La simulation probabiliste avec Excel

(2e version)

Emmanuel Grenier emmanuel.grenier@isab.fr

Relu par Kathy Chapelain et Henry P. Aubert

Incontournable lorsqu'il s'agit de gérer des phénomènes aléatoires complexes, la simulation probabiliste s'impose également dans l'enseignement des probabilités et de la statistique inductive parce qu'elle permet d'aborder ces disciplines, réputées théoriques et ardues, par la voie de l'expérimentation.

La simulation probabiliste repose sur des « séries de valeurs pseudo-aléatoires ». Nous verrons comment obtenir de telles séries avec Excel, puis comment les utiliser pour simuler une loi de probabilité. Nous appliquerons la méthode dite « des fractiles » parce qu'elle est simple à mettre en œuvre et qu'elle permet de simuler avec Excel la plupart des lois d'usage courant. Nous présenterons également des méthodes plus particulières, à usage pédagogique.

Un exemple d'application sera proposé dans le domaine de l'analyse du risque. Le lecteur pourra trouver d'autres exemples dans le manuel du groupe « Le Cercle d'Excel'Ense » [1] : vérification de propriétés probabilistes, caractérisation de la distribution de statistiques d'échantillonnage et étude des propriétés d'un estimateur.

1 Production de valeurs pseudo-aléatoires avec Excel

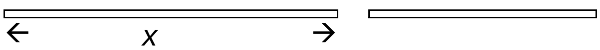
Le point de départ de la simulation probabiliste est la production de séries de valeurs aléatoires issues de la loi « uniforme entre 0 et 1 ». On parle de valeurs « pseudo-aléatoires » parce que le caractère aléatoire du processus n'est qu'apparent.

1.1 La loi uniforme entre 0 et 1

On colorie une ficelle sur le quart de sa longueur puis on casse la ficelle. Quelle est la probabilité que la rupture ait lieu dans la partie colorée ?

Sans information sur l'état de la ficelle, la probabilité ne dépend pas de l'emplacement de la partie colorée. Elle ne dépend que de sa longueur. Si la partie colorée correspond au quart de la longueur de la ficelle, la probabilité est égale à un quart.

Repérons le point de rupture par la longueur du bout de gauche.



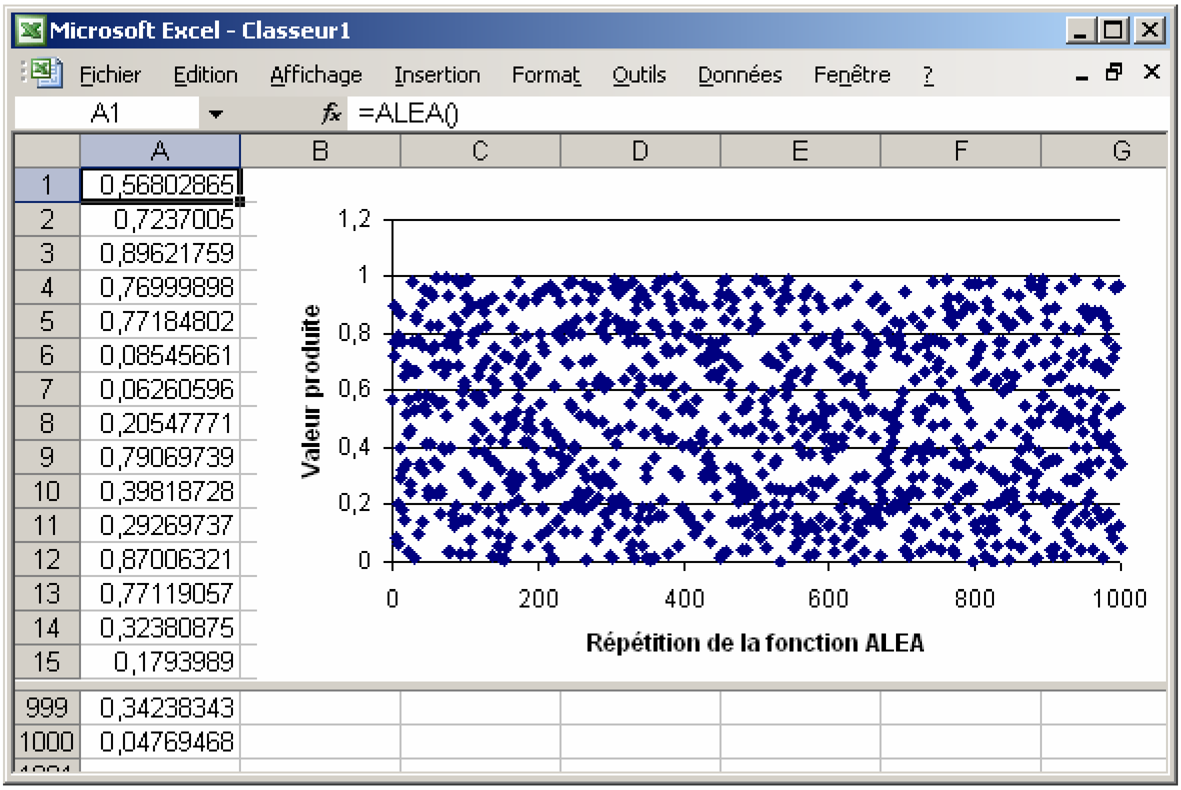
Pour simplifier, prenons une ficelle d'une unité (1 mètre) de long. Le point de rupture, X, varie alors entre 0 et 1.

Dire que la probabilité ne dépend que de la longueur de l'intervalle considéré (ici, la partie colorée) équivaut à poser une loi de probabilité « uniforme » sur la variable X.

1.2 La fonction ALEA

On peut simuler l'expérience avec Excel grâce à la fonction **ALEA** (voir dans le manuel [1] la fiche « Fonction » correspondante et le chapitre « Probabilités et jugement sur échantillon »).

Vérifions le en recopiant la formule **=ALEA()** sur 1000 cellules puis en représentant la distribution des valeurs obtenues par un nuage de points :



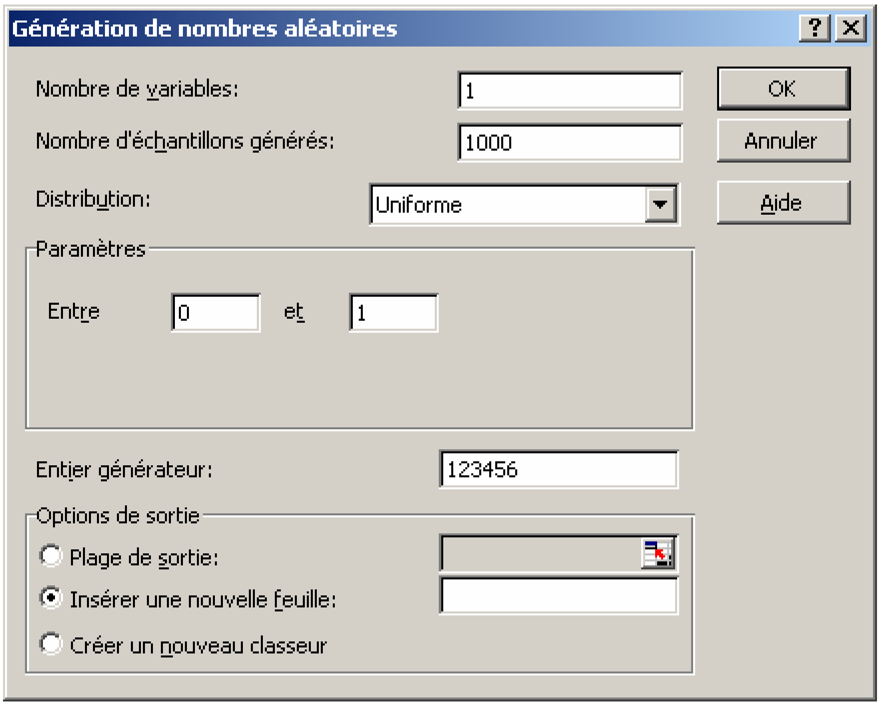
Les valeurs de la série changent quand on relance les calculs (touche fonction F9) mais elles se répartissent toujours de manière apparemment indépendante et uniforme entre 0 et 1.

