HEC MONTRĒAL

30-650-17
Introduction à l’analytique d’affaires

Thème 4 :
Outils d’aide à la décision en contexte d’incertitude

Partie 1 : Les bases du calcul en probabilités

Cas Québec 49

La loterie *Québec 49* de Loto-Québec est la petite cousine de la *lotto 6/49*. Les seules différences sont les suivantes:

- Les billets ne sont vendus qu'au Québec.
- Le coût d'une participation (d'un billet) est de 1$.
- Les lots sont prédéterminés. Le gros lot est de 2 000 000$. La distribution des gains est fournie dans le fichier « *lotto649.xlsx* » et est reproduite ici:

Résultat	Lot	Chances de gagner	Probabilité
6/6	2 000 000	1	7.1511E-08
5/6 +C	75 000	6	4.2907E-07
5/6	750	252	1.8021E-05
4/6	75	13545	0.00096862
3/6	10	246820	0.0176504
2/6 +C	5	172200	0.01231424
2 ou moins	0	13550992	0.96904822
	Somme	13983816	1

Résultats des tirages

RÉSULTATS DES TIRAGES

12 septembre 2015

Résultats détaillés

Québec 49

09 21 31 36 39 48

COMPLÉMENTAIRE (C) 04

DÉTAILS

Résultats détaillés pour le tirage régulier

Catégorie	Gagnants	Lots
6/6	0	2 000 000,00 $*
5/6+C	0	75 000,00 $**
5/6	8	750,00 $
4/6	475	75,00 $
3/6	9 424	10,00 $
2/6+C	8 101	5,00 $
TOTAL	18 008	

*Le gros lot est partageable.

**Le montant total des lots payables dans la catégorie 5/6 + complémentaire est limité à 450 000 $ par tirage. Si cette limite est atteinte, le montant des lots sera déterminé au prorata du nombre de sélections gagnantes de la catégorie.

VENTES TOTALES : 608 521 $

Québec 49: point de vue de Loto-Québec

- Comment Loto-Québec peut-elle s'assurer de **générer des profits** en vendant des billets de loterie (sachant que les résultats des tirages sont aléatoires)?
- Loto Québec ne peut pas connaitre a priori (d'avance) la valeur de son profit avec certitude.
- Les lots sont prédéterminés, ce qui expose Loto Québec à un risque. **Le profit de Loto Québec est donc une variable aléatoire!**
- À quel point Loto-Québec est-elle en mesure de faire des prévisions de ses profits résultant des ventes de billets de loterie?

Variable aléatoire (VA)

Une variable est **aléatoire**, si on ne peut pas déterminer, **a priori**, sa valeur avec certitude.

Il existe deux types de variables aléatoires:

- si l'ensemble des possibilités d'une variable aléatoire est **fini ou dénombrable**, on parle de **VA discrète**;
  - exemple: nombre d'enfants par ménage, nombre de pièces, nombre de clients servis…
- si l'ensemble des **valeurs possibles constitue un intervalle des réels** ($\Re$), on parle de **VA continue**.
  - Exemple :Âge, revenu, température, poids…

Notion de distribution

Une **distribution** peut être regardée comme un **modèle théorique** qui décrit le comportement probabiliste d'une variable.

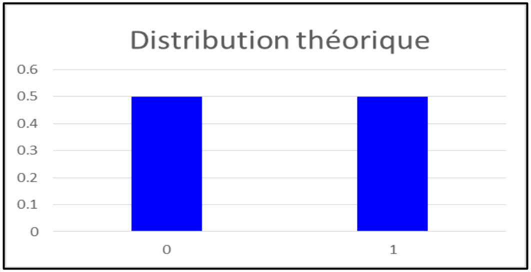
Cela signifie que si on observe en pratique un grand échantillon de valeurs de la variable (indépendamment les unes des autres), alors le graphe de la distribution épouse assez bien la forme de l'histogramme ou du diagramme en bâtonnets des données obtenues.

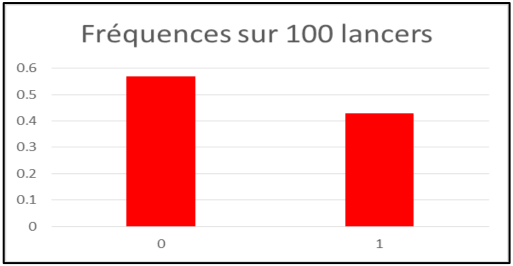
Remarque: Un histogramme ou un diagramme en bâtonnets peut donc être vu comme une **version empirique** du graphique de **la distribution** (le modèle théorique).

