3

Gestion des approvisionnements

Ce chapitre est consacré à l'étude de la gestion du stock d'un produit, une pièce de rechange automobile par exemple, sur plusieurs périodes de temps. Au début de chaque période, le gestionnaire du stock effectue une commande auprès de son fournisseur ; pendant la période, la quantité commandée est livrée et des clients formulent des demandes que le gestionnaire peut ou non satisfaire suivant le stock dont il dispose. On appelle stratégie du gestionnaire la manière dont il décide de la quantité commandée en fonction du passé. L'objectif du chapitre est de caractériser les stratégies optimales en termes de minimisation du coût total qui s'exprime comme la somme du coût d'achat du produit auprès du fournisseur, du coût de stockage et du coût associé aux demandes de clients qui n'ont pu être honorées faute de stock. Bien entendu, minimiser la somme du coût d'achat et du coût de stockage et minimiser le coût associé aux demandes non satisfaites sont des objectifs antagonistes.

Dans le premier paragraphe, nous commençons par résoudre le problème d'optimisation sur une seule période de temps avant de décrire précisément le problème dynamique de gestion de stock sur plusieurs périodes de temps.
Le second paragraphe est consacré à une introduction au contrôle de chaînes de Markov, dans un cadre qui englobe le problème dynamique de gestion de stock. L'objectif est de montrer comment le principe de la programmation dynamique introduit par Bellman [1] à la fin des années 1950 permet de ramener l'étude d'un problème d'optimisation à N périodes de temps à celle de N problèmes à une seule période de temps. Nous illustrons sur le problème dit de la secrétaire comment ce principe permet d'expliciter la stratégie optimale d'un recruteur. On suppose que le recruteur sait classer les N candidats à un poste qui se présentent successivement pour passer un entretien avec lui et que tout candidat qui ne reçoit pas de réponse positive lors de son entretien trouve un emploi ailleurs. Pour maximiser la probabilité de choisir le meilleur des N candidats lors de cette procédure, le recruteur doit d'abord observer une proportion proche de $1/e$ de candidats sans les recruter, puis choisir ensuite tout candidat meilleur que ceux qu'il observés (si aucun candidat meilleur ne se présente ensuite, il recrute le dernier candidat qu'il reçoit).

Enfin, dans le troisième paragraphe, nous utilisons les résultats du second paragraphe pour déterminer les stratégies optimales du gestionnaire de stock dans le modèle dynamique : à chaque instant, lorsque le stock est égal à x, le gestionnaire doit commander la quantité $\mathbf{1}_{\{x\leq s\}}(S-x)^+$ qui permet de se ramener au stock objectif S si le stock est inférieur au seuil s, où le couple (s,S) peut dépendre ou non du temps.

3.1 Le modèle probabiliste de gestion de stock

3.1.1 Le modèle à une période de temps

Ce modèle comporte un seul produit et une seule période de temps. Au début de la période, pour satisfaire les besoins de sa clientèle, un gestionnaire (par exemple un vendeur de journaux) commande une quantité q de produit qui lui est facturée au coût unitaire $c>0$ par le fournisseur. Il est souvent naturel de supposer que le produit se présente sous forme d'unités (cas des journaux par exemple), auquel cas $q\in\mathbb{N}$. Mais on peut aussi se placer dans un cadre continu (cas d'un liquide comme l'essence par exemple) et supposer $q\in\mathbb{R}_+$, ce qui simplifie parfois le problème d'optimisation. Comme la demande des clients sur la période de temps n'est pas connue du gestionnaire au moment où il effectue sa commande, il est naturel de la modéliser par une variable aléatoire D à valeurs dans $\mathbb{N}$ ou $\mathbb{R}_+$. On suppose que D est d'espérance finie $\mathbb{E}[D]=\mu$ et que l'on connaît sa loi au travers de sa fonction de répartition $F(x)=\mathbb{P}(D\leq x)$.

À l'issue de la période trois situations sont possibles :
- $q=D$: le gestionnaire a visé juste.
- $q>D$ i.e. il y a $(q-D)$ unités de produit en surplus. On associe à ce surplus un coût unitaire c_S qui correspond par exemple au coût de stockage sur la période. Notons que l'on peut supposer c_S négatif pour rendre compte de la possibilité de retourner le surplus au fournisseur (c'est le cas pour les journaux qui sont repris par les Nouvelles Messageries de la Presse Parisienne). Dans ce cas, il est naturel de supposer que $c>-c_S$ i.e. que le prix auquel le fournisseur reprend le surplus est plus petit que le coût unitaire c auquel le gestionnaire se fournit. Nous supposerons donc désormais que $\boxed{c>0 \text{ et } c>-c_S}$.
- $q<D$: on appelle manquants les demandes que le gestionnaire n'a pu satisfaire. On associe aux $D-q$ manquants un coût unitaire $\boxed{c_M\geq 0}$ qui rend compte de la détérioration de l'image du gestionnaire auprès des clients non servis.

L'objectif pour le gestionnaire est de trouver $q\geq 0$ qui minimise l'espérance du coût total

$$g(q)=cq+c_S\mathbb{E}[(q-D)^+]+c_M\mathbb{E}[(D-q)^+] \tag{3.1}$$

où pour $y \in \mathbb{R}$, $y^+ = \max(y, 0)$ désigne la partie positive de y. Même si la quantité commandée est positive, nous allons supposer que la fonction g est définie sur $\mathbb{R}$ car cela sera utile lorsque nous introduirons un coût fixe d'approvisionnement et un stock initial. Vérifions que g est continue. Comme la demande D est positive, pour $q \leq 0$, $(D-q)^+ = D-q$ et $\mathbb{E}[(D-q)^+] = \mu - q$. Donc $q \to \mathbb{E}[(D-q)^+]$ est continue sur $\mathbb{R}_-$. Pour $q \geq 0$, $0 \leq (D-q)^+ \leq D$. Comme $q \to (D-q)^+$ est continue, on en déduit par convergence dominée que $q \to \mathbb{E}[(D-q)^+]$ est continue sur $\mathbb{R}_+$ et donc sur $\mathbb{R}$. L'égalité $y^+ = y + (-y)^+$ entraîne que

$$\forall q \in \mathbb{R},\ \mathbb{E}[(q-D)^+] = q - \mu + \mathbb{E}[(D-q)^+]. \tag{3.2}$$

On en déduit la continuité de $q \to \mathbb{E}[(q-D)^+)]$ puis celle de g.
Pour assurer que $\inf_{q\geq 0} g(q)$ est atteint, il suffit maintenant de vérifier que $\lim_{q\to+\infty} g(q) = +\infty$. Pour montrer ce résultat, on distingue deux cas dans lesquels on utilise respectivement les inégalités $c > 0$ et $c > -c_S$:

- si $c_S \geq 0$, alors $g(q) \geq cq$,
- si $c_S < 0$, pour $q \geq 0$, $(q-D)^+ \leq q$ et donc $\mathbb{E}[(q-D)^+] \leq q$ ce qui implique $g(q) \geq (c + c_S)q$.

Ainsi $\inf_{q\geq 0} g(q)$ est atteint. La proposition suivante indique quelle quantité le gestionnaire doit commander pour minimiser le coût.

Proposition 3.1.1.

