


Probabilités 2


1. Variables aléatoires : compléments

 **Définition 1 : Rappels :** Lorsque X est une variable aléatoire :

- **Variance :**
$$\begin{aligned}\text{Var}(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2\end{aligned}$$
- **Écart-type :**
$$\sigma(X) = \sqrt{\text{Var}(X)}$$


 **Propriété 1 :** Si a et b sont deux constantes, et X et Y deux variables aléatoires :

- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- $X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$

 **Exercice 1 :** Montrer que si X est une variable aléatoire d'espérance μ et d'écart-type σ , alors $Z = \frac{X - \mu}{\sigma}$ est une variable aléatoire d'espérance 0 (on dit centrée) et d'écart-type 1 (on dit réduite).


 **Définition 2 :** X_1, \dots, X_n sont n variables aléatoires.


On dit que X_1, \dots, X_n sont **indépendantes** lorsque pour tous $1 \leq i < j \leq n$ entiers et pour tous les intervalles réels A, B , on a : $P(X_i \in A \text{ et } X_j \in B) = P(X_i \in A) \times P(X_j \in B)$.

 **Remarque 1 :** Pour des variables aléatoires discrètes, on peut utiliser :

X_1, \dots, X_n sont **indépendantes** lorsque pour tous $1 \leq i < j \leq n$ entiers et pour tous les réels a, b , on a :

$$P(X_i = a \text{ et } X_j = b) = P(X_i = a) \times P(X_j = b).$$


 **Exercice 2 :** On note X une variable aléatoire de 3 valeurs $\{-1 ; 0 ; 1\}$ équiprobables. On note Y la variable aléatoire qui vaut 1 lorsque $X = 0$ et 0 sinon. Montrer que X et Y ne sont pas indépendantes.

 **Propriété 2 :** X_1, \dots, X_n sont n variables aléatoires.

On note leur somme $S_n = X_1 + \dots + X_n$ et leur moyenne $M_n = \frac{1}{n}(X_1 + \dots + X_n)$

- $\mathbb{E}(S_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)$ et $\mathbb{E}(M_n) = \frac{1}{n}(\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n))$
- **Si les variables aléatoires sont indépendantes (non-corrélées est suffisant) :**
$$\text{Var}(S_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) \text{ et } \text{Var}(M_n) = \frac{1}{n^2}(\text{Var}(X_1) + \dots + \text{Var}(X_n))$$
- **Si les variables aléatoires sont indépendantes (non-corrélées est suffisant) et ont même espérance μ et même variance v (donc même écart-type σ) :**
$$\mathbb{E}(S_n) = n\mu \text{ et } \mathbb{E}(M_n) = \mu ; \text{Var}(S_n) = nv \text{ et } \text{Var}(M_n) = \frac{1}{n}v ; \sigma(S_n) = \sigma\sqrt{n} \text{ et } \sigma(M_n) = \frac{\sigma}{\sqrt{n}}$$

 **Exercice 3 :** Dédurre les propriétés précédentes à partir des deux premières égalités donnant l'espérance et la variance de S_n .

 **Exercice 4 :** Montrer que si les X_1, \dots, X_n sont n variables aléatoires indépendantes suivant une loi de Bernoulli de probabilité p ($x = 1$ avec une probabilité p et $x = 0$ avec une probabilité $1 - p$), alors S_n suit une loi binomiale de paramètres n et p . En déduire que $\mathbb{E}(S_n) = np$ et que $\sigma(S_n) = \sqrt{np(1 - p)}$.

 **Exercice 5 :** X et Y sont deux variables aléatoires indépendantes d'espérance respectivement 15 et 20 et de variance 10 et 5.

1. Déterminer les espérances, variances et écart-types de :
 - $Z_1 = 4X$
 - $Z_2 = -10Y$
 - $Z_3 = X + Y$
 - $Z_4 = 4X - 10Y$
2. Si X et Y mesurent la durée de deux tâches successives indépendantes composant un projet, quelle variable aléatoire Z_n mesure la durée du projet ?