Un exemple très simple

Considérons la variable $X =$ résultat d'une pièce de monnaie où la valeur 0 est associée à pile et 1 est associée à face.

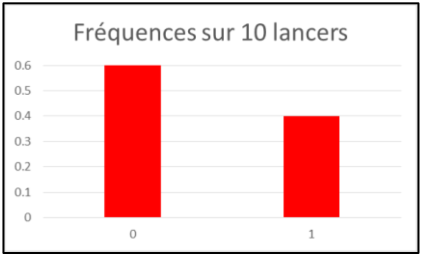
Distribution (théorique)	**Une expérience (pratique)**
La probabilité d'observer chacune des valeurs est la même (1 chance sur 2). La distribution de X est ainsi représentée par le graphique :	On a lancé 100 fois la pièce de monnaie et tracé un diagramme en bâtonnets des fréquences observées :

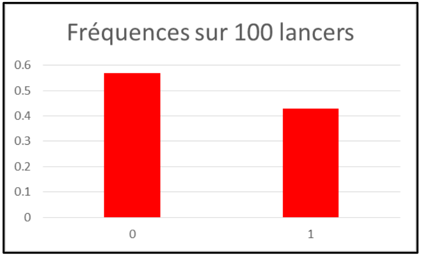


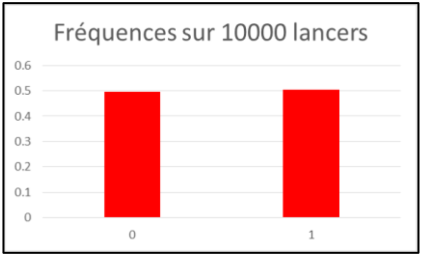


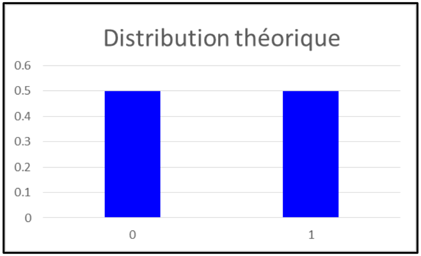
Un exemple très simple

On remarque que plus le nombre d'observations est grand, plus on a de chances que le diagramme en bâtonnets des fréquences d'observation ressemble au graphique de la distribution théorique :









Valeur espérée

Définition: On définit la valeur espérée (ou l'espérance) de la variable aléatoire X comme étant la valeur **moyenne** que l'on obtient si on l'observe un **très grand nombre (en fait infini) de fois**. Notons que les observations doivent ici être effectuées indépendamment les unes des autres.

Notation: On représente souvent la valeur espérée par $\mathbf{E[X]}$ ou par $\boldsymbol{\mu}$.

Calcul: Puisque la distribution décrit les fréquences d'observation lorsque le nombre d'observations est très grand (infini), alors on peut calculer cette moyenne en utilisant les probabilités spécifiées par la distribution. Précisément, en utilisant la notation :

$x_1, x_2, x_3, \ldots, x_k$ pour les k valeurs possibles de la variable X;
$p_1, p_2, p_3, \ldots, p_k$ pour les probabilités de ces k valeurs,

alors la valeur espérée est obtenue au moyen de

$$\boxed{E[X] = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k.}$$

Variance et écart-type

Définition: L'écart-type d'une distribution représente, grosso modo, l'écart moyen entre les observations et leur moyenne.
Plus précisément, il s'agit de la racine de la moyenne des écarts à la moyenne au carré:.

Notation: On représente souvent l'écart type par $\boldsymbol{\sigma}[\mathbf{X}]$ ou par $\boldsymbol{\sigma}$.

Calcul: la variance de X est définie par

$$V[X] = E[(X - E[X])^2] = E[X^2] - E^2[X]$$

On obtient alors que

$$\boxed{V[X] = (\, p_1\, x_1^2 + p_2 x_2^2 + \cdots + p_n x_n^2\,) - \; E^2[X]}$$

On calcule ensuite l'écart-type $\sigma[X]$ en prenant la racine carrée de la variance.