- *Si $c_M \leq c$ alors la fonction g est croissante et ne rien commander est optimal.*
- *Sinon, $(c_M - c)/(c_M + c_S) \in]0, 1[$ et si on pose*

$$S = \inf\{z \in \mathbb{R} : F(z) \geq (c_M - c)/(c_M + c_S)\} \tag{3.3}$$

alors $S \in \mathbb{R}_+$. En outre, si D est une variable aléatoire entière i.e. $\mathbb{P}(D \in \mathbb{N}) = 1$, alors $S \in \mathbb{N}$. Enfin, g est décroissante sur $]-\infty, S]$ et croissante sur $[S, +\infty[$, ce qui implique que commander S est optimal.

Remarque 3.1.2. L'hypothèse $c_M > c$ qui rend le problème d'optimisation intéressant peut se justifier par des considérations économiques. En effet, il est naturel de supposer que le prix de vente unitaire du produit par le gestionnaire à ses clients est supérieur au coût c auquel il s'approvisionne. Comme la perte c_M correspondant à une unité de manquants est égale à la somme de ce prix de vente unitaire et du coût en termes d'image, on a alors $c_M > c$. ◊

Démonstration. D'après (3.1) et (3.2),

$$g(q) = c_M\mu + (c - c_M)q + (c_M + c_S)\mathbb{E}[(q-D)^+]. \tag{3.4}$$

Pour $q \geq 0$, $(q-D)^+ = \int_0^q \mathbf{1}_{\{z\geq D\}} dz$ ce qui implique en utilisant le théorème de Fubini,

$$\mathbb{E}[(q-D)^+] = \mathbb{E}\left[\int_0^q \mathbf{1}_{\{z\geq D\}} dz\right] = \int_0^q \mathbb{E}\left[\mathbf{1}_{\{D\leq z\}}\right] dz = \int_0^q F(z)dz.$$

Comme D est une variable aléatoire positive, sa fonction de répartition F est nulle sur $]-\infty, 0[$ et l'égalité précédente reste vraie pour $q < 0$. Donc

$$\forall q \in \mathbb{R},\ g(q) = c_M \mu + \int_0^q ((c - c_M) + (c_M + c_S)F(z))\, dz. \tag{3.5}$$

- Si $c_M \leq c$ alors comme F est à valeurs dans $[0, 1]$, l'intégrande dans le membre de droite est minoré par $c - c_M + \min(0, c_M + c_S) = \min(c - c_M, c + c_S)$. Ce minorant est positif si bien que g est une fonction croissante et ne rien commander est optimal.
- Si $c_M > c$, comme $c > -c_S$, $0 < c_M - c < c_M + c_S$. D'où

$$0 < \frac{c_M - c}{c_M + c_S} < 1.$$

Comme F est nulle sur $]-\infty, 0[$, l'ensemble $\{z \in \mathbb{R} : F(z) \geq (c_M - c)/(c_M + c_S)\}$ est inclus dans $\mathbb{R}_+$. Puisque $\lim_{z \to +\infty} F(z) = 1$, il est non vide. Donc sa borne inférieure S définie par (3.3) est dans $\mathbb{R}_+$.
Lorsque la demande D est une variable entière, la fonction de répartition F est constante sur les intervalles $[n, n+1[$, $n \in \mathbb{N}$ et présente des sauts égaux à $\mathbb{P}(D = n)$ aux points $n \in \mathbb{N}$. On en déduit que $S = \min\{n \in \mathbb{N},\ \sum_{k=0}^{n} \mathbb{P}(D = k) \geq (c_M - c)/(c_M + c_S)\}$ et que $S \in \mathbb{N}$.
L'intégrande dans le membre de droite de (3.5) est croissant, négatif pour $z < S$ et positif pour $z \geq S$. Donc g est décroissante sur $]-\infty, S]$, croissante sur $[S, +\infty[$ et commander S est optimal.

□

Remarque 3.1.3.

- Lorsque $c_M > c$, si la fonction de répartition F de la demande D est continue (c'est le cas par exemple si la variable aléatoire D possède une densité), alors $F(S) = (c_M - c)/(c_M + c_S)$. Cette égalité peut se voir comme la condition d'optimalité du premier ordre $g'(S) = 0$ puisque $g'(q) = (c - c_M) + (c_M + c_S)F(q)$. On peut la récrire

$$c + c_S \mathbb{P}(D \leq S) = c_M \mathbb{P}(D > S).$$

Elle a donc l'interprétation économique suivante : le surcoût moyen $c + c_S \mathbb{P}(D \leq S)$ lié à la commande d'une unité supplémentaire est compensé par l'économie moyenne $c_M \mathbb{P}(D > S)$ réalisée grâce à cette unité supplémentaire.
- La fonction $q \to (q - D)^+$ est convexe sur $\mathbb{R}$. Lorsque $c_M + c_S \geq 0$, on en déduit en multipliant par $c_M + c_S$ et en prenant l'espérance que $q \to (c_M + c_S)\mathbb{E}[(q - D)^+]$ est convexe sur $\mathbb{R}$. Avec (3.4), on conclut que g est alors une fonction convexe sur $\mathbb{R}$.

◊

Choix de la loi de la demande D

Il est souvent raisonnable de considérer que la demande D provient d'un grand nombre n de clients indépendants qui ont chacun une probabilité p de

commander une unité du produit. C'est par exemple le cas pour une pièce de rechange automobile : les n clients potentiels sont les détenteurs de la voiture pour laquelle la pièce est conçue. Dans ces conditions, la demande suit la loi binomiale de paramètre (n,p) i.e. pour $0 \leq k \leq n$, la probabilité qu'elle vaille k est donnée par $\binom{n}{k}p^k(1-p)^{n-k}$. En particulier $\mu = \mathbb{E}[D] = np$. La loi binomiale n'étant pas d'une manipulation très agréable, on pourra préférer les deux lois obtenues dans les passages à la limite suivants :

- n grand ($n \to +\infty$) et p petit ($p \to 0$) avec $np \to \mu > 0$. Dans cette asymptotique, d'après l'exemple 6.3.1, la loi binomiale de paramètres (n,p) converge étroitement vers la loi de Poisson de paramètre μ. On peut donc modéliser la demande comme une variable de Poisson de paramètre μ.
- n grand ($n \to +\infty$) avec $p > 0$ fixé, alors le théorème central limite A.3.15 justifie l'utilisation de la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ (avec $\sigma^2 = \mu(1-\mu/n)$) de densité $\frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$ comme loi pour D. Une variable aléatoire gaussienne de variance $\sigma^2 > 0$ a toujours une probabilité strictement positive de prendre des valeurs négatives. Mais dans les conditions d'application du théorème central limite, cette probabilité est très faible pour la loi limite $\mathcal{N}(\mu, \sigma^2)$ et la commande de la quantité S donnée par la proposition 3.1.1 est une bonne stratégie.

Taux de manquants

On suppose $c_M > c$. Du point de vue du gestionnaire, le taux de manquants i.e. le taux de demandes de clients non satisfaites est un indicateur important. Lorsqu'il a commandé S, ce taux est égal à $\frac{(D-S)^+}{D}$. On a

$$\mathbb{E}\left[\frac{(D-S)^+}{D}\right] \leq \mathbb{E}\left[\mathbf{1}_{\{D>S\}}\right] = \mathbb{P}(D > S) = 1 - F(S) \leq \frac{c + c_S}{c_M + c_S},$$

la dernière inégalité étant une égalité si F est continue en S. En général, l'espérance du taux de manquants est même significativement plus petite que la probabilité pour qu'il y ait des manquants.

