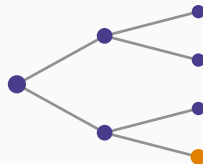


Probabilités

Chapitre 1 : Du dé aux probabilités



Prérequis

Le point de départ

Le dé, vu par le statisticien

Le dé, vu par le probabiliste

La passerelle

Ce qu'il faut retenir

Prérequis

Le cours qui précède

Statistique descriptive, dont ce cours reprend la moyenne, la variance et la covariance pour les transposer au hasard.

Les notions mobilisées

La moyenne pondérée et la variance d'une série. La lecture d'un tableau croisé. Du côté mathématique : les sommes, et pour les lois continues, l'intégrale simple.

Repris depuis le début

Le dénombrement, l'espace probabilisé, la variable aléatoire et les lois usuelles sont construits ici.

Le parallèle avec la statistique descriptive est explicite : l'espérance est une moyenne, la variance une variance. Ce qui change n'est pas la formule, c'est ce sur quoi elle porte — une population imaginée plutôt qu'un échantillon observé.

Le point de départ

Une bonne nouvelle, à prendre au sérieux

Les probabilités passent pour un domaine difficile. Elles ne le sont pas : vous en connaissez déjà l'essentiel, sous un autre nom.

Tout ce que vous avez appris en statistique descriptive — effectifs, fréquences, moyenne, variance, cumuls, tableaux croisés — se transpose ici **mot pour mot**. Un seul mot change de place. Le reste suit.

Le programme de ce chapitre

Prendre un objet que tout le monde connaît, un **dé**, et montrer qu'il peut se lire de deux façons : celle du statisticien qui observe, celle du probabiliste qui modélise.

Puis montrer que ces deux lectures produisent **les mêmes formules**.

Les deux moitiés d'une même discipline

La **statistique descriptive** part de données déjà là et les résume : elle décrit *ce qui s'est produit*.

Le **calcul des probabilités** part d'un modèle du hasard et en déduit ce qui devrait arriver : il décrit *ce qui devrait se produire*.

Pourquoi un gestionnaire a besoin des deux

Une banque qui constate qu'elle a subi 8 défauts l'an dernier fait de la statistique descriptive. Une banque qui doit **provisionner aujourd'hui** pour les défauts de l'an prochain fait du calcul des probabilités.

Le premier regard sert à rendre compte ; le second sert à décider. Aucune décision ne porte sur le passé.

Le dé, vu par le statisticien

Un dé est une variable statistique comme une autre

L'expérience

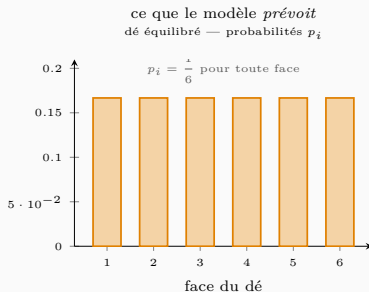
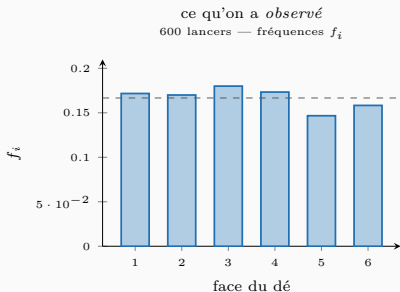
On lance un dé 600 fois et on relève la face obtenue. On dispose alors d'une série de 600 observations, exactement comme on disposerait de 600 salaires.

La **variable statistique** est « la face obtenue » ; ses **modalités** sont 1, 2, 3, 4, 5, 6.

Le tableau statistique

Face							
x_i	1	2	3	4	5	6	Total
Effectif	98	101	104	103	89	105	600
n_i							
Fréquence	0,163	0,168	0,173	0,172	0,148	0,175	1
f_i							

Ce que donne le diagramme



les deux diagrammes ont la même forme et la même somme, 1.
La probabilité est la fréquence que l'on obtiendrait si l'on lançait le dé une infinité de fois.