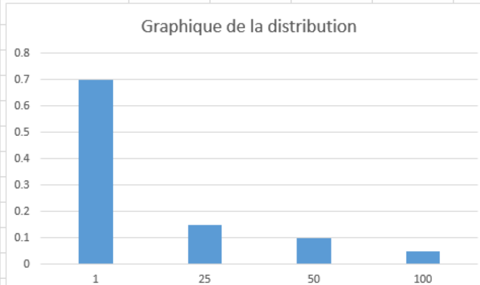
Calcul

Ces calculs se font bien avec la fonction Excel ***sommeprod*.**

Dans ce cours, vous pouvez également obtenir l'espérance, l'écart-type et le graphique d'une distribution en utilisant le gabarit Excel **« résumé_distribution.xlsx »**.

distribution		somme des probabilités
valeur	probabilité	
1	0.7	1
25	0.15	
50	0.1	
100	0.05	

Valeur espérée	14.45
Variance	635.648
Écart-type	25.2121



Entrez ici les valeurs possibles de la variable et leurs probabilités respectives.

Exemple d'une VA discrète

Une urne contient 20 boules dont

- 1 de couleur Rouge,
- 2 de couleur Bleu,
- 3 de couleur Orange,
- 14 de couleur Noir.

- La boule Rouge vaut 100$,
- La boule Bleu vaut 50$,
- La boule Orange vaut 25$,
- et la boule Noir vaut 1$.

Exemple d'une VA discrète

On peut définir la variable aléatoire discrète

X = « la valeur en $ de la boule tirée. »

Si on tire au hasard une boule de l'urne, alors l'ensemble des valeurs possibles de la variable aléatoire **X** est

S={1$, 25$, 50$, 100$}.

Questions :

1. Présentez la distribution de X.
2. Trouvez la valeur espérée du gain (le gain moyen attendu).

Cas Québec 49

Loto-Québec vend un très grand nombre de billets (à 1$ chacun).

Question 1 : Quel montant Loto-Québec doit-elle verser, **en moyenne**, à chacun des détenteurs de billet?

Question 2 : Le montant **moyen** versé aux joueurs excède-t-il les revenus générés par la vente des billets?

Question 3 : Sur le site web de Loto-Québec, on indique que le taux de remise de la loterie *Québec 49* est de 49,9%. Est-ce conforme à vos calculs?

Question 4 : Calculez les écart-types du gain brut du joueur et du profit par billet de Loto-Québec.

Connaissez-vous une autre industrie dont le modèle d'affaire repose principalement sur le principe de la valeur espérée?

Exemple d'une compagnie d'assurance

Une compagnie d'assurance offre une police d'annulation de voyage. La prime est de 144$ par personne.
Le coût pour la compagnie d'assurance est de 1600$ en cas d'annulation.
Supposons que la probabilité qu'un client annule son voyage est de 0.02 (d'après l'historique).

Soit X: le gain de la compagnie lorsqu'elle assure une personne.

1. Quelles sont les valeurs possibles de X?
2. Déterminer la distribution de probabilité de X.
3. Calculer l'espérance mathématique et l'écart type de X en utilisant le gabarit « Résumé_distribution.xlsx ».

Partie 2 : la loi binomiale

Le cas de la publicité de la Schlitz au XVe Super Bowl*

Bref historique de la Joseph Schlitz Brewing Company:

- Trouve ses origines dans une brasserie-taverne à Milwaukee, fondée en **1849** par un dénommé August Krug.
- Particulièrement florissante dans la première partie du 20e siècle. Elle rivalise avec Anheuser-Busch (qui produit la Budweiser et la Michelob) pour le titre de la **bière la plus vendue aux États-Unis**, jusqu'à la fin des années 50, où elle est définitivement surpassée par sa grande rivale.
- Afin d'augment la marge de profit, la recette est altérée en utilisant des additifs chimiques au début des années 70. Son image commence à se dégrader petit à petit auprès du public
- Est surpassée par Miller en 1977, puis par Pabst en 1980.
- Elle est **déficitaire** en **1979**.