Exemple 3.1.4. Dans le cas particulier où D suit la loi de Poisson de paramètre 50, $c = 10$, $c_M = 20$ et $c_S = 5$, on vérifie numériquement que le stock objectif vaut $S = 48$. Le rapport $\dfrac{c + c_S}{c_M + c_S}$ qui vaut 0.6 est légèrement supérieur à $\mathbb{P}(D > S) \simeq 0.575$ mais très sensiblement supérieur au taux de manquants qui est égal à 6.8 %. La fonction g correspondant à ce cas particulier est représentée sur la Fig. 3.1. ◊

Résolution avec stock initial et coût fixe d'approvisionnement

On suppose maintenant que si le gestionnaire commande une quantité q non nulle auprès de son fournisseur, il doit payer un coût fixe c_F positif en plus

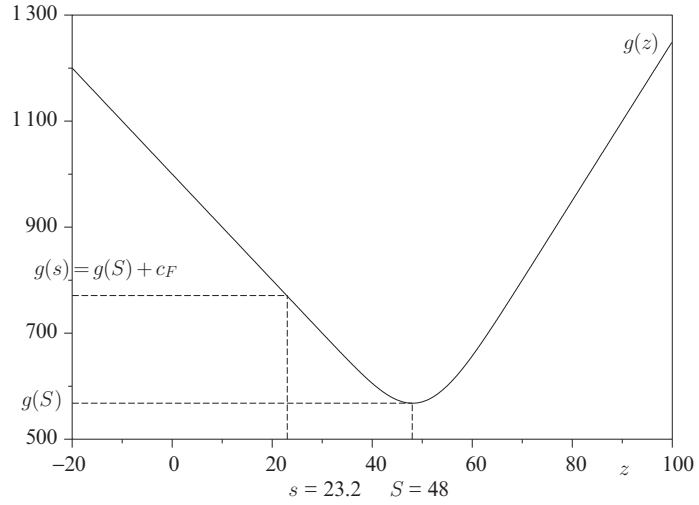

Fig. 3.1. Représentation de $g(z)$, s et S ($c = 10$, $c_M = 20$, $c_S = 5$, $c_F = 200$, D distribuée suivant la loi de Poisson de paramètre 50)

du coût unitaire c. On suppose également que le stock initial x n'est pas nécessairement nul. On autorise même la situation où x est négatif qui traduit le fait qu'une quantité égale à $-x$ de demandes formulées par des clients avant le début de la période n'ont pu être satisfaites et sont maintenues par ces clients. Le problème est maintenant de trouver la quantité commandée $q \geq 0$ qui minimise

$$c_F \mathbf{1}_{\{q>0\}} + cq + c_S \mathbb{E}[(x+q-D)^+] + c_M \mathbb{E}[(D-x-q)^+]. \tag{3.6}$$

Proposition 3.1.5. *On suppose $c_M > c$. Alors l'ensemble*

$$\{z \in]-\infty, S],\ g(z) \geq c_F + g(S)\}$$

est non vide. Si on note s sa borne supérieure, la stratégie (s, S) qui consiste à commander la quantité $q = (S - x)$ permettant d'atteindre le stock objectif S lorsque x est inférieur au stock seuil s et à ne rien faire ($q = 0$) sinon est optimale au sens où elle minimise (3.6).

La figure 3.1 représente la fonction g et fournit une interprétation graphique du couple (s, S) dans le cas particulier où D suit la loi de Poisson de paramètre 50, $c = 10$, $c_M = 20$, $c_S = 5$ et $c_F = 200$.

Démonstration. Lorsque $z \leq 0$, on a $(z - D)^+ = 0$ et $(D - z)^+ = D - z$, d'où pour la fonction g définie par (3.1),

$$\forall z \leq 0,\ g(z) = cz + c_M(\mathbb{E}[D] - z) = (c - c_M)z + c_M\mu.$$

Comme $c_M > c$, cela implique que $\lim_{z\to-\infty} g(z) = +\infty$. Ainsi l'ensemble $\{z \in]-\infty, S],\ g(z) \geq c_F + g(S)\}$ est non vide et sa borne supérieure s est dans $]-\infty, S]$.

On ne change rien à la quantité q optimale en ajoutant cx au coût (3.6). Le critère à minimiser devient $c_F\mathbf{1}_{\{q>0\}} + g(x+q)$. Comme d'après la proposition 3.1.1, g est croissante sur $[S, +\infty[$, il est clairement optimal de ne rien commander si $x \geq S$.
Pour $x < S$, comme g atteint son minimum en S, il est optimal de commander $S - x$ si $g(x) \geq c_F + g(S)$ et de ne rien commander sinon. Notons que lorsque le produit se présente sous forme d'unités, le stock initial x est entier tout comme le stock objectif S (voir la proposition 3.1.1), ce qui assure que le gestionnaire peut bien commander la quantité $S - x$ qui est entière. La décroissance et la continuité de g sur $]-\infty, S]$ entraînent que $\{x \in]-\infty, S],\ g(x) \geq c_F + g(S)\} =]-\infty, s]$, ce qui achève la démonstration. □

Remarque 3.1.6.

- Si $c_F = 0$, alors $s = S$. Comme la stratégie (S, S) consiste à commander $(S - x)$ lorsque $x \leq S$, on retrouve bien le résultat de la proposition 3.1.1.
- Le coût minimal associé à la stratégie (s, S) est

$$\begin{aligned} u(x) &= -cx + \inf_{q\geq 0}\left(c_F\mathbf{1}_{\{q>0\}} + g(x+q)\right) \\ &= -cx + (g(S) + c_F)\mathbf{1}_{\{x\leq s\}} + g(x)\mathbf{1}_{\{x>s\}}. \end{aligned}$$

  Dans la définition 3.3.3, nous introduirons une notion de convexité spécifique à la gestion des approvisionnements. Le coût minimal $u(x)$ fournira un exemple typique de fonction satisfaisant cette propriété de convexité.

  ◊

3.1.2 Le modèle dynamique de gestion de stock

Le modèle comporte toujours un seul produit mais N périodes de temps (typiquement des journées, des semaines ou des mois). Pour $t \in \{0, \ldots, N-1\}$, une demande D_{t+1} positive est formulée par les clients sur la période $[t, t+1]$. Les variables aléatoires $D_1, D_2, \ldots, D_N$ sont supposées indépendantes et identiquement distribuées de fonction de répartition F et d'espérance finie i.e. $\mathbb{E}[D_1] = \mu < +\infty$.
Au début de chaque période $[t, t+1]$, le gestionnaire décide de commander une quantité $Q_t \geq 0$ qui lui est livrée pendant la période. Plutôt que le stock physique qui est toujours positif ou nul la grandeur intéressante à considérer est le **stock système** qui peut prendre des valeurs négatives. Le stock système est défini comme le stock physique si celui-ci est strictement positif et comme moins la quantité de manquants (i.e. les demandes de clients qui n'ont pu être honorées faute de stock) sinon. On note X_t le stock système à l'instant $t \in \{0, \ldots, N\}$.