Attention ! Dans les versions d'Excel antérieures à 2003, les séries bouclent au bout d'un grand nombre d'appels de la fonction. Voir l'article en ligne de Microsoft [2].

1.3 *L'utilitaire Génération de nombres aléatoires*

Si on veut avoir la possibilité de reproduire une série, on peut utiliser la macro ***Génération de nombres aléatoires*** de l'***Utilitaire d'analyse*** (Allez dans le menu ***Outils***. Si l'***Utilitaire d'analyse*** n'apparaît pas, installez le en passant par ***Macros complémentaires***).

Avec la boîte de dialogue qui suit, on obtient une série équivalente à celles obtenues précédemment.



L'***Entier générateur*** (ici **123456**) est le « germe » de la série. Pour reproduire la série, il suffit de reprendre le même germe.

Note : La macro simule d'autres lois (voir le § 2.4 page 6) et produit des séries périodiques (non aléatoires).

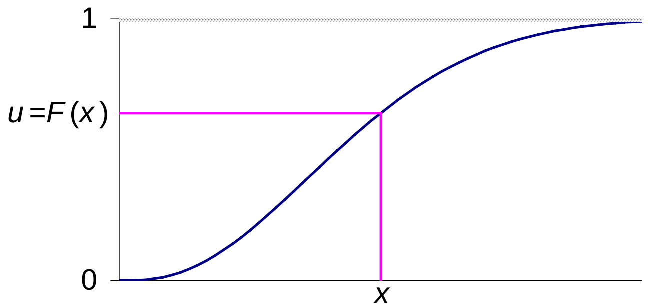
2 Simulation d'une loi par la méthode des fractiles

2.1 Domaine d'application et principe

Notons *X* la variable à simuler et *F* sa fonction de répartition. Par définition, *F*(*x*) est la probabilité cumulée jusqu'à la valeur *x*, c'est-à-dire la probabilité que la variable *X* prenne une valeur égale à *x* ou plus petite.

La méthode des fractiles s'applique à partir du moment où on peut calculer la réciproque F^{-1} de la fonction de répartition. Son principe est simple :

Prenons une réalisation possible *x* de la variable. Notons *u* la probabilité cumulée jusqu'à la valeur *x*. Par définition, *u* = *F*(*x*).



Si *x* est une réalisation quelconque de la variable *X*, on n'a pas d'information sur la valeur de *u*, mis à part qu'elle se situe nécessairement entre 0 et 1 puisque c'est une probabilité. Nous nous trouvons dans une situation analogue à celle du point de rupture de la ficelle (page 1) : la valeur *u* peut être considérée comme la réalisation d'une variable de loi uniforme entre 0 et 1.

Réciproquement, si u est la réalisation d'une variable de loi uniforme entre 0 et 1, le « fractile » correspondant, c'est-à-dire la valeur $x = F^{-1}(u)$, peut être considéré comme une réalisation de la variable X (pour une démonstration plus formelle voir par exemple l'ouvrage de G. Saporta [3]).

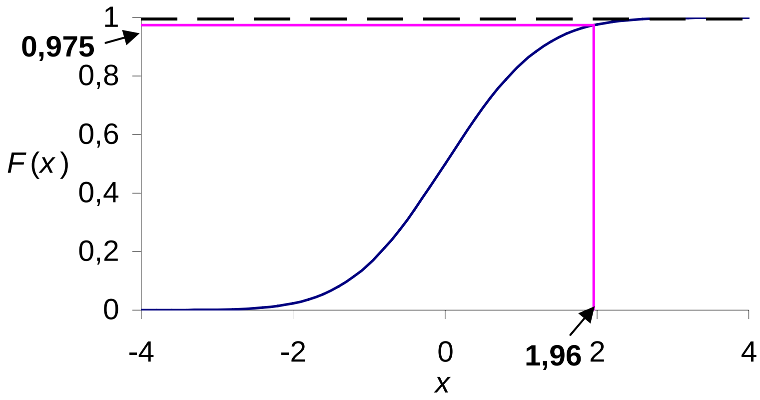
Nous avons vu au § 1 que la réalisation d'une variable de loi uniforme entre 0 et 1 peut être simulée avec la fonction **ALEA**. Pour simuler une réalisation de la variable X, il suffit donc d'appliquer la réciproque de sa fonction de répartition au résultat de la fonction **ALEA**.

2.2 *Vérification sur un exemple*

Prenons la loi normale standard (voir le chapitre « Probabilités et jugement sur échantillon » du manuel [1]).

La fonction de répartition est calculée par la fonction **LOI.NORMALE.STANDARD**, sa réciproque par la fonction **LOI.NORMALE.STANDARD.INVERSE**. (voir les fiches « Fonction » dans le manuel).

Par exemple, la formule **=LOI.NORMALE.STANDARD(1,96)** donne **0,975**, probabilité d'obtenir une valeur inférieure à 1,96.

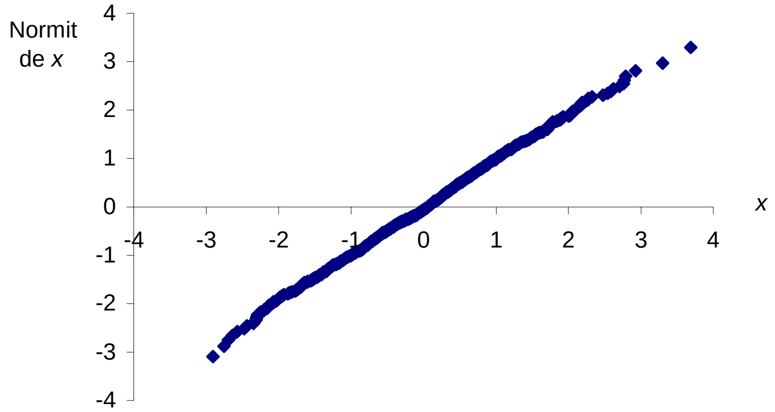


En appliquant la fonction réciproque sur le résultat, c'est-à-dire en tapant la formule **=LOI.NORMALE.STANDARD.INVERSE(0,975)**, on retrouve la valeur de départ, **1,96**.

Appliquons la réciproque de la fonction de répartition au résultat de la fonction **ALEA** :

=LOI.NORMALE.STANDARD.INVERSE(ALEA())

Recopiez la formule sur 1000 cellules et représentez la distribution des valeurs obtenues par un histogramme (voir la fiche « Comment faire » dans le manuel). La distribution des valeurs simulées semble correspondre à la loi normale standard. On peut le vérifier en construisant un diagramme d'Henry (voir le manuel ou Goldfarb et Pardoux [4]).



