

DÉCISION AVANCÉE

GEM, SDC-3A

Jean-Philippe Rennard

Table des matières

1	Modélisation et simulation	5
1.1	La démarche de modélisation d'un problème	6
1.1.1	Formalisation	6
1.1.2	Identification de solutions	13
1.1.3	Validation	14
1.2	Exemple 2 : Recherche d'une valeur cible	14
1.3	Exemple 3 : Tables et simulations	17
1.4	Exemple 4 : Calcul itératif	18
2	Prévision	22
2.1	Mesure de l'erreur	22
2.2	Les séries chronologiques	23
2.3	Lissage des séries chronologiques	24
2.3.1	Les moyennes mobiles	25
2.3.2	Le lissage exponentiel	26
2.4	Décomposition	28
2.4.1	Moyennes mobiles et coefficients saisonniers	29
2.4.2	Recherche de la tendance	30
2.4.3	Calcul des prévisions	31
2.4.4	Récapitulation de la méthode	33
2.4.5	Lissage exponentiel avec tendance et saisonnalité	33
2.4.6	Le modèle de Holt	33
2.4.7	Le modèle de Holt-Winters	34
2.4.8	Récapitulation de la méthode	36
2.5	Méthodes autorégressives	38

TABLE DES MATIÈRES

3	Décisions et arbres d'analyse	40
3.1	Décisions uniques	40
3.2	Méthodes probabilistes	42
3.3	Arbres de décision	43
3.3.1	Décisions successives	47
3.3.2	Fonctions d'utilité	53
4	Simulations	57
4.1	Exemple d'application	58
4.2	Tableur et séries aléatoires	61
4.3	Allocation d'un portefeuille d'investissement	62
4.4	Événements discrets	63
4.4.1	Simulation discrète par horloge	64
4.4.2	Simulation discrète par événements	66
4.5	Adéquation de la distribution	68
4.5.1	Adéquation à une loi de Poisson	70
4.5.2	Adéquation à une loi normale	71

TABLE DES MATIÈRES

AVEC L'EXPLOSION de la puissance informatique et l'effondrement des coûts associés, la simulation a envahi tous les domaines de la science et on la rencontre désormais au quotidien, consciemment ou non. Les prévisions météo sont ainsi basées sur l'utilisation de modèles mathématiques et informatiques simulant la dynamique des masses d'air ; l'État élabore son budget à partir de modèles permettant de simuler le comportement de l'économie nationale ¹ ; des simulateurs de trafic routiers permettent de limiter les bouchons ²...

Dans le domaine de la gestion, modélisation et simulation jouent également un rôle croissant. De la simple modélisation du calcul du résultat net à la simulation complexe des flux logistiques ou du trafic télécom, les outils informatiques permettent de prendre des décisions éclairées.

La décision est au cœur du quotidien du manager. Du choix du mode de commissionnement des commerciaux à la validation des décisions d'investissement en passant par l'organisation d'un système de distribution ou l'arbitrage make ou buy, il est confronté à des choix parfois difficiles, car complexes à évaluer. Même quand l'analyse est simple, l'incertitude de l'environnement opacifie souvent les effets potentiels des différentes options.

Le tableur, trop souvent utilisé comme simple outil de calcul et de présentation est un instrument puissant de modélisation et de simulation. Sa maîtrise est un avantage essentiel pour le décideur en ce qu'elle lui permet d'avoir une vision concrète des conséquences de ses décisions. Même si, fondamentalement, la décision restera dirigée par l'intuition, ces approches quantitatives permettront d'en éclairer les suites.

Les outils présentés dans ce cours sont simples et directement opérationnels. Nul besoin d'avoir fait des mathématiques avancées ou de maîtriser des langages de programmation complexes pour les comprendre et les mettre en œuvre. Au contraire, les méthodes présentées, outre leur apport général en termes de méthodologie d'analyse et de résolution de problèmes, ne nécessitent qu'un peu de réflexion associée à une connaissance de certains outils standards — mais peu utilisés — des tableurs. Notre insistance sur la simplicité des modèles n'a pas seulement pour objet de rassurer l'étudiant, elle est en effet intrinsèque à la logique même des instruments présentés, la règle d'or de la modélisation restant le fameux *KISS Keep It Simple Stupid*. Quelques points pour terminer :

- Les outils présentés pourront être mis en œuvre aussi bien avec Excel,

¹On peut ainsi tester ses capacités à exercer le ministère des finances à partir d'une simulation de l'économie anglaise : <http://www.bized.ac.uk/virtual/economy/index.htm>

²Par exemple : <http://www.i3s.unice.fr/ol/CD/Mascotte/node8.html>

TABLE DES MATIÈRES

qu'avec OpenOffice ou Gnumeric, ces deux derniers étant compatibles avec le premier. Nul impératif donc à investir dans un logiciel commercial onéreux pour aider à la décision.

- Ce texte est conçu pour pouvoir être imprimé afin d'être lu dans les meilleures conditions. Il s'accompagne toutefois de quelques vidéo et captures d'écran montrant l'utilisation de certaines fonctions. Il suffit de cliquer sur la caméra dans la marge pour voir les vidéo correspondantes.
- Un fichier tableur est associé à ce texte.

1 Modélisation et simulation

MODÉLISER c'est construire un artefact qui reproduit le fonctionnement du système étudié. Il ne s'agit pas de recopier au plus près l'original, mais de considérer les seuls éléments pertinents de la dynamique analysée. La simulation en est une extension. Basée sur le modèle construit, elle ne se contente pas des paramètres environnementaux courants, mais les étend à des situations diverses. En ce sens, elle autorise la mesure des conséquences de certaines évolutions ou de certaines décisions. La simulation permet :

- de réaliser des *expériences* impossible à réaliser sur le système réel. C'est le cas par exemple de la simulation des déplacements urbains ou de la simulation des comportements de foules. On a alors besoin d'une connaissance précise des mécanismes étudiés.
- de tester des hypothèses afin de valider un modèle théorique. C'est notamment le cas de toutes les simulations en sciences sociales.

Le développement de la simulation est intimement lié aux développements de l'informatique³. Elle connaît ainsi une prospérité inégalée depuis le début des années 1990 du fait de l'explosion de la puissance informatique. Dans les sciences, elle est désormais massivement utilisée pour l'analyse de ces *systèmes complexes* pour lesquels il n'existe pas de solutions analytiques aux équations les décrivant.

En sciences sociales, la simulation est utilisée tant à des fins théoriques que pratiques.

En gestion, hors les simulations dont l'objectif est strictement scientifique, modélisation et simulation sont utilisées :

- pour rechercher des solutions à des problèmes donnés.
- pour étudier la sensibilité des résultats aux différentes décisions.
- pour analyser les processus aléatoires.
- pour étudier les processus complexes (logistique, gestion de production...).

Nous verrons par la suite comment mettre en œuvre les outils associés avec un tableur.

³La conception du premier ordinateur moderne par John Von Neumann et ses collègues avait ainsi pour objectif de simuler certaines lois de la physique nucléaire dans le cadre du projet Manhattan. Votre ordinateur reste conçu selon l' *architecture de Von Neumann*. L'apport essentiel du chercheur d'origine hongroise a consisté à traiter les programmes au sein de la mémoire de la même manière que les données, alors qu'auparavant la programmation s'appuyait sur des câblages physiques.

1.1 La démarche de modélisation d'un problème

1.1 La démarche de modélisation d'un problème

La modélisation d'un problème passe par trois étapes principales :

- la formalisation.
- la recherche de solution.
- l'interprétation des résultats.

Chacune de ces étapes est elle-même formée de plusieurs sous-étapes (Figure 1, page 7) .

Nous allons reprendre chacune des étapes au travers d'un problème simple afin d'illustrer la démarche.

À l'heure du renouvellement du parc automobile, le responsable d'un service logistique souhaite vérifier l'impact de son choix sur le résultat. À partir des données commerciales, l'alternative porte sur quatre modèles dont les spécifications sont les suivantes (tableau 1) :

Modèle	Coût km	Capacité
1	0.16	18
2	0.08	8
3	0.14	15
4	0.17	22

TAB. 1 – Véhicules disponibles

1.1.1 Formalisation

Identification du problème : Première étape, évidente en apparence : *l'identification et la définition du problème.*

Il s'agit là d'identifier de manière stricte le type de problème posé. Cette phase peut sembler évidente, mais dans les faits, elle est souvent délicate. Elle nécessite notamment de prendre en compte les arbitrages entre objectifs contradictoires (les financiers souhaitent minimiser les stocks alors que les commerciaux souhaitent améliorer le taux de service ; la RH cherche à minimiser la masse salariale alors que les responsables de département cherchent à avoir le maximum de ressources humaines pour atteindre leurs objectifs, ...).

Il convient en outre de chercher à dépasser l'apparence pour cerner les critères fondamentaux du problème. Par exemple, dans le cadre d'une analyse de productivité, on pourrait être tenté de centrer la simulation sur l'organisation de la succession des tâches, alors que le problème de fond repose sur la productivité même de certains postes.

1.1 La démarche de modélisation d'un problème

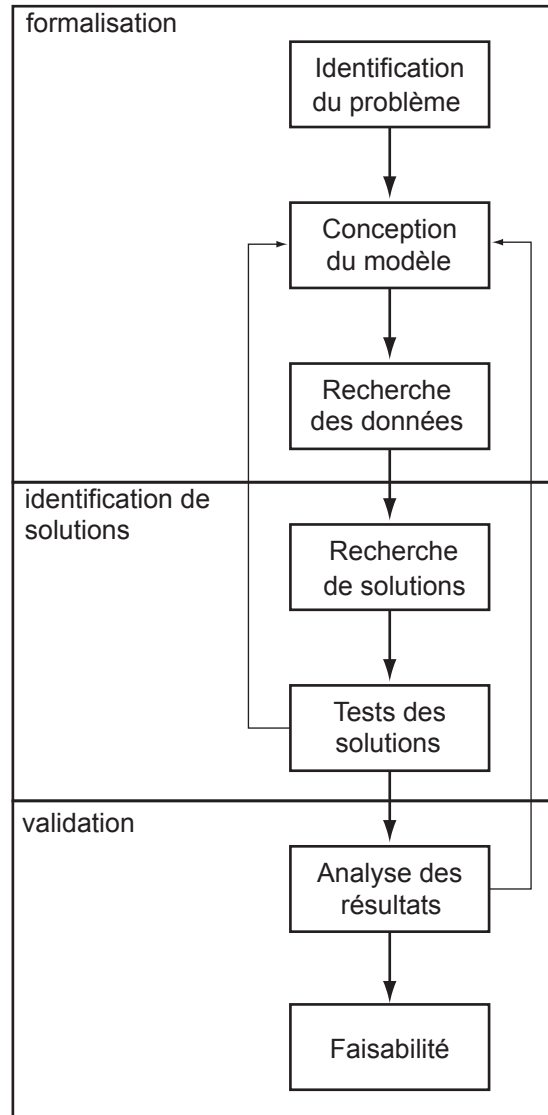


FIG. 1 – Étapes de la modélisation

Notre exemple est élémentaire, il relève du simple calcul et ne pose pas de problème d'identification : il s'agit de calculer le résultat net en fonction de l'incidence du type de véhicule utilisé.

1.1 La démarche de modélisation d'un problème

Conception du modèle : Cette étape consiste en la définition des différents liens et inférences qui dirigent le processus étudié et que l'on cherche à formaliser sous forme d'équations.

Équations, le grand mot est lâché... Il est fondamental de comprendre que le terme équation n'est aucunement lié à la maîtrise d'outils mathématiques avancés. Bien au contraire, on peut souvent formaliser les problèmes quotidiens à partir de ces équations élémentaires que l'on étudie au collège.



On conçoit le modèle en s'appuyant, outre sur l'expertise que l'on a des mécanismes en jeu, sur sa connaissance du contexte :

- Pour les modèles les plus simples, on peut construire les inférences à partir de son expertise empirique.
- Pour les modèles plus complexes, on fait souvent appel aux manuels traitant du domaine afin de disposer de règles rigoureuses. Là encore, on peut amender le modèle théorique en fonction de son expertise du contexte.

De manière générale, le manager est appelé à traiter de nombreux problèmes extrêmement variés. La modélisation doit alors s'adapter au besoin :

- Dans le cas de problèmes complexes et stratégiques, on recherchera une certaine précision. Le modèle sera alors plus difficile et réservé à ceux qui ont une plus grande familiarité avec les mathématiques.
- Dans la grande majorité des cas, le modèle aura essentiellement pour fonction d'orienter et/ou de valider la décision. On recherchera alors la simplicité au détriment de la précision. De manière très générale, il est de bonne pratique de commencer avec un modèle simplifiant fortement la réalité avant d'aller éventuellement au-delà. Dans le domaine de la modélisation comme dans beaucoup d'autres, le mieux est l'ennemi du bien. Partir sur un modèle tortueux oblige souvent à un long travail de recueil de données pour des résultats difficiles à obtenir et à interpréter.

On définit lors de la phase de conception :

- les **paramètres** qui représentent l'environnement ou le contexte.
- les **variables contrôlables** qui définissent les points d'actions, c'est-à-dire ceux qui feront l'objet de la décision (on parle également de variables de décision.).
- les **variables intermédiaires** qui permettent de relier les variables contrôlables aux critères.
- les **critères** qui définissent l'objectif.

1.1 La démarche de modélisation d'un problème

- les **équations du modèle** qui permettent de calculer les valeurs des critères en fonction des paramètres, des variables intermédiaires et des variables contrôlables.

Toutes ces définitions sont *quantitatives*. On verra comment traiter, à travers des variables quantitatives, des données qualitatives comme la région, le jour de la semaine ou le sexe.

Pour construire un modèle fiable, on peut utiliser la méthode des *graphes de dépendances*. Cet outil largement utilisé dans la conception des bases de données permet d'identifier clairement les différentes variables. On commence par construire un *arbre d'analyse*. La méthode est la suivante :

- On pose la racine de l'arbre qui correspond à la (ou les) variable associée à l'objectif (critère).
- On identifie les grandeurs qui déterminent l'objectif, c'est-à-dire celles dont l'objectif dépend.
- Pour chacune des grandeurs identifiées, on recherche celles qui les déterminent.
- On répète l'opération jusqu'à atteindre le niveau de détail souhaité. Ce niveau de détail dépend du niveau de détail que l'on s'est fixé, de la disponibilité des données et du niveau d'action souhaité.

Par exemple, si on travaille sur le chiffre d'affaires, on identifie deux grandeurs déterminantes : le prix de vente et la quantité vendue (Figure 2, page 9) .



FIG. 2 – Arbre d'analyse de calcul du chiffre d'affaires

Dans notre exemple, l'arbre d'analyse pourrait avoir la forme suivante (Figure 3, page 10) .

Le raisonnement est alors le suivant :

- Le *résultat net* est fonction du *coût total* et du *prix de vente*.
- Étant en présence d'un problème de logistique, on distingue dans le *coût total*, le *coût départ* et le *coût de livraison*.
- Le *coût de livraison* se décompose en *coût fixe* et *coût variable*.
- Chaque tournée engendre des coûts fixes associés aux *coûts de préparation* (chargement ...). Ces coûts fixes sont amortis sur le nombre de livraisons possibles à chaque tournée, soit la *capacité* du véhicule.

1.1 La démarche de modélisation d'un problème

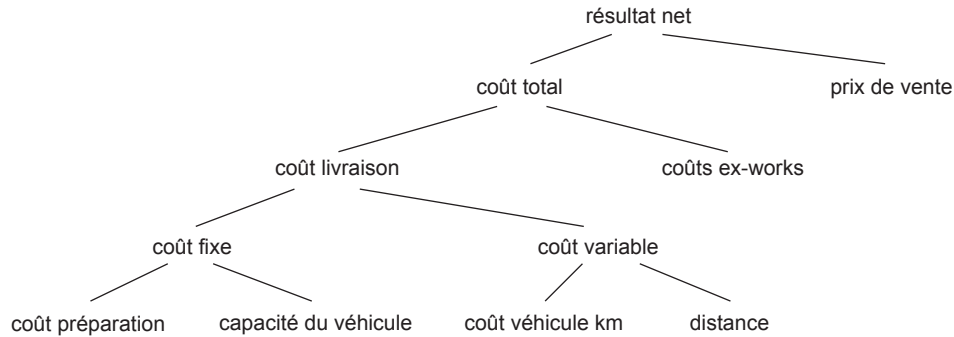


FIG. 3 – Arbre d'analyse de calcul des coûts de livraison

- Les coûts variables sont fonction de la *distance moyenne* et de du *coût au km*.

On pourrait évidemment décliner les sommets *prix* (fonction du prix de marché, de l'élasticité de la demande ...) et *coût départ* (fonction du prix des matières premières, de la valeur ajoutée...). Le niveau de détail choisi est fonction de la complexité désirée du modèle et des variables d'action que s'est donné le décideur. Ici, on doit choisir un véhicule, on ne peut donc agir que sur le coût au kilomètre et sur la capacité d'emport par tournée. On n'a pas à se poser le problème du prix de vente ou des coûts départs.

On construit ensuite le graphe de dépendance de la façon suivante :

- on oriente les branches de l'arbre en fonction du sens de la dépendance.
- on supprime les occurrences multiples pour ne conserver qu'une occurrence unique de chaque variable. Plusieurs arcs pourront alors partir d'une variable donnée.

L'exemple du calcul du résultat net est évident (Figure 4, page 11) .

À partir du graphe de dépendances, on identifie les variables et leurs types :

- **Paramètres** : Les paramètres définissent le contexte, on trouve ici : le *prix*, le *coût départ*, le *coût de préparation* et le *trajet moyen en kilomètre*.
- **Variables contrôlables** : C'est ici le coût au kilomètre et la capacité d'emport du véhicule.
- **Variables intermédiaires** : Ce sont les variables qui permettent de passer des paramètres et variables contrôlables au critères. On a ici : le *coût fixe d'une livraison*, le *coût variable d'une livraison*, le *coût d'une livraison* et le *coût total*.

1.1 La démarche de modélisation d'un problème

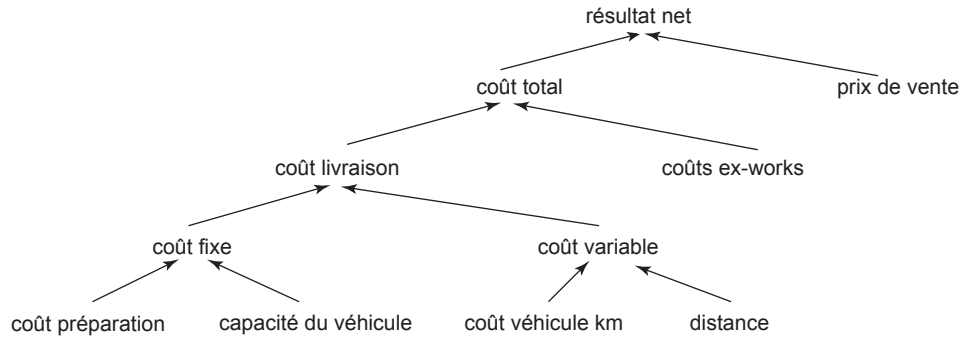


FIG. 4 – Graphe de dépendance du calcul des coûts de livraison

- **Critères** : C'est ce qui permet de mesurer l'objectif recherché, soit ici le *résultat net*.

Avant de passer à l'écriture des équations, on peut commencer à poser le problème sur le tableur. On pose d'abord les paramètres, puis les variables contrôlables, les variables intermédiaires⁴ et enfin le critère (Figure 5, page 12).

La première colonne contient les libellés des différentes variables. La seconde colonne est optionnelle, mais fortement recommandée. Elle contient les noms des variables qui seront utilisés dans les équations. Ceci facilite grandement la lecture du modèle. Exprimer le résultat net sous la forme : **Prix - CtTotal** est incontestablement plus lisible que « C5 - C18 ». Pour nommer les cellules, on sélectionne les colonnes B et C (ici B5 :C21) et sous Excel, on choisit **Insertion/Nom/Créer**, puis **colonne de gauche**.



On passe ensuite à l'écriture des équations en commençant par les variables intermédiaires :

- Le coût fixe d'une livraison est le coût de préparation divisé par la capacité d'emport (exprimée par souci de simplification comme la charge moyenne d'une livraison), soit :

$$\text{CtFixeLiv} = \text{CtPreparation} / \text{Capacite}.$$
- Le coût variable d'une livraison est la distance moyenne d'un trajet multiplié par le coût au kilomètre, soit :

$$\text{CtVarLiv} = \text{TrajetKm} * \text{CtKm}.$$
- Le coût d'une livraison est la somme des coûts fixes et variables, soit :

⁴Il est recommandé de poser les variables intermédiaires en fonction de leur profondeur dans l'arbre. On pose d'abord les variables les plus profondes et on remonte progressivement niveau par niveau.