Deux écritures de la même moyenne

$$\bar{x} = \frac{1}{n} \sum_i n_i x_i \quad \text{ou, en divisant par } n \text{ d'abord,}$$

$$\boxed{\bar{x} = \sum_i f_i x_i}$$

La seconde écriture est celle qui compte

Elle dit : **on multiplie chaque modalité par sa fréquence, et on additionne.**

C'est une *moyenne pondérée*, où le poids de chaque valeur est la part qu'elle occupe dans les données.
Retenez cette phrase : tout le chapitre en dépend.

La moyenne observée

$$\bar{x} = 1(0,163) + 2(0,168) + 3(0,173) + 4(0,172) + 5(0,148) + 6(0,175)$$

$$\bar{x} = 0,163 + 0,336 + 0,519 + 0,688 + 0,740 + 1,050 = 3,496 \approx 3,5$$

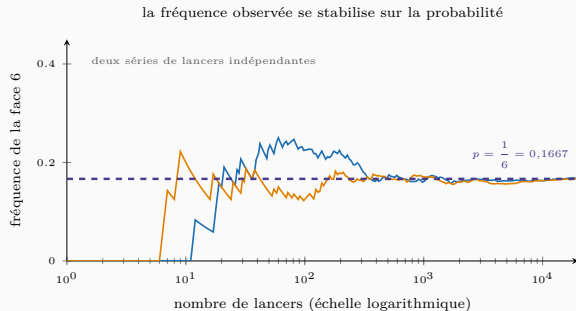
Une valeur qui n'est pas une face

3,5 n'est pas un résultat possible : aucune face ne porte ce nombre. Une moyenne n'a pas à être une valeur observable.

Ce détail, banal ici, sera exactement celui qui déroute les étudiants quand on l'appellera *espérance*.

Le dé, vu par le probabiliste

Ce que l'on observe quand on lance longtemps



Au début, tout bouge : 10 lancers ne disent rien.

Après quelques milliers, les deux séries se rejoignent sur la même valeur.

C'est cette valeur limite qu'on appelle la probabilité.

Ce fait a un nom et une démonstration : la **loi des grands nombres**, chapitre 20.

La fréquence se stabilise

Sur dix lancers, la fréquence de la face 6 ne vaut rien : elle saute de 0 à 0,3 sans raison.

Sur dix mille, elle s'est immobilisée autour d'un nombre — et **deux séries indépendantes s'immobilisent sur le même.**

La probabilité, définie par la fréquence

La **probabilité** d'un résultat est la valeur limite vers laquelle tend sa fréquence quand on répète l'expérience indéfiniment :

$$p_i = \lim_{n \rightarrow \infty} f_i$$

L'approche par symétrie, dite de Laplace

Si les issues sont **également probables**, la probabilité d'un événement est le rapport

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

Sur le dé

Un dé équilibré n'a aucune raison de privilégier une face : $p_i = \frac{1}{6} \approx 0,167$ pour chacune des six.

C'est exactement la valeur autour de laquelle la figure précédente s'était stabilisée.

Laplace propose, l'expérience dispose

La symétrie **fournit un modèle** sans avoir à lancer quoi que ce soit; elle suppose l'équiprobabilité, qui est une hypothèse et non un fait.

La fréquence **vérifie ce modèle**; elle ne suppose rien, mais exige de très nombreuses répétitions.

En gestion, la symétrie n'existe pas

Aucune raison de croire *a priori* qu'un crédit fera défaut avec telle probabilité plutôt que telle autre. Il n'y a pas de faces à compter.

La probabilité de défaut se lit donc dans les **fréquences historiques** : sur dix mille crédits du même profil, quatre cents ont fait défaut, d'où $p = 0,04$. C'est le chiffre que nous suivrons dans tout le cours.

La passerelle

Une seule idée : remplacer la fréquence par la probabilité, et toutes les formules de la statistique descriptive deviennent celles du calcul des probabilités.