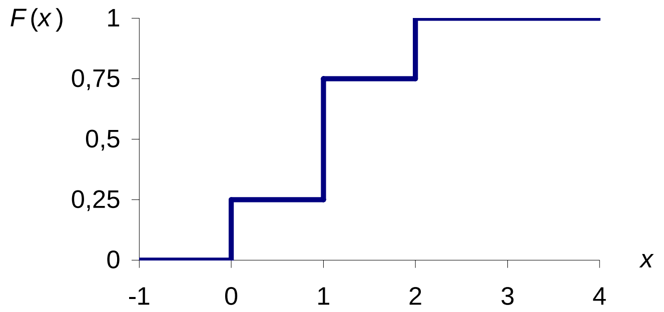
2.3 *Le cas particulier des variables discontinues*

2.3.1 Méthode générale

On fait 2 lancers d'une pièce et on note le nombre x de faces obtenu. La variable correspondante est distribuée de la manière suivante :

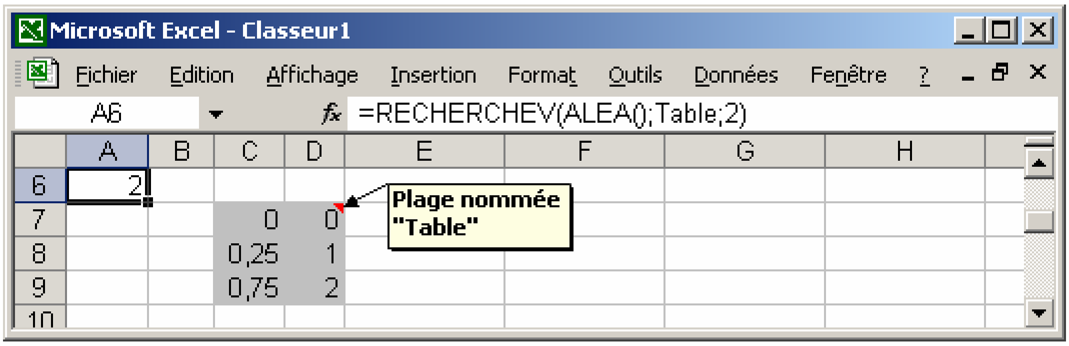
x	Probabilité	Probabilité cumulée
0	0,25	0,25
1	0,5	0,75
2	0,25	1

Représentons la distribution par la fonction de répartition :



Appelons u le résultat de la fonction **ALEA**. La valeur simulée de la variable X est la réciproque de la fonction de répartition en u, c'est-à-dire 0 si u est inférieur à 0,25, 1 si u est compris entre 0,25 et 0,75 et 2 si u est supérieur à 0,75.

La correspondance peut être faite sur Excel avec la fonction **RECHERCHEV** (voir la fiche « Fonction » dans le manuel [1]) :



Attention ! Les valeurs de la plage **Table** doivent être disposées comme le montre la capture d'écran et non comme dans le tableau de distribution plus haut.

On peut vérifier que les valeurs de la variable apparaissent avec des fréquences proches des probabilités.

2.3.2 Le cas de la loi discrète uniforme

On veut simuler le point marqué par un dé. La méthode des fractiles fait correspondre la valeur 1 aux valeurs de la fonction **ALEA** comprises entre 0 et 1/6, la valeur 2 aux valeurs entre 1/6 et 2/6, etc. Cela revient à multiplier le résultat de la fonction **ALEA** par 6, puis à éliminer les décimales de la valeur obtenue (fonction **ENT**) augmentée d'une unité :

=ENT(6*ALEA())+1

2.3.3 Le cas des variables binaires

Reprenons le lancer de pièces. Pour simuler le résultat sur une pièce, il suffit de dire qu'on a obtenu face si la résultat de la fonction **ALEA** est plus petit que 0,5 et pile sinon :

=SI(ALEA()<1/2;"Face";"Pile")

De manière générale, on simule la réalisation d'un événement A de probabilité *p* par la formule :

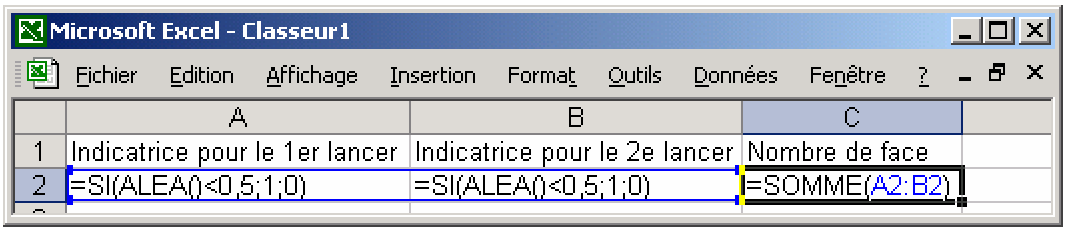
=SI(ALEA()<p;"A";"Non A")

et la variable indicatrice de l'événement, c'est-à-dire la variable qui prend la valeur 1 si l'événement est réalisé et 0 sinon (loi « de Bernouilli »), par :

=SI(ALEA()<p;1;0)

Remarque :

Pour simuler le résultat des deux lancers de la page 5, il suffit de copier la formule précédente dans 2 cellules et de faire la somme :



Nous reprendrons cette idée pour simuler la loi binomiale (p. 8).

2.4 Application aux lois d'usage courant

Le tableau qui suit regroupe deux catégories de lois :

1. les lois fondamentales de la statistique : de la loi continue uniforme à la loi de Fisher.
2. les lois exponentielle, bêta, gamma et log-normale, couramment utilisées en analyse du risque et en fiabilité.

Les 12 premières lois (de la loi continue uniforme à la loi exponentielle), sont introduites dans le manuel du groupe « Le Cercle d'Excel'Ense » [1]. Pour une présentation plus classique de ces lois, on peut se reporter au cours en ligne St@tNet [5], accessible également aux débutants.

Pour les trois dernières lois, on peut consulter l'ouvrage de G. Saporta [3] ou un cours en ligne comme Engineering Statistics Handbook (cliquez sur le lien pour accéder directement aux lois de probabilité) [6] ou SEL (plus sommaire mais en français) [7].

Les 6 premières lois (de la loi continue uniforme à la loi de Poisson) sont simulées directement par l'***Utilitaire d'analyse*** (voir page 2).

Loi	*Formule*
continue uniforme entre *a* et *b*	**=ALEA()*(b-a)+a**
de Bernouilli de paramètre *p* (voir page 6 le § 2.3.3)	**=SI(ALEA()<p;1;0)**
discrète uniforme sur les entiers de n_1 à n_2 (voir page 5)	**=ENT((n2-n1+1)*ALEA())+n1** **=ALEA.ENTRE.BORNES(n1;n2)**
normale de paramètres μ et σ	**=LOI.NORMALE.STANDARD.INVERSE(ALEA())*sigma +mu** **=LOI.NORMALE.INVERSE(ALEA();mu;sigma)**
binomiale de paramètres *n* et *p* (voir page 8)	**=CRITERE.LOI.BINOMIALE(n;p;ALEA())** la formule **=CRITERE.LOI.BINOMIALE(n;p;u)** donnant la plus petite valeur pour laquelle la probabilité cumulée est supérieure ou égale à *u*
de Poisson de paramètre λ	On peut utiliser la convergence de la loi binomiale vers la loi de Poisson et reprendre la formule du dessus en donnant une grande valeur à **n** et en remplaçant **p** par λ/n
hypergéométrique (voir page 10)	Utiliser **RECHERCHEV** (voir le § « Méthode générale » page 5) à partir des probabilités cumulées calculées avec la fonction **LOI.HYPERGEOMETRIQUE**
géométrique de paramètre *p* (voir page 8)	**=ENT(LN(ALEA())/LN(1-p))+1** (voir la note explicative page 8)
du khi-deux à ν degrés de liberté	**=KHIDEUX.INVERSE(ALEA();nu)**
de Student à ν degrés de liberté	**=LOI.STUDENT.INVERSE(ALEA();nu) *SIGNE(ALEA()-0,5)** Il faut affecter un signe de manière aléatoire parce que la fonction ne donne que des valeurs absolues
de Fisher à ν_1 et ν_2 degrés de liberté	**=INVERSE.LOI.F(ALEA();nu1;nu2)**
exponentielle de paramètre λ	**=-1/lambda*LN(ALEA())** La réciproque de la fonction de répartition est la fonction qui à *u* associe $-1/\lambda \ln(1-u)$ et simuler la variable $1-U$, quand *U* suit la loi continue uniforme entre 0 et 1, équivaut à simuler directement *U*
bêta de paramètres α et β entre les bornes *a* et *b* (voir le § 4.1)	**=BETA.INVERSE(ALEA();alpha;beta;a;b)**
gamma de paramètres α et β (Attention : le paramètre β peut être défini de 2 manières différentes. Voir l'aide d'Excel)	**=LOI.GAMMA.INVERSE(ALEA();alpha;beta)**
log-normale	**=LOI.LOGNORMALE.INVERSE(ALEA();mu;sigma)**

