

TRAVAUX DIRIGES SIMULATION

ANNEE 2006-2007

1 ANALYSE D'UN SYSTEME

1.1 Système, entités et variables

- Système = { Supermarché }
- Entités = { caisses, clients }

Les clients sont gérés comme des entités dans le modèle conceptuel, même si l'implémentation que nous avons utilisée ne crée pas d'objets « client ».

Supermarché

Les paramètres du système, qui ne varient pas pendant la simulation, sont

- son heure d'ouverture (10 h) ainsi que son heure de fermeture (20 h),
- le nombre de caisses minimum (3) et maximum (10) et leurs heures d'ouverture et de fermeture (stratégie à définir).

Variables d'état = { clients_caisse[10], nbre_clients, caisse_ouverte[10] }

Variables statistiques = { cumul_clients }

Caisse

Les états possibles sont { ouverte, fermée }

Les paramètres pour l'entité caisse sont :

- numéro de la caisse

Client

Les états possibles sont { courses, caisse } (sinon il n'existe pas)

Les paramètres pour le client sont :

- le type de chariot avec lequel il arrivera en caisse, c'est-à-dire petit, moyen, grand ou sans achat,
- son heure d'arrivée dans le magasin
- son arrivée en caisse (heure d'arrivée + vingt minutes en semaine et trente minutes le samedi, mais on le prend comme un paramètre pour évolution future vers un modèle plus réaliste).

Variables statistiques = { temps_attente_caisse }

1.2 Événements

- Événements = { ouverture_magasin, fermeture_magasin, fin, arrivée_client(id), passage_caisse(id), départ_client(id), ouverture_caisse(n), fermeture_caisse(n), fin }

1.3 Processus déterministes ou considérés comme tels

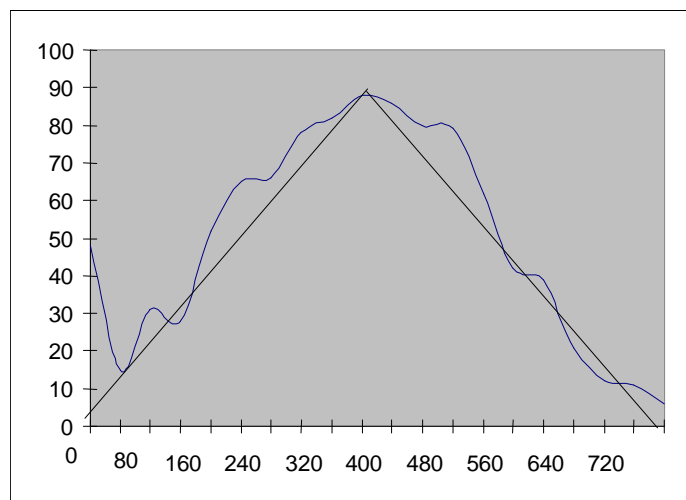
- le choix de la caisse par le client qui se fait en fonction de la longueur de la file d'attente à chaque caisse,
- le temps de remplissage du chariot qui vaut 20 minutes en semaine et 30 minutes le samedi,
- l'ouverture du magasin qui a toujours lieu à 10 heures,
- la fermeture du magasin qui a toujours lieu à 20 heures.
- l'ouverture et la fermeture des caisses s'ils suivent des horaires fixes (s'ils varient en temps réel suivant la fréquentation, alors il s'agit d'une fonction aléatoire très complexe).

1.4 Processus aléatoires

- L'arrivée des clients dans le supermarché suit un processus de Poisson, l'intervalle entre deux événements suit une loi exponentielle.
- Le type de chariot est un processus suivant une loi uniforme.

