Expériences répétées, échantillonnage

Dans tout le chapitre, E désigne l'ensemble des issues d'une expérience aléatoire et P désigne une loi de probabilité sur E.

I – Loi Binomiale :

1. Exemple :

Un escroc utilise une pièce truquée pour faire des paris. La probabilité d'obtenir pile est P(P)=0,6. Faire un arbre résumant la situation pour une série de 5 lancés, et obtenir la probabilité d'avoir 3 « pile », puis la probabilité d'avoir plus de pile que de face.

2. Définitions :

a) Si n est un entier naturel et si k est un entier compris entre 0 et n, on note $\binom{n}{k}$ et on lit « k parmi n « le nombre de chemins qui réalisent exactement k succès dans l'arbre à n étapes n'ayant que 2 issues. Ce sont les coefficients binomiaux.

b) La loi binomiale de paramètre n et p notée $\mathrm{B}(n;p)$ est la loi de la variable aléatoire X qui mesure le nombre de succès dans la répétition de façon identique et indépendante de n épreuves n'ayant que deux issues avec une probabilité de succès p.

Pour cette variable aléatoire X, on a $\mathrm{P}(\mathrm{X}=k)=\binom{n}{k}\times p^k\times(1-p)^{(n-k)}$

3. Propriétés :

a) L'espérance d'une variable aléatoire X suivant une loi binomiale $\mathrm{B}(n;p)$ est $\mathrm{E}(\mathrm{X})=np$. L'espérance peut être interprétée comme tendance à long terme de la valeur moyenne .

Exemple : reprendre l'exemple du 1.

b) La variance d'une variable aléatoire X suivant une loi binomiale $\mathrm{B}(n;p)$ est $\mathrm{V}(\mathrm{X})=np(1-p)$ et son écart type est $\sigma(\mathrm{X})=\sqrt{np(1-p)}$

Ce sont des indicateurs de répartition autour de la moyenne. Nous verrons des exemples en exercices.

c) Représentation graphique de l'exemple sur géogébra et sur tableur.

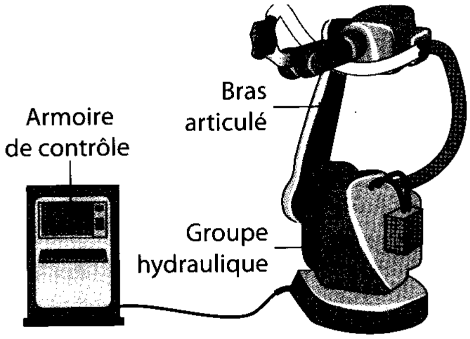
d) Coefficients binomiaux et triangle de Pascal :

$$si\,0\leq k\leq n,\ \binom{n}{k}=\binom{n}{n-k} \qquad \text{et} \qquad si\,1\leq k\leq n-1,\ \binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1}$$

(explication à partir d'un arbre, triangle de Pascal et formule du binôme)

4. utilisation de la calculatrice :

Défaillances des robots de peinture



Les ateliers de peinture d'un grand constructeur automobile fonctionnent à l'aide de robots permettant de positionner les pistolets autour de la carrosserie.

Les pannes mécaniques sur le bras articulé de ces robots, assez fréquentes, sont souvent sans gravité. Elles sont généralement dues à l'encrassage par la peinture, à du jeu ou à un blocage dans les articulations mécaniques.

Les ateliers de peinture comptent 100 robots équipés de bras articulés identiques dont les pannes mécaniques surviennent de façon indépendante. Pour chaque bras articulé, la probabilité, qu'une semaine choisie au hasard, ce type de panne se produise est 0,05.

Une semaine étant choisie au hasard, on réalise, pour chacun des 100 robots, la même expérience aléatoire consistant à observer si son bras connaît une défaillance mécanique.

On désigne par X la variable aléatoire qui, à toute semaine, associe le nombre de robots dont le bras a connu une panne mécanique.

1. Justifier que la variable aléatoire X suit une loi binomiale, dont on déterminera les paramètres.
2. Quelle est l'espérance de la variable aléatoire X ? Interpréter le résultat.
3. La plupart des modèles de calculatrices permettent d'obtenir, pour tout entier k compris entre 0 et 100, les probabilités $P(X = k)$ et $P(X \leq k)$. Utiliser une telle calculatrice pour déterminer les valeurs approchées arrondies à 10^{-3} des probabilités suivantes :

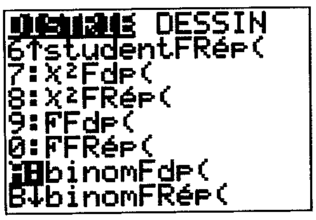
a. $P(X = 2)$;

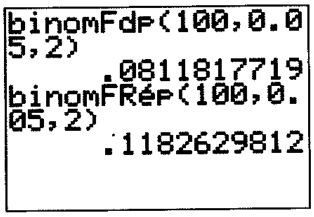
b. $P(X \leq 2)$;

c. $P(E)$ où E est l'événement : « en une semaine, strictement plus de 10 robots ont connu une panne mécanique sur leur bras articulé ».