3 Méthodes à usage pédagogique

3.1 La loi binomiale à partir du processus de Bernouilli

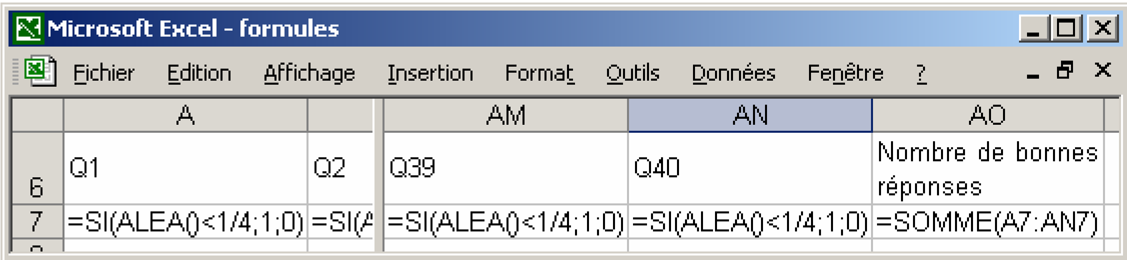
Un examen consiste en une série de 40 questions indépendantes comptant chacune pour le même nombre de points. A chaque question, 4 réponses sont proposées dont une seule est exacte. Un étudiant qui répond au hasard a-t-il des chances d'avoir plus de la moyenne ?

Le nombre de bonnes réponses suit une loi binomiale, qu'on peut simuler par la formule :

=CRITERE.LOI.BINOMIALE(n;p;ALEA())

avec n = 40 (nombre de questions) et p = 0,25 (4 choix possibles par question). Voir le tableau de la page 7.

Mais le nombre de bonnes réponses est aussi la somme des indicatrices de succès à chaque question (on note 1 si la réponse à la question est bonne, 0 sinon). On a affaire à un « processus de Bernouilli », qu'on simule par la ligne de formule suivante (voir page 6, « Le cas des variables binaires ») :



Microsoft Excel - formules

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

	A		AM	AN	AO
6	Q1	Q2	Q39	Q40	Nombre de bonnes réponses
7	=SI(ALEA()<1/4;1;0)	=SI(A	=SI(ALEA()<1/4;1;0)	=SI(ALEA()<1/4;1;0)	=SOMME(A7:AN7)

Recopiez les formules (plage **A7:AO**) sur un grand nombre de lignes. Représentez la distribution du nombre de bonnes réponses par un diagramme en bâtons (voir la fiche « Comment faire »). Comparez la à la distribution de probabilités (calculées avec la fonction **LOI.BINOMIALE**). L'étudiant a-t-il des chances d'avoir coché la bonne réponse à plus de la moitié des questions ?

3.2 La loi géométrique par simulation de tirages

On tire une bille dans un sac contenant des billes blanches et des billes rouges. Si la bille est blanche, on la remet dans le sac. On répète l'opération jusqu'à ce qu'on obtienne une bille rouge. Combien faut-il de tirages ?

Notons p la proportion de billes rouges dans le sac.

La probabilité d'avoir une boule rouge au premier tirage est p. En faisant appel aux probabilités conditionnelles, on montre que la probabilité de s'arrêter au 2^{e} tirage est $p(1-p)$, au 3^{e} tirage $p(1-p)^2$ et, de manière générale, $p(1-p)^{x-1}$ au bout de x tirages.

Note explicative de la formule proposée dans le tableau de la page 7

La probabilité cumulée en *x*, *F*(*x*), est égale à $p + p(1-p) + \cdots p(1-p)^{x-1}$

En posant *q* = 1 – *p*, on a $F(x) = p(1 + q + \cdots q^{x-1})$

Or $1 + q + \cdots q^{x-1} = \dfrac{1-q^x}{1-q}$ donc $F(x) = 1 - q^x = 1-(1-p)^x$

Cette formule nous permet de calculer la réciproque de la fonction de répartition :

Posons *u* = *F*(*x*). On a $u = 1-(1-p)^x$, d'où $(1-p)^x = 1-u$ et $x = \dfrac{\ln(1-u)}{\ln(1-p)}$

Appliquons maintenant la méthode des fractiles :

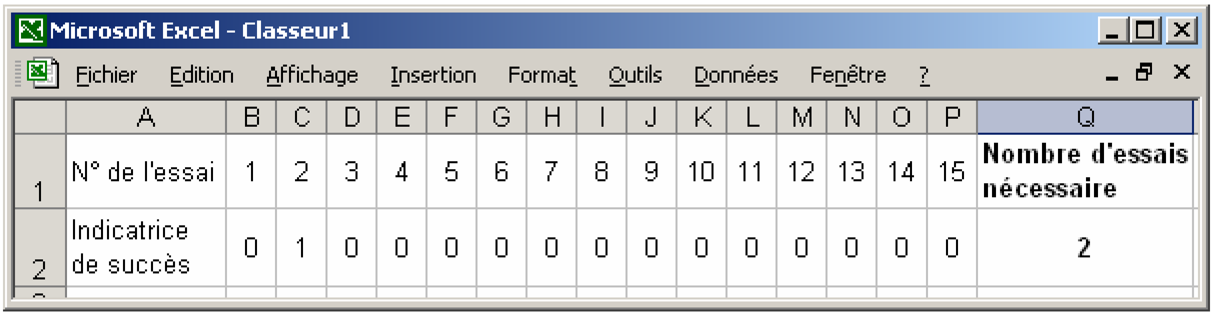
A partir de la valeur simulée *u* d'une variable de loi continue uniforme entre 0 et 1, on recherche la valeur *x* la plus petite vérifiant *F*(*x*) > *u* (voir page 5, le § « Méthode générale »), c'est-à-dire supérieure à $\alpha = \dfrac{\ln(1-u)}{\ln(1-p)}$

La variable pouvant prendre toute valeur entière (à partir de 1), on prend l'entier le plus petit supérieur à α, c'est-à-dire la partie entière de α + 1. En remarquant d'autre part que, si *U* suit la loi continue uniforme entre 0 et 1, simuler 1 – *U* équivaut à simuler directement la variable *U*, on aboutit à la formule proposée au § 2.4.

Pour vérifier la loi de probabilité, nous allons simuler l'épreuve des billes.

