gain algébrique (les pertes sont considérées comme des gains négatifs) ?
   (b) Commment modifier ces valeurs pour que le jeu devienne équitable ?

4. Deux frères s'associent pour fonder une P.M.E.. Au bout d'une année, celle-ci emploie 10 personnes : les 2 patrons, 3 employés et 5 ouvriers. Un journaliste s'intéressant à cette P.M.E. décide d'en interviewer trois personnes au hasard. Soit X le nombre de patrons et Y le nombre d'ouvriers parmi ces 3 personnes interrogées.
   (a) Déterminer les lois de X et de Y ; calculer leurs espérances et écarts-types.
   (b) Déterminer (directement) la loi du couple (X, Y), et retrouver les lois de X et Y.
   (c) Calculer la covariance de (X, Y). Ces variables aléatoires sont-elles indépendantes ?

Travail personnel :

On suppose que le nombre de personnes se présentant à une station-service pendant une période de 15 minutes est une variable aléatoire X dont la loi de probabilité est donnée par :

x_i	0	1	2	3	4
$p(X = x_i)$	0,1	0,2	0,4	0,2	0,1

De plus, la probabilité pour qu'une personne se présentant à la station prenne du gazole est de 0,4. Soit Y la variable aléatoire : "nombre de personnes ayant demandé du gazole pendant la période de 15 minutes".

1. Pour chaque valeur x_i de X, préciser les valeurs que peut prendre Y et donner la loi de probabilité conditionnelle de Y sachant que $X = x_i$.
2. En déduire la loi du couple (X, Y) et la loi de Y. Les deux variables sont-elles indépendantes ?
3. Quelle est la probabilité qu'il y ait au moins deux personnes qui se présentent à la station et qu'aucune ne prenne du gazole ?
4. Quelle est la probabilité qu'il y ait une personne demandant du gazole sachant qu'au moins deux se sont présentées ?