	Statistique descriptive	Probabilités
Ce qu'on a L'objet	600 lancers <i>observés</i> modalité x_i	un dé <i>équilibré</i> valeur x_i de la v.a. X
Le poids	fréquence $f_i = \frac{n_i}{n}$	probabilité $p_i = P(X = x_i)$
Somme des poids	$\sum f_i = 1$	$\sum p_i = 1$
Centre	$\bar{x} = \sum x_i f_i$	$E(X) = \sum x_i p_i$
Dispersion	$V(x) = \sum x_i^2 f_i - \bar{x}^2$	$V(X) = \sum x_i^2 p_i - E(X)^2$
Cumul	$F_i = \sum_{j < i} f_j$	$F(x) = P(X \leq x)$
Sur le dé	$\bar{x} = 3,428$	$E(X) = 3,5$

La colonne de gauche décrit *ce qui s'est produit*. Celle de droite décrit *ce qui devrait se produire*.

Une seule substitution

Partout où la statistique descriptive écrit une **fréquence** f_i , le calcul des probabilités écrit une **probabilité** p_i .

$$\bar{x} = \sum_i f_i x_i \quad \Rightarrow \quad \boxed{E(X) = \sum_i p_i x_i}$$

L'espérance mathématique n'est rien d'autre

$E(X)$ est la moyenne arithmétique dans laquelle on a remplacé ce que l'on a observé par ce que l'on prévoit.

Même formule, même calcul, même interprétation. Seule la source des poids change.

L'espérance du dé

$$E(X) = \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{6}(3) + \frac{1}{6}(4) + \frac{1}{6}(5) + \frac{1}{6}(6) = \frac{21}{6} = 3,5$$

Comparez les deux résultats

Moyenne observée sur 600 lancers : $\bar{x} = 3,496$. Espérance du modèle : $E(X) = 3,5$.

L'écart de 0,004 n'est pas une erreur : c'est le hasard d'échantillonnage. Il se réduira encore si l'on lance davantage — et le chapitre 20 démontrera qu'il tend vers zéro.

Ce que la substitution emporte avec elle

Toutes les formules suivent

Statistique descriptive

$$\sum_i f_i = 1$$

$$\bar{x} = \sum_i f_i x_i$$

$$V(x) = \sum_i f_i x_i^2 - \bar{x}^2$$

$$F_i = \sum_{j \leq i} f_j$$

tableau de contingence

$$\text{cov}(x, y)$$

Probabilités

$$\sum_i p_i = 1$$

$$E(X) = \sum_i p_i x_i$$

$$V(X) = \sum_i p_i x_i^2 - [E(X)]^2$$

$$F(x) = P(X \leq x)$$

loi jointe

$$\text{cov}(X, Y)$$

Il n'y a donc pas de formules nouvelles à apprendre

Il y a des formules **connues, appliquées à des poids nouveaux**.

Ce qui est nouveau dans ce cours n'est pas le calcul : c'est la question de savoir **d'où viennent les p_i** .
Les répondre occupe les chapitres 2 à 8; les exploiter occupe le reste.

Ce qu'il faut retenir

Le résumé du chapitre

Cinq points

1. La statistique descriptive décrit **ce qui s'est produit**; les probabilités décrivent **ce qui devrait se produire**.
2. Une probabilité est la **limite d'une fréquence** quand on répète l'expérience indéfiniment.
3. Quand les issues sont symétriques, elle se calcule aussi comme le rapport *cas favorables sur cas possibles*.
4. Remplacer f_i par p_i dans la moyenne arithmétique donne l'**espérance mathématique**
$$E(X) = \sum_i p_i x_i.$$
5. Cette substitution vaut pour **toutes** les formules du cours précédent : le calcul des probabilités est l'autre face de la statistique descriptive.

La suite

Une seule difficulté reste : obtenir les p_i quand l'expérience est plus riche qu'un dé — trois tirages, cinq candidats, deux cents crédits.

Il faut alors **compter les cas possibles**, et les compter sans les énumérer. C'est le dénombrement, et c'est l'objet des chapitres 3 à 5. Le chapitre 2 pose d'abord le vocabulaire.