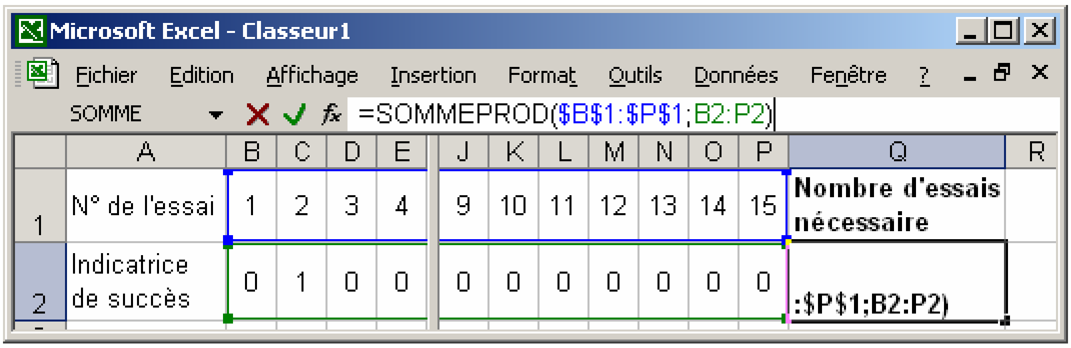
Prenons une proportion *p* égale à ½.. On peut alors fixer le nombre maximum d'essais à 15, la probabilité de dépasser cette valeur étant négligeable (en reprenant les formules de la note précédente, on a trouve ½ à la puissance 15, soit 3 pour cent mille !).

Nous repèrerons le nombre de tirages par une indicatrice. Dans la capture d'écran ci-dessous, il a fallu 2 tirages.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	N° de l'essai	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**Nombre d'essais nécessaire**
2	Indicatrice de succès	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**

Le nombre de tirages se calcule en faisant la somme des produits des valeurs de la ligne des numéros et de celle des indicatrices (fonction **SOMMEPROD**).



	A	B	C	D	E	J	K	L	M	N	O	P	Q	R
1	N° de l'essai	1	2	3	4	9	10	11	12	13	14	15	**Nombre d'essais nécessaire**	
2	Indicatrice de succès	0	1	0	0	0	0	0	0	0	0	0	**:P1;B2:P2)**	

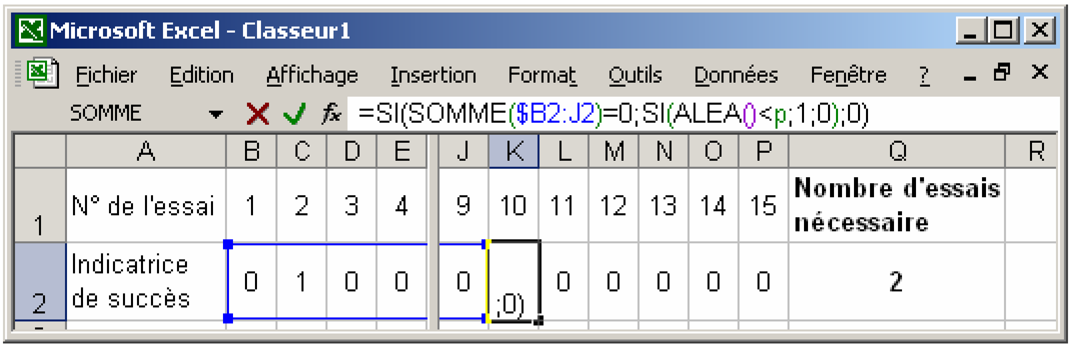
Au premier tirage, l'indicatrice de succès peut être simulée avec la formule suivante (voir page 6) :

=SI(ALEA()<p;1;0)

Pour les tirages suivants, de deux choses l'une :

- Soit l'événement n'a pas été réalisé. Dans ce cas, la probabilité *p* de réussite à chaque tirage restant constante, on reprend la formule du dessus
- Soit l'événement a déjà été réalisé. Dans ce cas, le tirage n'a pas lieu.

Pour voir si l'événement a été réalisé ou non aux tentatives précédentes, il suffit de faire la somme des indicatrices.



Microsoft Excel - Classeur1

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

SOMME =SI(SOMME($B2:J2)=0;SI(ALEA()<p;1;0);0)

	A	B	C	D	E	J	K	L	M	N	O	P	Q	R
1	N° de l'essai	1	2	3	4	9	10	11	12	13	14	15	**Nombre d'essais nécessaire**	
2	Indicatrice de succès	0	1	0	0	0	;0)	0	0	0	0	0	**2**	

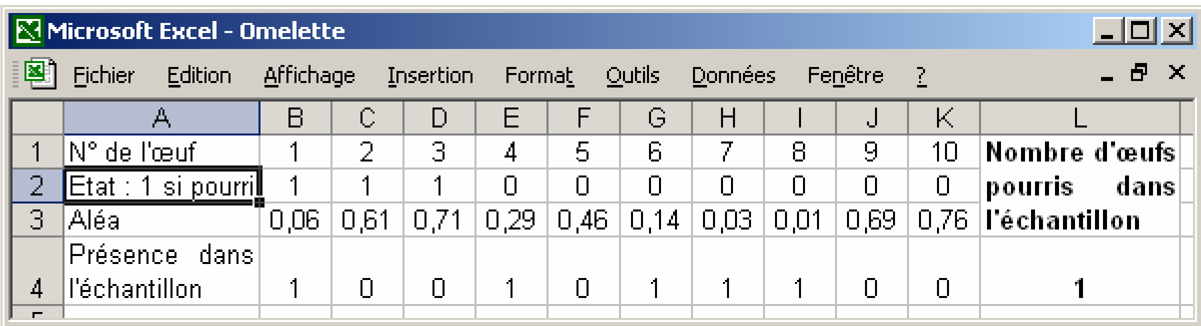
Ces formules sont reprises dans le document ***Billes***. Ouvrez le. Recopiez la plage **A2:Q2** sur un grand nombre de lignes. Représentez la distribution des valeurs observées et vérifiez qu'elle correspond à la loi définie plus haut.

3.3 La loi hypergéométrique à partir d'un tirage sans remise

Dans un panier de 10 œufs, 3 sont pourris. On en prend 5. Combien sont pourris ?

Les initiés ont reconnu la loi « du tirage sans remise », encore appelée « loi hypergéométrique ».

Pour simuler l'exemple, ouvrez le document ***Omelette***.

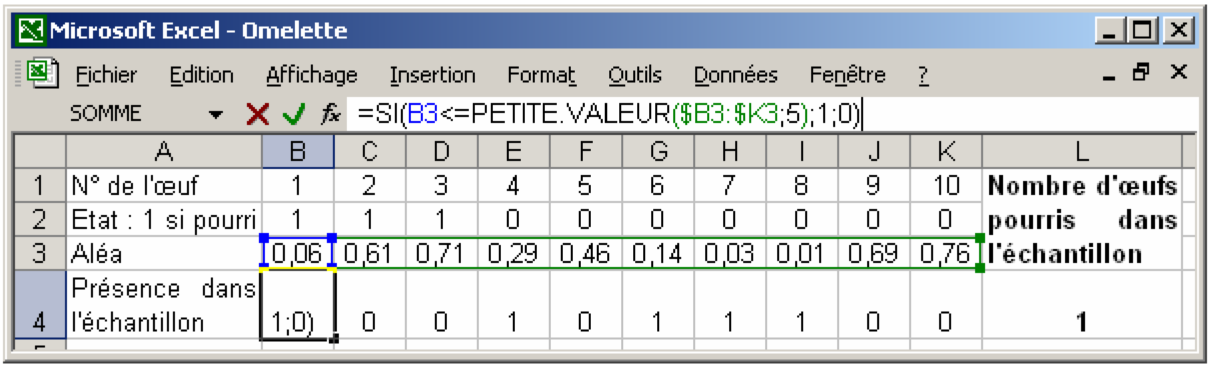


Microsoft Excel - Omelette

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

	A	B	C	D	E	F	G	H	I	J	K	L
1	N° de l'œuf	1	2	3	4	5	6	7	8	9	10	**Nombre d'œufs pourris dans l'échantillon**
2	Etat : 1 si pourri	1	1	1	0	0	0	0	0	0	0	
3	Aléa	0,06	0,61	0,71	0,29	0,46	0,14	0,03	0,01	0,69	0,76	
4	Présence dans l'échantillon	1	0	0	1	0	1	1	1	0	0	**1**

On affecte à chacun des œufs du panier une valeur aléatoire (fonction **ALEA**) et on choisit les 5 premiers.