2. Échantillons, loi des grands nombres


Définition 3 :

- Une **population** de taille N est modélisée par un ensemble de variables aléatoires $\{Y_1, \dots, Y_N\}$, qui représentent le caractère mesuré sur chaque individu.
On suppose, sauf cas particulier, que ces variables aléatoires suivent la même loi et sont indépendantes.
- On sélectionne n individus, qui forment un **échantillon** de taille $n < N$; on a donc un ensemble de variables aléatoires $\{X_1, \dots, X_n\}$, tel que, par exemple, $X_1 = Y_3, X_2 = Y_{14}, \dots$
- Une **réalisation** de cette échantillon consiste à donner à chaque variable aléatoire de l'échantillon une valeur réelle, selon la loi suivie par cette variable aléatoire : on mesure le caractère étudié sur l'échantillon.
Une réalisation d'un n -échantillon se traduit donc par l'obtention de n valeurs : x_1, \dots, x_n (notées en minuscule).

Propriété 3 : Inégalité de Bienaymé-Tchebychev

X est une variable aléatoire d'espérance μ et de variance v .

Pour tout réel strictement positif δ , on a : $P(|X - \mu| \geq \delta) \leq \frac{\text{Var}(X)}{\delta^2}$

 **Remarque 2 :** $|X - \mu| \geq \delta$ équivaut à $X \leq \mu - \delta$ ou $X \geq \mu + \delta$, c'est à dire X est à l'extérieur de l'intervalle $]\mu - \delta; \mu + \delta[$.

Exercice 6 : Démonstration :

On définit la variable aléatoire $Z = \begin{cases} \delta^2 & \text{si } (X - \mu)^2 \geq \delta^2 \\ 0 & \text{sinon} \end{cases}$

En écrivant la loi de Z , montrer que $\mathbb{E}(Z) = \delta^2 P(|X - \mu| \geq \delta)$

Justifier que $0 \leq Z \leq (X - \mu)^2$. En déduire que $0 \leq \mathbb{E}(Z) \leq \mathbb{E}((X - \mu)^2) = \text{Var}(X)$ et conclure.

Propriété 4 : Inégalité de concentration

M_n est une variable aléatoire moyenne d'un échantillon de taille n de variables aléatoires indépendantes d'espérance μ et de variance v .

Pour tout réel strictement positif δ , on a : $P(|M_n - \mu| \geq \delta) \leq \frac{v}{n\delta^2}$

Exercice 7 : Démonstration :

Appliquer l'inégalité de Bienaymé-Tchebychev à M_n (quelle est sa variance ?).
Que se passe-t-il lorsque n est très grand ?

Théorème 1 : Loi faible des grands nombres

M_n est une variable aléatoire moyenne d'un échantillon de taille n de variables aléatoires indépendantes d'espérance μ et de variance v . On a : $\lim_{n \rightarrow +\infty} P(|M_n - \mu| \geq \delta) = 0$

 **Remarque 3 :** Cette loi affirme que parmi tous les échantillons de valeurs possibles, ceux dont la moyenne s'éloigne de l'espérance sont rares, et que cette rareté s'accroît avec la taille de l'échantillon.

Méthode 1 : Taille d'un échantillon

Pour exploiter l'inégalité de concentration, on peut selon les cas agir sur n (agrandir l'échantillon) ou δ (s'éloigner de l'espérance), ou les deux, en les choisissant assez grands pour que $\frac{v}{n\delta^2}$ soit inférieur à une probabilité donnée.

 **Exercice 8 :** On lance n fois un dé en notant les résultats $(X_i)_{1 \leq i \leq n}$ inscrits sur la face supérieure pour chaque lancer.

1. Calculer les espérances, variances et écarts-types de X_1 , de S_n et M_n ($\text{Var}(X) \approx 2,917$).
2. Appliquer l'inégalité de Bienaymé-Tchebychev à X_1 , pour $\delta = 2$, comparer le résultat obtenu à un calcul direct de la probabilité.
3. Si $n = 10$, déterminer δ pour que $P(|M_n - \mu| \geq \delta) \leq 5\%$
4. Déterminer la taille de l'échantillon pour que la probabilité que M_n soit inférieure à 2 ou supérieure à 5 soit inférieure à 1%.