On suppose que
- le gestionnaire ne renvoie pas de produit à son fournisseur,
- les demandes correspondant aux manquants sont maintenues par les clients jusqu'à ce que le gestionnaire puisse les servir.

Ainsi la dynamique du stock système est donnée par

$$X_{t+1} = X_t + Q_t - D_{t+1}. \tag{3.7}$$

Au cours de la période $[t, t+1]$, le coût est donné par une formule analogue à celle du modèle à une seule période : $c_F \mathbf{1}_{\{Q_t>0\}} + cQ_t + c_S(X_{t+1})^+ + c_M(X_{t+1})^-$ où $y^- = \max(-y, 0)$. On associe au stock système final X_N un coût $u_N(X_N)$ (par exemple, estimer que ce stock final a une valeur unitaire égale à c consiste à poser $u_N(x) = -cx$). Enfin pour traduire la préférence du gestionnaire pour 1 Euro à la date t par rapport à 1 Euro à la date $t+1$, on utilise un facteur d'actualisation $\alpha \in [0, 1]$. L'espérance du coût total actualisé est donnée par

$$\mathbb{E}\left[\sum_{t=0}^{N-1} \alpha^t \left(c_F \mathbf{1}_{\{Q_t>0\}} + cQ_t + c_S(X_{t+1})^+ + c_M(X_{t+1})^-\right) + \alpha^N u_N(X_N)\right]. \tag{3.8}$$

L'objectif du gestionnaire est de choisir les quantités Q_t au vu du passé jusqu'à l'instant t de manière à minimiser l'espérance du coût total. Le but du paragraphe suivant est de montrer que la résolution d'un tel problème d'optimisation à N périodes de temps peut se ramener à la résolution de N problèmes à une période de temps définis par récurrence descendante.

3.2 Éléments de contrôle de chaînes de Markov

Nous allons dans ce paragraphe décrire puis résoudre un problème de contrôle de chaînes de Markov qui englobe le problème dynamique de gestion de stock que nous venons de présenter.

3.2.1 Description du modèle

On considère un modèle d'évolution d'un système commandé par un gestionnaire sur N périodes de temps. On note $\mathcal{E}$ l'ensemble des états possibles du système et $\mathcal{A}$ l'ensemble des actions possibles du gestionnaire. Ces deux ensembles sont supposés discrets. L'état du système à l'instant $t \in \{0, \ldots, N\}$ est noté X_t tandis que l'action choisie par le gestionnaire à l'instant $t \in \{0, \ldots, N-1\}$ est notée A_t. Pour $t \in \{0, \ldots, N-1\}$, le gestionnaire choisit l'action A_t au vu de l'histoire $H_t = (X_0, A_0, X_1, A_1, \ldots, X_{t-1}, A_{t-1}, X_t) \in (\mathcal{E} \times \mathcal{A})^t \times \mathcal{E}$ jusqu'à l'instant t. On suppose que l'état X_{t+1} du système à l'instant $t+1$ ne dépend de l'histoire H_t et de l'action A_t qu'au travers du couple (X_t, A_t) : $\forall t \in \{0, \ldots, N-1\}$,

$$\forall h_t = (x_0, a_0, x_1, a_1, \ldots, x_t) \in (\mathcal{E} \times \mathcal{A})^t \times \mathcal{E},\ \forall a_t \in \mathcal{A},\ \forall x_{t+1} \in \mathcal{E},$$
$$\begin{aligned}\mathbb{P}(X_{t+1} = x_{t+1} | H_t = h_t, A_t = a_t) &= \mathbb{P}(X_{t+1} = x_{t+1} | X_t = x_t, A_t = a_t) \\ &= p_t((x_t, a_t), x_{t+1}),\end{aligned} \tag{3.9}$$

où la matrice $p_t : (\mathcal{E} \times \mathcal{A}) \times \mathcal{E} \to [0,1]$ vérifie $\sum_{y \in \mathcal{E}} p_t((x,a),y) = 1$ pour tout $(x,a) \in \mathcal{E} \times \mathcal{A}$.

Pour $t \in \{0, \ldots, N-1\}$, le choix de l'action A_t induit un coût égal à $\varphi_t(X_t, A_t, X_{t+1})$ sur la période $[t, t+1]$ où $\varphi_t : \mathcal{E} \times \mathcal{A} \times \mathcal{E} \to \mathbb{R}$. A l'instant final, un coût $u_N(X_N)$ est associé à l'état final X_N avec $u_N : \mathcal{E} \to \mathbb{R}$. L'espérance du coût total actualisé avec le facteur d'actualisation $\alpha \in [0,1]$ est donnée par

$$\mathbb{E}\left[\sum_{n=0}^{N-1} \alpha^n \varphi_n(X_n, A_n, X_{n+1}) + \alpha^N u_N(X_N)\right].$$

Exemple 3.2.1. Le problème dynamique de gestion de stock décrit au paragraphe 3.1.2 entre dans ce cadre lorsque le produit se présente sous forme d'unités. L'espace d'état est $\mathcal{E} = \mathbb{Z}$ pour le stock système X_t et l'ensemble d'actions $\mathcal{A} = \mathbb{N}$ pour la quantité Q_t de produit commandée. On a

$$X_{t+1} = X_t + Q_t - D_{t+1}$$

où les demandes $(D_1, \ldots, D_N)$ sont des variables aléatoires positives indépendantes et identiquement distribuées que l'on suppose à valeurs dans $\mathbb{N}$. Donc pour $t \in \{0, \ldots, N-1\}$ et $h_t = (x_0, q_0, x_1, q_1, \ldots, x_t) \in (\mathbb{Z} \times \mathbb{N})^t \times \mathbb{Z}$, on a

$$\begin{aligned}&\mathbb{P}(X_{t+1} = x_{t+1} | H_t = h_t, Q_t = q_t) \\ &\qquad = \mathbb{P}(X_t + Q_t - D_{t+1} = x_{t+1} | H_t = h_t, Q_t = q_t) \\ &\qquad = \frac{\mathbb{P}(D_{t+1} = x_t + q_t - x_{t+1}, H_t = h_t, Q_t = q_t)}{\mathbb{P}(H_t = h_t, Q_t = q_t)} \\ &\qquad = \mathbb{P}(D_{t+1} = x_t + q_t - x_{t+1}) = \mathbb{P}(D_1 = x_t + q_t - x_{t+1}),\end{aligned}$$

car les variables H_t et Q_t ne dépendent que de $D_1, D_2, \ldots, D_t$ et sont indépendantes de D_{t+1}. Ainsi (3.9) est vérifié pour $p_t((x,q),y) = \mathbb{P}(D_1 = x + q - y)$. Enfin la fonction φ_t qui intervient dans l'expression de l'espérance du coût total actualisé est donnée par : $\forall (t,x,q,y) \in \{0, \ldots, N-1\} \times \mathbb{Z} \times \mathbb{N} \times \mathbb{Z}$,