**Librement inspiré du livre Naked Statistics, Stripping the Dread from the Data, de Charles Wheelan.*

Pub durant les séries de la NFL de 1980-81

Dans l'espoir de regagner des parts de marché, Schlitz lance une importante campagne de publicité durant les séries de fin de saison de la NFL. Celle-ci consiste à :

- Mener cinq **tests à l'aveugle en direct** pendant la télédiffusion de matchs de football, culminant au Super Bowl.

- Chaque test oppose Schlitz à l'une de ses plus grandes concurrentes : Michelob, Budweiser et Miller.

- Schlitz **pousse l'audace jusqu'à utiliser 100 amateurs de la bière adverse** lors de chacun des tests.

- Les tests eux-mêmes jouissent d'un important battage publicitaire pour mousser l'intérêt du public.

La publicité du Super Bowl (janvier 1981)

- **Cent loyaux buveurs** de Michelob sont recrutés.
- Chacun de ces 100 goûteurs est attablé devant deux verres de bière non identifiés (l'un de Michelob et l'autre de Schlitz).
- Au signal, en direct devant un gigantesque auditoire télé, chacun doit incliner un levier pour indiquer sa bière préférée. Le résultat s'affiche sur un tableau indicateur.
- Si x (par exemple 40) buveurs choisissent la Schlitz, le message publicitaire est

« x% (exemple 40%) des buveurs de Michelob préfèrent la Schlitz ».

Quelle campagne!

À l'époque, cette campagne

- est novatrice;
- coûte 4 millions (la pub du Super Bowl coûte 1,7 million à elle seule);
- suscite plusieurs réactions (on en parle; elle fait couler de l'encre dans les journaux).

Que pensez-vous de cette tactique publicitaire?

Ce plan a-t-il une probabilité élevée de succès?

Les risques sont-ils bien sous contrôle?

Épreuve de Bernoulli

On appelle **épreuve de Bernoulli** toute expérience aléatoire pour laquelle il n’y a que **deux résultats possibles**, qu’on désignera par **« succès »** et **« échec »**.

Remarque: Le terme **«succès»** est en fait un **nom choisi arbitrairement**. Par succès, nous voulons simplement désigner «le résultat qui nous intéresse». Des exemples de «succès» peuvent donc être «On obtient pile», «Le goûteur choisit la Schlitz», aussi bien que «L’entreprise fait faillite », «La personne meurt du cancer», etc.
(un « succès » peut donc en réalité désigner un événement fâcheux, voire dramatique, s’il s’agit du résultat qui est l’objet de l’étude).

Loi binomiale

- *Considérons une suite de $\boldsymbol{n}$ épreuves de Bernoulli **indépendantes** ayant toute la **même probabilité** de produire un « succès », dénotons-la $\boldsymbol{p}$.*

- *On dit alors que la variable aléatoire*

$$\boldsymbol{X = \text{nombre de succès parmi ces } n \text{ épreuves}}$$

*obéit à la **loi binomiale de paramètres $\boldsymbol{n}$ et $\boldsymbol{p}$**.*

- *On utilise la notation $\boldsymbol{X \sim Bin(n, p)}$.*

- *Les valeurs possibles de cette variable sont: $0,1,2, \dots, n$.*

- *On peut calculer les probabilités de la forme $P(X = x)$ ou $P(X \leq x)$ en utilisant la fonction Excel **LOI.BINOMIALE.N()**. Toutes les autres probabilités utiles peuvent alors être déduites.*

Loi binomiale

Exemples de variables de loi binomiale:

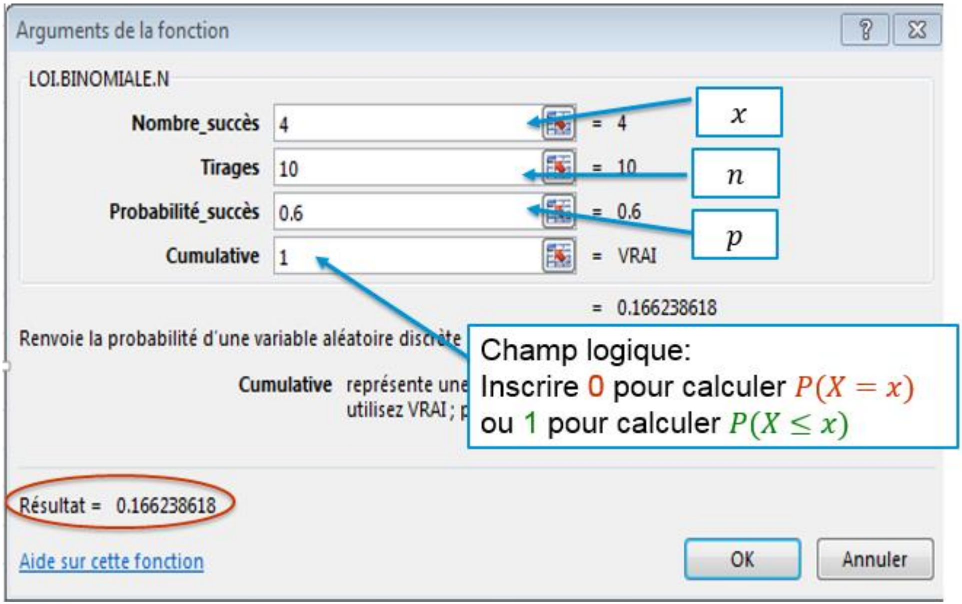
- **Le nombre** de faces sur un certain nombre de lancers d'une pièce de monnaie.
- **Le nombre de goûteurs** qui préfèrent la Schlitz sur un échantillon de 100 buveurs.
- **Le nombre** d'items défectueux dans une boîte de 50 items.
- **Le nombre de personnes** dans une population favorable à une idée, à un produit (sondages, études de marché).
- **Le nombre de clients** qui ne renouvelleront pas leur contrat de téléphonie cellulaire parmi les clients actuels.

Lorsqu'il n'y a qu'une seule épreuve de Bernoulli **($n = 1$)**, on dit que X est une **variable de Bernoulli de paramètre p**, ou encore une **variable indicatrice**. Remarquons effectivement qu'elle *indique* si oui ou non on a observé un événement donné (correspondant au « succès »):

$$X = \begin{cases} 1 & \text{si le résultat est un « succès »} \\ 0 & \text{sinon} \end{cases}$$

Calcul avec Excel

Avec la fonction LOI.BINOMIALE.N, on peut calculer $P(X = x)$ ou $P(X \leq x)$ lorsque $X \sim \text{Bin}(n, p)$.



Moyenne et écart-type de la binomiale

Supposons que $X \sim \text{Bin}(n, p)$.

- La **valeur espérée** de X est

$$\boxed{\boldsymbol{\mu = np.}}$$

Exemple: Si X représente le nombre de goûteurs qui choisissent la Schlitz, alors $X \sim \text{Bin}(n = 100,\ p = 0{,}5)$. Puisque $\boldsymbol{p = 0,5}$, on peut espérer que la moitié des goûteurs choisissent la Schlitz, ce qui représente 50 goûteurs. La formule corrobore ce raisonnement intuitif:

$$\text{valeur espérée} = \mu = np = 100 \times 0{,}5 = 50.$$

- L'**écart-type** de X est

$$\boxed{\boldsymbol{\sigma = \sqrt{np(1-p)}.}}$$

Le cas Schlitz

Question 1 :

On s'intéresse à la variable aléatoire X, **le nombre de buveurs** choisissant la Schlitz parmi les 100 buveurs dans la salle.

Hypothèses : - On suppose que le choix des buveurs est indépendant l'un de l'autre.
- La probabilité qu'un buveur choisisse la Schlitz est de 0,5.

a) Quelle est la loi de probabilité de X?

b) Quelle est la probabilité qu'exactement la moitié des 100 buveurs choisissent la Schlitz?

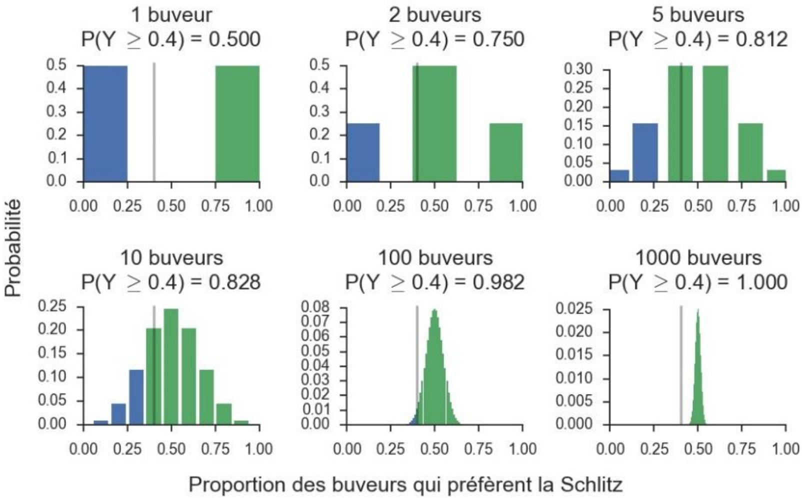
c) Quelle est la probabilité que 50 buveurs ou moins choisissent la Schlitz?