1.1 La démarche de modélisation d'un problème

Calcul résultat net

Paramètres

Prix de vente	Prix	
Coût départ	CtDepart	
Coût de préparation	CtPreparation	
Trajet moyen en kilomètres	TrajetKm	

Variables contrôlables

Capacité d'emport	Capacité	
Coût au kilomètre	CtKm	

Variables intermédiaires

Coût fixe d'une livraison	CtFixeLiv	
Coût variable d'une livraison	CtVarLiv	
Coût d'une livraison	CtLivraison	
Coût total	CtTotal	

Critères

Résultat net	RN	
--------------	----	--

FIG. 5 – Mise en forme du modèle

$$\text{CtLivraison} = \text{CtFixeLiv} + \text{CtVarLiv}.$$

- Le coût total est la somme du coût départ et du coût de livraison, soit :

$$\text{CtTotal} = \text{CtLivraison} + \text{CtDepart}.$$

- Le résultat net est la différence entre le prix de vente et le coût total, soit :

$$\text{RN} = \text{Prix} - \text{CtTotal}.$$

Recueil des données : Les variables et autres paramètres étant identifiés, il convient ensuite de rechercher les données qui permettront de faire fonctionner le modèle. Selon le type de problème, le recueil de ces données peut être plus ou moins difficile, mais les ERP centralisant l'essentiel de l'information, quelques requêtes (Query) spécifiques – éventuellement avec l'aide du service compétent de la DSI – permettent souvent d'obtenir un volume important de données. On devra alors porter la plus grande attention à la qualité des données obtenues. La qualité du résultat en étant évidemment fonction directe ⁵.

Dans notre exemple, on a :

⁵C'est là un problème générique aux systèmes d'information. On parle en anglais de GIGO, "Garbage In, Garbage out".

1.1 La démarche de modélisation d'un problème

- Prix de vente = 10
- Coût départ = 7
- Coût de préparation = 2.5
- Distance moyenne = 20

1.1.2 Identification de solutions

Recherche de solutions : Au cours de cette étapes, on fait fonctionner le modèle de manière à identifier la ou les solutions les meilleures. On parle là de problèmes d'*optimisation*. Le tableur va nous être ici d'un grand secours. Même s'ils restent limités face aux logiciels dédiés, le tableur d'Open Office ou Excel permettent de poser, au moins d'une manière simplifiée, une infini diversité de problèmes quotidiens.

Souvent, au lieu de "la" solution optimum, on recherchera un ensemble d'alternatives. Dans la pratique, c'est l'intuition qui éclairera la décision finale du manager. Hors les cas purement techniques, rechercher une solution unique c'est conférer une confiance excessive au pouvoir de la modélisation qui est une aide à la décision et non un substitut.

On va utiliser ici le gestionnaire de scénarios d'Excel pour comparer les différentes solutions possibles.

Excel permet d'enregistrer des ensembles de valeurs sous forme de scénarios. La méthode est la suivante :

- On accède aux scénarios par le menu **Outils/Gestionnaire de scénarios**.
- Pour ajouter le premier scénario, on sélectionne le bouton **Ajouter**.
- On choisit un nom pour le scénario. On prendra ici simplement le numéro du modèle.
- On choisit les cellules variables. Ici, il s'agit de la capacité et du coût au kilomètre. On valide ensuite par **OK**.
- On fixe les valeurs des cellules variables. Pour le scénario 1, on aura 18 et 0.16 (tableau 1, page 6). On valide par **OK** si on a terminé, sinon le bouton **Ajouter** permet de construire un second scénario.

Grâce au bouton **Afficher** du gestionnaire de scénarios, on peut maintenant afficher le résultat net pour chacune des quatre hypothèses. Mieux encore, le bouton **Synthèse** construit automatiquement un tableau synthétisant les résultats, éventuellement sous forme de tableau croisé dynamique (Figure 6, page 14) . On vérifie ainsi aisément que le véhicule 2 est le meilleur choix.



1.2 Exemple 2 : Recherche d'une valeur cible

	A	B	C	D	E	F	G	H
1								
2		Synthèse de scénarios						
3		Valeurs actuelles :						
4					1	2	3	4
5		Cellules variables :						
6		Capacité	22.00	18.00	8.00	15.00	22.00	
7		CtKm	0.17	0.16	0.08	0.14	0.17	
8		Cellules résultantes :						
9		RN	-	0.51	-	0.34	1.09	0.03
10								
11								
12								

La colonne Valeurs actuelles affiche les valeurs des cellules variables au moment de la création du rapport de synthèse. Les cellules variables de chaque scénario se situent dans les colonnes grisées.

FIG. 6 – Synthèse des scénarios

Test des solutions : C'est la phase de validation des résultats. Avant leur mise en œuvre, il convient de s'assurer que les solutions trouvées sont réellement efficaces et ne proviennent pas d'un biais du modèle, auquel cas, on doit en revoir la conception.

1.1.3 Validation

On s'interroge ici sur les résultats et sur leur mise œuvre.

Interprétation : Au cours de cette phase, on va plus particulièrement analyser la *sensibilité des résultats*. Il s'agit de vérifier dans quelle mesure une modification mineure d'une variable contrôlable influe sur le résultat final. Si la sensibilité est forte, la mise en œuvre des décisions en sera d'autant plus difficiles.

Faisabilité : Avant de valider la décision dictée par le modèle, il convient de s'assurer de sa faisabilité. Est-ce que la mise en œuvre nécessite des changements acceptables ? Quel est l'impact sur les autres fonctions de l'entreprise ?...

La démarche présentée ici est générique. Quelques exemples complémentaires vont nous permettre de nous y familiariser tout en introduisant à certains outils spécifiques du tableur.

1.2 Exemple 2 : Recherche d'une valeur cible

Souhaitant lancer un nouveau produit, l'entreprise *X* veut en déterminer le seuil de rentabilité.

L'identification est évidente : il s'agit de trouver la quantité qui permet de couvrir coûts fixes et variables.

1.2 Exemple 2 : Recherche d'une valeur cible

On fera l'économie du graphe en posant les paramètres suivants :

- Prix = 10
- Coût des matières premières = 2
- Coût d'utilisation des équipements = 5
- Frais de siège = 25 000
- Frais de structure = 10 000

Pour trouver la solution, on peut utiliser ici la fonction **Valeur cible** que l'on obtient sous Excel avec **Outils/Valeur cible**. La cellule à définir est le résultat net (C21), la valeur à atteindre est 0 et la cellule à modifier est le volume de vente (C12). Quand on valide, le tableau recherche la valeur cible qui permet d'atteindre l'objectif (Figure 7, page 15) .



	A	B	C	D	E	F
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						

Recherche du point mort		
Paramètres		
Prix	Prix	10
Coût des matières premières	CtMP	2
Coût d'utilisation des équipements	CtEqpt	5
Frais de siège	FraisSiege	25 000
Frais de structure	FraisStruct	10 000
Variables contrôlables		
Volume des ventes	Ventes	
Variables intermédiaires		
Chiffre d'affaires	CA	-
Charges variables	ChVar	-
Marge sur coûts variables	MargeCV	-
Charges fixes	ChFixe	35 000
Critères		
Résultat net	RN	- 35 000

Valeur cible

Cellule à définir :

\$C\$21

Valeur à atteindre :

0

Cellule à modifier :

\$C\$12

OK

Annuler

FIG. 7 – Recherche de la valeur cible

Le résultat est : 11 667 (Figure 8, page 16) .

1.2 Exemple 2 : Recherche d'une valeur cible

	A	B	C
1	Calcul du résultat net		
2			
3			
4	Paramètres		
5	Prix	Prix	10
6	Coût des matières premières	CtMP	2
7	Coût d'utilisation des équipements	CtEqpt	5
8	Frais de siège	FraisSiege	25 000
9	Frais de structure	FraisStruct	10 000
10			
11	Variables contrôlables		
12	Volume des ventes	Ventes	11 667
13			
14	Variables intermédiaires		
15	Chiffre d'affaires	CA	116 667
16	Charges variables	ChVar	81 667
17	Marge sur coûts variables	MargeCV	35 000
18	Charges fixes	ChFixe	35 000
19			
20	Critères		
21	Résultat net	RN	-

FIG. 8 – Résultat

1.3 Exemple 3 : Tables et simulations

1.3 Exemple 3 : Tables et simulations

L'outil scénarios rencontré plus haut reste limité, il est en effet beaucoup trop lourd dès que le nombre de possibilités devient important. Les tableurs implémentent la fonction *Table* qui permet de simuler aisément un grand nombre de configurations possibles.

Considérons un nouvel exemple de calcul de marge. On a un coût unitaire de transformation de 15,35 €, un coût unitaire des matières premières de 8,32 € et des frais fixes de 3500 €. La demande pour le produit varie en fonction du prix. Une analyse historique montre l'approximation linéaire suivante : demande = 135000 - 4000 × prix de vente.

La variation de la demande en fonction de la variation de prix est *l'élasticité prix de la demande*. Elle se mesure comme $\epsilon = \frac{dy/y}{dx/x}$ soit le rapport de la variation de la demande et de la variation du prix. On utilise souvent un modèle, dit doublement logarithmique, de la forme $\log y = a \log x + b$. On a alors : $(\log y)' dy = (a \log x + b)' dx$, d'où $\frac{dy}{y} = a \frac{dx}{x}$ dont on déduit directement : $\epsilon = a$. C'est un *modèle à élasticité constante*.

Ce modèle est très facile à traiter à partir de données brutes, il suffit en effet de calculer les logarithmes de ces données et de leur appliquer une simple régression linéaire. On aura alors la relation $\log y = a \log x + B$ exprimant la relation $y = b.x^a$ avec $b = \exp^B$.

Le modèle est évident (Figure 9, page 18) , notons simplement que la demande est calculée à partir de la fonction linéaire évoquée.

Quelles seront les conséquences d'une variation de prix ? Le calcul n'est en effet par direct, car une modification de prix a une incidence à la fois sur la marge unitaire et sur le volume de la demande. Pour en visualiser les conséquences, on peut utiliser une *Table* de simulation⁶. Pour construire une table simple, les étapes sont les suivantes (Figure 10, page 19) :

- On met en colonne la liste des différentes valeurs du paramètre que l'on souhaite tester. Ici on fera varier le prix de 10 à 50 € par tranches de 2.
- Juste au dessus de la colonne suivante (soit par exemple en B1 si la colonne des paramètres commence en A2), on met une référence à la cellule du modèle qui contient le résultat que l'on souhaite analyser.
- On sélectionne l'ensemble des deux colonnes et on active l'outil *Table* dans le menu données.

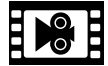
⁶C'est l'outil OPERATIONS.MULTIPLES sous OpenOffice.

1.4 Exemple 4 : Calcul itératif

Paramètres		
Coût unitaire de transformation	CtUPrep	15.25
Coût unitaire des matières premières	CtMP	8.32
Frais fixes	FF	3 500.00
Demande de base	Demande	135 000.00
Variation demande selon prix		- 4 000.00
Variables de décision		
Prix de vente	PxVente	32.00
Variables intermédiaires		
Demande	Demande	7 000.00
Coût de transformation	CtPrep	106 750.00
Coût des matières premières	CtIngredient	58 240.00
Charges variables	ChVar	164 990.00
Charges fixes	ChFixes	3 523.57
Chiffre d'affaires	CA	224 000.00
Marge sur coûts variables	MgCtVar	59 010.00
Taux de marge sur coûts variables	TxMgCtVar	0.26
Critères		
Seuil de rentabilité	SeuilRentabilite	13 375.35
Marge brute	MargeBrute	55 486.43

FIG. 9 – Calcul de la marge

- Dans la boîte de dialogue qui s'ouvre, dans cellule d'entrée en colonne, on fait référence à la cellule qui dans le modèle contient le paramètre que l'on souhaite tester.
- Il suffit de valider pour que la table se remplisse.



On peut de la même manière créer une table à deux entrées avec par exemple le prix en colonne et la capacité de production de base en ligne. Dans ce cas, on met en haut à gauche une référence à la cellule qui contient le résultat. On sélectionne ensuite l'ensemble du tableau et on indique à l'outil table que la référence en ligne est la capacité, la référence en colonne le prix. On obtient ainsi un tableau à double entrée donnant la valeur de la marge pour l'ensemble des combinaisons prix/capacité.

1.4 Exemple 4 : Calcul itératif

Le directeur commercial de l'entreprise X souhaite remettre en cause le système actuel d'encouragement traditionnellement basé sur une commission sur les ventes. Une étude lui a en effet montré que les commerciaux ont tendance à mettre l'accent sur les produits chers ou aisément vendables au détriment des produits qui dégagent une marge importante. Il souhaiterait calculer les primes des commerciaux en fonction non plus de leur chiffre d'affaires, mais plutôt de la marge qu'ils ont engendrés. Pour valider sa

1.4 Exemple 4 : Calcul itératif

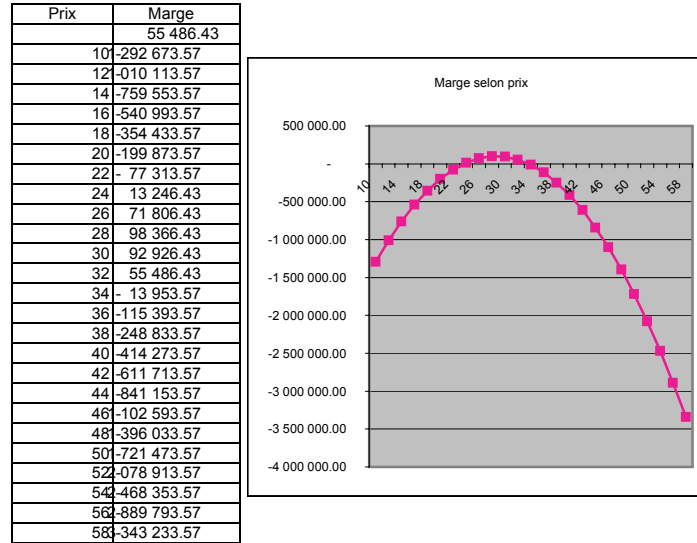


FIG. 10 – Évolution de la marge en fonction du prix

décision, il doit au préalable modéliser la situation actuelle.

Une étude statistique simple sur l'historique des ventes montre qu'elles sont de la forme :

$$\text{Ventes} = 1\,000 + 10 \times \text{Commissions}$$

On dispose également des données suivantes :

- Le taux de marge brute sur les ventes est de 15%.
- Le taux actuel de commissions est de 2.5% du chiffre d'affaires.

Avant de mettre en œuvre le nouveau système, le directeur commercial veut en vérifier l'impact sur la marge nette.

La réalisation du graphe de dépendance ne pose pas de problème. En revanche, le graphe met en évidence un phénomène appelant un traitement spécifique : deux variables dépendent l'une de l'autre. En effet, les commissions sont fonction du chiffre d'affaires, mais celui-ci est fonction des commissions.

Partant du résultat net qui est l'objectif, on constate que les ventes dépendent des commissions qui elles-mêmes dépendent des ventes.

Pour résoudre ce problème, il faut faire appel au **calcul itératif**.

Avec un taux de marge brute de 15% et un taux de commission de 2.5%, le problème se présente sous la forme de la figure 12, page 20. On peut y voir :

1.4 Exemple 4 : Calcul itératif

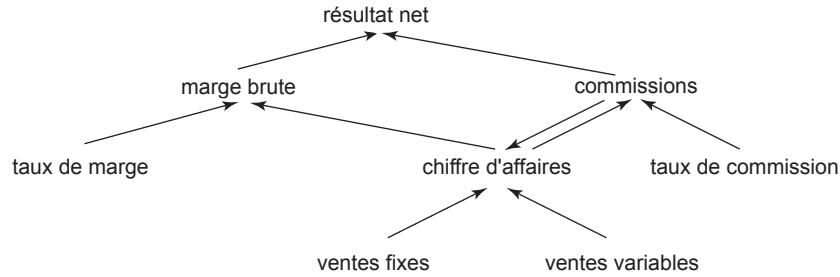


FIG. 11 – Graphe de dépendance du commissionnement

$$- CA = VentFixe + VentVar * Comm$$

$$- Comm = CA * TxComm$$

Le chiffre d'affaire dépend de la commission qui elle-même dépend du chiffre d'affaires.

Commissions : situation actuelle

Paramètres

Ventes fixes	VentFixe	1000
Ventes variables	VentVar	10
Taux de marge	TxMarge	0.15

Variables contrôlables

Taux de commission	TxComm	0.025
--------------------	--------	-------

Variables intermédiaire:

Chiffre d'affaires	CA	=VentFixe+VentVar*Comm
Marge brute	MargeBr	=CA*TxMarge
Commissions	Comm	=CA*TxComm

Critères

Résultat net	RN	=MargeBr-Comm
--------------	----	---------------

FIG. 12 – Commissions : situation actuelle

Quand on entre la formule de calcul de la commission, le tableur émet un message d'erreur informant de l'existence d'une *référence circulaire*, c'est-à-dire qu'une cellule fait référence à son propre résultat. Pour calculer ces cellules, on fait appel aux capacités de calcul itératif du tableur. Sous Excel, on sélectionne **Outils/Options**, puis l'onglet **Calcul** de la boîte qui s'affiche alors. Dans la zone supérieure, on sélectionne **Sur ordre**, puis on coche **Itération**. Dans ce mode, le tableur calcule plusieurs fois la feuille en

1.4 Exemple 4 : Calcul itératif

prenant à chaque fois les résultats du calcul précédent. La zone **Nb maximal d'itération** permet de fixer une limite au nombre de calculs successifs. La zone **Écart maximal** permet d'arrêter le calcul si l'écart entre deux calculs successifs est inférieur à cet écart. On conservera ici les valeurs 100 et 0.01. On valide ensuite par OK et on lance le calcul de la feuille avec la touche F9. On obtient alors un résultat net de 166.67 (Figure 13, page 21)).



Commissions : situation actuelle		
Paramètres		
Ventes fixes	VentFixe	1 000.00
Ventes variables	VentVar	10.00
Taux de marge	TxMarge	0.15
Variables contrôlables		
Taux de commission	TxComm	0.025
Variables intermédiaires		
Chiffre d'affaires	CA	1 333.33
Marge brute	MargeBr	200.00
Commissions	Comm	33.33
Critères		
Résultat net	RN	166.67

FIG. 13 – Commissions : situation actuelle

2 Prévvision

À des échelles diverses, tous les êtres vivants sont coutumiers des prévisions. Le moindre insecte « sait » qu’une ombre qui s’approche rapidement signifie l’arrivée prochaine d’un prédateur et réagit en conséquence. Le plus ahuri des chiens anticipe parfaitement le lieu où va tomber la balle qui vient d’être lancée et tous les enfants (occidentaux du moins) savent que Noël va arriver un jour ou l’autre.

Dans le domaine de la décision les prévisions sont essentielles. Prendre une décision revient en effet à faire un pari sur l’avenir et les outils de la prévision permettent d’en minimiser les risques en fixant un cadre au sein duquel la probabilité d’évolution d’une grandeur quelconque est maximale.

On se contentera ici de présenter deux grandes méthodes quantitatives de prévision : le modèle de Holt-Winter et une brève introduction aux modèles autorégressifs. En première approche, on peut considérer que le premier est plus particulièrement adapté à l’analyse de la demande, le second aux prévisions financières.

2.1 Mesure de l’erreur

Avant de présenter les techniques d’analyse en tant que telles, revenons rapidement sur la mesure de l’erreur. Le meilleur moyen de valider une méthode de prévision consiste en effet à l’appliquer au passé. Quand on conçoit un nouveau modèle météo par exemple, on vérifie si, en fonction des paramètres d’une période donnée, les résultats correspondent bien à ce qui s’est passé à ce moment là. On fait de même quand on analyse une chronique. On vérifie donc que lorsqu’on l’applique au passé, le modèle donne des réponses proches de la réalité.

À partir d’une série $y_1 \dots y_n$ connue, le modèle doit permettre de calculer $\hat{y}_{n+i, i>0}$, l’estimation de y_{n+i} . Pour tout $t \leq n$, la mesure de l’écart entre y_t et \hat{y}_t , donne des indications sur la qualité du modèle. On définit :

- *L’erreur* : $e_t = \hat{y}_t - y_t$ qui est simplement l’écart entre la réalité et la valeur calculée par le modèle.
- *L’erreur absolue* : $|e_t| = |\hat{y}_t - y_t|$ est la valeur absolue de l’erreur.
- *L’erreur moyenne carrée* (MSE, Mean Squared Error⁷) : $MSE = \frac{1}{n} \sum e_t^2$. Comme toutes les mesures basées sur le carré, elle a l’avantage d’être simple à calculer, mais l’inconvénient d’amplifier les écarts.
- *L’erreur absolue moyenne* (MAD, Mean Absolute Error) : $MAD = \frac{1}{n} \sum |e_t|$. D’un calcul moins direct que la MSE, la MAD a l’avantage de

⁷On utilisera les acronymes anglais qui sont maintenant de loin les plus communs.

2.2 Les séries chronologiques

ne pas accentuer les écarts. C'est un indicateur très largement utilisé, les tableurs en permettant un calcul aisé. L'un des intérêts de cet indicateur est que, pour n grand, si l'erreur est normalement distribuée, on peut estimer l'écart type de l'erreur : $\sigma_e = \sqrt{\pi/2}MAD \approx 1,25MAD$ (et en conséquence au moins approximer l'intervalle de confiance sur la prévision).

- *L'erreur absolue moyenne en pourcentage* (MAPE, Mean Absolute Percent Error) : $MAPE = \frac{1}{n} \sum \left| \frac{e_t}{y_t} \right|$. Cet indicateur permet, à la différence de la MAD, de s'accommoder des changements d'échelle. C'est le cas par exemple si on intègre dans la prévision des données en provenance d'un petit centre commercial et d'un centre beaucoup plus important.