$$\varphi_t(x,q,y) = c_F \mathbf{1}_{\{q>0\}} + cq + c_S y^+ + c_M y^-.$$

Elle ne dépend que des deux dernières variables. ◊

Le fait que le gestionnaire décide de l'action A_t au vu de l'histoire H_t se traduit par l'existence d'une application d_t de l'ensemble $(\mathcal{E} \times \mathcal{A})^t \times \mathcal{E}$ des histoires possibles jusqu'à l'instant t dans l'ensemble $\mathcal{A}$ des actions telle que $A_t = d_t(H_t)$. Cette application d_t est appelée règle de décision du gestionnaire à l'instant t. La stratégie $\pi = (d_0, \ldots, d_{N-1})$ du gestionnaire est constituée

de l'ensemble de ses règles de décision. Bien sûr, la dynamique du système $X_t,\ t \in \{0,\dots,N\}$ dépend de la stratégie du gestionnaire. Pour expliciter cette dépendance on note désormais X_t^π et H_t^π l'état du système et l'histoire à l'instant $t \in \{0,\dots,N\}$ qui correspondent à la stratégie π. Sachant que l'état initial du système est x_0, le coût moyen associé à la stratégie π est donné par :

$$v_0^\pi(x_0) = \mathbb{E}\left[\sum_{n=0}^{N-1} \alpha^n \varphi_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) + \alpha^N u_N(X_N^\pi)\middle| X_0^\pi = x_0\right].$$

On note

$$v_0^*(x_0) = \inf_\pi v_0^\pi(x_0).$$

L'objectif du gestionnaire est de trouver à l'instant initial une stratégie optimale π^* qui réalise l'infimum lorsque celui-ci est atteint ou une stratégie ε-optimale π^ε vérifiant $\forall x_0 \in \mathcal{E},\ v_0^{\pi^\varepsilon}(x_0) \leq v_0^*(x_0) + \varepsilon$ avec $\varepsilon > 0$ arbitraire lorsque l'infimum n'est pas atteint.

Dans le paragraphe 3.2.2 nous allons montrer comment évaluer le coût associé à une stratégie par récurrence descendante. Puis dans le paragraphe 3.2.3, nous en déduirons les équations d'optimalité et nous montrerons que leur résolution permet de construire des stratégies optimales ou ε-optimales. Enfin, dans le paragraphe 3.2.4, nous appliquerons la théorie développée pour déterminer la stratégie optimale d'un recruteur dans « le problème de la secrétaire ».

3.2.2 Évaluation du coût associé à une stratégie

Nous allons montrer que le coût moyen associé à la stratégie π peut être évalué par récurrence descendante. On introduit pour cela le coût à venir à l'instant $t \in \{1,\dots,N\}$ sachant que l'histoire est $h_t \in (\mathcal{E}\times\mathcal{A})^t \times \mathcal{E}$:

$$v_t^\pi(h_t) = \mathbb{E}\left[\sum_{n=t}^{N-1} \alpha^{n-t} \varphi_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) + \alpha^{N-t} u_N(X_N^\pi)\middle| H_t^\pi = h_t\right].$$

Cette notation est bien compatible avec la définition de v_0^π au paragraphe précédent. La proposition suivante indique comment exprimer v_t^π en fonction de v_{t+1}^π :

Proposition 3.2.2.

$$\forall h_N = (x_0, a_0, x_1, a_1, \dots, x_N) \in (\mathcal{E}\times\mathcal{A})^N \times \mathcal{E},\ v_N^\pi(h_N) = u_N(x_N).$$

En outre, pour $t \in \{0,\dots,N-1\}$, $\forall h_t = (x_0, a_0, x_1, a_1, \dots, x_t) \in (\mathcal{E}\times\mathcal{A})^t \times \mathcal{E}$,

$$v_t^\pi(h_t) = \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, d_t(h_t)), x_{t+1}))$$
$$\left[\varphi_t(x_t, d_t(h_t), x_{t+1}) + \alpha v_{t+1}^\pi((h_t, d_t(h_t), x_{t+1}))\right].$$

La formule précédente peut s'interpréter de la façon suivante : lorsque l'histoire jusqu'à l'instant t est $h_t \in (\mathcal{E} \times \mathcal{A})^t \times \mathcal{E}$, sous la stratégie π, l'action du gestionnaire en t est $d_t(h_t)$. Pour tout $x_{t+1} \in \mathcal{E}$, l'état X_{t+1}^π du système à l'instant $t+1$ et l'histoire H_{t+1}^π jusqu'à l'instant $t+1$ sont donc respectivement égaux à x_{t+1} et à $(h_t, d_t(h_t), x_{t+1}) \in (\mathcal{E} \times \mathcal{A})^{t+1} \times \mathcal{E}$ avec probabilité $p_t((x_t, d_t(h_t)), x_{t+1})$. Le coût à venir à l'instant t sachant que l'histoire est h_t se décompose donc comme la somme sur les états possibles $x_{t+1} \in \mathcal{E}$ du système à l'instant $t+1$ pondérée par $p_t((x_t, d_t(h_t)), x_{t+1})$ du coût $\varphi_t(x_t, d_t(h_t), x_{t+1})$ sur la période $[t, t+1]$ plus le coût actualisé $\alpha v_{t+1}^\pi((h_t, d_t(h_t), x_{t+1}))$ à venir à l'instant $t+1$ sachant que l'histoire est $(h_t, d_t(h_t), x_{t+1})$.

Démonstration. Par définition,

$$v_N^\pi(h_N) = \mathbb{E}[u_N(X_N^\pi) | H_N^\pi = h_N] = \frac{\mathbb{E}\left[u_N(X_N^\pi)\mathbf{1}_{\{H_N^\pi = h_N\}}\right]}{\mathbb{P}(H_N^\pi = h_N)}.$$

Comme $H_N^\pi = h_N = (x_0, a_0, x_1, a_1, \ldots, x_N)$ implique $X_N^\pi = x_N$, on a

$$\mathbb{E}\left[u_N(X_N^\pi)\mathbf{1}_{\{H_N^\pi = h_N\}}\right] = \mathbb{E}\left[u_N(x_N)\mathbf{1}_{\{H_N^\pi = h_N\}}\right] = u_N(x_N)\mathbb{P}(H_N^\pi = h_N),$$

et on conclut que $v_N^\pi(h_N) = u_N(x_N)$.
Soit $t \in \{0, \ldots, N-1\}$. Par linéarité de l'espérance,

$$\begin{aligned}
\mathbb{P}(H_t^\pi = h_t) v_t^\pi(h_t) &= \mathbb{E}\Bigg[\mathbf{1}_{\{H_t^\pi = h_t\}}\Bigg(\sum_{n=t}^{N-1} \alpha^{n-t}\varphi_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) \\
&\qquad + \alpha^{N-t}u_N(X_N^\pi)\Bigg)\Bigg] \\
&= \sum_{x_{t+1} \in \mathcal{E}} \mathbb{E}\Bigg[\mathbf{1}_{\{X_{t+1}^\pi = x_{t+1}, H_t^\pi = h_t\}}\Bigg(\sum_{n=t}^{N-1} \alpha^{n-t}\varphi_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) \\
&\qquad + \alpha^{N-t}u_N(X_N^\pi)\Bigg)\Bigg]
\end{aligned}$$