Type de chariot	%
Petit	32
Moyen	42
Grand	15
Sans achat	11

- La durée de traitement des chariots peut être approchée par une loi triangulaire (a,b,c) :



Type de chariot	a	b	c
Petit	15	40	95
Moyen	45	115	216
Grand	101	210	335

- Les autres événements suivent des lois déterministes

1.5 Fréquentation moyenne des caisses

- du lundi au vendredi

Plage horaire	Clients / heure
10H – 12H	120
12H – 13H	200
13H – 18H	170
18H – 20H	300

- le samedi

Plage horaire	Clients / heure
10H – 12H	180
12H – 13H	300
13H – 18H	255
18H – 20H	450

plage	clients/h	Nombre de caisses ouvertes							
		3	4	5	6	7	8	9	10
10H-12H	150	72,0	96,0	120,0	144,0	168,0	192,0	216,0	240,0
12H-13H	250	43,2	57,6	72,0	86,4	100,8	115,2	129,6	144,0
13H-18H	200	54,0	72,0	90,0	108,0	126,0	144,0	162,0	180,0
18H-20H	350	30,9	41,1	51,4	61,7	72,0	82,3	92,6	102,9

2 PREMIER EXERCICE

Une usine fabrique des rouleaux de pellicules photographiques. On mesure l'épaisseur d des couches sensibles de ces pellicules. Pour cela, on utilise un échantillon

de $N = 620$ pellicules. On répartit les épaisseurs par tranche de $1 \mu\text{m}$ et on désigne par x_i l'épaisseur moyenne de la classe i .

Épaisseur des couches	Nombre de pellicules
$d < 2,5 \mu\text{m}$	8
$2,5 \mu\text{m} \leq d < 3,5 \mu\text{m}$	47
$3,5 \mu\text{m} \leq d < 4,5 \mu\text{m}$	137
$4,5 \mu\text{m} \leq d < 5,5 \mu\text{m}$	211
$5,5 \mu\text{m} \leq d < 6,5 \mu\text{m}$	154
$6,5 \mu\text{m} \leq d < 7,5 \mu\text{m}$	51
$d > 7,5 \mu\text{m}$	12

2.1 Loi a priori

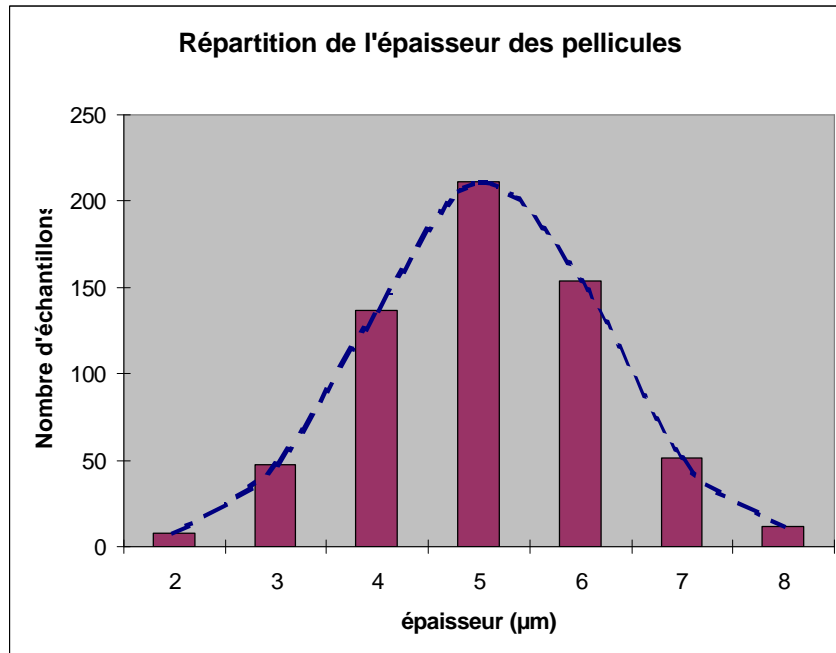
Il s'agit vraisemblablement d'une loi normale. En effet, les variations d'épaisseurs sont le résultat de l'addition des effets de plusieurs facteurs aléatoires indépendants (imprécision de la machine, variation de la composition de l'émulsion, température du local, etc.) dont aucun n'est vraiment prépondérant.

Attention, la lecture des chiffres ou la forme de la courbe est un indice, non une justification : certaines distributions peuvent avoir des courbes similaires.