	A	B	C	D	E	F	G	H	I	J	K	L
1	N° de l'œuf	1	2	3	4	5	6	7	8	9	10	**Nombre d'œufs pourris dans l'échantillon**
2	Etat : 1 si pourri	1	1	1	0	0	0	0	0	0	0	
3	Aléa	0,06	0,61	0,71	0,29	0,46	0,14	0,03	0,01	0,69	0,76	
4	Présence dans l'échantillon	1;0)	0	0	1	0	1	1	1	0	0	**1**

Les 5 premières valeurs sont **0,01** (œuf n°8), **0,03** (n°7), **0,06** (n°1), **0,14** (n°6) et **0,29** (n°4).

Par conséquent, on casse les œufs n° 8, 7, 1, 6 et 4, parmi lesquels se trouve un œuf pourri : le 1.

Le nombre d'œufs pourris est calculé en faisant la somme des produits (fonction **SOMMEPROD**) des indicatrices de l'état et de celles de la présence dans l'échantillon.

Simulez un grand nombre de tirages en suivant la petite astuce de la feuille ***Recopie***. Vérifiez que les fréquences sont proches des probabilités données par la fonction **LOI.HYPERGEOMETRIQUE**.

3.4 Autres lois

- la loi de Laplace-Gauss à partir du théorème de la limite centrée (voir dans le manuel [1] le chapitre « Probabilités et jugement sur échantillon »)
- la loi du khi-deux à partir de la somme des carrés de lois normales standard indépendantes (idem)
- etc.

4 Une application en analyse du risque

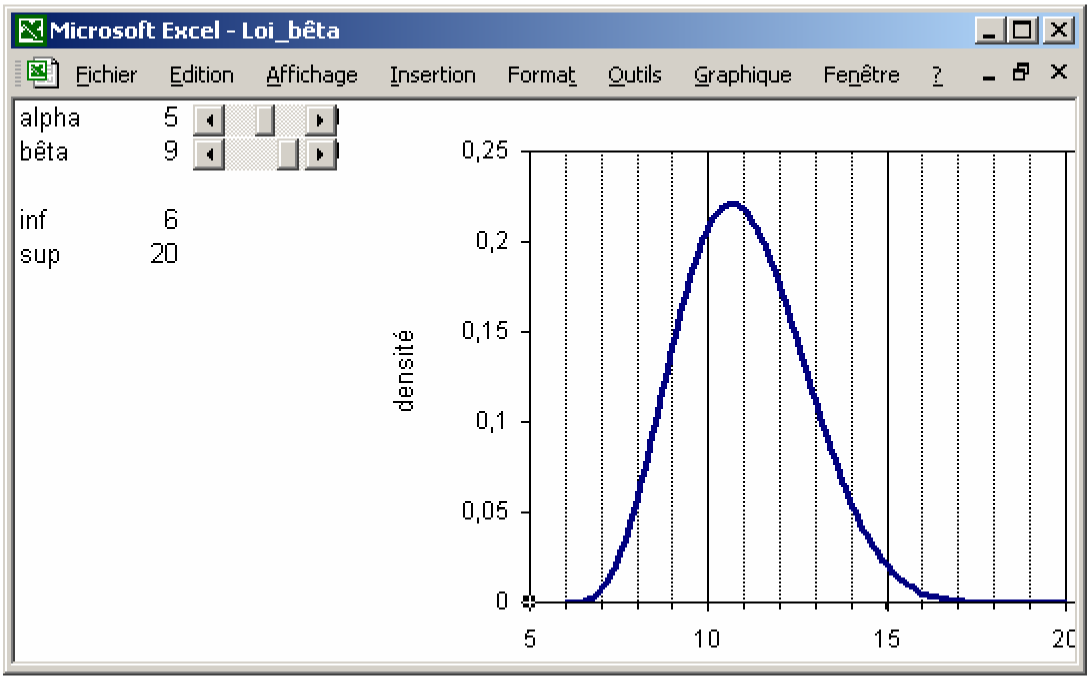
4.1 Le problème – modélisation par la loi bêta

M. Martin doit déposer ses enfants à l'école le matin. Les grilles de l'école ferment à 8 h 30, avec une tolérance de 2 minutes pour les retardataires, c'est-à-dire à 8 h 32 précisément. Quelle est la probabilité de trouver les grilles fermées et, en conséquence, d'avoir à s'expliquer avec la directrice de l'école ?

Cela dépend bien entendu de l'heure de départ. M. Martin pense qu'il a de fortes chances de partir entre 8 h 08 et 8 h 15. En tout cas, il ne partira jamais avant 8 h 06 ni après 8 h 20.

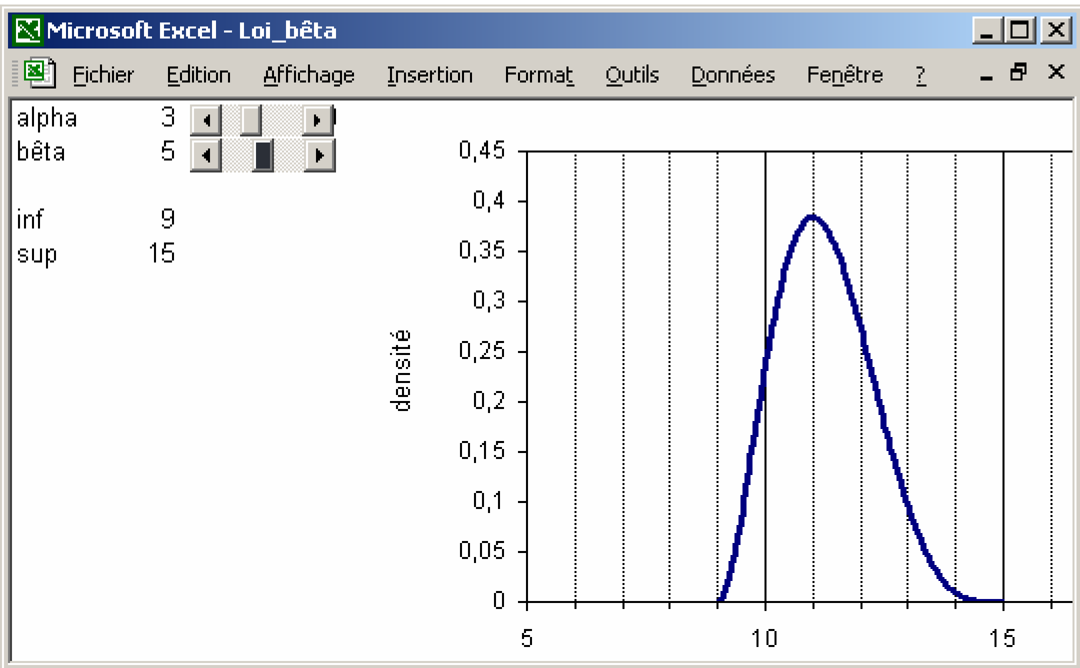
On propose à M. Martin de modéliser son idée sur l'heure de départ par une densité de probabilité.