Comme sous la stratégie π, $H_t^\pi = h_t$ implique $A_t = d_t(h_t)$, $(H_t^\pi = h_t, X_{t+1}^\pi = x_{t+1})$ implique $H_{t+1}^\pi = (h_t, d_t(h_t), x_{t+1})$. Donc

$$\begin{aligned}
\mathbb{P}(H_t^\pi = h_t) v_t^\pi(h_t) &= \sum_{x_{t+1} \in \mathcal{E}} \mathbb{E}\Bigg[\mathbf{1}_{\{H_{t+1}^\pi = (h_t, d_t(h_t), x_{t+1})\}}\Bigg(\varphi_t(x_t, d_t(h_t), x_{t+1}) \\
&+ \alpha\Bigg\{\sum_{n=t+1}^{N-1} \alpha^{n-t-1}\varphi_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) + \alpha^{N-t-1}u_N(X_N^\pi)\Bigg\}\Bigg)\Bigg] \\
&= \sum_{x_{t+1} \in \mathcal{E}} \mathbb{P}(H_{t+1}^\pi = (h_t, d_t(h_t), x_{t+1}))\Big[\varphi_t(h_t, d_t(h_t), x_{t+1}) \\
&\qquad + \alpha v_{t+1}^\pi((h_t, d_t(h_t), x_{t+1}))\Big].
\end{aligned}$$

Il suffit de remarquer que

$$\begin{aligned}\frac{\mathbb{P}(H_{t+1}^\pi = (h_t, d_t(h_t), x_{t+1}))}{\mathbb{P}(H_t^\pi = h_t)} &= \frac{\mathbb{P}(X_{t+1}^\pi = x_{t+1}, H_t^\pi = h_t, A_t = d_t(h_t))}{\mathbb{P}(H_t^\pi = h_t, A_t = d_t(h_t))} \\ &= \mathbb{P}(X_{t+1}^\pi = x_{t+1} | H_t^\pi = h_t, A_t = d_t(h_t)) \\ &= p_t((x_t, d_t(h_t)), x_{t+1})\end{aligned}$$

d'après (3.9) pour conclure la démonstration de la formule de récurrence. □

3.2.3 Équations d'optimalité

On définit $v_t^*(h_t) = \inf_\pi v_t^\pi(h_t)$. Le théorème suivant explique comment évaluer v_t^* par récurrence descendante :

Théorème 3.2.3. *Pour tout $t \in \{0, \dots, N\}$,*

$$\forall h_t = (x_0, a_0, x_1, a_1, \dots, x_t) \in (\mathcal{E} \times \mathcal{A})^t \times \mathcal{E},\ v_t^*(h_t) = u_t(x_t)$$

où les fonctions u_t sont définies par récurrence descendante à partir du coût terminal u_N par **les équations d'optimalité** *: pour $t \in \{0, \dots, N-1\}$,*

$$\forall x_t \in \mathcal{E},\ u_t(x_t) = \inf_{a \in \mathcal{A}} \sum_{x_{t+1} \in \mathcal{E}} p_t((x_t, a), x_{t+1}) \left[\varphi_t(x_t, a, x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right]. \tag{3.10}$$

Démonstration. D'après la proposition 3.2.2, $\forall \pi,\ v_N^\pi(h_N) = u_N(x_N)$. Donc $v_\pi^*(h_N) = u_N(x_N)$, ce qui permet d'initialiser la démonstration par récurrence descendante de l'assertion $\forall t \in \{0, \dots, N\},\ v_t^*(h_t) \geq u_t(x_t)$. Supposons que l'hypothèse de récurrence est vraie à l'instant $t+1$.
Soit π une stratégie. On a $v_{t+1}^\pi(h_{t+1}) \geq v_{t+1}^*(h_{t+1}) \geq u_{t+1}(x_{t+1})$. En insérant cette inégalité dans la formule de récurrence de la proposition 3.2.2, on obtient

$$\begin{aligned}v_t^\pi(h_t) &\geq \sum_{x_{t+1} \in \mathcal{E}} p_t((x_t, d_t(h_t)), x_{t+1}) \left[\varphi_t(x_t, d_t(h_t), x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right] \\ &\geq \inf_{a \in \mathcal{A}} \sum_{x_{t+1} \in \mathcal{E}} p_t((x_t, a), x_{t+1}) \left[\varphi_t(x_t, a, x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right] = u_t(x_t).\end{aligned}$$

Comme la stratégie π est arbitraire, on conclut que $v_t^*(h_t) = \inf_\pi v_t^\pi(h_t) \geq u_t(x_t)$ i.e. que l'hypothèse de récurrence est vérifiée au rang t.
Pour montrer l'inégalité inverse, on fixe $\gamma > 0$. Pour $t \in \{0, \dots, N-1\}$ et $x_t \in \mathcal{E}$, d'après (3.10), il existe $\delta_t(x_t) \in \mathcal{A}$ t.q.

$$\sum_{x_{t+1}\in\mathcal{E}} p_t((x_t,\delta_t(x_t)),x_{t+1})\left[\varphi_t(x_t,\delta_t(x_t),x_{t+1})+\alpha u_{t+1}(x_{t+1})\right]\leq u_t(x_t)+\gamma.$$

Notons π^γ la stratégie telle que la règle de décision à chaque instant t dans $\{0,\dots,N-1\}$ est $d_t(h_t)=\delta_t(x_t)$ et montrons par récurrence descendante que

$$\forall t\in\{0,\dots,N\},\ \forall h_t=(x_0,a_0,x_1,a_1,\dots,x_t),\ v_t^{\pi^\gamma}(h_t)\leq u_t(x_t)+(N-t)\gamma.$$

On a $v_N^{\pi^\gamma}(h_N)=u_N(x_N)$. Supposons que l'hypothèse est vérifiée pour n dans $\{t+1,\dots,N\}$. En insérant l'inégalité $v_{t+1}^{\pi^\gamma}(h_{t+1})\leq u_{t+1}(x_{t+1})+(N-t-1)\gamma$ dans la formule de récurrence de la proposition 3.2.2, on obtient

$$\begin{aligned}v_t^{\pi^\gamma}(h_t)\leq \sum_{x_{t+1}\in\mathcal{E}} & p_t((x_t,d_t(h_t)),x_{t+1}))\\ & \left[\varphi_t(x_t,d_t(h_t),x_{t+1})+\alpha\left(u_{t+1}(x_{t+1})+(N-t-1)\gamma\right)\right]\end{aligned}$$

En utilisant successivement $\sum_{x_{t+1}\in\mathcal{E}} p_t((x_t,d_t(h_t)),x_{t+1}))=1$, la définition de d_t et $\alpha\leq 1$, on en déduit que

$$\begin{aligned}v_t^{\pi^\gamma}(h_t)&\leq \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t,\delta_t(x_t)),x_{t+1}))\left[\varphi_t(x_t,\delta_t(x_t),x_{t+1})+\alpha u_{t+1}(x_{t+1})\right]\\ &\quad+\alpha(N-t-1)\gamma\\ &\leq u_t(x_t)+\gamma+(N-t-1)\gamma.\end{aligned}$$

Donc l'hypothèse de récurrence est vérifiée au rang t.
Pour $t\in\{0,\dots,N\}$ et $h_t\in(\mathcal{E}\times\mathcal{A})^t\times\mathcal{E}$, on conclut que

$$v_t^*(h_t)\leq\inf_{\gamma>0} v_t^{\pi^\gamma}(h_t)\leq\inf_{\gamma>0}\left(u_t(x_t)+(N-t)\gamma\right)=u_t(x_t).$$

□

Définition 3.2.4.