2.2 Les séries chronologiques

Particularité des séries chronologiques : elles ne portent pas toujours sur des périodes homogènes. En effet, les mois n'ont pas tous le même nombre de jours, ni calendaires, ni ouvrables ; le jeu des jours fériés notamment tend à hétérogénéiser les semaines ; les périodes d'analyse ne comportent pas toujours le même nombre de lundi, de mardi ... ce qui peut avoir une incidence significative dans certains domaines.

De manière générale, il est donc nécessaire de redresser les chroniques avant de les analyser, c'est-à-dire de faire en sorte que les données portent sur des périodes homogènes. Le principe en est simple (même si l'application pratique peut être délicate). On définit par exemple des mois standards (nombre de jours ouvrables dans l'année divisé par 12) et on corrige les données du mois courant en fonction de son écart au mois standard. Si l'année comporte 252 jours ouvrables, le mois standard est de 21 jours ; si un mois comporte 22 jours ouvrables, on corrigera la donnée associée (par exemple le chiffre d'affaires) du facteur $\frac{21}{22} = 0,955$. Pour le redressement du nombre de jours de la semaine (nombre de lundi ...), on calcule à partir de l'historique la part moyenne de chaque jour dans la grandeur étudiée afin d'obtenir des jours standards. L'utilité du redressement dépend naturellement tant de la précision recherchée que de l'importance des variations associées aux périodes. La question du redressement et de son éventuelle nécessité est un préalable essentiel à toute analyse de série chronologique.

Seconde particularité des chroniques : la composition. Une série chronologique associe souvent (mais ce n'est pas systématique) différentes composantes. On retient généralement :

- La *tendance* (ou le *trend*), noté t , représente la composante lourde

2.3 Lissage des séries chronologiques

de la série ; c'est un élément structurel qui incarne l'évolution de long terme. On la modélise généralement avec une fonction dérivable dépendant seulement du temps. Les plus courantes sont les fonctions polynomiales, classiquement de degré 1 (linéaire) ou 2 (quadratique). Pour les phénomènes économiques, on utilise également fréquemment des fonctions exponentielles et logarithmiques.

- Le *cycle*, noté c , représente les évolutions périodiques liées aux activités économiques. Une chronique peut superposer différents cycles de périodes variées. Le plus évident est le *cycle saisonnier* de périodicité annuelle. On dissocie généralement la composante saisonnière s du cycle long c .
- La *composante aléatoire*, noté ϵ correspond au *bruit* associé à l'évolution de la grandeur.

Les principales compositions utilisées sont :

- Le modèle additif : $y_t = t_t + c_t + \epsilon_t$ pose que les variations cycliques et aléatoires sont indépendantes de la tendance et de somme nulle. Exprimé autrement, l'ampleur du cycle est constante quel que soit le niveau.
- Le modèle multiplicatif : $y_t = t_t \cdot c_t \cdot \epsilon_t$ pose que les variations cycliques et aléatoires sont proportionnelles à la tendance. Ce modèle se ramenant au précédent par transformation logarithmique, on peut dire également que la somme des logarithmes des cycles est nulle.
- Le modèle mixte : $y_t = (t_t \cdot c_t) + \epsilon_t$ où le cycle est proportionnel à la tendance alors que la composante aléatoire en est indépendante.

L'identification des composantes (et leur décomposition) est l'une des difficultés de l'analyse des chroniques. Nous verrons dans les sections suivantes comment la réaliser.

2.3 Lissage des séries chronologiques

Les méthodes de prévision que nous allons voir sont basées sur l'hypothèse selon laquelle le passé, et lui seul, permet de prévoir l'avenir. Ceci suppose naturellement qu'il y ait continuité, une rupture ne sera pas perçue par ces méthodes. Les méthodes de lissage sont utilisées pour les séries considérées comme stable sur le court terme. Elles peuvent également servir de support à la décomposition.

2.3 Lissage des séries chronologiques

2.3.1 Les moyennes mobiles

La méthode des moyennes mobiles considère que les n dernières périodes sont significatives et représentatives de l'avenir immédiat ; c'est une forme de *prévision naïve*. Elle reporte sur la (ou les) période suivante, le niveau actuel N vu comme la moyenne des n valeurs précédentes, soit : $N_t = \frac{1}{n} \sum_{i=t-n+1}^t y_i$, $\hat{y}_{t+1} = N_t = S_t$ (S étant la valeur lissée — *smoothed*). Par exemple, la prévision de la période 5, pour une moyenne mobile de période 3 sera : $S_4 = \frac{y_2 + y_3 + y_4}{3}$, $\hat{y}_5 = S_4$. Dans la pratique, on fait le calcul sur une partie significative de l'historique afin d'estimer la validité du modèle.

Considérons un marchand de glace qui, à partir des ventes des douze derniers mois, souhaite prévoir les ventes du mois suivant (Figure 14, page 25) . Pour ce faire, il décide de travailler sur la base d'une moyenne mobile de période trois⁸. Il en résulte que les ventes prévisionnelles sont d'environ 56 (cellule D15), avec une erreur moyenne absolue de 3,74 (cellule G17).

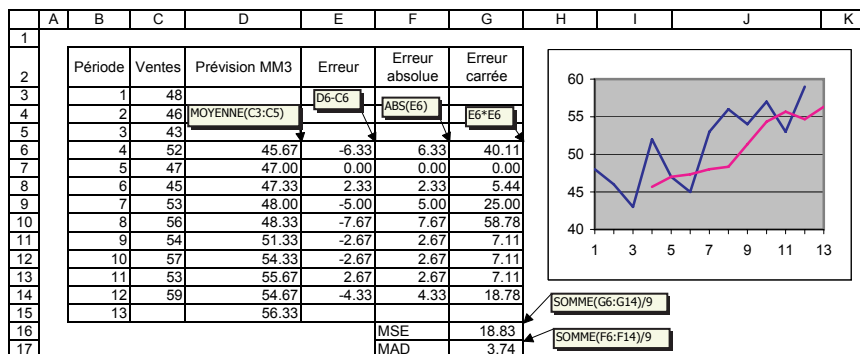


FIG. 14 – Moyenne mobile de période trois

Comment choisir la période des moyennes mobiles ? Il n'y a pas de règle absolue. En l'absence de changement de structure et de tendance, plus la période est longue, plus le bruit sera atténué. Inversement, si le niveau change (existence d'une tendance ou évolution structurelle), les données récentes seront plus significatives. Plus encore, si ce changement est récent par rapport à l'historique utilisé, il faudra d'autant plus de temps pour l'intégrer que la période choisie sera longue.

La méthode présentée donne le même poids aux p dernières observations

⁸On l'a calculée ici en utilisant directement la formule de la moyenne, mais Excel dispose d'un utilitaire d'analyse dédié ; il en va de même pour Gnumeric qui intègre le calcul des moyennes mobiles dans les outils de prévision.

2.3 Lissage des séries chronologiques

pour des moyennes mobiles de période p . Il est possible d'accorder un poids supérieur aux données récentes en utilisant des moyennes pondérées. On aura alors $S_t = \sum_{i=t-n+1}^t \alpha_i x_i$ avec $\sum \alpha_i = 1$. Sur une période trois, on aura par exemple : $S_4 = 0,2y_2 + 0,3y_3 + 0,5y_4$. Le choix des poids est délicat. On pourra utiliser le solveur pour déterminer les poids optimaux, on cherchera alors à minimiser l'erreur (par exemple la MAD) en faisant varier les poids sous contrainte que leur somme soit égale à 1. Le problème n'étant pas linéaire, on veillera dans les options du solveur Excel à ce que la case **modèle supposé linéaire** ne soit pas cochée.

Insistons enfin sur le fait que les moyennes mobiles, forme de prévision naïve, sont très utilisées, car efficaces malgré leur simplicité. On les retrouve notamment très couramment dans les prévisions de stockage ou de vente des produits de faible importance (rang C). Elles sont en effet très faciles à calculer et à mettre à jour.

2.3.2 Le lissage exponentiel

La méthode des moyennes mobiles présente l'avantage de la simplicité, mais, sauf à utiliser les moyennes pondérées, elle accorde au passé récent le même poids qu'au passé lointain. En outre, elle est incapable de s'ajuster en cas de dérive de la qualité de la prévision. Le lissage exponentiel pallie au moins partiellement ces deux limites. La valeur lissée en t est : $S_t = \alpha y_t + (1 - \alpha)S_{t-1}$, avec $0 < \alpha < 1$. On a donc une forme de moyenne pondérée où les poids décroissent avec le temps. Comme pour les moyennes mobiles, la prévision est vue comme la continuation du passé proche, soit : $\hat{y}_{t+1} = S_t$.

2.3 Lissage des séries chronologiques

On a :

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}$$

soit :

$$S_{t-1} = \alpha y_{t-1} + (1 - \alpha)S_{t-2}$$

Si on substitue dans la première équation, on obtient :

$$S_t = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + (1 - \alpha)^2 S_{t-2}$$

Soit on continuant la substitution :

$$S_t = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots + \alpha(1 - \alpha)^{t-1} y_1$$

La valeur α étant inférieure à 1, le terme $\alpha(1 - \alpha)^i$ décroît avec la puissance. On accorde donc aux observations un poids décroissant avec le temps.



Une simple transformation permet d'exprimer le lissage exponentiel sous la forme : $S_t = S_{t-1} + \alpha(y_t - S_{t-1})$ ou encore, puisque $\hat{y}_t = S_{t-1}$, $S_t = S_{t-1} + \alpha(y_t - \hat{y}_t) = S_{t-1} + \alpha e_t$. Soit : valeur lissée en t égale valeur lissée en $t - 1$ moins une fraction de l'écart entre la valeur réelle en t et la valeur prévue en t , soit une fraction de l'erreur de prévision. On va donc beaucoup plus loin que la moyenne mobile, puisque l'on pondère les valeurs passées et que l'on tient compte de l'erreur d'estimation.

Comment fixer la valeur du coefficient α ? Le lissage exponentiel est à la fois une procédure de lissage sur longue période puisqu'elle tient compte de l'ensemble du passé ; en ce sens, elle tend à éliminer les variations aléatoires ; mais c'est aussi une méthode qui privilégie le passé immédiat du fait des pondérations décroissantes dans le temps. Le coefficient α a précisément pour objet de « positionner le curseur » entre stabilité et réactivité. Un coefficient proche de 0 privilégiera la stabilité ; proche de 1 on préférera la réactivité. Si l'on considère par exemple la valeur y_{t-4} , un coefficient de 0,2 donnera une pondération de 0,1024 ($0,2 \times (1 - 0,2)^3$), alors que pour 0,8, on aura une pondération de 0,0064. Là encore le solveur peut aider au choix de la valeur.

Reprenons l'exemple précédent (Figure 15, page 28) . On peut calculer le lissage avec la formule présentée, mais aussi directement avec l'utilitaire d'analyse **Lissage exponentiel** d'Excel ou de Gnumeric⁹. L'utilisation du solveur montre que, pour cet exemple, le α optimal est d'environ 0,62, pour

⁹Dans les deux cas, le paramètre demandé n'est pas α mais $1 - \alpha$

2.4 Décomposition

une MAD de 3,48 contre 3,74 pour les moyennes mobiles de période trois.

	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								

Période	Ventes	Prévision	Erreur	Erreur absolue	Erreur carrée
1	48	48.00	0.00	0.00	0.00
2	46	48.00	2.00	2.00	4.00
3	43	46.75	3.75	3.75	14.04
4	52	44.40	-7.60	7.60	57.75
5	47	49.16	2.16	2.16	4.66
6	45	47.81	2.81	2.81	7.88
7	53	46.05	-6.95	6.95	48.31
8	56	50.40	-5.60	5.60	31.34
9	54	53.91	-0.09	0.09	0.01
10	57	53.97	-3.03	3.03	9.21
11	53	55.87	2.87	2.87	8.21
12	59	54.07	-4.93	4.93	24.29
Prév. 13		57.16			
				MSE	17.48
				MAD	3.48
	Alpha	0.6262			

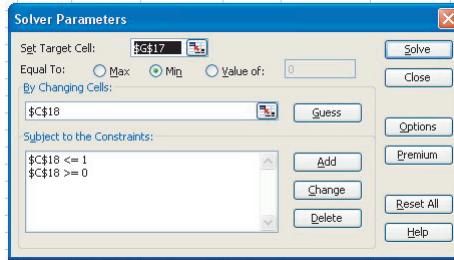


FIG. 15 – Lissage exponentiel et utilisation du solveur

2.4 Décomposition

Les chroniques sont rarement stationnaires; elles intègrent de manière très générale des composantes tendanciellles, saisonnières¹⁰ et cycliques.

Une fois encore le plus simple lors de la première étape est de tracer le graphique des données. On pourra y repérer l'existence de composantes saisonnières voire cycliques ainsi que l'éventuelle tendance. Le graphique met également en évidence les éventuels points aberrants qui s'écartent fortement des autres (*outliers*). Ces points doivent être éliminés ou normalisés pour ne pas nuire à la prévision. Il est naturellement important de trouver l'explication de ces aberrations (qui peuvent être aussi variées qu'un problème météo,

¹⁰Insistons ici sur le fait que ce terme ne doit pas être considéré dans son sens littéral. On l'appliquera à tout cycle lié à des périodicités infra-annuelles.

2.4 Décomposition

une grève ou une panne d'équipement).

Retrouvons notre marchand de glaces quelques années après son établissement. Il dispose désormais d'un historique sur trois ans et souhaite faire des prévisions mensuelles sur l'année à venir. Afin de pallier les écarts de durées entre les différents mois, on va travailler sur treize périodes de quatre semaines ; on homogénéise ainsi simplement les périodes.

Le graphe (Figure 16, page 29) montre clairement l'existence d'une composante saisonnière. On constate en effet que le même motif d'évolution se reproduit sur 13 périodes. On constate de même que cette répétition n'est pas horizontale ce qui traduit l'existence d'une tendance, en l'occurrence croissante. Naturellement, les choses ne sont pas toujours aussi évidentes. La vente de glaces est en effet une activité fortement saisonnières ; le repérage peut être plus difficile pour les activités moins sujettes aux évolutions météorologiques.

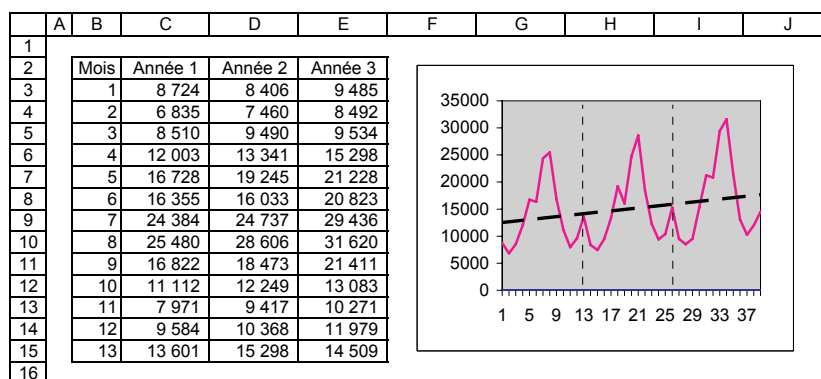


FIG. 16 – Ventes par période

2.4.1 Moyennes mobiles et coefficients saisonniers

On peut considérer les coefficients saisonniers comme représentatifs des écarts par rapport à la moyenne de la périodicité, soit ici par rapport à la moyenne annuelle. On calcule donc pour chacune des périodes (mois, trimestres, ...) la moyenne centrée annuelle glissante, soit si le nombre de périodes p est pair : $S_t = (\frac{y_{t-(p/2)}}{2} + \frac{y_{t+(p/2)}}{2} + \sum_{i=t+1-(p/2)}^{i=t-1+(p/2)} y_i) / 2$. Par exemple pour des données trimestrielles, on aura pour le troisième trimestre : $S_3 = (\frac{y_1}{2} + \frac{y_5}{2} + y_2 + y_3 + y_4) / 4$. Si le nombre de périodes est impair, on n'a évidemment pas besoin des demi-sommes pour centrer la moyenne. Si on prend par exemple les 3 quadrimestres, on aura : $S_2 = \frac{Q_1+Q_2+Q_3}{3}$.

2.4 Décomposition

Le calcul des coefficients saisonniers à partir des moyennes mobiles centrées suit les étapes suivantes (Figure 17, page 31) :

1. Calcul des moyennes centrées (colonne E). On perd alors une demi période en début et en fin de série si la périodicité est paire.
2. Si l'on considère un modèle multiplicatif : calcul des rapports entre les valeurs de la série et les moyennes centrées correspondantes (colonne F). Dans le cas d'un modèle additif, on fait la différence entre les deux valeurs (colonne G).
3. Calcul des rapports moyens dans le cas d'un modèle multiplicatif, soit somme des rapports d'une période donnée (tous les mois de janvier par exemple) divisé par le nombre d'occurrence (colonne H). Le calcul est similaire pour les différences moyennes (colonne J).
4. La somme des rapports doit être égale au nombre de périodes constitutifs de la périodicité (12 pour les mois si la périodicité est annuelle, dans notre cas 13). On normalise les rapports afin d'obtenir un total conforme. Il suffit de multiplier le rapport par le nombre de périodes et de diviser le tout par la somme des rapports (dans notre cas, on multiplie par 13 et on divise par 13,02, colonne I ; la somme des rapports normalisés est bien de 13). Dans le cas des différences, la somme doit être nulle. On peut normaliser simplement (mais c'est une approximation) en faisant la moyenne des différences et en la soustrayant aux différences brutes. La somme des différences normalisées est alors nulle (colonne K).

Les coefficients saisonniers sont les rapports normalisés (ou les différences). Dans notre exemple, on constate que les ventes de la période 1 ne sont que de 59% des ventes moyennes d'une période (cellule I7) ; en revanche elles sont de 189% sur la période 8 (cellule I14).

2.4.2 Recherche de la tendance

Connaissant les coefficients saisonniers, il est maintenant possible de calculer une série désaisonnalisée, c'est-à-dire d'exprimer les ventes « hors variations saisonnières ». Le calcul est évident puisqu'il s'agit simplement d'utiliser les coefficients pour corriger les effets des saisons. On divise les ventes réelles par le rapport saisonnier correspondant dans le cas d'un modèle multiplicatif ; on soustrait les différences pour le modèle additif.

À partir de là, on peut déterminer la tendance par régression sur la série désaisonnalisée, on pourra utiliser la fonction `DROITEREG` ou l'utilitaire d'analyse `régression linéaire`.

2.4 Décomposition

	B	C	D	E	F	G	H	I	J	K
3					Somme	Somme	Somme	Somme	Moyenne	Somme
4					27.77	10 989.08	13.02	13.00	86.54	0.00
5										
6	Mois	Période	Ventes	MM centrées	Rapports	Différences	Rapports moyens	Rapports normalisés	Différences moyennes	Différences normalisées
7	1	1	8 724				0.59	0.59	-6 131.00	-6 217.54
8	2	2	6 835		MOYENNE(F20;F33)		0.52	0.52	-7 336.65	-7 423.19
9	3	3	8 510				0.62	0.62	-5 977.15	-6 063.69
10	4	4	12 003		D13/E13		0.92	0.92	-1 245.46	-1 332.00
11	5	5	16 728	MOYENNE(D7:D19)		D13-E13	1.29	1.29	4 583.08	4 496.54
12	6	6	16 355				1.17	1.16	2 682.46	2 595.92
13	7	7	24 384	13 701	1.78	10 683.31	1.74	1.73	11 306.04	11 219.50
14	8	8	25 480	13 676	1.86	11 803.77	1.89	1.89	13 667.38	13 580.85
15	9	9	16 822	13 724	1.23	3 097.69	1.23	1.23	3 455.00	3 368.46
16	10	10	11 112	13 800	0.81	-2 687.69	0.81	0.81	-2 772.38	-2 858.92
17	11	11	7 971	13 903	0.57	-5 931.62	0.60	0.60	-5 754.92	-5 841.46
18	12	12	9 584	14 096	0.68	-4 512.23	0.68	0.68	-4 956.46	-5 043.00
19	13	13	13 601	14 071	0.97	-470.46	0.97	0.97	-394.92	-481.46
20	1	14	8 406	14 099	0.60	-5 692.62				
21	2	15	7 460	14 339	0.52	-6 879.08	H19*I13/\$H\$4			I19-\$I\$4
22	3	16	9 490	14 466	0.66	-4 976.08				
23	4	17	13 341	14 554	0.92	-1 212.54		MOYENNE(G32;G45)		
24	5	18	19 245	14 665	1.31	4 580.23				
25	6	19	16 033	14 725	1.09	1 307.92				
44	12	38	11 979							
45	13	39	14 509							

FIG. 17 – Recherche des coefficients saisonniers

Dans notre exemple (Figure 18, page 31) , on obtient pour le modèle additif : $\hat{y}_t = 105,2t + 12984$ avec un coefficient de corrélation de 0,606. Le modèle multiplicatif donne : $\hat{y}_t = 105,1t + 12977$ avec un coefficient de corrélation de 0,758.