2.2 Histogramme

Pour la suite, on prendra le milieu de chaque intervalle comme valeur de l'épaisseur x_i .

Nous assimilerons la classe « $< 2,5 \mu\text{m}$ » à une classe « $1,5$ à $2,5 \mu\text{m}$ », et la classe « $> 7,5 \mu\text{m}$ » à la classe « $7,5$ à $8,5 \mu\text{m}$ ». Nous vérifierons plus tard le bien fondé de cette approximation.



L'histogramme *ressemble* bien à une distribution normale.

2.3 Calcul des paramètres

Une loi normale a deux paramètres : l'espérance m et l'écart-type s .

La fonction MOYENNE() d'EXCEL ne permettant pas de calculer une moyenne pondérée, nous devons entrer les calculs :

$$m = \frac{\sum_{i=1}^N n_i x_i}{N}$$

$$s^2 = \frac{\sum_{i=1}^N n_i x_i^2}{N} - m^2$$

$N = \text{population} = 620$

$n_i = \text{taille de la classe } x_i$

Ce qui nous donne :

$$m = 5,06 \text{ et } s = 1,18$$

Dans notre tableau EXCEL, nous nommerons les cellules correspondantes moyenne et sigma.

Inutile de donner 10 décimales ! Ici, 2 c'est suffisant.

2.4 Courbe théorique

Nous utilisons la fonction EXCEL `LOI.NORMALE()` qui renvoie la fonction de densité de probabilité $f(x_i)$ ou la fonction de répartition $F(x_i)$ suivant que le quatrième paramètre (de type booléen) est respectivement à `faux` ou `vrai`.

Pour chaque classe x_i la valeur théorique de n_i est :

$$n_i = N * \text{LOI.NORMALE}(x_i; m; \mathbf{S}; \text{faux})$$

Dans le tableau ci-dessous, nous avons :

x : classe

N : effectif observé de la classe

pT : densité de probabilité théorique $f(X)$

T : effectif théorique

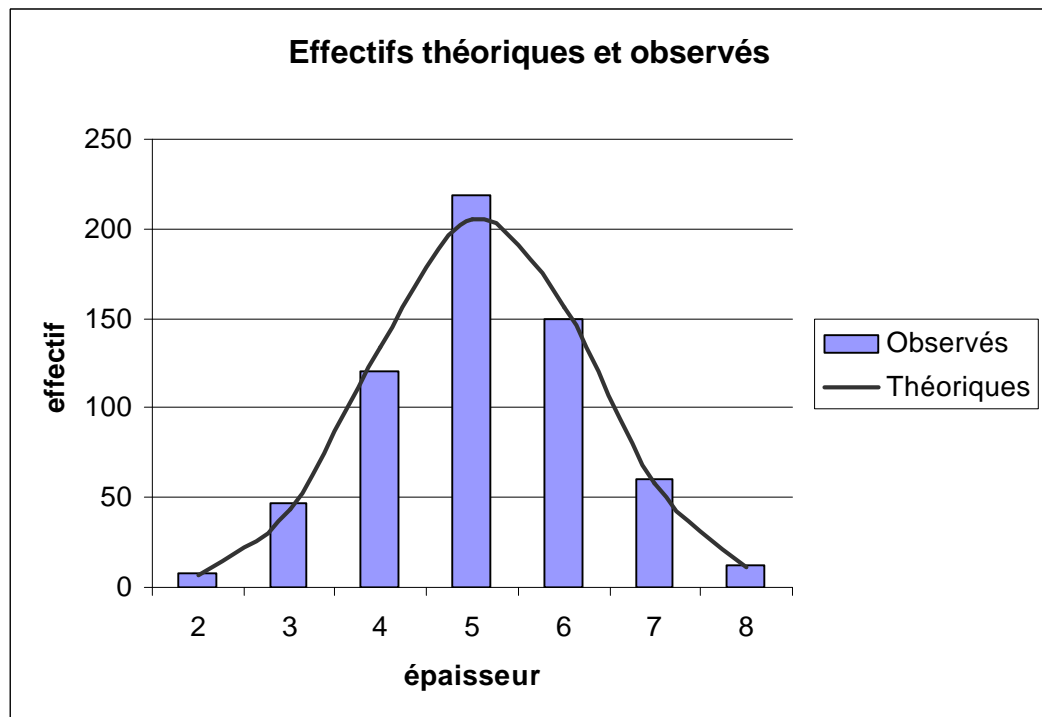
x	N	pT	T
2	8	0,01162683	7
3	47	0,07347654	46
4	137	0,22595827	140
5	211	0,33814253	210
6	154	0,24624271	153
7	51	0,08726076	54
8	12	0,01504758	9