Procédure à suivre sur une calculatrice TI

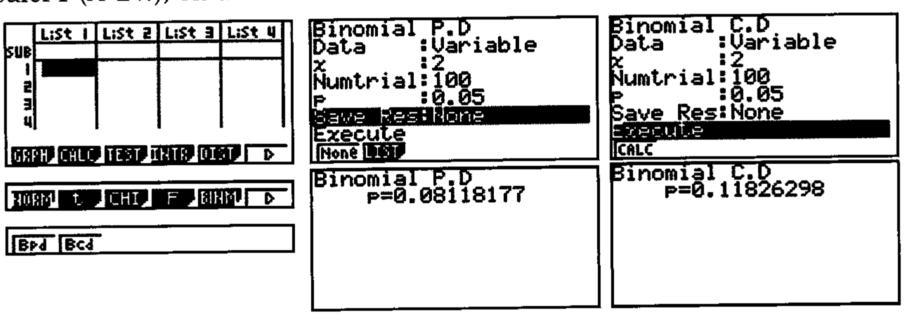
On accède au menu « distribution » par 2nde / Distrib (ou DISTR).
Pour calculer $P(X = k)$, on utilise l'instruction binomFdp ou binompdf.
Pour calculer $P(X \leq k)$, on utilise l'instruction binomFRép.





Procédure à suivre sur une calculatrice CASIO

On se place dans le Menu STAT.
On accède à la loi binomiale par DIST / BINM.
Pour calculer $P(X = k)$, on utilise l'instruction Bpd.
Pour calculer $P(X \leq k)$, on utilise l'instruction Bcd.



Remarque : on peut faire un tableau de valeur sur la calculatrice en utilisant la loi binomiale comme fonction, on remplace le k par la variable X de la calculatrice en n'oubliant pas de ne prendre que des valeurs entières (de 0 à n) pour X.

II – Intervalles de confiance et de fluctuation :

Dans ce paragraphe, on va étudier deux domaines des statistiques qu'il faut savoir distinguer :

Échantillonnage ou test d'hypothèse	**Estimation**
Une urne contient un très grand nombre de boules blanches et de boules noires dont **on connaît la proportion** *p* de boules blanches. On tire avec remise *n* boules (échantillon) et on observe la fréquence d'apparition des boules blanches. Cette fréquence observée appartient à un intervalle, appelé **intervalle de fluctuation.** Dans le cas où on ne connaît pas la proportion *p* mais on est capable de faire une hypothèse sur sa valeur, on parle de **prise de décision**.	Une urne contient un très grand nombre de boules blanches et de boules noires dont **on ignore la proportion** *p* de boules blanches. On tire avec remise *n* boules dans le but d'estimer la proportion *p* de boules blanches. Cette estimation s'obtient à l'aide d'un **intervalle de confiance** construit selon un niveau de confiance que l'on attribue à l'estimation.

1) Échantillonnage

a) Intervalle de fluctuation :

Soit $\alpha \in]0;1[$ et X une variable aléatoire qui suit une loi binomiale $B(n;p)$.

La probabilité que la fréquence F prenne ses valeurs dans l'intervalle $I=[\frac{k_1}{n};\frac{k_2}{n}]$ est $1-\alpha$, avec k_1 et k_2 les deux plus petits entiers tels que $P(X \leq k_1) > \frac{\alpha}{2}$ et $P(X \leq k_2) \geq 1-\frac{\alpha}{2}$.

I est appelé intervalle de fluctuation de la fréquence F au seuil 1-α.