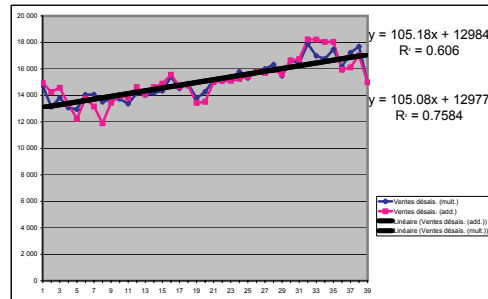


FIG. 18 – Calcul de la tendance

2.4.3 Calcul des prévisions

Avant de passer aux prévisions, il convient de valider le modèle et de choisir entre additif et multiplicatif (Figure 19, page 32) . On commence donc classiquement par appliquer le modèle au passé. Dans une première

2.4 Décomposition

étape, on calcule la tendance, soit $\hat{y}_t = b_0 + b_1t$ (colonnes N et O pour chacun des deux modèles). À partir de là, il faut « resaisonnaliser » la série obtenue. Dans le cas du modèle multiplicatif, on multiplie les valeurs de la tendance par le coefficient saisonnier correspondant (colonne P) ; pour le modèle additif, on ajoute la différence saisonnière (colonne Q). On peut ensuite calculer les erreurs absolues (colonnes R et S) et la MAD. On obtient une MAD de 485 pour le modèle multiplicatif (cellule R4) et de 701 pour le modèle additif (cellule S4). On aurait de même pu calculer la MSE et l'estimation de l'écart type de l'erreur $\hat{\sigma}_e = \sqrt{\sum(\hat{y}_i - y_i)^2/n - 1}$ qui permet de calculer un intervalle de confiance sur les prévisions (on aura environ 95% de chances que la prévision obtenue se situe dans un intervalle de $\pm 2\hat{\sigma}_e$). Dans notre cas, on trouve $\hat{\sigma}_e = 687$ pour le modèle multiplicatif et $\hat{\sigma}_e = 967$ pour le modèle additif (les calculs ne sont pas présentés).

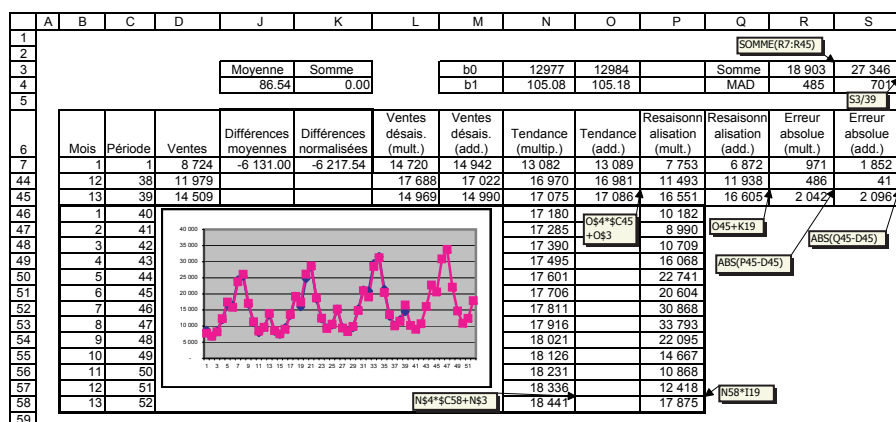


FIG. 19 – Validation du modèle et prévisions

Le modèle multiplicatif étant en l'occurrence le plus précis, c'est celui que l'on va choisir pour les prévisions. On calcule donc la tendance pour les périodes 40 à 52 (cellule N46 à cellule N58) et on les resaisonnalise (cellule P46 à cellule P58). On trouve ainsi que les ventes de la première période de l'année suivante seront de 10182 (cellule P46). L'écart type sur l'erreur étant de 687, on peut dire que l'intervalle de confiance à 95% des ventes à venir est grossièrement de $10182 - 2 \times 687$ à $10182 + 2 \times 687$, soit [8808 - 11556].

2.4 Décomposition

2.4.4 Récapitulation de la méthode

Le calcul des prévision sur la base des moyennes mobiles suit toujours les mêmes étapes. Reprenons le cas d'un modèle multiplicatif :

- Numérotation des périodes connues par incrément de 1.
- Calcul des moyennes mobiles centrées.
- Calcul des rapports saisonniers (valeur réelle divisée par valeur lissée).
- Calcul des rapports moyens, soit la moyenne des rapports d'une période donnée.
- Normalisation des rapports, soit rapport multiplié par le nombre de période d'un cycle; le tout divisé par la somme des rapports moyens.
- Calcul des valeurs désaisonnalisées, soit valeur réelle divisée par le rapport saisonnier de la période.
- Calcul de l'équation de la droite de tendance sur la série désaisonnalisée avec par exemple la fonction `DROITEREG`. Si la tendance n'est pas linéaire, on pourra passer par les transformations adéquates.
- Calcul de la tendance, soit application de l'équation de tendance en fonction des numéros de périodes.
- Resaisonnalisation, soit valeur de la tendance multipliée par le rapport saisonnier correspondant.
- Calcul de l'erreur, par exemple erreur absolue et MAD.
- Calcul des prévisions, soit poursuite du calcul des valeurs resaisonnalisées.



2.4.5 Lissage exponentiel avec tendance et saisonnalité

La méthode que l'on vient de présenter souffre des mêmes limitations que le lissage par moyennes mobiles évoqué plus haut. Là aussi, il est possible d'y pallier en partie en y substituant un lissage exponentiel.

2.4.6 Le modèle de Holt

Le lissage exponentiel tel qu'on l'a présenté plus haut est théoriquement limité aux séries quasi-stationnaires, donc ne présentant ni tendance, ni saisonnalité. En cas d'existence d'une tendance, le lissage exponentiel simple va systématiquement sous-estimer la prévision (ou la sur-estimer si la tendance

2.4 Décomposition

est décroissante). À partir d'une tendance linéaire ($x_t = b_0 + b_1 t$), le *modèle de Holt* propose le lissage suivant :

$$\begin{aligned} S_t &= \alpha y_t + (1 - \alpha)(S_{t-1} + T_{t-1}) \\ T_t &= \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \\ \hat{y}_{t+1} &= S_t + T_t \end{aligned}$$

La première équation est une simple transformation du lissage exponentiel où l'on tient compte du fait que le niveau doit être corrigé de la tendance T . La seconde applique le modèle de lissage à la composante tendancielle. La prévision immédiate est calculée comme la somme de la dernière valeur connue lissée et de la tendance. Si l'on veut poursuivre la prévision sur les périodes ultérieures, on applique simplement la dernière tendance connue $\hat{y}_{t+p} = \hat{y}_t + pT_t$. Comme pour le lissage exponentiel simple, on peut chercher les valeurs optimum de α et β grâce au solveur.

L'exemple suivant (Figure 20, page 35) montre comment mettre en œuvre le modèle de Holt. Dans un souci de simplicité, il est basé sur des données trimestrielles, mais on peut évidemment aisément l'étendre à toute périodicité souhaitée. La question se pose de l'initialisation du modèle. De manière courante, on prend pour niveau initial la première valeur de la série (x_0) et pour tendance initiale, soit 0, soit la différence entre les deux premières valeurs connues ($x_1 - x_0$), dans ce cas le calcul ne peut commencer qu'en $t = 2$.

2.4.7 Le modèle de Holt-Winters

Le modèle de Holt-Winters étend le lissage exponentiel aux séries présentant à la fois une composante tendancielle et une composante cyclique. Dans le cas d'un modèle additif ($y_t = N + T_t + C_t + \epsilon_t$ où N est le niveau de base — représenté par la valeur lissée S —, T la tendance, C le cycle et ϵ la composante aléatoire), on utilisera les fonctions suivantes, avec p le nombre de périodes du cycle (par exemple 12 dans le cas d'une périodicité annuelle et de données mensuelles) :

$$\begin{aligned} S_t &= \alpha(y_t - C_{t-p}) + (1 - \alpha)(S_{t-1} + T_{t-1}) \\ T_t &= \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \\ C_t &= \gamma(y_t - S_t) + (1 - \gamma)C_{t-p} \\ \hat{y}_{t+1} &= S_t + T_t + C_{t-p+1} \end{aligned}$$

2.4 Décomposition

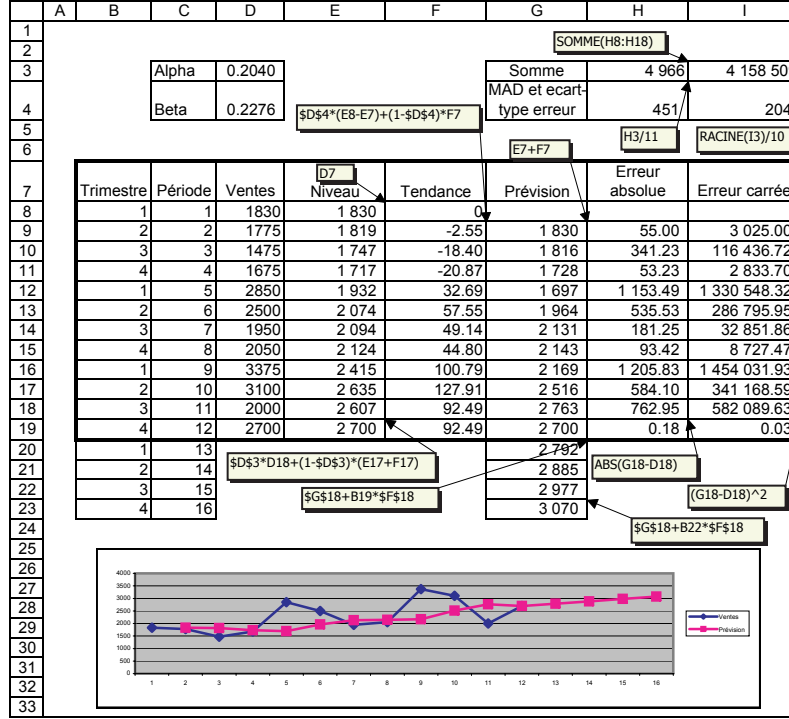


FIG. 20 – Lissage exponentiel double (Holt)

Pour les prévisions ultérieures, on suit la même logique que pour le modèle de Holt c'est-à-dire que l'on se base sur les derniers éléments connus : $\hat{y}_{t+k} = S_t + kT_t + C_{t-p+k}$.

Plus courant, le modèle multiplicatif ($y_t = (N + T_t)C_t + \epsilon_t$) conduit aux fonctions suivantes :

$$S_t = \alpha(y_t/C_{t-p}) + (1 - \alpha)(S_{t-1} + T_{t-1})$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

$$C_t = \gamma(y_t/S_t) + (1 - \gamma)C_{t-p}$$

$$\hat{y}_{t+1} = (S_t + T_t)C_{t-p+1}$$

avec pour les prévisions ultérieures : $\hat{y}_{t+k} = (S_t + kT_t)C_{t-p+k}$. Là encore, le solveur permet de trouver les valeurs de α, β, γ .

L'exemple suivant (Figure 21, page 36) montre la mise en œuvre du modèle de Holt-Winters dans le cas multiplicatif. La série étant saisonnière,

2.4 Décomposition

on peut vérifier que, sur les mêmes données, les résultats sont beaucoup plus précis qu'avec le lissage exponentiel double de Holt. Là encore se pose le problème de l'initialisation. Nous avons retenu ici la méthode la plus courante, soit :

- Le niveau initial est la moyenne de la périodicité traitée (ici l'année).
- Le trend initial est la différence moyenne entre les valeurs de fin et de début de période.
- Les coefficients cycliques (saisonniers ici) sont la différence — pour le modèle additif — ou le rapport — modèle multiplicatif — entre la valeur courante et le niveau initial.

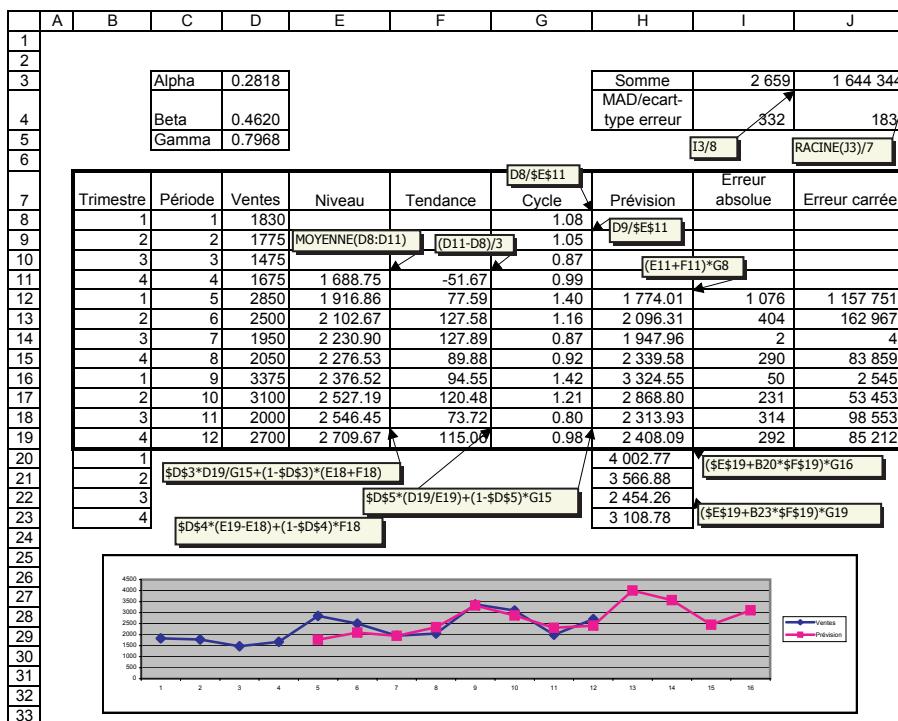


FIG. 21 – Lissage exponentiel triple (Holt-Winters multiplicatif)

2.4.8 Récapitulation de la méthode

L'utilisation du lissage exponentiel triple de Holt-Winter n'est pas très compliqué. Il nécessite simplement une certaine rigueur. Reprenons le cas d'un modèle multiplicatif.

2.4 Décomposition

- Initialisation du niveau, soit moyenne sur l'ensemble de la périodicité. Par exemple si on travaille sur des mois, le niveau initial sera calculé en décembre comme la moyenne de l'ensemble de l'année.
- Calcul des niveaux suivants, soit α multiplié par le rapport entre la valeur réelle et la valeur du cycle correspondant à la période courante (valeur non encore calculée). À ceci, on ajoute $(1 - \alpha)$ multiplié par la somme du niveau précédent et de la tendance précédente (non encore calculée).
- Initialisation de la tendance, soit valeur de fin de période moins valeur initiale multipliée par le nombre de périodes constitutives d'une périodicité moins un. Par exemple pour une série mensuelle, on initialisera la tendance en décembre par valeur décembre moins valeur janvier, le tout divisé par 11.
- Calcul des tendances suivantes, soit β multiplié par niveau précédent moins niveau courant ; le tout additionné de $(1 - \beta)$ fois la tendance précédente.
- Initialisation des cycles, soit pour chacune des périodes constitutives de la première périodicité, valeur réelle divisée par niveau initial. Sur 12 mois, on calculera successivement valeur janvier divisée par niveau initial, valeur février divisée par niveau initial ...
- Calcul des cycles, soit γ multiplié par le rapport entre la valeur courante réelle et le niveau correspondant. On y ajoute $(1 - \gamma)$ fois la valeur du cycle correspondant précédent (par exemple pour janvier, on prendra la valeur du cycle pour janvier de l'année précédente).
- Calcul des valeurs théoriquement prévues sur les périodes connues, soit niveau précédent plus tendance précédente, le tout multiplié par la valeur du cycle précédent correspondant.
- Calcul de l'erreur, par exemple erreur absolue et MAD.
- Calcul des prévisions. On commence par numéroter de 1 à n chacune des périodes constitutives de la périodicité suivante. Les prévisions sont alors égales à dernier niveau calculé plus dernière tendance multiplié par numéro de la période courante ; le tout multiplié par la valeur du cycle précédent correspondant.



2.5 Méthodes autorégressives

2.5 Méthodes autorégressives

Les méthodes autorégressives considèrent que le futur immédiat ne dépend que du passé proche. Elles sont très diverses et complexes et on se contentera ici de la plus simple d'entre elles qui pose un ensemble de modèles linéaires de la forme : $\hat{y}_t = b_0 + \sum_{i=1}^{i=p} b_i y_{t-i} + \epsilon_t$, où p est l'ordre de l'autorégression. On aura ainsi pour un modèle d'ordre 1 : $\hat{y}_t = a_0 + a_1 y_{t-1} + \epsilon_t$, et pour un modèle d'ordre 2 : $\hat{y}_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t$.

Puisque l'on pose une relation linéaire, le modèle d'ordre 1 peut se résoudre avec une régression linéaire simple. Au delà de 1, on utilisera la régression linéaire multiple. Construire un modèle autorégressif signifie donc simplement appliquer les outils de la régression linéaire à des variables qui sont en fait les valeurs successives de la variable observée dans le passé. La difficulté réside dans le choix de l'ordre, sachant que plus il est élevé, plus le calcul sera lourd et surtout plus de données seront perdues (on perd les p premières valeurs pour un modèle d'ordre p ; il est par exemple impossible de calculer \hat{y}_2 avec $p = 2$ puisqu'il faudrait disposer de y_0 qui n'existe pas.)

Illustrons la démarche par un exemple. On dispose des statistiques de résultats sur les 25 dernières périodes. La première étape consiste à remplir les colonnes correspondant aux différents décalages. On prendra dans notre exemple des décalages de 1, 2 et 3 (Figure 22, page 38) .

	A	B	C	D	E	F
1						
2		Période	Résultats	Décalage 1	Décalage 2	Décalage 3
3		1	1 161.88			
4		2	1 191.58	1 161.88	=C3	
5		3	1 236.61	1 191.58		=D4
6		4	1 340.56	1 236.61	1 161.88	=E5
7		5	1 382.71	1 340.56	1 236.61	1 191.58
8		6	1 474.37	1 382.71	1 340.56	1 236.61
9		7	1 419.88	1 474.37	1 382.71	1 340.56
10		8	1 404.50	1 419.88	1 474.37	1 382.71
11		9	1 287.75	1 404.50	1 419.88	1 474.37
12		10	1 275.81	1 287.75	1 404.50	1 419.88
13		11	1 232.30	1 275.81	1 287.75	1 404.50
14		12	1 312.04	1 232.30	1 275.81	1 287.75
15		13	1 468.93	1 312.04	1 232.30	1 275.81
16		14	1 791.54	1 468.93	1 312.04	1 232.30
17		15	1 857.29	1 791.54	1 468.93	1 312.04
18		16	1 812.88	1 857.29	1 791.54	1 468.93
19		17	1 788.45	1 812.88	1 857.29	1 791.54
20		18	1 801.59	1 788.45	1 812.88	1 857.29
21		19	1 414.98	1 801.59	1 788.45	1 812.88
22		20	1 158.01	1 414.98	1 801.59	1 788.45
23		21	1 258.28	1 158.01	1 414.98	1 801.59
24		22	1 301.72	1 258.28	1 158.01	1 414.98
25		23	1 132.54	1 301.72	1 258.28	1 158.01
26		24	1 030.41	1 132.54	1 301.72	1 258.28
27		25	1 059.68	1 030.41	1 132.54	1 301.72

FIG. 22 – Données de ventes et décalages

2.5 Méthodes autorégressives

Dans une seconde étape, on utilise l'utilitaire d'analyse de régression linéaire du tableur, pour réaliser trois régressions différentes avec y_t comme variable dépendante, soit :

1. y_{t-1} comme seule variable indépendante.
2. y_{t-1}, y_{t-2} donc deux variables indépendantes.
3. $y_{t-1}, y_{t-2}, y_{t-3}$ donc trois variables indépendantes.

On peut voir (Figure 23, page 39) la synthèse des résultats. Dans le cas d'une régression simple, la constante n'est pas significative (p-valeur de 0,25). La régression par rapport à y_{t-1}, y_{t-2} donne un ensemble de variables significatifs. La régression par rapport à $y_{t-1}, y_{t-2}, y_{t-3}$ donne une constante et un y_{t-3} non significatifs. On retiendra donc, en première approche la régression par rapport à y_{t-1}, y_{t-2} . L'équation de régression est alors : $\hat{y}_t = 358,18 + 1.276y_{t-1} - 0.532y_{t-2}$. Il suffit d'appliquer cette équation pour $t > 25$ pour obtenir les prévisions souhaitées.

	Coefficients	p-valeur	Erreur type
Constante	199.96	0.25	138.67
Variable X 1	0.85	0.00	
Constante	358.18	0.04	124.68
Variable X 1	1.28	0.00	
Variable X 2	-0.53	0.01	
Constante	333.65	0.12	130.85
Variable X 1	1.32	0.00	
Variable X 2	-0.64	0.09	
Variable X 3	0.09	0.73	

FIG. 23 – Régressions linéaires sur les trois décalages

3 Décisions et arbres d'analyse

Face à un ensemble d'alternatives, quelle décision prendre ? Évidemment en situation de certitude, il suffira de sélectionner celle qui maximise (ou minimise) le critère. Si j'ai le choix entre deux décisions dont j'ai la certitude que la première rapportera x alors que la seconde rapportera $x+a$ avec $a > 0$, je choisirai naturellement la seconde.

La difficulté provient ainsi du fait qu'en réalité, on évolue en univers incertain, c'est-à-dire que l'on est au mieux capable d'estimer les probabilités d'occurrence des différentes alternatives, il n'y a aucune garantie de succès. Dans une première partie, nous allons voir les méthodes élémentaires qui permettent de faire un choix sur une période en fonction de son appétence au risque. La seconde partie sera consacrée aux arbres de décision qui permettent de sélectionner le meilleur enchaînement de décisions en fonction de ses objectifs propres.