d) Quelle est la probabilité que plus de 50 buveurs choisissent la Schlitz?

e) Supposons que la compagnie Schlitz estime que le message est suffisamment fort pour que la publicité soit efficace si au moins 40 buveurs choisissent la Schlitz. Quelle est la probabilité que cela se produise dans une salle pendant le Super Bowl?

Pourquoi 100 buveurs?

Les graphiques suivants représentent la distribution de la proportion de goûteurs qui choisissent la Schlitz, pour différents nombres de goûteurs:



Le cas Schlitz

Question 2 :

La campagne comporte au total cinq tests (salles) utilisant 100 buveurs chacun.
Soit Y : **le nombre de salles** procurant un succès (une publicité efficace avec au moins 40 buveurs choisissent la Schlitz) parmi les 5 salles.
Hypothèse : on suppose que le résultat dans les 5 salles est indépendant l'un de l'autre.

a) Quelle est la loi de probabilité de Y?

b) Quelle est la probabilité que les cinq tests soient un succès?

c) Quelle est la probabilité que 4 tests ou plus mènent à un succès?

Le cas Schlitz

Question 3 :

Rappelons que toute ***la stratégie de Schlitz repose sur <u>l'hypothèse</u> que les bières sont impossibles à distinguer au goût***. Or, en rétrospective, l'expérience s'est soldée par 2 publicités où moins de 40 buveurs ont préféré la Schlitz, et 3 publicités où 40 buveurs ou plus ont choisi la Schlitz. Voici le détail des résultats:

Budweiser		Miller		Michelob
46	50	38	37	50

Y a-t-il lieu de remettre en question l'hypothèse selon laquelle les bières sont identiques au goût?

Retour sur le cas Loto-Québec

Question 1 :

Une différence majeure entre la 6/49 et la *Québec 49* est **le nombre de billets vendus à chaque tirage**. Supposons que tous les joueurs choisissent leur sélection indépendamment les uns des autres et qu'ils choisissent chacun leur sélection au hasard. En supposant que Loto-Québec vende à chaque tirage 600 000 billets et 5.5 millions de billets respectivement pour Québec 49 et Lotto 6/49.

On s'intéresse à
N_Q : nombre de gagnants du gros lot de Québec 49 lors d'un tirage
N_L : nombre de gagnants du gros lot de Lotto 6/49 lors d'un tirage

Quelles sont les lois de probabilité de N_Q et N_L respectivement?

Rappel : la probabilité qu'une personne gagne le gros lot est d'environ 1 chance sur 14 millions ≈ 7.15112E-08.

Retour sur le cas Loto-Québec

Question 2 :

Déterminez les chances que le gros lot soit remporté par un unique gagnant ou qu'il soit partagé entre de multiples gagnants, en complétant le tableau suivant, portant sur des tirages où respectivement 600 000 et 5 500 000 billets sont vendus pour la *Québec 49* et la *lotto 6/49*:

		Nombre de gagnants du gros lot		
Loterie	Nombre de joueurs	0	1	>=2
Québec 49	600 000			
Lotto 6/49	5 500 000			

Retour sur le cas Loto-Québec

Question 3 :

En pratique, les joueurs ne choisissent pas nécessairement leur sélection au hasard… Lorsque l'on joue à la lotto 6/49, y a-t-il certaines sélections qu'il est préférable de jouer à d'autres?

Par exemple, considérons les sélections

A. 1-2-3-4-5-6;
B. 33-34-38-40-43-45.

Ces deux sélections sont-elles en tous points similaires, ou a-t-on avantage à en préférer une à l'autre?