3. Approfondissements : Covariance

Définition 4 : On nomme **Covariance** de X et de Y le nombre : $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$
 X et Y sont dites **non corrélées** lorsque leur covariance est nulle.

Exercice 9 :

- Démontrer que $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- En développant, montrer que $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- En reprenant les variables aléatoires données à l'ex 2, montrer qu'elles sont non corrélées mais non indépendantes.
- Démontrer que $\text{Cov}(X + Y, X - Y) = 0$.
- Le coefficient de corrélation linéaire est donné par $r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$.
 En étudiant $f(t) = \text{Var}(-tX + Y)$, démontrer que $-1 \leq r \leq 1$.
- Démontrer que $r = \pm 1$ équivaut à Y est fonction affine de X (indication : une loi de variance nulle est ...).

Définition 5 : La **droite d'ajustement linéaire** calculée par les outils numériques passe par le point moyen $(\mathbb{E}(X); \mathbb{E}(Y))$, et a pour coefficient directeur $\frac{\text{Cov}(X, Y)}{\text{Var}(X)}$.

4. Approfondissement : Estimation

Les scientifiques et les outils numériques utilisent «deux» «écarts-types» : σ et s_n . On dit que s_n est un estimateur «sans biais». Ce paragraphe permet d'éclaircir cette notion.

Définition 6 :

- On appelle **estimateur** sur un échantillon de taille n fonction (à valeurs réelles) de n variables $h(x_1, \dots, x_n)$;
 par exemple $h(x_1, \dots, x_n) = \frac{1}{n}(x_1 + \dots + x_n)$ (formule de la moyenne, notée \bar{x} en statistiques), ou bien $h(x_1, \dots, x_n) = x_1 \times x_2$ (peu d'utilité).
- $g(X_1, \dots, X_n)$ étant une variable aléatoire, on peut noter $\mathbb{E}(g)$, si elle existe, l'**espérance** de l'estimateur g .
- Par essence, l'estimateur vise à approcher un paramètre de la population (par exemple la moyenne du caractère observé sur la population, une proportion, ...).
 Ainsi, lorsque k est la notation du paramètre approché, on **peut noter** l'estimateur \hat{k} (avec un accent circonflexe), lorsqu'il n'y a pas d'ambiguïté.
- Pour mesurer l'erreur entre l'estimation et la réalité, on utilise le **biais** \mathbb{B} , défini par $\mathbb{B}(\hat{k}) = \mathbb{E}(\hat{k}) - k$.
 Lorsque son biais est nul, on dit que l'estimateur est **sans biais**.

Propriété 5 : L'estimateur $\hat{\bar{x}} = \frac{1}{n}(x_1 + \dots + x_n)$ est sans biais.

Démonstration : On note μ la moyenne sur la population entière.

$$\mathbb{B}(\hat{\bar{x}}) = \mathbb{E}\left(\frac{1}{n}(X_1 + \dots + X_n)\right) - \mu = \frac{1}{n}(\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)) - \mu = \frac{1}{n}(\mu + \dots + \mu) - \mu = \frac{1}{n}n\mu - \mu = \mu - \mu = 0$$

Exercice 10 : On note \bar{X} la variable aléatoire définie par $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ et on note V la variable aléatoire définie par $V = \bar{X}^2 - \bar{X}^2$.

On suppose que X_1, \dots, X_n ont toutes pour espérance μ et pour variance v .

Démontrer que $\mathbb{E}(V) = \frac{n-1}{n}v$; (utiliser $\mathbb{E}(T^2) = \text{Var}(T) + \mathbb{E}(T)^2$) ;

en déduire que le calcul de la variance sur l'échantillon ne fournit pas un estimateur sans biais de la variance sur la population ;

en déduire qu'un estimateur sans biais de la variance sur la population est donné par $\frac{n}{n-1}V$.

Définition 7 : Écart-type ponctuel (ou corrigé)

On utilise l'**écart-type ponctuel ou corrigé** $s_n = \sqrt{\frac{n}{n-1}}\sigma_n$ comme estimateur de l'écart-type sur la population globale ; σ_n étant l'écart-type calculé sur un échantillon de taille n .