3.1 Décisions uniques

Pour introduire à la théorie de la décision, adaptons l'un des exemples les plus célèbres de la littérature : *l'omelette de Savage*¹¹.

Souhaitant faire une omelette, vous disposez de 6 œufs, mais vous avez des doutes sur la fraîcheur de l'un d'entre eux. Vous avez trois choix :

- Mettre l'œuf douteux dans l'omelette (décision $D1$).
- Le casser à part pour en vérifier la fraîcheur (décision $D2$).
- Le jeter (décision $D3$).

En fonction de l'état réel de l'œuf (état de la nature), à savoir sain (état $S1$ ¹²) ou pourri (état $S2$), il vous est possible de classer vos préférences selon la décision (Table 2, page 41) . Profitons en pour noter que les utilités sont subjectives. Pour une personne marquant une très forte allergie au nettoyage des plats, on pourrait avoir par exemple une utilité de 70 au fait de casser à part un œuf sain parce qu'on ne veut vraiment pas avoir un plat à laver. Dans ce cas, quel que soit l'état de la nature, jeter l'œuf ($D3$) sera préférable au fait de le casser à part ($D2$) (utilité de 75 contre 70 en situation $S1$ et utilité de 85 contre 50 pour $S2$). On dit que $D3$ domine $D2$. La dernière n'aurait donc pas de sens et pourrait d'emblée être éliminée de l'analyse.

L'utilité (ou le gain) des différentes alternatives étant fixé, comment arriver à une décision ?

¹¹Savage, L., *Foundations of Statistics*, Wiley, 1954.

¹²On utilise classiquement la lettre S pour les états de la nature (*State of Nature*).

3.1 Décisions uniques

	œuf sain ($S1$)	œuf pourri ($S2$)
Mis dans l'omelette ($D1$)	On dispose d'une omelette de 6 œufs. utilité : 100	5 œufs fichus. utilité : 0
Cassé à part ($D2$)	On dispose d'une omelette de 6 œufs et on doit laver le plat. utilité : 80	On dispose d'une omelette de 5 œufs et on doit laver le plat. utilité : 50
Jeté ($D3$)	On dispose d'une omelette de 5 œufs et on a perdu un œuf sain. utilité : 75	On dispose d'une omelette de 5 œufs. utilité : 85

TAB. 2 – Synthèse des résultats

Le Maximax : C'est le mode de décision optimiste : le choix correspondra au *maximum* des *maxima* de chacune des décisions. Pour $D1$, $D2$ et $D3$, les maxima sont respectivement de 100, 80 et 85. On choisira donc $D1$, la décision qui rapporte le plus dans la meilleure des situations.

Le Maximin : C'est le mode de décision pessimiste : le choix correspondra au *maximum* des *minima* de chacune des décisions. Pour $D1$, $D2$ et $D3$, les minima sont respectivement de 0, 50 et 75. On choisira donc $D3$, la décision qui rapporte le plus dans la pire des situations.

Le Minimax Regret : On va chercher ici à minimiser le *regret* (la perte d'opportunité) maximum c'est-à-dire que l'on va faire en sorte d'être le moins déçu possible. Dans notre exemple, si l'œuf est sain et que l'on a choisi $D1$, le regret est nul, pour $D2$, il est de $100 - 80 = 20$, et pour $D3$ de $100 - 75 = 25$. Si l'œuf est pourri, on a pour $D1$ un regret de $85 - 0 = 85$, pour $D2$ de $85 - 50 = 35$ et pour $D3$ de 0. Pour $D1$, on a donc un regret maximum de 85 et respectivement pour $D2$ et $D3$ des regrets de 35 et de 25. Le minimum de ces trois valeurs 25 correspond à la décision $D3$ qui est donc celle que l'on adoptera si l'on suit le Minimax Regret.

3.2 Méthodes probabilistes

Le critère de Laplace : C'est la méthode la plus simple : on sélectionne la décision qui donne le meilleur résultat moyen. On a ici : $L_{D1} = (100+0)/2 = 50$, $L_{D2} = (80 + 50)/2 = 65$, $L_{D3} = (75 + 85)/2 = 80$. On choisira donc $D3$.

Les critères de Hurwitz : Cette méthode permet de s'affranchir de la dimension « tout ou rien » des critères optimistes et pessimistes. On définit un coefficient α qui, d'une certaine manière, positionne le curseur de l'optimisme. La valeur étudiée sera égale à α fois l'utilité optimiste plus $1 - \alpha$ fois l'utilité pessimiste. Dans le cas de l'omelette, si on pose par exemple $\alpha = 0,4$, soit un certain pessimisme, on aura : $H_{D1} = 0.4 \times 100 + 0.6 \times 0 = 40$, $H_{D2} = 0.4 \times 80 + 0.6 \times 50 = 62$, $H_{D3} = 0.4 \times 85 + 0.6 \times 75 = 79$. On prendra donc la décision $D3$.

3.2 Méthodes probabilistes

Les méthodes que l'on vient de voir supposent de fait que l'on considère que les probabilités d'occurrence des différents états de la nature sont égales. Dans les faits, on a tendance à assigner différentes valeurs à ces probabilités qu'elles soient le résultat d'un calcul, de l'expérience ou de la simple intuition. Dans ce cas, on applique le critère de la *valeur espérée*.

Considérons la matrice suivante (Table 3, page 42) :

	$S1$	$S2$	$S3$
$D1$	8	4	6
Regret $D1$	12	0	2
$D2$	2	3	8
Regret $D2$	18	1	0
$D3$	20	1	-30
Regret $D3$	0	3	38
Proba.	0.25	0.6	0.15

TAB. 3 – Synthèse des résultats

La valeur espérée est la somme des produits des résultats par les probabilités. On a :

- $E_{D1} = 0.25 \times 8 + 0.6 \times 4 + 0.15 \times 6 = 5.3$
- $E_{D2} = 0.25 \times 2 + 0.6 \times 3 + 0.15 \times 8 = 3.5$
- $E_{D3} = 0.25 \times 20 + 0.6 \times 1 + 0.15 \times -30 = 1.1$

On prendra donc la décision $D1$. Si on avait fait le même calcul sur les regrets (critère du regret espéré), on aurait pris la même décision. Les deux

3.3 Arbres de décision

méthodes sont équivalentes en ce qu'elles mènent à la même décision si l'on veut maximiser le résultat espéré ou minimiser la perte espérée.

Quel intérêt peut on trouver à minimiser l'incertitude? Exprimée autrement, combien pouvons nous accepter de payer pour connaître l'avenir? La *valeur espérée de l'information parfaite* (EVPI¹³) est la différence entre ce que l'on pourrait espérer gagner si on levait l'incertitude et ce que l'on pourrait gagner en situation d'incertitude, soit la valeur espérée calculée plus haut.

Si on pouvait lever l'incertitude, on pourrait espérer gagner la somme des résultats maximum de chacun des états de la nature par leurs probabilités. Soit dans notre exemple : 0.25×20 pour $S1$, plus 0.6×4 pour $S2$, plus 0.15×8 pour $S3$, soit au total 8.6.

On a calculé plus haut que la valeur espérée en situation d'incertitude est de 5.3, on a donc : $EVPI = 8.6 - 5.3 = 3.3$. Ce qui signifie que l'on peut investir à concurrence de 3.3 pour acquérir l'information.

3.3 Arbres de décision

Les arbres de décision sont généralement utilisés pour les décisions successives, ils permettent ainsi de déterminer une *stratégie*. On va en présenter les bases à travers un petit exemple.

L'entreprise *NanoBio* met au point un nouveau produit utilisant une technologie encore unique. Elle doit décider du moment de son introduction sur le marché. On peut décider d'un lancement immédiat, mais la technologie n'étant pas encore complètement mature, on estime qu'il y a 25% de chances que la demande soit forte (on aura alors un profit de 70M€) et 75% qu'elle soit faible (perte 15M€). L'alternative est dans le report de la sortie au semestre suivant, le produit sera alors plus mature — au prix d'un investissement supplémentaire et R&D de 4M€ — et les risques de rejets du marché plus faibles. Dans ce cas, il existe toutefois un risque que le principal concurrent, *BioNano*, soit désormais prêt et sorte un produit équivalent. *NanoBio* estime la probabilité d'entrée du concurrent à 40%. Si le concurrent sort son produit, on estime qu'il y a alors 40% de chances que la demande soit forte (profit 50M€) et 60% qu'elle soit faible (perte 80M€). Sinon, il y a 75% de chances que la demande soit forte (profit 100M€) et 25% qu'elle soit faible (profit 10M€).

On peut synthétiser le problème avec la matrice de résultats suivante (Table 4, page 44) .

¹³On utilisera l'acronyme anglais de Expected Value of Perfect Information

3.3 Arbres de décision

Entrée immédiate					
Demande	Probabilité	Profit			
Haute	0.25	70			
Faible	0.75	-15			
Entrée reportée					
Demande	<i>BioNano</i> entre		<i>BioNano</i> n'entre pas		
	Probabilité	Profit	Probabilité	Profit	
	Haute	0.4	50	0.75	100
	Basse	0.6	-80	0.25	10
Proba. entrée conc.	0.4	Coûts de R&D	4		

TAB. 4 – Conséquences de la décision de sortie du produit

On peut représenter le problème avec un arbre de décision. La symbolisation est élémentaire : le trait vertical représente un nœud terminal, un cercle représente un événement et un carré représente une décision. Pour chacun des arcs, on peut poser le libellé et le résultat (coût/profit) correspondant. Pour les nœuds terminaux, on peut également poser les probabilités d'occurrence (Figure 24, page 44) .

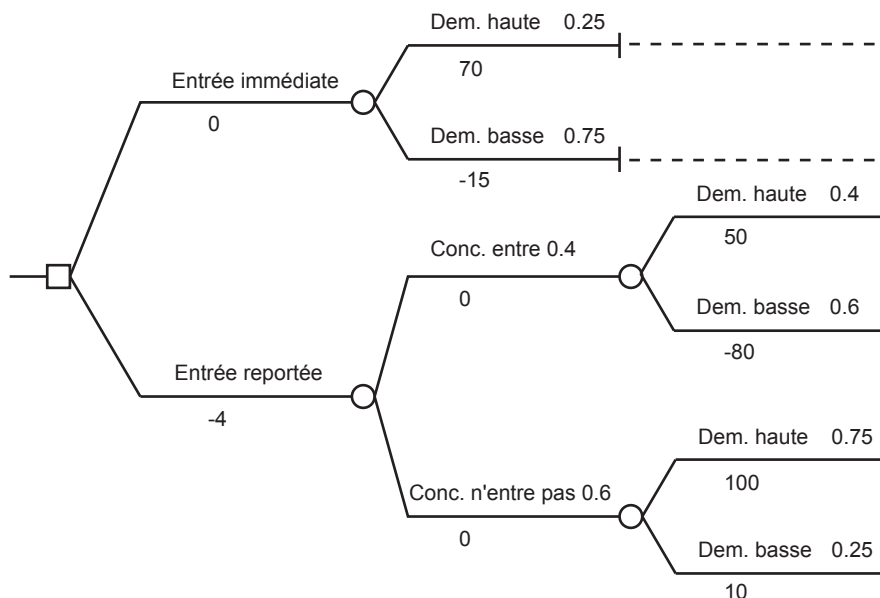


FIG. 24 – Arbre d'analyse de la décision de *NanoBio*

3.3 Arbres de décision

Pour calculer l'arbre on procède en remontant des sommets vers la racine. On distingue deux méthodes : le paiement immédiat et le paiement à terme. Dans le premier cas, les paiements ont lieu à chaque événement et on mesure le résultat maximum à partir de ce point quoi qu'il se soit passé en amont ; dans le second ils n'ont lieu qu'à la fin et on mesure le résultat maximum en fin de process en tenant compte des coûts et résultats associés au chemin menant au point courant.

Paiement immédiat (Figure 25, page 46)

- On donne la valeur 0 aux nœuds terminaux.
- Pour chaque arc associé à un événement, on cumule les valeurs des nœuds amonts et on multiplie par la probabilité de l'arc en cours. Par exemple, pour l'arc « Conc. entre », on aura $(50 + 0) \times 0.4 + (-80 + 0) \times 0.6 = -28$. Pour « Conc. n'entre pas », on aura $(100 + 0) \times 0.75 + (10 + 0) \times 0.25 = 77.5$. Pour « Entrée immédiate », on a $(70 + 0) \times 0.25 + (-15 + 0) \times 0.75 = 6.25$. Pour « Entrée reportée », on a $(0 - 28) \times 0.4 + (0 + 77.5) \times 0.6 = 35.3$.
- Pour les arcs associés à une décision, on prend la plus grande des valeurs liées aux différentes alternatives. Soit ici $0 + 6.25 = 6.25$ pour l'entrée immédiate et $-4 + 35.3 = 31.3$ pour l'entrée reportée. C'est cette dernière valeur qui sera donc retenue. On décidera de reporter l'entrée sur le marché.

3.3 Arbres de décision

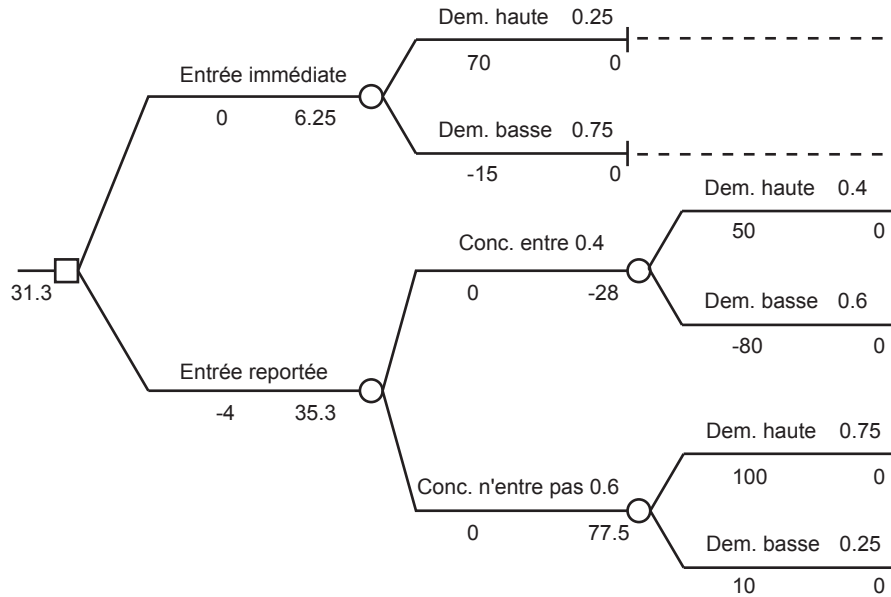


FIG. 25 – Calcul avec paiement immédiat

Paieement à terme (Figure 26, page 47)

- La valeur des nœuds terminaux est la somme des paiements et des coûts associés au chemin qui mène au nœud. Dans l'hypothèse d'une entrée immédiate, pour « Dem. haute », on aura $0 + 70 = 70$; pour « Dem. basse », on aura $0 - 15 = -15$. Si l'entrée est reportée et que la concurrence entre, pour « Dem. haute », on aura $-4 + 0 + 50 = 46$; pour « Dem. basse », on aura $-4 + 0 - 80 = -84$. Enfin, si le concurrent n'entre pas, pour « Dem. haute », on aura $-4 + 0 + 100 = 96$; pour « Dem. basse », on aura $-4 + 0 + 10 = 6$.
- Pour les arcs associés à des évènements, on cumule le produit de la valeur de chacun des arcs amonts par leur probabilité. Pour « Conc. entre » on aura $(0.4 \times 46) + (0.6 \times -84) = -32$. Pour « Conc. n'entre pas » on aura $(0.75 \times 96) + (0.25 \times 6) = 73.5$. Pour « Entrée immédiate », on aura $(0.25 \times 70) + (0.75 \times -15) = 6.25$ et pour « Entrée reportée », $(0.4 \times -32) + (0.6 \times 73.5) = 31.3$.
- La valeur du nœud associé à la décision est la plus grande des valeurs des arcs qui en émergent, soit ici 31.3 qui correspond à la décision de report d'entrée.

3.3 Arbres de décision

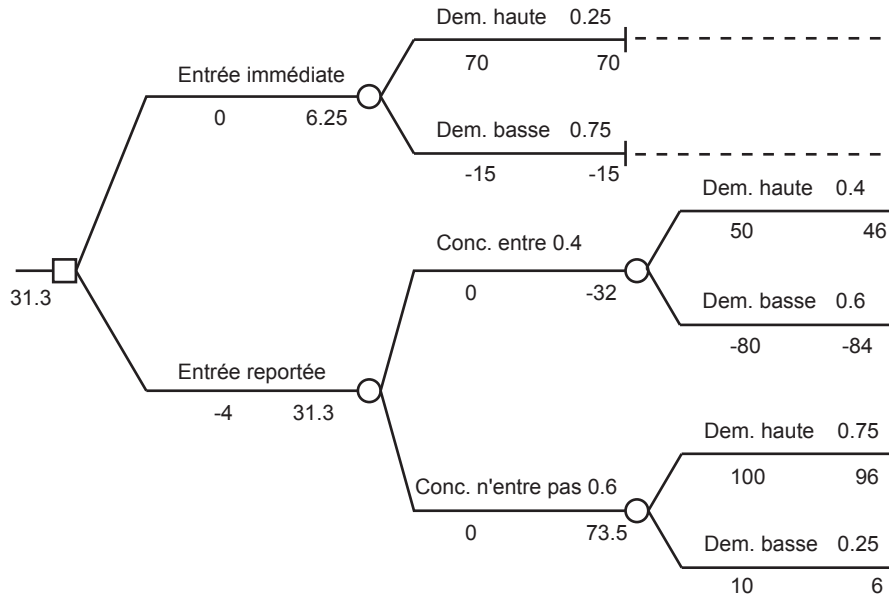


FIG. 26 – Calcul avec paiement à terme

3.3.1 Décisions successives

Face à l'ampleur de l'enjeu, *NanoBio* envisage de faire appel à un cabinet d'étude spécialiste du domaine afin d'avoir une vision aussi précise que possible de la probabilité d'entrée de son concurrent. Elle consulte le cabinet *GMPK* — dont une division est spécialisée dans l'analyse du marché des biotechnologies — qui lui remet un devis et, afin de crédibiliser son offre, une présentation de la pertinence de ses études antérieures concernant l'entrée de nouveaux arrivants sur le marché (Table 5, page 47) .

Entrée d'un nouveau	Prévision du cabinet	
	Entrée	Non entrée
Oui	85 sur 100 (85%)	15 sur 100 (15%)
Non	8 sur 32 (25%)	24 sur 32 (75%)

TAB. 5 – Pertinence des analyses de *GMPK*

NanoBio doit maintenant gérer deux décisions successives : Commander l'étude ou non, puis entrer maintenant ou plus tard. Pour compléter le nouvel arbre, on doit faire appel aux *probabilités conditionnelles* et au *théorème de Bayes*. Si on a deux évènements A et B , on note $P(A.B)$ la probabilité de

3.3 Arbres de décision

réalisation simultanée de A et de B ; on note $P(A|B)$ la probabilité de A sachant B . On a :

$$P(A|B) = \frac{P(A.B)}{P(B)}$$

On peut alors écrire :

$$P(A|B)P(B) = P(A.B) = P(B|A)P(A)$$

soit :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

On doit d'abord mesurer le gain d'information lié à l'achat de l'étude. Pour ce faire, on doit connaître les probabilités d'obtenir la réponse « Entrée ». On note C_E la réponse du cabinet selon laquelle le concurrent va entrer, C_R la réponse inverse, E l'entrée du concurrent et R le report de l'entrée du concurrent. On a :

$$P(C_E|E) = \frac{P(C_E.E)}{P(E)} \Leftrightarrow P(C_E.E) = P(C_E|E)P(E) = 0.85 \times 0.4 = 0.34$$

$$P(C_R.E) = P(C_R|E)P(E) = 0.15 \times 0.4 = 0.06$$

$$P(C_E.R) = P(C_E|R)P(R) = 0.25 \times 0.6 = 0.15$$

$$P(C_R.R) = P(C_R|R)P(R) = 0.75 \times 0.6 = 0.45$$

À partir de là, on peut calculer la probabilité de recevoir un rapport prévoyant l'entrée : $P(C_E) = P(C_E.E) + P(C_E.R) = 0.34 + 0.15 = 0.49$; celle d'obtenir un rapport de report : $P(C_R) = P(C_R.R) + P(C_R.E) = 0.45 + 0.06 = 0.51$; celle d'avoir une entrée : $P(E) = P(E.C_E) + P(E.C_R) = 0.34 + 0.06 = 0.4$ et celle d'avoir un report : $P(R) = P(R.C_R) + P(R.C_E) = 0.45 + 0.15 = 0.6$.