1. *Une stratégie π^* est dite optimale si*

$$\forall t\in\{0,\dots,N\},\ \forall h_t\in(\mathcal{E}\times\mathcal{A})^t\times\mathcal{E},\ v_t^{\pi^*}(h_t)=\inf_\pi v_t^\pi(h_t)=v_t^*(h_t).$$

2. *Une stratégie π^ε est dite ε-optimale si*

$$\forall t\in\{0,\dots,N\},\ \forall h_t\in(\mathcal{E}\times\mathcal{A})^t\times\mathcal{E},\ v_t^{\pi^\varepsilon}(h_t)\leq v_t^*(h_t)+\varepsilon.$$

3. *Une stratégie $\pi_M=(d_0,\dots,d_{N-1})$ est dite markovienne si pour tout t dans $\{0,\dots,N-1\}$, la règle de décision à l'instant t ne dépend du passé qu'au travers de l'état du système à l'instant t i.e. il existe $\delta_t:\mathcal{E}\to\mathcal{A}$ t.q. $\forall h_t=(x_0,a_0,x_1,a_1,\dots,x_t),\ d_t(h_t)=\delta_t(x_t)$.*

Remarque 3.2.5.

- Les équations d'optimalité (3.10) expriment que pour qu'une stratégie soit optimale sur la période $[t, N]$, il faut que la décision prise en t soit optimale mais aussi que toutes les décisions ultérieures le soient. Ainsi, lorsqu'une stratégie π est optimale au sens introduit à la fin du paragraphe 3.2.1, à savoir $v_0^\pi(x_0) = v_0^*(x_0)$, elle est aussi optimale au sens du point 1 de la définition précédente (qui peut sembler plus exigeant en première lecture).
- Les stratégies markoviennes sont particulièrement intéressantes car il n'est pas nécessaire de garder en mémoire tout le passé pour les appliquer. La terminologie markovienne provient de ce que sous une telle stratégie $\pi_M = (\delta_0, \ldots, \delta_{N-1})$, $H_t^{\pi_M} = h_t$ entraîne $A_t = \delta_t(x_t)$ et donc

$$\begin{aligned}&\mathbb{P}(X_{t+1}^{\pi_M} = x_{t+1} | H_t^{\pi_M} = h_t)\\ &\quad = \mathbb{P}(X_{t+1}^{\pi_M} = x_{t+1} | H_t^{\pi_M} = h_t, A_t = \delta_t(x_t)) = p_t((x_t, \delta_t(x_t)), x_{t+1})\end{aligned}$$

  d'après (3.9). Ainsi l'état $X_{t+1}^{\pi_M}$ du système à l'instant $t+1$ ne dépend du passé $H_t^{\pi_M}$ qu'au travers de l'état $X_t^{\pi_M}$ du système à l'instant t i.e. la suite $(X_t^{\pi_M})_{t\in\{0,\ldots,N\}}$ est une chaîne de Markov.

$\Diamond$

Il faut noter que pour $\gamma = \varepsilon/N$, la stratégie π^γ qui a été construite dans la démonstration du théorème 3.2.3 est ε-optimale et markovienne, ce qui fournit la première assertion du corollaire suivant :

Corollaire 3.2.6.

1. *Pour tout $\varepsilon > 0$, il existe une stratégie markovienne ε-optimale.*
2. *Si $\mathcal{A}$ est fini ou bien s'il existe une stratégie $\pi^* = (d_0, \ldots, d_{N-1})$ optimale, alors il existe une stratégie markovienne optimale.*

Démonstration. Si $\mathcal{A}$ est fini, alors pour tout $t \in \{0, \ldots, N-1\}$ et tout $x_t \in \mathcal{E}$, l'infimum de l'application

$$a \in \mathcal{A} \to \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, a), x_{t+1}) \left[\varphi_t(x_t, a, x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right] \tag{3.11}$$

est un minimum i.e. il est atteint. Vérifions maintenant que cette propriété reste vraie s'il existe une stratégie $\pi^* = (d_0, \ldots, d_{N-1})$ optimale. Pour $h_t = (x_0, a_0, x_1, a_1, \ldots, x_t)$, d'après le théorème 3.2.3, $u_t(x_t) = v_t^*(h_t) = v_t^{\pi^*}(h_t)$. En utilisant l'équation d'optimalité (3.10) et la proposition 3.2.2, on en déduit que

$$\begin{aligned}&\inf_{a\in\mathcal{A}} \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, a), x_{t+1}) \left[\varphi_t(x_t, a, x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right] = u_t(x_t)\\ &\quad = v_t^{\pi^*}(h_t) = \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, d_t(h_t)), x_{t+1}) \Big[\varphi_t(x_t, d_t(h_t), x_{t+1})\\ &\qquad\qquad + \alpha v_{t+1}^{\pi^*}((h_t, d_t(h_t), x_{t+1}))\Big].\end{aligned}$$

Comme $v_{t+1}^{\pi^*}((h_t, d_t(h_t)), x_{t+1})) = v_{t+1}^*((h_t, d_t(h_t)), x_{t+1})) = u_{t+1}(x_{t+1})$, on en déduit que

$$\inf_{a\in\mathcal{A}} \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, a), x_{t+1}) \left[\varphi_t(x_t, a, x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right]$$

$$= \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, d_t(h_t)), x_{t+1}) \left[\varphi_t(x_t, d_t(h_t), x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right].$$

Puisque $d_t(h_t) \in \mathcal{A}$, l'infimum de (3.11) est atteint pour $a = d_t(h_t)$.
La quantité minimisée (3.11) ne dépend que de x_t et a. Donc il existe δ_t^* : $\mathcal{E} \to \mathcal{A}$ t.q. l'infimum est atteint pour $a = \delta_t(x_t)$. On en déduit que

$$u_t(x_t) = \sum_{x_{t+1}\in\mathcal{E}} p_t((x_t, \delta_t^*(x_t)), x_{t+1}) \left[\varphi_t(x_t, \delta_t^*(x_t), x_{t+1}) + \alpha u_{t+1}(x_{t+1})\right].$$

Il suffit de reprendre l'argument de récurrence descendante de la fin de la démonstration du théorème 3.2.3 avec $\gamma = 0$ pour voir que la stratégie markovienne $\pi_M^* = (\delta_0^*, \ldots, \delta_{N-1}^*)$ est optimale. □

Les équations d'optimalité (3.10) sont aussi appelées équations de Bellman ou équations de la programmation dynamique. Elles permettent de ramener la résolution du problème d'optimisation à N périodes de temps à celle plus simple de N problèmes d'optimisation à une période. Leur résolution par récurrence descendante fournit une procédure effective qui porte le nom de programmation dynamique pour calculer le coût minimal et déterminer les stratégies markoviennes optimales (ou ε-optimales).

La théorie développée dans ce paragraphe peut être étendue à un cadre beaucoup plus général. Par exemple, on peut considérer des espaces d'états $\mathcal{E}$ et d'actions $\mathcal{A}$ non discrets mais pour cela il faut utiliser la notion d'espérance conditionnelle sachant une tribu qui dépasse le cadre de ce livre. Néanmoins, les conclusions sont analogues à celles que nous avons obtenues.