Notons qu'ici le fait de prendre la valeur de la densité de probabilité au milieu de chaque intervalle est une approximation : pour être rigoureux, nous devrions calculer la probabilité d'avoir l'épaisseur de l'émulsion compris dans l'intervalle donné (en fait, nous assimilons la fonction $f(x)$ à un segment de droite à l'intérieur de chaque classe).

Note sur l'interprétation de pT : On a par exemple 33,8% de chances qu'une pellicule ait une épaisseur comprise entre 4,5 μm et 5,5 μm . Il s'agit du produit de la fonction de distribution par la largeur de l'intervalle. Notons que cela revient à intégrer $f(x)$ entre les valeurs 4,5 et 5,5. On aura pu également calculer $F(5,5)-F(4,5)$ (on a à peu près les mêmes valeurs).

Nous avons fait précédemment une approximation sur les classes extrêmes. Si nous calculons la somme des pT nous obtenons 0,998. Les 0,2% qui manquent sont les effectifs théoriques des classes de moins de 1,5 μm ou de plus de 8,5 μm , ce qui est négligeable et fonde notre approximation. Si cette proportion n'avait pas été aussi faible, nous aurions dû reprendre le problème, par exemple en rejetant ces classes extrêmes, et en ne travaillant que sur les 5 autres.

L'histogramme est ainsi :



Note : ici la courbe des effectifs théoriques a été lissée.

2.5 Test du χ^2

EXCEL a une fonction `TEST.KHIDEUX()` qui compare directement deux séries de données.

Comparons les effectifs théoriques et observés.

EXCEL nous renvoie `KHI2 = 0,979`

Le test du χ^2 d'EXCEL est un peu élémentaire, par exemple il prend comme degrés de liberté le nombre de ligne moins 1, donc ici $7-1 = 6$. Or, les paramètres de la loi normale lient les variables entre elles, il n'y a donc que $6-2 = 4$ degrés de liberté.

Effectuons le test du χ^2 « à la main » :
$$c^2 = \sum_{i=1}^N \frac{(n_i - Np_i)^2}{Np_i}$$

On trouve : $\chi^2 = 1,16$ avec $v = 4$ degrés de liberté.

La table de la distribution χ^2 vue en cours nous donne les valeurs de χ^2 ayant la probabilité \mathbf{a} d'être dépassées (\mathbf{a} est appelé seuil de signification). Pour $\mathbf{a} = 0,01$, on a $\chi_{0,01}^2 = 13,277$.

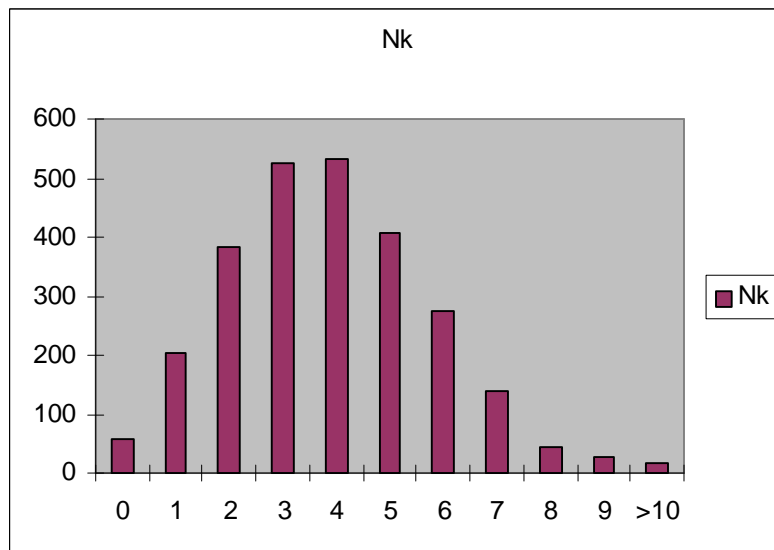
$\chi^2 < \chi_{0,01}^2$, donc la différence entre la distribution observée et la distribution théorique n'est pas significative.