	Réponse du cabinet		Proba. marginale
	C_E	C_R	
E	0.34	0.06	0.4
R	0.15	0.45	0.6
Proba. Marginale	0.49	0.51	1

TAB. 6 – Probabilités conjointes

3.3 Arbres de décision

Sur ces bases, on peut commencer à mesurer l'apport du cabinet. Si on reçoit un rapport d'entrée, quelle est la probabilité *a posteriori* que le concurrent entre ; exprimé différemment, en quoi notre mesure de la probabilité d'entrée (qui était de 40%) est-elle modifiée par le rapport ? On cherche $P(E|C_E) = \frac{P(E.C_E)}{P(C_E)} = \frac{0.34}{0.49} = 0.6939$ ou encore $P(E|C_E) = \frac{P(C_E|E).P(E)}{P(C_E)} = \frac{0.85 \times 0.4}{0.49} = 0.6939$. Le fait de recevoir un rapport positif fait passer la probabilité d'entrée qui était *a priori* de 0.4 à 0.6939. De même, pour la probabilité de report, on a : $P(R|C_R) = \frac{P(R.C_R)}{P(C_R)} = \frac{0.45}{0.51} = 0.8824$ ou encore $P(R|C_R) = \frac{P(C_R|R).P(R)}{P(C_R)} = \frac{0.75 \times 0.6}{0.51} = 0.8824$. Le fait de recevoir un rapport négatif fait passer la probabilité de report qui était *a priori* de 0.6 à 0.8824. On voit ainsi que le recours au cabinet réduit l'incertitude. On peut maintenant calculer le nouvel arbre de décision en y incluant les probabilités *a posteriori*. On fera le calcul sur les paiements à terme (Figure 27, page 52) .

Dans le cas où l'on ne consulte pas le cabinet, l'arbre est évidemment le même qu'au cas précédent. Plaçons nous maintenant dans le cas où on consulte (coût de 1.5) et où on obtient un rapport prévoyant l'entrée du concurrent. On traite d'abord des nœuds terminaux :

1. Si on entre immédiatement et que la demande est haute, on aura :
 $-1.5 + 0 + 0 + 70 = 68.5$.
2. Dans la même situation avec une demande basse, on aura :
 $-1.5 + 0 + 0 - 15 = -16.5$.
3. Si on décide de reporter l'entrée et que la concurrence entre, alors :
 - (a) Si la demande est haute, on aura :
 $-1.5 + 0 - 4 + 0 + 50 = 45.5$.
 - (b) Sinon on a :
 $-1.5 + 0 - 4 + 0 - 80 = -85.5$.
4. Si on décide de reporter l'entrée et que la concurrence n'entre pas, alors :
 - (a) Si la demande est haute, on aura :
 $-1.5 + 0 - 4 + 0 + 100 = 94.5$.
 - (b) Sinon on a :
 $-1.5 + 0 - 42 + 0 + 10 = 4.5$.

Dans le cas où on reçoit un rapport prévoyant le report de l'entrée, les résultats sont naturellement identiques, les coûts terminaux ne dépendant pas des probabilités d'occurrence des événements.

3.3 Arbres de décision

On remonte maintenant l'arbre du recrutement du cabinet d'un niveau. Le cas de l'introduction immédiate étant déjà traité (ce sont des nœuds terminaux), il ne reste que les cas de report de l'introduction.

Si on a reçu un rapport prévoyant l'entrée du concurrent, alors dans l'hypothèse où on reporte sa propre entrée, on a :

1. Si le concurrent entre :
$$0.4 \times 44.5 + 0.6 \times -85.5 = -33.5$$
2. Si le concurrent n'entre pas :
$$0.75 \times 94.5 + 0.25 \times 4.5 = 72$$

Si le rapport prévoit le report de l'entrée du concurrent, les résultats sont les mêmes puisqu'on ne fait ici appel qu'aux probabilités concernant le niveau de la demande, lesquelles sont indépendantes du résultat du rapport.

On remonte maintenant d'un niveau supplémentaire. Dans le cas où on reçoit un rapport prévoyant l'entrée du concurrent, alors :

1. Si on décide d'introduire maintenant, on a :
$$0.25 \times 68.5 + 0.75 \times -16.5 = 4.75$$
2. Si on décide de reporter l'entrée, on a :
$$0.6939 \times -33.5 + 0.3061 \times 72 = -1.2065$$

Si le rapport prévoit le report de l'entrée du concurrent, alors :

1. Si on décide d'introduire maintenant, on a :
$$0.25 \times 68.5 + 0.75 \times -16.5 = 4.75$$
2. Si on décide de reporter l'entrée, on a :
$$0.1176 \times -33.5 + 0.8824 \times 72 = 59.5932$$

On est maintenant à un nœud de décision dans le cas où l'on a consulté le cabinet. Si on a reçu un rapport prévoyant l'entrée, la valeur est le maximum entre 4.75 et -1.2065, donc 4.75. Dans le cas contraire, la valeur est le maximum entre 4.75 et 59.5932, donc 59.5932.

Il reste à calculer la valeur de la branche de recrutement du cabinet, soit : $0.49 \times 4.75 + 0.51 \times 59.5932 = 32.72$. On calcule enfin la valeur de la décision initiale de recrutement qui est le maximum entre 31.3 (que l'on a calculé à l'exemple précédent) et 32.72, donc 32.72.

La stratégie à suivre est donc la suivante :

- On recrute le cabinet.
- Si on reçoit un rapport prévoyant l'entrée du concurrent, on entre immédiatement sur le marché.
- Sinon, on reporte l'introduction.

3.3 Arbres de décision

On a ainsi validé des décisions *séquentielles* et exploré la valeur de l'ensemble des possibles.

La question se pose enfin de la pertinence du recours au cabinet. On va calculer la « Expected Value of Sample Information » (EVSI¹⁴). Celle-ci est la différence entre la valeur espérée pour une information imparfaite gratuite (soit $32.72 + 1.5 = 34.22$ car il faut supprimer le coût de 1.5 demandé par le cabinet) et la valeur espérée sans information 31.3, soit dans notre cas : $34.22 - 31.3 = 2.92$. Il sera donc utile de recruter le cabinet tant que ses honoraires seront inférieurs à cette valeur.

On peut de même calculer la valeur espérée de l'information parfaite (*Expected Value of Perfect Information EVPI*). Supposons que l'on ait la certitude que le concurrent va entrer, les alternatives sont alors les suivantes :

1. On ne commande pas d'étude et on entre immédiatement, la valeur est alors de 6.25.
2. On ne commande pas d'étude et on reporte l'entrée. Dans ce cas, sachant que la concurrence va entrer, on peut espérer gagner -32 .
3. Si on commande l'étude, on a :
 - (a) En cas de rapport prévoyant l'entrée on peut entrer maintenant et gagner 4.75 ou reporter et gagner -33.5 .
 - (b) En cas de rapport prévoyant le report de l'entrée on peut entrer maintenant et gagner 4.75 ou reporter et gagner -33.5 . Le résultat est évidemment similaire au précédent puisque l'on ne tient pas compte des résultats de l'étude.

Au final, si on sait que le concurrent va entrer, la meilleure stratégie consiste à ne pas commander l'étude et à entrer immédiatement pour un gain espéré de 6.25.

Si on a maintenant la certitude que le concurrent ne va pas entrer, la stratégie optimale consiste à ne pas recruter le cabinet et à reporter l'entrée pour un gain espéré de 73.5.

Sachant qu'il y a 40% de chance que l'on ait la certitude d'une entrée et 60% de chances que l'on ait la certitude d'un report de l'entrée du concurrent, le résultat espéré d'une information parfaite est de : $0.4 \times 6.25 + 0.6 \times 73.5 = 46.6$

Puisque le maximum que l'on puisse espérer sans information est de 31.3, la valeur espérée de l'information parfaite est : $EVPI = 46.6 - 31.3 = 15.3$. Cette valeur est évidemment supérieure à celle de l'information imparfaite.

¹⁴Soit la valeur espérée de l'information échantillonnale qui est en fait une valeur espérée d'information imparfaite.

3.3 Arbres de décision

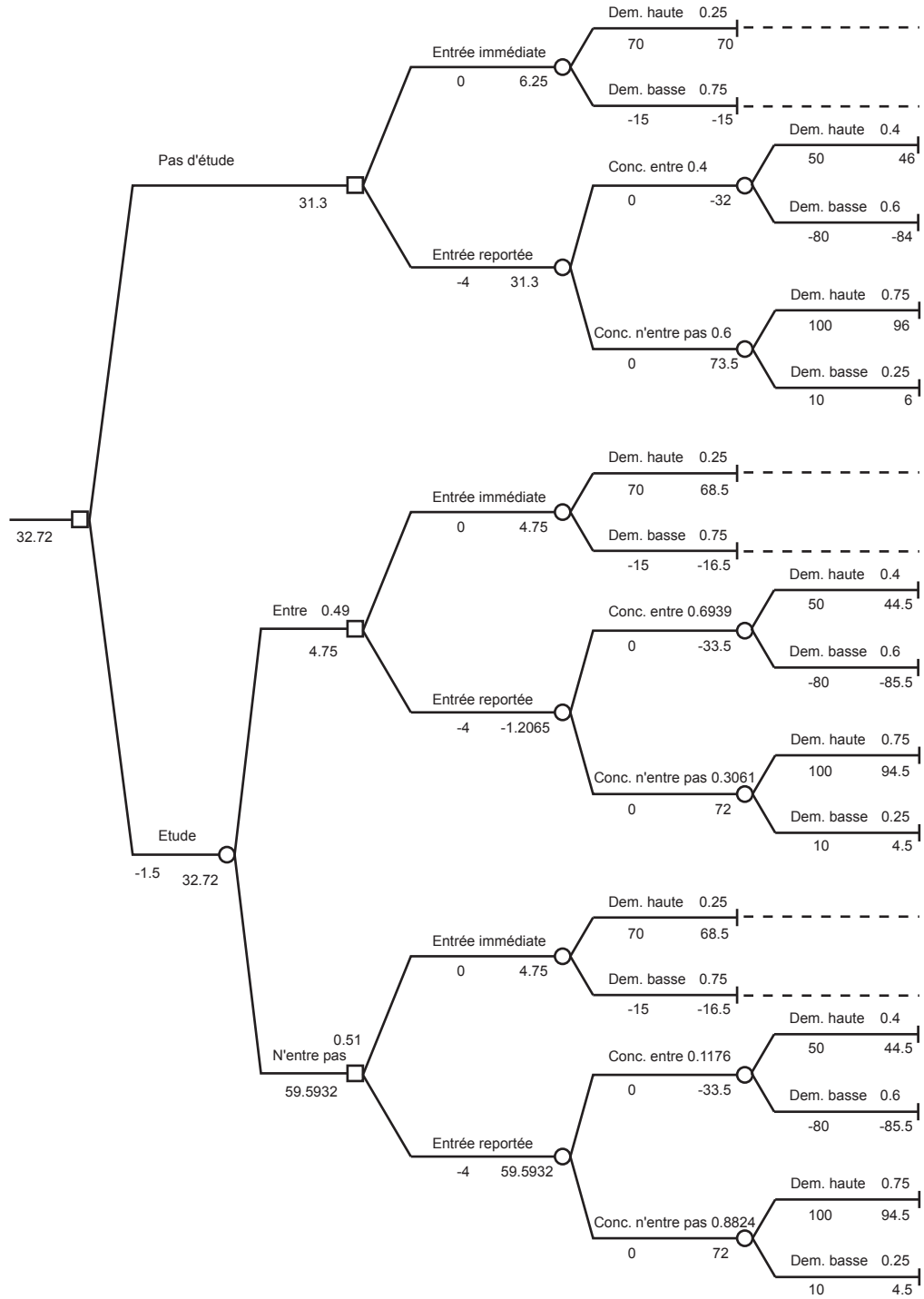


FIG. 27 – Arbre d'analyse des deux décisions successives

3.3 Arbres de décision

3.3.2 Fonctions d'utilité

Tous les raisonnements précédents ont été basés sur une hypothèse de neutralité vis-à-vis du risque, c'est-à-dire que le décideur était prêt à risquer l'espérance de gain. Dans la pratique, les choses sont souvent moins simples. Si on propose de miser 5 € pour gagner à pile ou face 10 €, la valeur espérée est égale à la mise et le jeu est assez facilement accepté. Imaginons maintenant que l'on vous propose un gain de 10.000 € à pile ou face pour une mise de 5.000 €. Là encore, la valeur espérée est égale à la mise, pourtant, rares sont ceux qui vont accepter de participer. C'est là la manifestation de l'*aversion au risque*.

Autre exemple, l'entreprise peut avoir à décider d'un investissement de 1 million d'euros, avec une probabilité de 40% d'un gain de 5 millions et une probabilité de 60% d'une perte de 2 millions. Si elle ne fait rien, elle reste sur sa lancée actuelle et peut gagner 500.000 €. La valeur espérée de la décision d'investissement est de 800.000 €, contre 500.000 sinon. Pourtant, il est loin d'être évident que c'est bien la décision d'investissement qui sera prise.

De manière générale, plus le risque est élevé, moins on a tendance à accepter de miser. On appelle *équivalent certain* le montant que l'on préférerait encaisser avec certitude plutôt que de prendre le risque du jeu. On peut ainsi préférer encaisser 100 plutôt que de participer à une loterie où l'on a 50% de perdre 50 et 50% de gagner 300, soit une valeur espérée de 125. On appelle *prime de risque* l'écart entre l'équivalent certain et le gain espéré. Plus l'aversion au risque sera importante, plus la prime de risque sera élevée (Figure 28, page 54) .

Pour tenir compte de l'attitude de l'individu face au risque, plutôt que de travailler sur les valeurs espérées, on travaille ainsi l'utilité associée au gain. La fonction d'utilité permet de tenir compte de l'aversion (ou du goût) face au risque, mais aussi du caractère généralement décroissant de l'utilité des gains croissants (on trouve moins d'utilité au passage d'un gain de 100 à 101 qu'au passage d'un gain de 10 à 11).

On se contentera ici de présenter la plus commune des fonctions d'utilité : l'utilité exponentielle définie comme :

$$U(x) = \alpha(1 - e^{-\lambda x})$$

On peut aisément calculer l'équivalent certain Z d'un jeu dont la valeur espérée est v avec une fonction d'utilité exponentielle :

$$v = U(Z) = \alpha(1 - e^{-\lambda Z}) \Rightarrow 1 - \frac{v}{\alpha} = e^{-\lambda Z} \Rightarrow Z = \frac{-1}{\lambda} \ln \frac{\alpha - v}{\alpha}$$

3.3 Arbres de décision

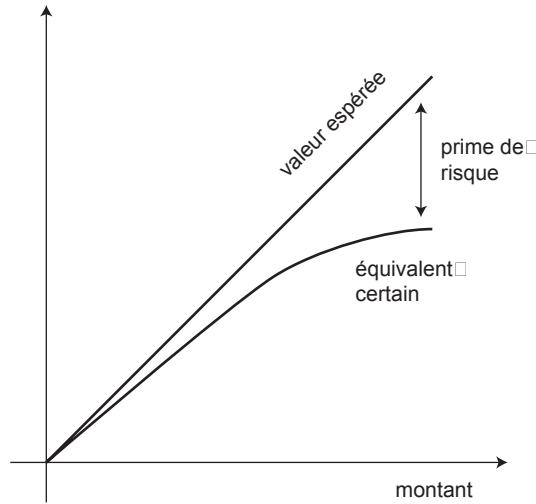


FIG. 28 – Équivalent certain et valeur espérée

On voit alors aisément que tout accroissement de λ se traduit par une réduction de l'équivalent certain Z .

Cette fonction appelle les commentaires suivants :

- α est un facteur d'échelle, il marque l'utilité maximum puisque la fonction est asymptotique à α .
- La fonction montre une utilité marginale décroissante, c'est-à-dire qu'à chaque accroissement du risque correspond une utilité de plus en plus faible.
- λ marque l'appétence au risque. Un λ faible correspond à une moindre crainte du risque. De fait quand λ tend vers 0, la fonction d'utilité tend à être linéaire, ce qui signifie que l'utilité tend vers la valeur espérée. Inversement, un λ élevé marque une aversion au risque (Figure 29, page 55) .

Le calcul de la fonction d'utilité et de l'équivalent certain s'applique à toutes les procédures de prise de décision, notamment aux arbres de décision. On les construits alors de la même manière que d'habitude, mais on substitue aux valeurs espérées classiques leur équivalent certain. Reprenons le cas de *NanoBio* en posant $\alpha = 10$ et $\lambda = 0.005$. S'il n'y a pas d'études et entrée immédiate, pour une demande haute, on a une utilité de : $v = 10 \times (1 -$

3.3 Arbres de décision

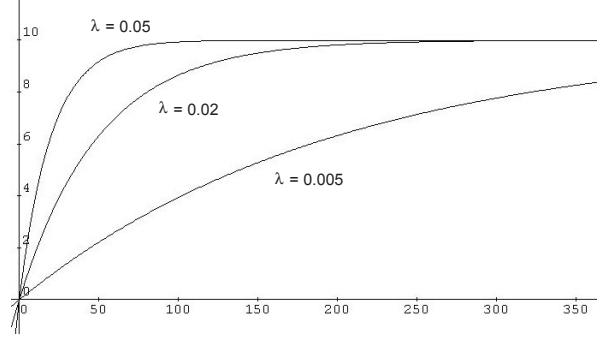
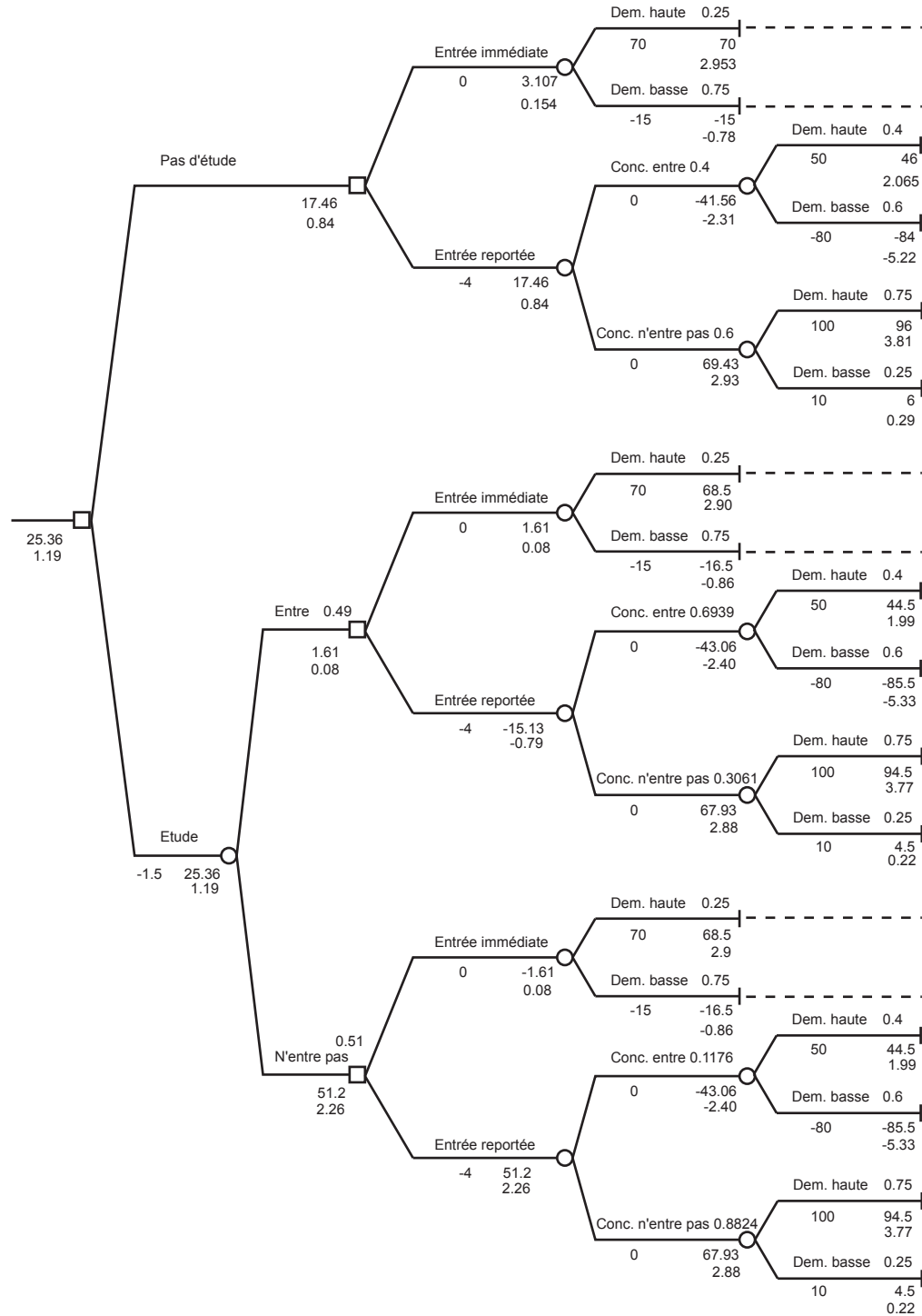


FIG. 29 – Fonction d'utilité exponentielle $U = \alpha(1 - e^{-\lambda x})$ pour $\alpha = 10$ et différentes valeurs de λ .

$e^{-0.005 \times 70} = 2.953$. Pour une basse, elle est de : $v = 10 \times (1 - e^{-0.005 \times -15} = -0.0779$. L'arc d'entrée immédiate a donc une valeur de $0.25 \times 2.953 + 0.75 \times -0.0779 = 0.1542$, soit un équivalent certain de $\frac{-1}{0.005} \ln \frac{10 - (0.1542)}{10} = 3.107$ (Figure 30, page 56) .