Exemple :

On dispose d'une urne contenant un grand nombre de boules blanches et noires. La proportion de boules blanches contenues dans l'urne est p = 0,3. On tire successivement avec remise n = 50 boules.
Soit X la variable aléatoire dénombrant le nombre de boules blanches tirées.
Elle suit la loi binomiale $B(50\,;0,3)$. Pour avoir l'intervalle de fluctuation au seuil 0,95 (ou 95%), $\frac{\alpha}{2}=0{,}025$ et $1-\frac{\alpha}{2}=0{,}975$. D'après le tableau de valeurs, $k_1=9$ et $k_2=22$.
Donc en effectuant 50 tirages dans cette urne, la fréquence d'apparition d'une boule blanche est comprise dans l'intervalle [0,18 ; 0,44] avec une probabilité de 0,95.

b) Prise de décision

Règle de décision : Soit f la fréquence du caractère étudié d'un échantillon de taille n.
Soit l'hypothèse : "La proportion de ce caractère dans la population est p."
Soit I l'intervalle de fluctuation au seuil 0,95.
- Si $f \in I$, alors on accepte l'hypothèse faite sur la proportion p.
- Si $f \in I$, alors on rejette l'hypothèse faite sur la proportion p.

Remarque : On peut interpréter cette propriété par le fait que la probabilité qu'on rejette à tord l'hypothèse sur p sachant qu'elle est vraie est approximativement égale à 5%.

Méthode : Prendre une décision à l'aide d'un intervalle de fluctuation

Dans l'exemple précédent, on automatise le tirage grâce à une machine à air comprimé. Pour vérifier que la machine fonctionne bien, on fait un tirage de 50 boules pour lequel on obtient 10 boules blanches.
La fréquence $f=\frac{10}{50}=0{,}2$ est bien dans l'intervalle de confiance, on peut supposer que c'est fiable.

II. Estimation

X est une variable aléatoire qui suit une loi binomiale $B(n;p)$. Soit f une fréquence observée du caractère étudié sur un échantillon de taille n. On appelle intervalle de confiance de la proportion p au niveau de confiance 0,95, l'intervalle $\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$.

Remarques :
- Un niveau de confiance 0,95 signifie que dans 95 cas sur 100, on affirme à juste titre que p appartient à l'intervalle de confiance.
- Il n'est pas vrai d'affirmer que p est égal au centre de l'intervalle de confiance. Il n'est pas possible d'évaluer la position de p dans l'intervalle de confiance.

Exemple :
On dispose d'une urne contenant un grand nombre de boules blanches et noires. La proportion de boules blanches contenues dans l'urne n'est pas connue. On réalise un tirage de 100 boules et on obtient 54 boules blanches. La fréquence observée est donc $f=0,54$.
L'intervalle de confiance de la proportion de boule blanche dans l'urne au niveau de confiance 95% est
$\left[0,54-\frac{1}{\sqrt{100}};0,54+\frac{1}{\sqrt{100}}\right]=\left[0,44;0,64\right]$.

Méthode : Estimer une proportion inconnue par un intervalle de confiance

Un institut de sondage interroge 1052 personnes entre les deux tours de l'élection présidentielle sur leur intention de vote. 614 déclarent avoir l'intention de voter pour Martine Phinon.
En supposant que les votes seront conformes aux intentions, la candidate a-t-elle raison de croire qu'elle sera élue ?

- La proportion p des électeurs de Martine Phinon est inconnue. La taille de l'échantillon est $n=1052$.
La fréquence observée est $f=\frac{614}{1052}\approx 0,5837$.
- L'intervalle de confiance de la proportion p au niveau de confiance 0,95 est :
$\left[0,5837-\frac{1}{\sqrt{1052}}\ ;\ 0,5837+\frac{1}{\sqrt{1052}}\right]\approx\left[0,553\ ;\ 0,615\right]$.
Pour être élue, la proportion p doit être strictement supérieure à 0,5. Selon ce sondage, il est envisageable que Martine Phinon soit élue.

Méthode : Déterminer une taille d'échantillon suffisante pour obtenir une estimation d'une proportion

Un constructeur automobile fait appel à un institut de sondage afin de mesurer le degré de satisfaction du service après-vente.
L'institut souhaite estimer la proportion de clients satisfaits au niveau de confiance 0,95 avec une amplitude d'au plus 5 centièmes. Combien de personnes au minimum faut-il interroger ?

On appelle p la proportion de clients satisfaits. Cette proportion est inconnue.
Une estimation de cette proportion peut être obtenue à l'aide de l'intervalle de confiance au niveau de confiance 0,95 : $\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$, où f est la fréquence observée. Cette intervalle a pour longueur $\frac{2}{\sqrt{n}}$.
Donc $\frac{2}{\sqrt{n}}\leq 0,05$ soit $n\geq\frac{4}{0,05^2}=1600$.
L'institut de sondage devra donc interroger au moins 1600 personnes.