La loi que nous avons trouvée « colle » donc bien à l'expérimentation.

3 DEUXIEME EXERCICE

$$E(x) = 3,87$$

$$\sigma = 4,31$$



Les conditions d'un processus de Poisson sont réalisées : la probabilité d'arrivée d'une particule durant une petite période est proportionnelle à la durée de cette période, les arrivées sont indépendantes, la probabilité de survenue de deux événements durant cet intervalle est négligeable.

$$\Pr\{X = x\} = \frac{e^{-I} I^x}{x!} \quad \text{et} \quad E(X) = I \quad \text{et} \quad V(X) = I$$

Classe	Probabilité	Effectifs Théoriques
0	0,02085837	54
1	0,08072189	211
2	0,15619686	407
3	0,20149394	525
4	0,19494539	508
5	0,15088773	394
6	0,09732259	254
7	0,05380549	140
8	0,0260284	68
9	0,01119221	29
10	0,00433139	11

4 TROISIEME PROBLEME

M. Charles Attand possède un magasin de matériel vidéo. L'un des produits phares de son commerce est le lecteur de DVD Shitsu KRX-2000 (prix : 5 000 F), dont il vend un exemplaire tous les deux jours environ.

4.1 Approvisionnement fixe

M. Attand est réapprovisionné tous les mois (un mois = 30 jours) d'une quantité correspondant aux ventes moyennes, quelque soient les ventes du mois précédent. Le stock initial est de 15 lecteurs.

Le problème n'était pas formulé de façon très détaillée. Ceci devait vous pousser soit à me poser des questions durant les séances, soit à formuler explicitement des hypothèses... pas trop farfelues !

4.1.1 Analyse

Nous nous intéressons ici au stock, qui sera notre système. Il est décrit par une variable d'état unique, qui est la quantité de lecteurs restant, laquelle ne peut être inférieure à zéro. L'objectif est de simuler les évolutions de ce stock au cours du temps, afin d'évaluer différentes stratégies de gestion du stock et d'optimiser celle-ci.

On remarque que dans l'énoncé du problème les week-ends et les nuits ont été négligés. On suppose donc que le magasin est ouvert $7j/7$. Pour les nuits, on les négligera simplement, cela ne devrait pas nous gêner, le temps n'étant pas une variable d'état du système.

L'arrivée des clients est un processus stochastique, qu'il va nous falloir modéliser. Nous avons deux solutions : soit modéliser globalement le nombre de clients par mois (typiquement par une loi de Poisson), soit modéliser le temps qui s'écoule entre deux clients (loi exponentielle). La première est plus simple, mais compte tenu de la question suivante, nous choisirons la seconde.

Nous allons donc devoir construire une simulation à événements discrets.

Le modèle est simple à construire : nous allons générer une série d'événements séparés par un temps t régi par une loi exponentielle de paramètre I :

$$I = \frac{1}{E(X)} = \frac{1}{2} = 0,5 \quad (E(X) \text{ est l'intervalle moyen entre deux clients, soit 2 jours})$$

La densité de probabilité d'une loi exponentielle est :

$$f(x) = I e^{-Ix} \quad (x \geq 0, I > 0)$$

$$\text{soit } f(x) = 0,5 e^{-0,5x}$$

J'ai pu voir pendant les séances de BE que certains étaient tentés de simuler l'arrivée des clients à l'aide d'une v.a. uniforme. Il suffit de calculer quelques valeurs pour constater l'écart important entre les lois uniformes et aléatoires ! Ce n'est donc absolument pas satisfaisant. En revanche, dans la question suivante, l'approximation de la loi normale par une loi triangulaire, beaucoup plus simple à calculer, est tout à fait valable.