On constate que dans cette configuration où l'aversion au risque est faible, la stratégie reste identique. Ce ne sera pas le cas pour toutes les valeurs de α . Le calcul manuel est trop lourd pour que l'on puisse faire des simulations, mais l'utilisation d'un logiciel de calcul des arbres permet par exemple de voir qu'aux environs de $\lambda = 0.025$ — soit une aversion au risque certaine — le décideur considérera qu'en aucun cas le jeu n'en vaut la chandelle ...

3.3 Arbres de décision

FIG. 30 – *NanoBio* et utilité exponentielle ($\alpha = 10, \lambda = 0.005$)

4 Simulations

La simulation a longtemps été le parent pauvre de la science. Les succès de l'analyse mathématique à partir du *XVII^{ème}* siècle ont en effet donné l'illusion que les phénomènes physiques pouvaient se décrire à partir de la seule méthode analytique. Le triomphe de la physique en ayant fait le modèle central de la méthode scientifique, cette illusion s'est diffusée aux sciences sociales et humaines. Pourtant, y compris au siècle des Lumières, certains s'intéressaient à la simulation. L'exemple le plus célèbre en est les *aiguilles de Buffon* où le célèbre naturaliste montrait comment il suffisait de faire tomber des aiguilles sur un parquet, puis de réaliser certains calculs simples, pour trouver la valeur de π . Ce n'était là rien d'autre qu'une simulation. Afin d'épargner au lecteur un laborieux rappel de certaines notions de trigonométrie, on peut illustrer le raisonnement de Buffon et l'apport de la simulation à partir d'un exemple plus simple.

Prenons une grande feuille de papier et traçons y un carré. On posera que les côtés sont de longueur 1, la surface du carré est donc de 1. Traçons maintenant un cercle de diamètre 1 à l'intérieur de ce carré, sa surface est de $\pi r^2 = \pi \times 0,5^2 = 0,25\pi = \frac{\pi}{4}$. Le rapport de la surface du cercle à la surface du carré est donc de $\frac{\pi}{4}$. Prenons maintenant une pile de n pièces de 1 centime et laissons les tomber une à une au hasard sur la feuille. Comptons enfin les m pièces qui se trouvent à l'intérieur du cercle. Le rapport $\frac{m}{n}$ doit tendre vers $\frac{\pi}{4}$ dont on peut déduire directement que $\pi \approx \frac{4m}{n}$. Il suffit bien de laisser tomber des pièces pour approximer π .

Cette méthode, maintenant connue sous le nom de *méthode de Monte Carlo* en référence au rôle qu'y joue le hasard, a réellement décollé au cours des années 1940 au laboratoire de Los Alamos travaillant sur la bombe atomique, grâce notamment à Stanislaw Ulam : « [La méthode de Monte Carlo] m'a été suggéré par une question que je me suis posé en 1946 alors que j'étais en convalescence à la suite d'une maladie et que je faisais des réussites. La question était : quelles sont les chances qu'une main de 52 cartes à la réussite Canfield¹⁵ soit gagnante ? Après avoir passé beaucoup de temps à essayer de l'estimer par des calculs de probabilités, je me suis demandé si une méthode plus pratique que la pensée abstraite ne pourrait pas être de tirer disons une centaine de mains et compter combien étaient gagnantes. Ceci était déjà envisageable avec l'arrivée d'une nouvelle ère d'ordinateurs rapides et j'ai immédiatement pensé au problème de la diffusion des neutrons et à

¹⁵C'est une forme de réussite où la probabilité de gagner est très faible. On en trouvera facilement les règles sur Internet.

4.1 Exemple d'application

d'autres questions de physique mathématique et, plus généralement à la façon de transformer certains processus décrits par des équations différentielles dans une forme équivalente interprétable comme une succession d'opérations aléatoires. [...] J'ai décrit cette idée à John von Neumann et nous avons commencé à planifier les calculs actuels¹⁶. » Son appropriation par Stanislaw Ulam, John von Neumann et Enrico Fermi¹⁷, parmi les plus prestigieux physiciens de l'époque, et ses résultats pratiques ont rapidement donné ses lettres de noblesse à la méthode de Monte Carlo. La simulation stochastique est ainsi passée du statut de jeu intellectuel à celui d'outil scientifique de pointe. Plus encore, le récent développement des théories de la complexité a montré l'omniprésence des phénomènes non linéaires, précisément ceux que la méthode analytique a tant de mal à résoudre, mais aussi précisément ceux qui semblent présider plus encore qu'ailleurs aux phénomènes sociaux.

4.1 Exemple d'application

Un petit exemple va nous permettre d'illustrer tant l'usage que la mise en œuvre de la méthode de Monte Carlo dans le cadre de la décision. Un responsable d'entreprise s'interroge sur l'opportunité d'un investissement de 300.000 € pour ouvrir une nouvelle ligne dans son unité d'embouteillage de lait. Le problème est complexe notamment parce que la production n'est pas constante et dépend de celles des fermiers environnants. D'après son expérience, il sait ainsi qu'en cas d'ouverture de la nouvelle ligne, il peut compter sur un volume qui suit une loi normale de moyenne 30.000 litres et d'écart-type 4.000. De même, la demande n'est pas constante et suit une loi normale de moyenne 30.000 litres et d'écart-type 2.000. Enfin, le prix de vente lui-même varie entre 0.40 et 0.47 € par litre suivant une distribution uniforme. Notons enfin que le coût variable est de 0.09 € le litre et que le traitement des volumes non vendus revient à 0.15 € par litre.

Comment estimer l'opportunité de l'investissement ? En l'état, un laborieux calcul de probabilité devrait permettre d'estimer la VAN correspondante et sa distribution, mais le problème n'est pas simple. Plus encore, il est des cas où les mathématiques s'avèrent d'un piètre secours car la solution analytique n'est tout simplement pas utilisable. C'est dans ce contexte que la simulation montre toute sa puissance. Simple d'application elle peut ainsi venir au secours aussi bien de celui qui, peu familier des mathématiques, ne veut ou ne peut trouver une solution, que de celui qui, malgré sa maîtrise

¹⁶Cité dans Eckhart, R., Stan Ulam, John von Neumann and the Monte Carlo Method, *Los Alamos Science*, Special Issue, 1987, p.131

¹⁷Fermi, E. et al., Studies of Nonlinear Problems I, *Los Alamos, Report* LA-1940

4.1 Exemple d'application

des mathématiques, cale devant un problème complexe¹⁸.

Le principe est élémentaire : puisque les calculs sont complexes, on va simuler la situation considérée un grand nombre de fois et analyser les résultats. C'est exactement la démarche suggérée par Ulam avec la réussite de Canfield.

Souhaitant un retour sur investissement de 3 ans maximum, on va travailler sur 36 mois. On commence par poser les paramètres dans une feuille de calcul (cellules B4 à J9). Sur cette base, on peut poser les calculs associés à chacune des 36 périodes, notons simplement que la fonction `ALEA()` renvoie une valeur aléatoire entre 0 et 1. Reprenons chacune des colonnes (Figure 31, page 60) .

- La production suit une loi normale de moyenne 30.000 et d'écart-type 4.000 ; on obtient une telle série en posant sur les 36 périodes la fonction `LOI.NORMALE.INVERSE(ALEA() ; μ ; σ)`, avec naturellement $\mu = 30.000, \sigma = 4.000$ (colonne C).
- On utilise la même méthode pour la demande aux paramètres prêts (colonne D).
- Le reliquat est le maximum entre 0 (on ne peut pas avoir un reliquat négatif) et la différence entre la production et la demande (colonne E).
- Le prix suit une loi uniforme que l'on peut représenter par la fonction : `Prix minimum + ALEA() * (prix maximum - prix minimum)` (colonne F).
- Les autres calculs sont élémentaires et n'appellent pas de commentaires.

On dispose maintenant de calculs sur 36 périodes qui, consolidés, permettent de vérifier l'opportunité de l'investissement. Dans un souci de simplicité, on va se contenter de juger l'opportunité à partir de la seule VAN. On calcule le solde (somme des flux, cellule J11) et la VAN (cellule J12). On constate ainsi dans l'exemple de la figure que la VAN est nettement positive puisqu'elle approche 14.000 €. L'investissement serait donc rentable.

Le résultat précédent est évidemment très insuffisant, il se trouve simplement que les valeurs des variables aléatoires ont conduit à une VAN positive, mais rien ne garantit que ce résultat soit représentatif.

On va faire ce calcul un grand nombre de fois (par exemple 100 fois) et analyser la distribution de la VAN sur ces 100 simulations. Pour multiplier les calculs, il suffit d'utiliser une table de simulation. On va construire 100 lignes numérotées de 1 à 100 (en fait peu importe leur contenu), mettre en

¹⁸Comme l'ont clairement montré les pères de la méthode dont von Neumann qui fut l'un des plus grands mathématiciens du siècle dernier.

4.1 Exemple d'application

	A	B	C	D	E	F	G	H	I	J	K
2											
3			Paramètres								
4			Investissement -	300 000		Coûts variables	0.09		Prix minimum	0.40	
5			Coût reliquat	0.15					Prix maximum	0.47	
6			Production moyenne	30 000		Demande moyenne	30 000		Taux d'actualisation	5%	
7			Ecart-type	4 000		Ecart-type	2 000				
8											
9											
10											
11											
12			Variables intermédiaires						Critères		
13			Mois	Production	Demande	Reliquat	Prix	CA	Coût de traitement du reliquat	Coûts variables	Marge
14			1	30 531	30 086	446	0.4647	13 982.24	66.84	2 747.82	11 167.57
15			2	28 812	29 807	-	0.4480	12 907.16	-	2 593.11	10 314.05
16			3	32 711	30 996	1 714	0.4163	12 904.04	257.13	2 943.95	9 702.96
17			36	30 954	31 877	-	0.4667	14 445.94	-	2 785.86	11 660.08
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											

FIG. 31 – Modèle de l'opérations d'embouteillage

en-tête de la colonne immédiatement à droite une référence vers la VAN. On appelle ensuite l'outil **Table** à qui on indique comme paramètre en colonne une référence vers une colonne vide (il n'y a en effet pas de valeur d'entrée, il s'agit simplement de faire en sorte que le calcul soit lancé 100 fois). On obtient ainsi 100 VAN correspondant à ce qui aurait pu se passer dans 100 situations différentes conformes aux lois de distribution des différents paramètres.

On doit maintenant analyser les résultats. On peut commencer par calculer les statistiques descriptives de la série des résultats. Sous Excel, on peut utiliser l'utilitaire d'analyse « **Statistiques descriptives** » qui donne notamment la moyenne, l'écart-type et l'étendue de la série. On peut ensuite utiliser l'utilitaire d'analyse « **Histogramme** » pour visualiser les résultats (Figure 32, page 61) .

On trouve une moyenne de 10.900 € avec un écart-type de 6.400. À partir de là, il est aisé de calculer les probabilités de résultat. Si on veut connaître la probabilité d'obtenir une VAN négative, on pourra tout simplement utiliser la fonction $LOI.NORMALE.INVERSE(ALEA();\$C\$6; \$G\$7; \$G\$8) = 0.044$. Il y a donc (à l'erreur d'échantillonnage près, d'où l'importance de prendre des échantillons suffisamment grand) seulement 4.4% de chances que l'investissement soit perdant.

4.2 Tableur et séries aléatoires

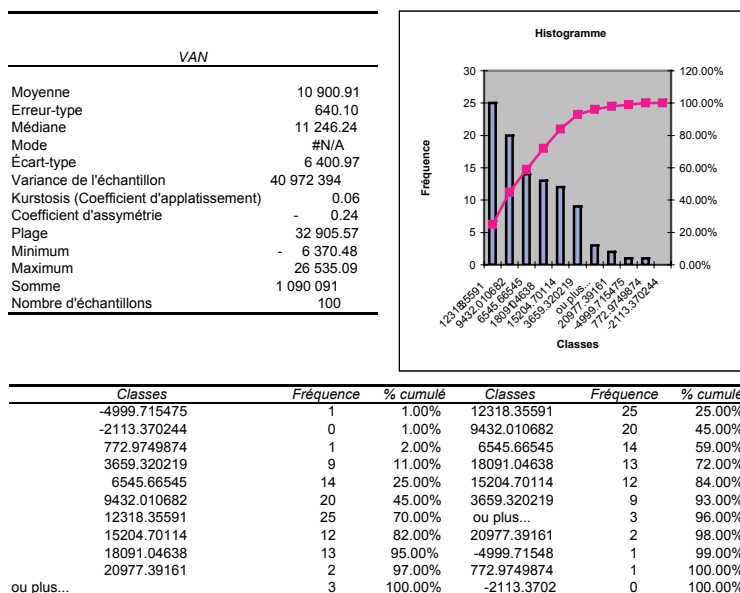


FIG. 32 – Synthèse de la simulation

4.2 Tableur et séries aléatoires

La méthode de Monte Carlo est basée sur la génération de séries aléatoires suivant une loi donnée. Tous les tableurs sont dotés d'au moins une fonction permettant de générer une série aléatoire¹⁹. La fonction `ALEA()` (ou `RAND()` en anglais) renvoie ainsi un chiffre aléatoire entre 0 et 1 à chaque fois qu'on l'appelle. Grâce à elle, on peut construire des séries suivant les principales lois de distribution :

- Distribution normale de moyenne μ et d'écart-type σ :
`LOI.NORMALE.INVERSE(ALEA() ; μ ; σ)`
- Distribution uniforme de minimum m et de maximum M :
 `$m + (M - m) * \text{ALEA}()$`
- Distribution exponentielle de moyenne μ :
 `$-(1/\mu) * \text{LN}(\text{ALEA}())$`
- Distribution triangulaire symétrique de minimum m et de maximum M :

¹⁹Plus précisément pseudo-aléatoire car pour reprendre l'expression de von Neumann « quiconque tente de générer du hasard à partir d'une méthode arithmétique — ce qui est obligatoirement le cas pour les ordinateurs qui sont des systèmes déterministes — est en état de péché. »

4.3 Allocation d'un portefeuille d'investissement

- $m + (M - m) * (\text{ALEA}() + \text{ALEA}()) / 2$
- Distribution uniforme discrète de n valeurs avec un minimum m :
 $\text{ENT}(n * \text{ALEA}()) + m$
 - Distribution discrète quelconque ; cas par exemple où l'on a 20% de A, 30% de B et 50% de C :
 Il suffit d'utiliser la fonction $\text{ALEA}()$ et de sélectionner le résultat avec des fonctions $\text{SI}()$ ou avec un $\text{RECHERCHEV}()$. Pour l'exemple ci-dessus, on pourrait mettre $\text{ALEA}()$ dans la cellule A1, le résultat en A2 serait :
 $=\text{SI}(A1 < 0.2 ; "A" ; \text{SI}(A1 < 0.5 ; "B" ; "C"))$ ²⁰.

4.3 Allocation d'un portefeuille d'investissement

Disposant d'un capital de 100.000 € on s'interroge sur l'équilibre de ses placements. Quatre types de supports sont possibles, dont les rendements sont les suivants (Table 7, page 62) .

	Moyenne	Écart-type
Sicav	2.50%	0.10%
Sérénité	4.50%	0.50%
Équilibre	6.00%	1.00%
Agressif	10.00%	5.00%

TAB. 7 – Rendement des différents supports

Quel résultat attendre d'un placement de 25.000 € dans chacun des supports ? On commence par construire le modèle avec le tableur, ce qui ne pose pas de difficulté (Figure 33, page 63) .

À partir de là, on peut simuler le comportement du portefeuille pour en estimer le rendement et le risque. On construit une table de 100 entrées faisant référence à la cellule « Résultat » et, pour rester simple, on se contente de calculer la moyenne et l'écart-type des différents résultats (cellules H3 et H4). On note que la table fait référence à la colonne J, laquelle est vide. Comme dans le premier exemple ci-dessus, l'objectif n'est pas de recalculer en fonction d'une valeur d'entrée, mais de recalculer tout court. Quand on crée la table, on sélectionne donc dans **cellule d'entrée en colonne** une colonne vierge. On trouve dans notre cas une moyenne de 5.759 € avec un écart-type de 1.281 €.

²⁰ Attention, on ne peut pas utiliser la fonction $\text{ALEA}()$ à l'intérieur des $\text{SI}()$, sinon elle sera calculée plusieurs fois et le résultat ne sera évidemment pas valide.

4.4 Événements discrets

	A	B	C	D	E	F	G	H	I
1									
2									
3		Paramètres		Décision			Simulation		
4			Moyenne	Ecart-type	Investissement				MOYENNE(H7:H106)
5		Sicav	2.50%	0.10%	25 000				
6		Sérénité	4.50%	0.50%	25 000				
7		Equilibre	6.00%	1.00%	25 000				
8		Agressif	10.00%	5.00%	25 000				
9				Total	100 000				
10									
11		Variables intermédiaires							
12		Sicav	2.28%						
13		Faible	4.25%						
14		Moyenne	7.62%						
15		Haute	9.45%						
16									
17		Critère							
18		Résultat	5 898.94						
19									
106									
107									

1	5 898.94
2	5 190.27
3	2 101.26
4	4 889.05
5	6 464.15
6	5 822.76
7	4 739.39
8	5 343.51
9	7 265.41
10	4 812.44
11	3 551.93
12	4 704.54
13	2 210.54
106	6 042.68
107	4 976.76

FIG. 33 – Calcul du rendement du portefeuille

Se pose naturellement la question de l'allocation optimum. A-t-on intérêt à se concentrer sur les SICAV où sur le support dynamique ? La réponse à cette question, même si elle n'est pas très complexe, sort du cadre de ce cours. Outre le problème classique en finance de la définition du risque, elle nécessite en effet dans le cadre de la simulation stochastique, soit l'utilisation d'outils dédiés²¹, soit l'écriture d'une macro qui suppose des bases élémentaires de programmation²².

On peut toutefois avancer vers l'analyse des résultats en fonction de l'allocation en utilisant le gestionnaire de scénarios. On a testé ici trois configurations possibles : prudente, médiane et agressive. On peut voir les rendements moyens ainsi que leur variation en fonction de la stratégie adoptée (Figure 34, page 64) .

4.4 Événements discrets

Combien de personnes mettre au guichet ? Combien de télé-opérateurs recruter ? Ce type de problème est extrêmement courant. Là encore la simulation peut être d'un grand secours pour quiconque ne souhaite se plonger de manière approfondie dans les mathématiques des files d'attente²³. Illustrons

²¹ Avec notamment l'incontournable Crystall Ball, dont le seul vrai défaut est de ne pas être libre ...

²² Profitons en pour noter à l'attention de ceux qui souhaitent approfondir que l'une des meilleures classes d'algorithmes pour ce type de problème (les algorithmes évolutionnaires) est basée sur une forme de simulation de l'évolution darwinienne. Elle permet en effet d'obtenir directement le front de Pareto correspondant.

²³ Dans les situations de base, l'outil analytique reste simple et est évidemment porteur de beaucoup plus d'information que la simulation.

4.4 Événements discrets

Synthèse de scénarios				
	Valeurs actuelles :	Prudent	Médian	Agressif
Cellules variables :				
Sicav	25 000	40 000	20 000	10 000
Sérénité	25 000	30 000	30 000	20 000
Equilibre	25 000	20 000	30 000	30 000
Agressif	25 000	10 000	20 000	40 000
Cellules résultantes :				
Moyenne	5 737.94	4 403.25	5 604.81	6 903.76
Ecart-type	1 270.59	489.06	1 083.65	2 101.94

La colonne Valeurs actuelles affiche les valeurs des cellules variables au moment de la création du rapport de synthèse. Les cellules variables de chaque scénario se situent dans les colonnes grisées.

FIG. 34 – Scénarios et rendement du portefeuille

la démarche à travers un exemple.

4.4.1 Simulation discrète par horloge

Afin d'améliorer la qualité de service, on étudie la fréquentation d'une hot-line. Considérant que le temps de traitement d'un appel est de 5 minutes²⁴ (au-delà on transfère l'appel à un centre spécialisé), on a comptabilisé les fréquences d'appels moyennes au cours d'une tranche de 5 mn (Figure 35, page 65) .

On va simuler la fréquentation et le fonctionnement de la hot-line sur une semaine, c'est-à-dire sur 1.152 périodes de 5 minutes correspondant aux 6 fois 16 heures d'ouverture hebdomadaire. On commence par reporter sur le tableur les chiffres de fréquentation. On en profite pour calculer qu'il y a en moyenne 4.05 appels par tranches de 5 minutes. On construit dans le même temps le tableau des probabilités cumulées qui est nécessaire à la fonction `RECHERCHEV()` que l'on va utiliser ensuite (Figure 35, page 65) .

La simulation a la forme suivante (Figure 36, page 66) :

- On fixe le nombre de télé-opérateurs qui correspond à la décision à prendre (cellule J13).
- On numérote les 1.152 périodes de la simulation (colonne H). Notons que la première ligne ne contient que des zéros. C'est le cas général quand on simule des événements discrets, il est en effet nécessaire d'initialiser « le déroulement du temps. »
- On simule ensuite le nombre d'appels (colonne I) avec la fonction `ALEA()`. On recherche avec la fonction `RECHERCHEV()` le nombre d'ap-

²⁴Plus encore, ce modèle suppose que la durée de traitement est précisément et systématiquement de 5 mn, c'est une simplification lourde, mais elle rend la modélisation beaucoup plus aisée. On verra plus bas comment s'en affranchir.