Ouvrez le document ***loi_beta***. La loi bêta est couramment utilisée en analyse du risque pour modéliser des opinions d'experts en raison de sa souplesse et de sa simplicité. Sa forme dépend de deux paramètres, notés alpha et bêta dans le document. Donnez des valeurs de votre choix aux bornes inférieures et supérieures puis faites varier les paramètres à l'aide des barres de défilement. Observez l'évolution de la courbe de densité de probabilité.

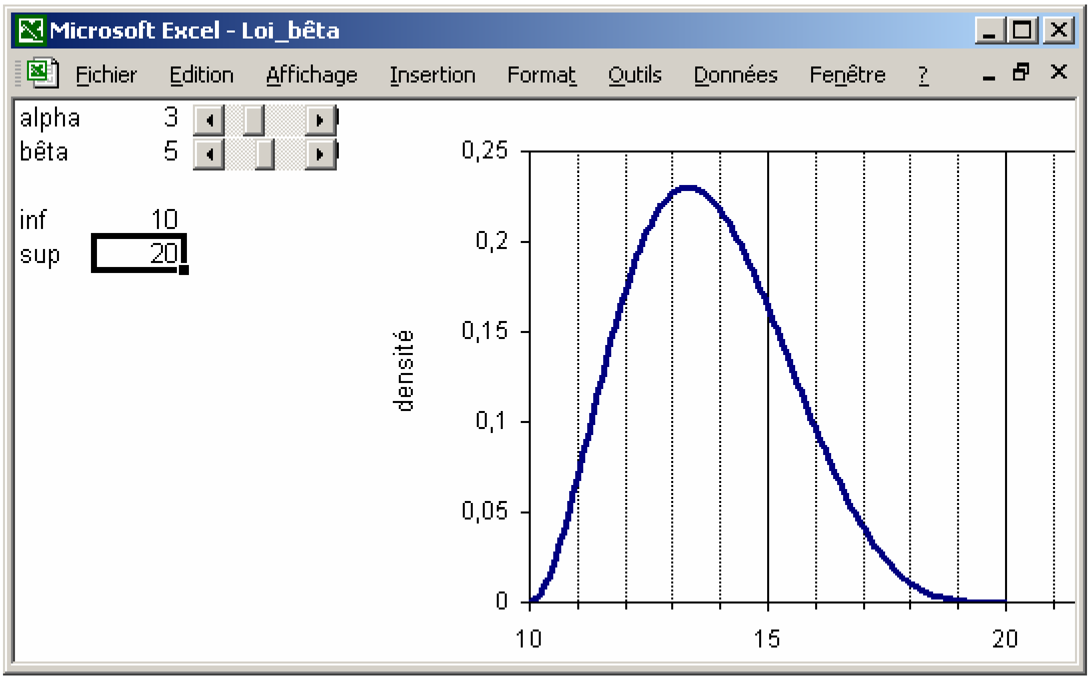


Nous dirons que la courbe du dessus traduit l'opinion de M. Martin : l'heure de départ, exprimée en minutes après 8 h, est comprise en 6 et 20 avec l'essentiel de la probabilité entre 8 et 15. Il s'agit de la densité de la loi bêta de paramètres alpha = 5 et bêta = 9 entre les bornes 6 et 20.

La durée du trajet dépend de l'heure de départ. M. Martin estime que, s'il arrive à partir avant 8 h 13, le trajet doit se faire entre 10 et 13 minutes, au mieux 9 minutes, au pire 15, selon la loi bêta de paramètres 3 et 5 entre les bornes 9 et 15 :



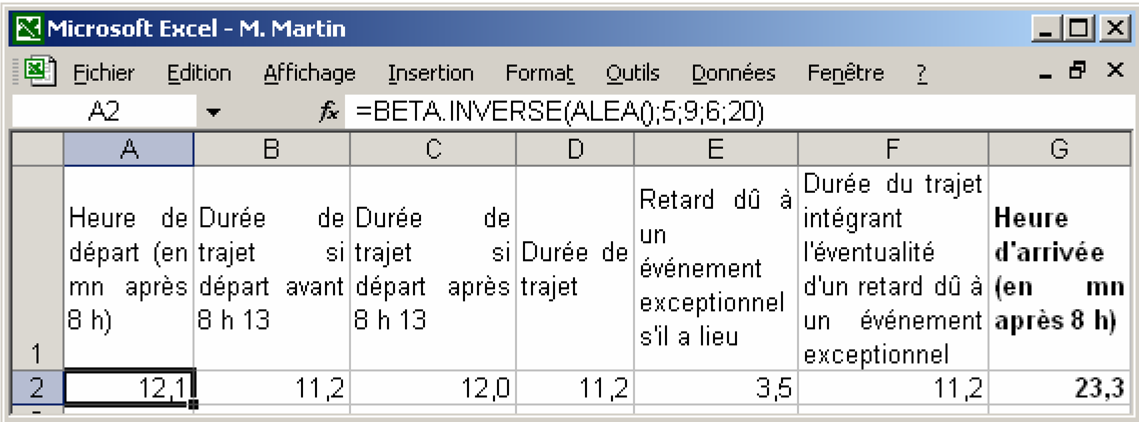
Après 8 h 13, la circulation ralentit. Selon M. Martin, la durée du trajet peut alors fluctuer entre 10 et 20 minutes avec une même répartition de la probabilité (c'est-à-dire les mêmes paramètres, alpha = 3 et bêta = 5, mais entre les bornes 10 et 20) :



Pour tout intégrer, il faut aussi tenir compte d'événements rares mais qui peuvent bloquer la circulation : livreur, déficience des feux de circulation, dispositif de sécurité à l'occasion du passage d'un ministre, etc. M. Martin évalue la probabilité d'un tel événement à 2% et estime que le retard qu'il peut occasionner se situe autour de 6 minutes avec un maximum de 15 minutes, selon une loi de bêta de paramètres 3 et 4 entre 0 et 15.

4.2 Evaluation du risque par simulation

Ouvrez le document ***M. Martin***.

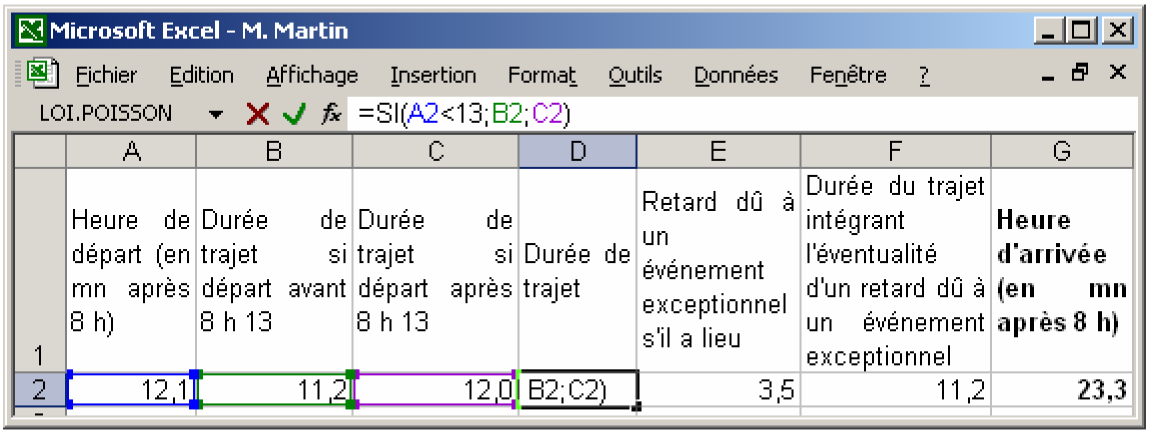

Microsoft Excel - M. Martin
Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?
A2 =BETA.INVERSE(ALEA();5;9;6;20)