Dans l'ensemble, l'analyse et la justification de la modélisation étaient légères. Peu ont pensé à la validation du modèle. Je n'en ai pas tenu compte, dans la mesure où celui-ci était tout de même fort simple.

4.1.2 Simulation d'un an de vente

4.1.2.1 Modélisation

Je donne ici la méthode que j'ai utilisée, mais il est bien évident que, comme souvent en programmation, il peut y avoir plus chemin pour parvenir à la solution.

Il nous faut maintenant construire un générateur aléatoire exponentiel. Or, l'outil que nous souhaitons utiliser, EXCEL, n'a qu'un générateur uniforme $RU(0,1)$ avec la fonction `ALEA()`.

Nous allons pour cela utiliser la méthode de la transformée inverse vue en cours :

$$\text{Fonction de répartition } F(x) = \int_{-\infty}^x f(z) dz$$

$$\text{Soit } U = RU(0,1)$$

$$\Pr\{X \leq x\} = \Pr\{F^{-1}(U) \leq x\} = \Pr\{U \leq F(x)\}$$

Le terme de droite est la fonction de répartition d'une distribution uniforme évaluée à $F(x)$.

Pour une distribution exponentielle :

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & (x > 0) \\ 0 & \text{sinon} \end{cases}$$

On cherche à générer x tel que $u = F(x)$

$$x = F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$$

$$X = -\frac{1}{\lambda} \ln(1 - U)$$

$$\boxed{t = -2 \ln(1 - U)}$$

Nous pouvons donc générer de proche en proche, avec EXCEL ou un programme en C par exemple, les dates des événements correspondant à l'arrivée d'un client et chaque ligne du tableau correspondant à l'un de ces événements. Nous nous arrêterons lorsque le temps $\{t = \sum t\} > \{12 \times 30 = 360\}$.

Le temps de chaque événement est :

$$t_n = t_{n-1} - 2 \ln[1 - RU(0,1)]$$

(pour le temps t_1 du premier client, nous prendrons $t_0 = 0$).

Nous allons maintenant voir comment faire évoluer les stocks. Ceux-ci sont régis par les règles suivantes :

- Si on vient de changer de mois, prendre en compte une livraison.
- Lorsqu'un client arrive, décrémenter le stock
- Sauf si le stock est nul, auquel cas il y a un « client insatisfait ».

Les possibilités de tests et de calculs étant limités dans EXCEL, et ne souhaitant pas avoir de formules trop compliquées, j'ai créé, outre les colonnes « date » et « stock », une colonne pour le mois (partie entière de la date) et une colonne « livraison », qui vaut 15 si le mois de la ligne courante et le mois de la ligne précédente sont différents, 0 sinon.

Pour avoir le stock, il suffit d'une simple condition :

SI (stock précédent + livraison) > 0

ALORS stock := stock précédent + livraison - 1

SINON stock := stock précédent

Pour pouvoir comptabiliser automatiquement les clients insatisfaits, j'ai créé une colonne « clients insatisfaits » :

SI (stock = stock précédent) ET (date < 360)

ALORS client insatisfait = 1

SINON client insatisfait = 0

Il me suffit alors de faire la somme de cette colonne pour obtenir le nombre de clients insatisfait durant l'année.

4.1.2.2 Exécution

On force le recalcul de la feuille en appuyant sur F9. On relève à chaque fois le nombre de clients insatisfaits (analyse quantitative) et on jette un œil aux stocks au cours de l'année (analyse qualitative).

4.1.2.3 Analyse de la simulation :

On remarque qu'entre deux simulations les variations sont importantes, de même que les stocks, qui atteignent parfois 40 unités...

Le nombre *NI* de clients insatisfaits est souvent important. Exécutons 20 fois la simulation.