4.4 Événements discrets

Paramètres			
Nombre d'arrivants	Probabilité	Probabilités cumulées	Nombre d'arrivants
0	0.02	0%	0
1	0.07	2%	1
2	0.15	9%	2
3	0.18	24%	3
4	0.2	42%	4
5	0.16	62%	5
6	0.1	78%	6
7	0.06	88%	7
8	0.03	94%	8
9	0.02	97%	9
10	0.01	99%	10
Moyenne	4.05		

FIG. 35 – Fréquences d'appels

pels dans la table des fréquences que l'on a construit plus haut (Figure 35, page 65) .

- On calcule le nombre de clients dans la file d'attente (colonne J) qui est le cumul du nombre d'appels au cours de la période et des clients en attente de la période précédente.
- Le nombre de clients servis (colonne K) est la plus petite valeur entre le nombre de clients en attente et le nombre de télé-opérateurs.
- Le nombre de clients en attente (colonne L) est le nombre de clients dans le système moins le nombre de clients servis.

Les critères sont sur les lignes 1170 à 1172. On a calculé ici pour chacune des valeurs simulées le maximum, le total et la moyenne. Le point essentiel est évidemment ici le nombre de clients dans la queue, soit en moyenne 78 clients. Sur une base de 4 opérateurs et d'un traitement en 5 minutes, cela signifie une attente de $78 \times 5/4 = 97.5$ minutes donc plus d'une heure et demie. La queue maximum de 181 clients correspond à une attente pour le dernier de 226 minutes²⁵. Pourtant, il y a en moyenne 4 appels par heure, les 4 opérateurs devraient donc suffire.

Une table de simulation va nous permettre d'étudier l'évolution de la file d'attente en fonction du nombre d'opérateurs (Figure 37, page 66) . On peut constater que le passage à 4, puis à 5 opérateurs se traduit par une baisse de la longueur de la queue d'un ordre de grandeur. La simulation a mis en

²⁵On comprend mieux la réticence des hot-line vis-à-vis de la gratuité des temps d'attente !

4.4 Événements discrets

	G	H	I	J	K	L	M
12							
13			Décision				
14			Nombre d'opérateurs	4	K18 : MIN(J18;\$J\$13)		
15			Variables intermédiaires		J18 : I18+L17		
16			Bloc de 5 mn	Nombre d'appels	Nombre de clients dans la queue	Nombre de clients servis	Nombre de clients mis en attente
17		I18 : RECHERCHEV(ALEA();\$D\$4:\$E\$14;2)	0	0	0	0	L18 : J18-K18
18			1	4	4	4	
19			2	0	0	0	
20			3	2	2	2	
21			4	6	6	4	2
22			5	2	4	4	0
23			6	2	2	2	0
24			7	1	1	1	0
1169			1152	5	144	4	140
		Critères					
1170			Maximum	10	181	4	177
1171			Total	4677	89797	4537	85260
1172			Moyenne	4.0599	77.9488	3.9384	74.0104

FIG. 36 – Simulation d'appels

évidence une phénomène parfaitement contre-intuitif.

		1	2	3	4	5	6	7	8	9	10
Queue maximum	103.00	3 537.00	2 238.00	1 246.00	191.00	19.00	15.00	12.00	11.00	10.00	10.00
Queue moyenne	39.83	1 763.17	1 123.16	639.73	88.23	5.40	4.40	4.30	4.06	4.03	4.03

FIG. 37 – Évolution de la queue en fonction du nombre d'opérateurs

4.4.2 Simulation discrète par événements

L'exemple précédent est élémentaire, on a notamment considéré que le temps de traitement d'un appel était précisément de 5 minutes. Un second exemple va nous permettre d'aller un peu plus loin dans la simulation, mais aussi de présenter une autre façon de simuler les processus discrets. Dans le premier cas, on a utilisé une méthode de simulation, dite simulation par horloge, basée sur la probabilité d'occurrence des événements par tranches de temps, c'est-à-dire que l'on a compté les nombres d'événements sur ces tranches et on a basé la simulation sur ce nombre. La méthode que nous allons utiliser ici — simulation par événements — repose, non sur le décompte du nombre d'événements par périodes, mais bien sur la simulation des événements eux-mêmes. Exprimé autrement, au lieu de compter le nombre d'événements, on compte le temps qui s'écoule entre eux. Cette méthode est plus

4.4 Événements discrets

lourde que la précédente, mais elle est plus précise et donc plus efficace.

Constatant la dégradation de sa qualité de service, le responsable de l'agence d'Annecy Bonlieu de l'opérateur téléphonique national souhaite optimiser le nombre de conseillers clientèle afin d'éviter les attentes qui parfois dépassent une heure. Le problème n'est pas aussi simple que le client grommelant sur sa chaise après 90 minutes d'attente ne semble le penser²⁶. Premier travail, le responsable note pour chaque conseiller — polyvalents ils sont tous équivalents — les arrivées et le temps de traitement des clients durant quelque temps afin d'obtenir quelques statistiques fiables. Il en résulte une table de fréquence des arrivées auprès des différents conseillers et une distribution du temps de traitement (Figure 38, page 67) .

Paramètres					
Durée entre deux arrivées	Probabilité	Probabilités cumulées			
0	0.01	0			
1	0.01	0.01			
2	0.01	0.02			
3	0.02	0.03			
4	0.02	0.05			
5	0.02	0.07			
6	0.03	0.09			
7	0.03	0.12			
8	0.03	0.15			
9	0.04	0.18			
10	0.05	0.22			
11	0.05	0.27			
12	0.06	0.32			
13	0.08	0.38			
14	0.1	0.46			
15	0.17	0.56			
16	0.12	0.73			
17	0.07	0.85			
18	0.03	0.92			
19	0.03	0.95			
20	0.02	0.98			

Durée de service	
Moyenne	14
Ecart-type	4

FIG. 38 – Statistiques des flux d'arrivants

La simulation a la forme suivante (Figure 39, page 69) :

- On fixe le numéro du client (colonne I) par simple incrément pour chacun des 100 clients que reçoit quotidiennement chaque conseiller. Comme précédemment, la première ligne sert à initialiser la simulation (en l'occurrence le déroulement du temps).
- On calcule ensuite le temps qui s'écoule avant l'arrivée du client cou-

²⁶En fait cette complexité n'est pas son problème, ce qu'il veut c'est être servi rapidement pour pouvoir enfin aller se rafraîchir dans les eaux du lac voisin.

4.5 Adéquation de la distribution

- rant (colonne J). La méthode est similaire au cas précédent²⁷.
- La minute d'arrivée (colonne K) est la somme entre l'arrivée du client précédent et la durée qui s'est écoulée avant l'arrivée du client courant.
 - On commence à s'occuper d'un client (colonne L) au plus tard quand il arrive ou quand on fini avec le client précédent.
 - On calcule ensuite la durée de traitement en simulant une distribution normale comme on l'a déjà vu au début de cette partie, soit : `LOI.NORMALE.INVERSE(ALEA() ; μ ; σ)`. On note que l'on tronque la distribution pour éviter les éventuels durées négatives. On a considéré ici que le temps de traitement minimum est de 1 minute.
 - La minute de fin de traitement (colonne N) est la somme entre le moment de l'arrivée et la durée de traitement.
 - Le temps d'attente (colonne O) est la différence entre le début de traitement et le moment de l'arrivée.
 - La longueur de la queue est le nombre de clients dont le début de traitement est postérieur au moment d'arrivée du client courant. On utilise ici la fonction `EQUIV()` qui renvoie le numéro de ligne d'un vecteur trié (ici le début de traitement) dont la valeur est immédiatement inférieure à la valeur cherchée²⁸.

On analyse l'efficacité du processus (critères) en mesurant le temps d'attente moyen et la longueur moyenne de la queue. Après avoir simulé l'arrivée des 100 clients quotidiens d'un conseiller, on construit une table de simulation qui fait fonctionner le modèle un certain nombre de fois de manière à obtenir une statistique sur les critères et non une simple donnée ponctuelle (Figure 40, page 69) . On mesure alors l'efficacité globale du système où l'on constate que la queue moyenne est de 6 clients par conseiller (cellule T116), soit une attente d'environ 90 minutes.

4.5 Adéquation de la distribution

Quand on veut utiliser la méthode de Monte Carlo, le choix de la distribution des différentes variables est l'une des principales difficultés rencontrées. Il existe de nombreux tests d'adéquation d'une sophistication variable ; on se contentera ici de revenir rapidement sur la méthode du Khi 2 (χ^2) qui, à la simplicité, conjugue une certaine généralité.

²⁷À ceci près que l'on a utilisé la fonction `RECHERCHE()` pour éviter de recopier la colonne des durées et pour varier les plaisirs . . .

²⁸L'interprétation en est un peu délicate ici. On peut naturellement faire le même calcul plus simplement avec par exemple des fonctions de recherche, mais, si l'interprétation en est plus facile, la mise en œuvre en est plus laborieuse.

4.5 Adéquation de la distribution

The diagram illustrates the structure of the simulation model, showing the flow of data and calculations between various components. The components are represented by boxes, and the flow is indicated by arrows.

Variables Intermédiaires (Intermediate Variables):

- J14:** RECHERCHE(ALEA();\$D\$13:\$D\$33;\$B\$13:\$B\$33)
- K14:** K13+J14
- L14:** MAX(K14;N13)
- M14:** MAX(L1;LOI.NORMALE.INVERSE(ALEA());\$G\$13;\$G\$14)
- N14:** L14+M14
- O14:** L14-K14
- P14:** I14-EQJIV(K14;\$L\$14:L14)

Tableau de données (Data Table):

Client	Durée entre deux arrivées	Arrivée	Début de traitement	Durée de traitement	Fin de traitement	Attente	Longueur de la queue
0	0	0	-	-	-	-	0
1	15	15	15.00	8.52	23.52	-	0
2	6	21	23.52	20.11	43.63	2.52	1
100	16	1332	1 413.48	10.23	1 423.71	81.48	7
Maximum				23.97		122.50	9.00
Total				1 384.08		5 464.51	436.00
Moyenne				13.84		54.65	4.36

Création de la file d'attente (Queue Creation):

The flow of data and calculations is as follows:

- J14** feeds into **K14**.
- K14** feeds into **L14**.
- L14** feeds into **M14**.
- M14** feeds into **N14**.
- N14** feeds into **O14**.
- O14** feeds into **P14**.
- P14** feeds into the **Longueur de la queue** column of the data table.
- The **Longueur de la queue** column of the data table feeds into the **Attente** column of the data table.
- The **Attente** column of the data table feeds into the **Fin de traitement** column of the data table.
- The **Fin de traitement** column of the data table feeds into the **Durée de traitement** column of the data table.
- The **Durée de traitement** column of the data table feeds into the **Début de traitement** column of the data table.
- The **Début de traitement** column of the data table feeds into the **Arrivée** column of the data table.
- The **Arrivée** column of the data table feeds into the **Durée entre deux arrivées** column of the data table.
- The **Durée entre deux arrivées** column of the data table feeds into the **Client** column of the data table.
- The **Client** column of the data table feeds into the **Maximum** row of the data table.
- The **Maximum** row of the data table feeds into the **Total** row of the data table.
- The **Total** row of the data table feeds into the **Moyenne** row of the data table.

FIG. 39 – Modèle de gestion des files d'attente par évènement

[illegible]

FIG. 40 – Monte Carlo sur la file d'attente par événement

Soit f_i la fréquence de la modalité M_i au sein d'un échantillon aléatoire de taille n . Soit p_i la fréquence de la modalité M_i au sein de la population. Alors $n \sum_{i=1}^k \frac{(f_i - p_i)^2}{p_i}$ converge vers une loi en χ^2 à $k - 1$ degrés de liberté.

Ce test est basé sur le fait que la distance entre les valeurs observées et les valeurs théoriques (celles qui correspondent à la distribution testée) converge vers une loi en χ^2 . Il est théoriquement adapté aux seules distributions discrètes, on verra plus loin comment le généraliser aux fonctions continues.

4.5 Adéquation de la distribution

4.5.1 Adéquation à une loi de Poisson

On relève le nombre de personnes arrivant à un guichet sur 2.000 périodes de 2 minutes et on trouve le résultat suivant²⁹ (Figure 41, page 70) . Sur 2.000 périodes, 163 n'ont vu aucun visiteur, 8 ont vu 8 visiteurs ...

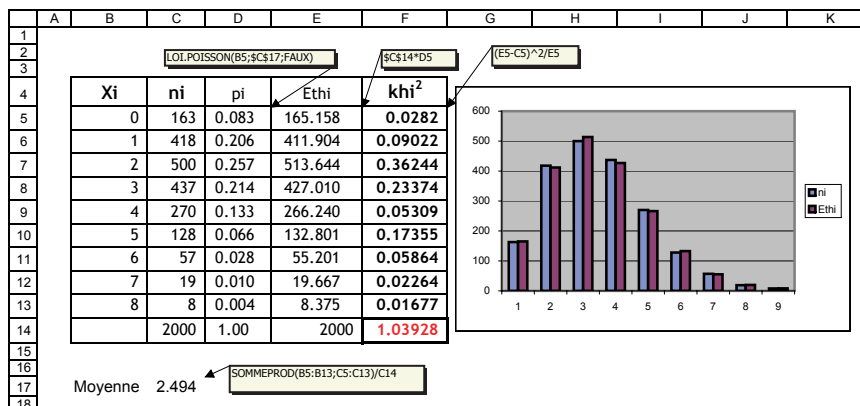


FIG. 41 – Adéquation à une loi de Poisson

On souhaite tester l'adéquation à une loi de Poisson. On pose donc l'hypothèse nulle H_0 correspondante et la contre hypothèse H_1 selon laquelle cette adéquation n'existe pas. On rappelle que cette loi nécessite un paramètre unique qui correspond à la moyenne. On calcule donc la valeur moyenne des visites (cellule C17, on note qu'il s'agit bien d'une moyenne pondérée). Dans la colonne D, on simule une distribution de Poisson de moyenne 2.49. On voit ainsi qu'il y a environ 8% de chances de n'avoir aucune visite et 4 chances sur 1.000 d'en avoir 8. À partir de cette distribution en fréquence, on passe au calcul de la distribution théorique en volume (colonne E). On dispose ainsi de la distribution théorique des visites si elles se conformaient précisément à une loi de Poisson de moyenne 2.49. On calcule ensuite les distances (colonne F), soit (Effectifs théoriques - Effectifs réels) au carré, divisé par Effectifs théoriques : $\frac{(Eth_i - n_i)^2}{Eth_i}$.

Le χ^2 calculé est la somme des χ^2 partiels (distances) soit ici 1.039 (cellule F14). Cette valeur doit être comparée à un χ^2 limite. Si on prend un seuil de signification $\alpha = 0.05$, on doit calculer : `KHIDEUX.INVERSE(0.05 ; 7) = 14.07` ; on a en effet 7 degrés de libertés, soit les 9 observations - 1 auxquelles

²⁹Ceux qui ont suivi le cours MQAD ne manqueront pas de reconnaître cet exemple !

4.5 Adéquation de la distribution

il faut encore retrancher 1 car on a estimé un paramètre supplémentaire, en l'occurrence la moyenne. Ici, le χ^2 calculé (1.039) est inférieur au χ^2 limite (14.07), on ne peut donc pas rejeter l'hypothèse d'adéquation à une loi de Poisson.

4.5.2 Adéquation à une loi normale

La logique est similaire au cas précédent à la différence près que, puisque la loi normale est continue, il convient de regrouper les données en classes. Le choix du nombre de classes est surtout affaire de pratique. On retiendra l'heuristique classique selon laquelle le nombre de classe k est le plus petit nombre tel que $2^k > n$ pour n observations. Illustrons la méthode à travers un exemple.

On dispose des relevés de 1.000 opérations bancaires d'un même type³⁰, le montant de ces opérations suit-il une distribution normale ? On commence par calculer la moyenne m et l'écart-type s associés à ces 1.000 valeurs. On trouve par exemple $m = 2.700$ et $s = 400$ (Figure 42, page 72) . On pose donc l'hypothèse H_0 selon laquelle les opérations étudiées suivent une loi normale de moyenne 2.700 et d'écart-type 400, ainsi que la contre hypothèse H_1 . On construit ensuite les classes :

- Puisque l'on a 1.000 valeurs, on devra construire 10 classes ($2^9 = 512$; $2^{10} = 1024$). Ce nombre est théorique, on pourra l'ajuster notamment en fonction des contraintes d'effectif minimum que l'on verra un peu plus loin.
- On calcule l'étendue (maximum - minimum) que l'on divise ensuite par 9 (10 - 1, il y a un écart de moins que de nombre de classes) pour avoir la dimension d'une classe. On trouve par exemple un minimum de 1.650 et un maximum de 3.550, soit une étendue de 1.950 qui donne des classes de 217. On arrondi cette valeur à un écart plus significatif de 200. Pour chacune des 10 classes obtenues (de < 1.800 à > 3.400) on note l'effectif correspondant (on peut utiliser ici la fonction `FREQUENCE()` du tableur).

On doit ensuite calculer les effectifs théoriques, c'est-à-dire les effectifs que l'on aurait au sein de chaque classe si la distribution était normale.

- On calcule d'abord la proportion des effectifs dans chaque classe (colonne D). Pour la première valeur, on cherche la proportion de valeurs inférieures à 1.800 dans une distribution normale de moyenne 2.700 et d'écart-type 400, soit : `LOI.NORMALE(1800 ; 2700 ; 400 ; VRAI)` (cellule

³⁰On n'en donnera pas le détail par souci de simplicité

4.5 Adéquation de la distribution

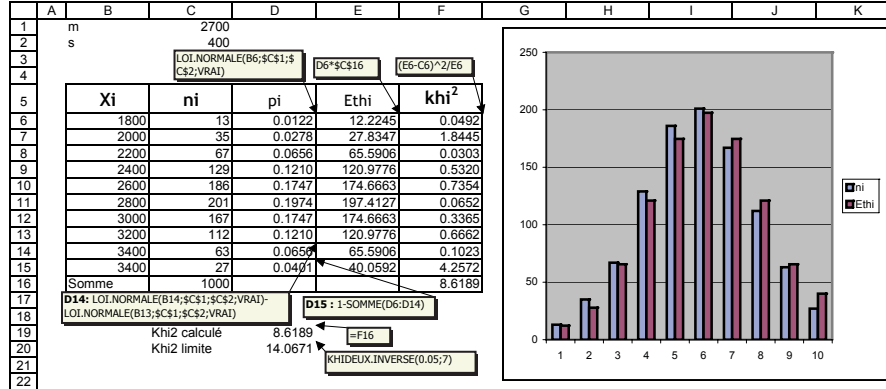


FIG. 42 – Adéquation à une loi Normale

D6). On cherche ensuite la proportion se situant entre 1.800 et 2.000 qui est la proportion des valeurs inférieures à 2.000 moins celle des valeurs inférieures à 1.800, soit : `LOI.NORMALE(2000 ; 2700 ; 400 ; VRAI) - LOI.NORMALE(1800 ; 2700 ; 400 ; VRAI)` (cellule D7). Le calcul est similaire pour les autres classes à l'exception de la dernière pour laquelle on prend le complément à 1 de la somme des valeurs antérieures, soit dans notre exemple `1-SOMME(D6 : D14)` (cellule D15).

- On calcule ensuite les effectifs théoriques qui sont simplement le produit de l'effectif total par la proportion de la classe (colonne E). Il est important que les effectifs théoriques soient au moins égaux à 5 (ceci évite que les variations relatives prennent trop d'importance par rapport aux variations absolues). Si ce n'est pas le cas, on pourra regrouper deux classes (en général cela concerne les classes extrêmes).

Disposant des effectifs réels et théoriques, on peut calculer les χ^2 partiels (colonne F), soit par exemple pour la tranche < 1.800 (cellule F6), $(E6-C6)^2/E6$. Le χ^2 calculé est la somme des χ^2 partiels, soit ici 8.6189 (cellule F16).

On doit enfin déterminer le χ^2 limite : `KHIDEUX.INVERSE(0.05 ; 7) = 14.07` (cellule D20). Le premier paramètre est le seuil de signification ($\alpha = 0.05$), le second est le nombre de degrés de liberté, soit ici $10 - 1 - 2$. On doit ôter 2 au nombre de degrés de liberté ($k - 1$) car on a estimé les deux paramètres supplémentaires que sont la moyenne et l'écart-type. Il y a au plus 5% de chances, si la série suit la loi testée, que le χ^2 soit supérieur à 14.07. On constate alors que le χ^2 limite est supérieur au χ^2 calculé ($14.07 > 8.62$). On ne peut donc pas rejeter l'hypothèse H_0 . Exprimé différemment,

4.5 Adéquation de la distribution

on a beaucoup plus de chance de se tromper en considérant que la série ne suit pas la distribution testée que l'inverse³¹.

On peut généraliser la démarche à toute distribution souhaitée, sachant qu'on reste limité pour les distributions continues par la nécessité d'avoir des effectifs de classes suffisant. D'autres tests sont mieux adaptés, notamment aux effectifs faibles, mais ils sont d'un abord un peu plus complexe. Notons simplement que les tests de normalité les plus courants sont les tests dit de *Kolmogorov-Smirnov* et de *Lilliefors* qui sont implémentés dans l'ensemble des logiciels statistiques.

³¹Les plus familiers avec les tests auront noté que l'on ne travaille pas ici, comme on le fait généralement, sur le risque de seconde espèce. En l'état, on ne connaît donc pas le risque d'erreur sur l'acceptation de l'hypothèse.