	A	B	C	D	E	F	G
1	Heure de départ (en mn après 8 h)	Durée de trajet si départ avant 8 h 13	Durée de trajet si départ après 8 h 13	Durée de trajet	Retard dû à un événement exceptionnel s'il a lieu	Durée du trajet intégrant l'éventualité d'un retard dû à un événement exceptionnel	**Heure d'arrivée (en mn après 8 h)**
2	12,1	11,2	12,0	11,2	3,5	11,2	**23,3**

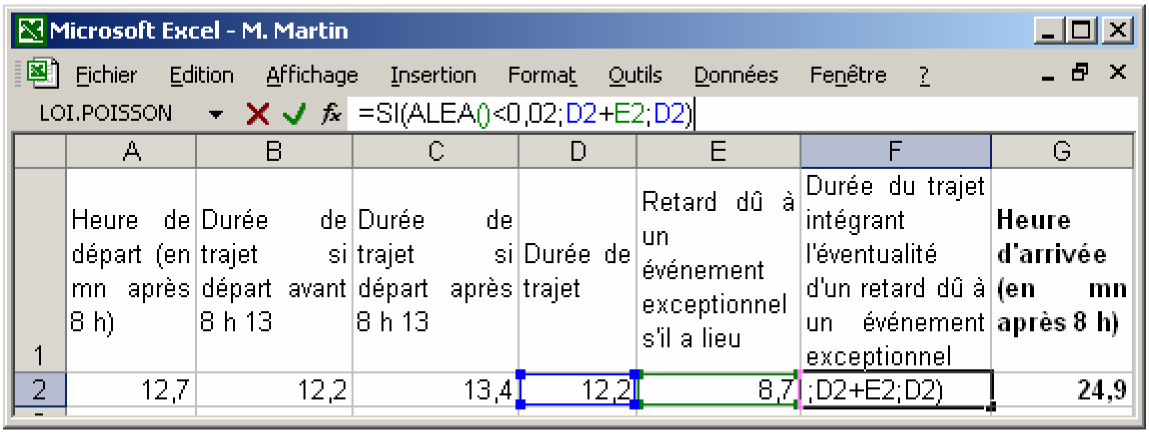
L'heure de départ (colonne **A**) est simulée par la méthode des fractiles : modélisée par la loi bêta de paramètres 5 et 9 entre les bornes 6 et 20, on la simule en utilisant la réciproque de la fonction de répartition (fonction **BETA.INVERSE**) appliquée au résultat de la fonction **ALEA** (voir page 7 et la barre de formule dans la copie d'écran du dessus).

La durée de trajet selon que le départ se fait avant ou après 8 h 13 (colonnes **B** et **C**) est simulée de la même manière.

La durée de trajet retenue (colonne **D**) est celle de la colonne **B** ou celle de la colonne **C**, selon que l'heure de départ est inférieure ou supérieure à 13 minutes après 8 h. Pour les valeurs simulées, elle est égale à la durée de la colonne **C**, 11,2, puisque l'heure de départ, 12,1, est inférieure à 13. Le choix se fait à l'aide de la fonction **SI** :

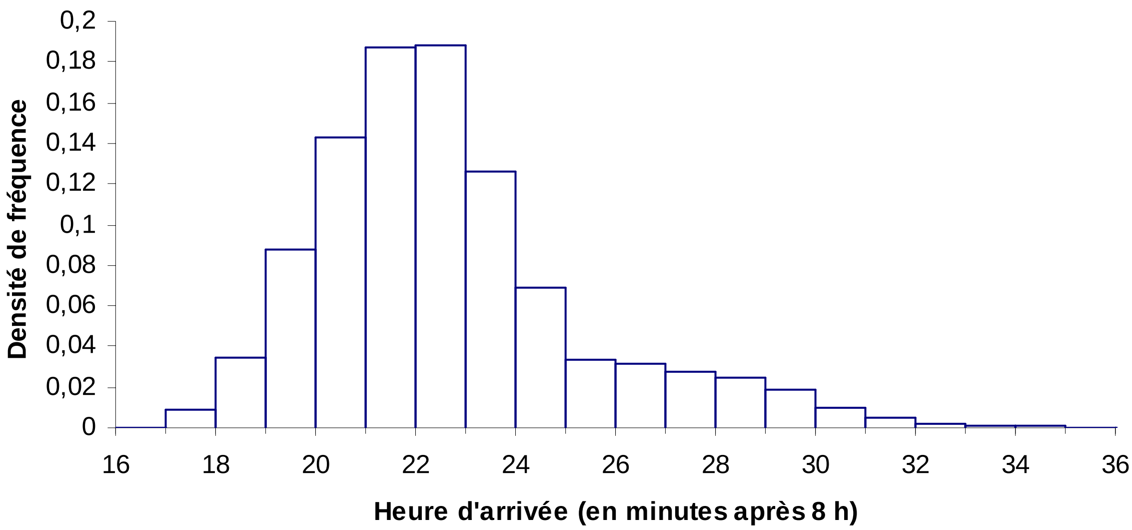


Le retard dû à un événement exceptionnel s’il a lieu est simulé comme l’heure de départ et les durées des colonnes **B** et **C**. On l’ajoute à la durée du trajet si l’événement à lieu, ce qu’on simule comme une variable binaire (voir page 6) :



Enfin, on obtient l’heure d’arrivée en ajoutant l’heure de départ et la durée du trajet.

Simulez un grand nombre de trajets en recopiant la ligne des formules (plage **A2:G2**). Représentez l’heure d’arrivée par un nuage de points ou par un histogramme.



Nous pouvons rassurer M. Martin : la probabilité de déposer ses enfants après 8 h 32 n’est pas nulle mais elle est faible.

5 Références

[1] Le manuel du groupe « Le Cercle d'Excel'ense » : Morineau A., Chatelin Y.-M. (Coordinateurs) – L'analyse statistique des données. Apprendre, comprendre et réaliser avec Excel. Editions Ellipses, 2005. Voir la page Excel'ense sur le site de MODULAD : www.modulad.fr

[2] Description of the RAND function in Excel 2003. Aide et support en ligne de Microsoft. http://support.microsoft.com/default.aspx?scid=kb;en-us;828795&Product=OFF2003

[3] Saporta G. – Probabilités, Analyse des Données et Statistique. Editions Technip, 1990

[4] Goldfarb B., Pardoux C. - Méthodes d'ajustements graphiques : Diagramme Quantile – Quantile. Excel'ense – Modulad n°33, 2005. En ligne sur www.modulad.fr

[5] St@tNet. Cours en ligne créé par le CNAM, l'ENSA de Montpellier et l'Université de Montpellier II. http://www.agro-montpellier.fr/cnam-lr/statnet/

[6] Engineering Statistics Handbook. NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/index.htm

[7] SEL (Statistique En Ligne). Cours en ligne créé par l'INRIA Rhône-Alpes et l'Université René Descartes Paris 5. http://www.inrialpes.fr/sel/index.html