18 ; 20 ; 5 ; 5 ; 8 ; 20 ; 0 ; 9 ; 5 ; 6 ; 16 ; 19 ; 5 ; 11 ; 19 ; 13 ; 13 ; 1 ; 13 ; 4

Moyenne *NI* = 10,5.

Outre des stocks souvent excessifs, le commerçant perd en moyenne 50 000F de chiffre d'affaire par an, ce qui indique que sa gestion de stocks est médiocre.

4.2 Amélioration de la gestion du stock

Mêmes questions, mais cette fois on effectue une commande de 12 unités dès que le stock tombe en-dessous de 4. Le délai de livraison est en principe de 7 jours, mais une variation d'environ ± 2 jours est à attendre en pratique.

Cette nouvelle stratégie est-elle payante ?

4.2.1 Analyse

Ayant constaté la médiocrité de la gestion de stock, on décide d'assouplir celle-ci en commandant les lecteurs à meilleur escient.

On remarque une ambiguïté potentielle : « quand le stock tombe en-dessous de 4 ». Nous l'interpréterons comme « quand le stock est strictement inférieur à 4 ».

On rajoute ici une nouvelle variable aléatoire, qui est le délai de livraison t_d . Celui-ci est variable de 5 à 9 jours, avec une moyenne de 7 jours.

Ce délai de livraison varie suite à l'accumulation de variables aléatoires indépendantes entre elles : temps de traitement, niveau des stocks du fournisseurs, embouteillages éventuels, disponibilité du livreur, etc. C'est un cas typiquement régi par une **loi normale**.

4.2.2 Simulation

4.2.2.1 Modélisation

Nous devons donc créer un générateur aléatoire permettant de modéliser les variations du délai de livraison.

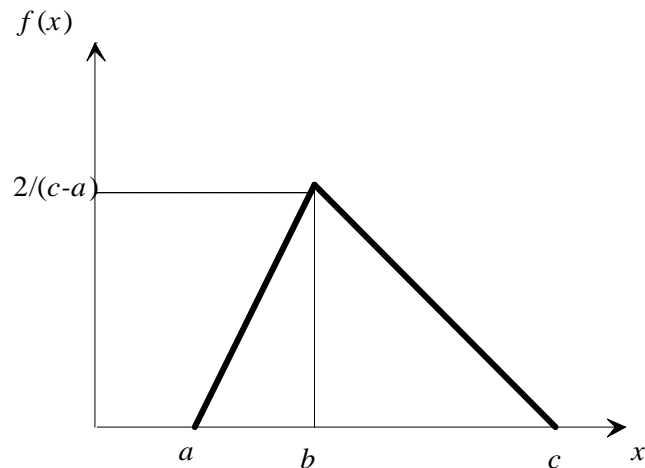
Compte tenu de l'intervalle possible des valeurs, des paramètres de la loi et de la (faible) précision requise dans cette étude, nous nous permettons d'approcher la loi normale par une **loi triangulaire**, nettement plus simple à calculer.

Nous déterminons ce générateur comme précédemment, par la méthode de la transformée inverse :

Densité :

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq b \\ \frac{2(c-x)}{(c-a)(c-b)} & b \leq x \leq c \\ 0 & \text{sinon} \end{cases}$$

La courbe de la densité de probabilité donne :



La fonction de distribution est :

$$F(x) = \begin{cases} 0 & (x < a) \\ \frac{(x-a)^2}{(b-a)(c-a)} & (a \leq x \leq b) \\ 1 - \frac{(c-x)^2}{(c-a)(c-b)} & (b \leq x \leq c) \\ 1 & (x > c) \end{cases}$$

Inversons F et calculons X en fonction de U :

$$\mathbf{b} = \frac{b-a}{c-a}$$

$$U = \text{rnd}(0;1)$$

$$\text{si } (U < \mathbf{b}) \quad T = \sqrt{\mathbf{b}U}$$

$$\text{sinon } T = 1 - \sqrt{(1-\mathbf{b})(1-U)}$$

$$X = a + (c-a)T$$

Application numérique :

$$a = 5 ; b = 7 ; c = 9$$

$$\mathbf{b} = 0,5$$

$$\text{Si } (U < \mathbf{b}) \quad X = 5 + \sqrt{8U}$$

$$\text{Sinon } X = 9 - \sqrt{8(1-U)}$$

Il est très compliqué avec EXCEL de superposer les deux variables aléatoires. Aussi, nous prendrons une solution « semi-automatique », en générant une série de valeurs pour les délais de livraison, et en reportant ces délais « à la main » dans un tableau EXCEL similaire à celui de la première question, mais dont les valeurs aléatoires auront été « figées » (en copiant / collant uniquement les valeurs).

Ceci est un peu fastidieux, mais pour une exécution sur six ou douze mois, c'est probablement moins coûteux que de programmer EXCEL. C'est là que nous y aurions gagné d'utiliser un programme C par exemple.

J'ai bien entendu utilisé une programmation en C pour cela, avec une gestion de files d'attente d'événements, ce qui n'est pas beaucoup plus long à mettre en œuvre et infiniment plus rapide et plus souple, notamment pour tester des hypothèses.

Notons que certains ont eu l'excellente idée d'utiliser les mêmes dates d'arrivée des clients pour les deux parties du problème, ce qui permettait une comparaison plus facile des deux méthodes.

4.2.2.2 Exécution

Pour pouvoir comparer, je prends une « photo » d'une simulation avec les paramètres de la première question, et, en utilisant les mêmes dates d'arrivée des clients, je détermine les dates de livraison et les stocks suivant les hypothèses de la question 2 et je compare l'efficacité.

La simulation de la question 1 que j'ai étudiée est un cas moyen : 11 clients insatisfaits, et des stocks d'en moyenne 10 unités (maximum 21). On a vu bien pire !

Avec les hypothèses de la question 2 et une série de valeurs aléatoires pour les délais de livraison proches de la valeur moyenne (la série utilisée a une moyenne de 7,09 jours), on a un stock moyen de 5,4 (et maxi de 15 bien entendu), ce qui facilite la vie de notre commerçant (après tout, il n'a qu'une capacité de stockage limité, et a probablement de nombreux autres modèles à vendre). Mais le nombre de clients insatisfaits est de... 29 ! Soit une perte de chiffre d'affaire de 145 000 F, dur dur !

En fait, en regardant le tableau, on voit que c'est la quantité commandée qui est insuffisante. Il faut soit commander lorsque le stock tombe en-dessous de 5 ou 6, ou alors commander plus. Comme on peut supposer que chaque livraison entraîne des frais de transporteur, on va augmenter le nombre d'unités livrées. Ça tombe bien, c'est un des paramètres de mon tableau : il suffit de changer la valeur dans la case correspondante.

Avec 13 unités commandées, on tombe à 14 clients insatisfaits. Trop grand.

Avec 14 unités commandées, on est à 1 client insatisfait (mais un stock moyen de 9,9 unités avec un maximum à 21 !).

Au-delà, le stockage devient trop problématique.

Conclusion : il faut commander 14 unités.

On pourrait de la même manière voir ce que donne l'augmentation du seuil de passage des commandes, puis faire varier ces deux paramètres pour trouver les valeurs idéales, mais ceci dépasse un peu le cadre de la question posée et est assez difficile sans programmation.

Les résultats de ce problème montrent bien qu'il faut se méfier de l'intuition, puisque la plupart d'entre vous ont été très surpris des résultats de cette seconde partie ! Comme certains d'entre vous l'ont fait très justement remarquer, cette dernière méthode a toutefois un avantage : la dispersion des résultats est beaucoup plus réduite d'une exécution de la simulation à l'autre, ce qui signifie qu'il est plus facile d'en prévoir le résultat. M. Attand peut trouver cela avantageux.

Globalement, j'attendais à la fois des chiffres (le client veut toujours du concret, et notamment le coût en chiffre d'affaire) et des commentaires, je n'ai pas toujours eu les deux à la fois.