Il est également intéressant de considérer le problème à horizon infini dans le cas où les fonctions p_t et φ_t ne dépendent pas du temps et sont notées respectivement p et φ. On suppose que le facteur d'actualisation α est strictement plus petit que 1 pour pouvoir définir le coût total. Le passage à la limite formel $t \to +\infty$ dans (3.10) explique pourquoi on travaille alors avec l'équation d'optimalité :

$$u(x) = \inf_{a\in\mathcal{A}} \sum_{y\in\mathcal{E}} p((x, a), y) \left[\varphi(x, a, y) + \alpha u(y)\right].$$

Nous renvoyons par exemple aux livres de Puterman [8], de Bertsekas [2, 3], de Bertsekas et Shreve [4], de White [10] et de Whittle [11] pour plus de détails concernant les généralisations possibles.

3.2.4 Application au recrutement : le problème de la secrétaire

L'objectif du problème que nous allons énoncer dans ce paragraphe est de voir comment un problème d'arrêt optimal peut être traité comme un problème de contrôle de chaîne de Markov. Nous renvoyons au Chap. 2 du livre de Lamberton et Lapeyre [5] pour une autre résolution du problème d'arrêt optimal basée sur la technique de l'enveloppe de Snell. L'exemple considéré ici est intéressant parce que suffisamment simple pour qu'il soit possible d'expliciter la stratégie optimale.

Problème 3.2.7. Un nombre N supérieur à 2 de candidats que le recruteur sait classer postule pour un emploi. Pour $t \in \{1, \ldots, N\}$, on note Θ_t le rang, dans l'ordre de préférence du recruteur, du t-ième candidat qui se présente : $\Theta_t = 1$ signifie que le t-ième candidat est le meilleur tandis que $\Theta_t = N$ signifie que c'est le moins bon. Les candidats se présentent dans un ordre aléatoire, ce que l'on modélise en supposant que la permutation aléatoire Θ est distribuée suivant la loi uniforme sur le groupe $\mathcal{S}_N$ des permutations de $\{1, \ldots, N\}$.

On suppose que le recrutement a lieu dans une période de plein emploi et on considère que les candidats qui ne reçoivent pas de réponse positive lors de leur entretien trouvent un emploi ailleurs avant que le recruteur ne puisse les recontacter. Le recruteur reçoit donc successivement les candidats jusqu'à l'instant $\tau \leq N$ où il décide de prendre le candidat qu'il a en face de lui, ce qui achève la procédure de recrutement. Notons que si aucun des $N-1$ premiers candidats n'a été choisi, le N-ième l'est forcément.

Le recruteur souhaite choisir un bon candidat et si possible le meilleur des N candidats, ce que l'on traduit en introduisant le coût $\beta\Theta_\tau + \gamma \mathbf{1}_{\{\Theta_\tau > 1\}}$ fonction du rang Θ_τ du candidat retenu. En outre, son temps est précieux et on affecte le coût $\delta\tau$ à la durée τ de la procédure de recrutement. Les trois constantes β, γ et δ sont supposées positives avec $\beta+\gamma+\delta > 0$. Finalement, le recruteur souhaite choisir l'instant τ qui met fin à la procédure de recrutement (problème d'arrêt optimal) de façon à minimiser

$$\mathbb{E}[\beta\Theta_\tau + \gamma \mathbf{1}_{\{\Theta_\tau > 1\}} + \delta\tau].$$

Dans le cas particulier où le recruteur veut maximiser la probabilité de recruter le meilleur des candidats ($\beta = \delta = 0$), la question 11 permettra de vérifier que sa stratégie optimale est la suivante : observer une proportion explicite proche de $1/e$ des candidats sans les recruter puis retenir le premier candidat meilleur que ceux qu'il a observés.

1. Montrer qu'en termes de stratégie optimale, il revient au même de minimiser $\mathbb{E}\left[\beta\Theta_\tau - \gamma \mathbf{1}_{\{\Theta_\tau = 1\}} + \delta\tau\right]$, problème qui va être traité dans la suite.

Pour $t \in \{1, \ldots, N\}$, on note $R_t \in \{1, \ldots, t\}$ le rang relatif du t-ième candidat parmi les t premiers. Si par exemple $N = 5$ et $(\Theta_1, \ldots, \Theta_5) = (2, 5, 3, 1, 4)$, alors $(R_1, \ldots, R_5) = (1, 2, 2, 1, 4)$.

2. Remarquer que l'ensemble des vecteurs de rang relatifs $(r_1, \ldots, r_N) \in \{1\} \times \{1, 2\} \times \ldots \times \{1, \ldots, N\}$ est en bijection avec celui des permutations

σ de $\mathcal{S}_N$. En déduire que le vecteur $(R_1, \ldots, R_N)$ suit la loi uniforme sur $\{1\} \times \{1,2\} \times \ldots \times \{1, \ldots, N\}$. Puis pour $t \in \{1, \ldots, N\}$ et $(r_1, \ldots, r_t) \in \{1\} \times \{1,2\} \times \ldots \times \{1, \ldots, t\}$ donner $\mathbb{P}(R_t = r_t)$ et $\mathbb{P}(R_1 = r_1, \ldots, R_t = r_t)$.

Désormais, on note $(r_1^\sigma, \ldots, r_N^\sigma)$ le vecteur des rangs relatifs associé à $\sigma \in \mathcal{S}_N$. L'objectif des questions qui suivent est de vérifier que le problème considéré peut être traité comme un problème de contrôle de chaîne de Markov puis de résoudre ce problème de contrôle.
À l'instant $t \in \{1, \ldots, N-1\}$, si le recruteur n'a retenu aucun des $t-1$ premiers candidats, alors il observe $X_t = R_t$ le rang partiel du t-ième candidat parmi les t premiers. Au vu de l'information $(X_1, \ldots, X_t)$ dont il dispose, il a deux choix possibles : soit il refuse ce candidat ce que l'on traduit par $A_t = 0$, soit il recrute ce candidat ce que l'on traduit par $A_t = 1$. Dans ce dernier cas, la procédure de recrutement est achevée ce que l'on traduit par $X_s = \Delta$ (Δ est l'état stoppé) pour $s \in \{t+1, \ldots, N\}$. À l'instant N, s'il n'a retenu aucun des $N-1$ premiers candidats, il observe $X_N = R_N = \Theta_N$ et recrute forcément le N-ième candidat.

3. Montrer que le critère à minimiser se met sous la forme

$$\mathbb{E}\Bigg[\sum_{t=1}^{N-1} \mathbf{1}_{\{X_t \neq \Delta\}} A_t \left(\beta\Theta_t - \gamma \mathbf{1}_{\{\Theta_t = 1\}} + \delta t\right) + \mathbf{1}_{\{X_N \neq \Delta\}} \left(\beta\Theta_N - \gamma \mathbf{1}_{\{\Theta_N = 1\}} + \delta N\right)\Bigg].$$

4. Soit $t \in \{1, \ldots, N\}$ et $(r_1, \ldots, r_t) \in \{1\} \times \ldots \times \{1, \ldots, t\}$.
Vérifier que la loi conditionnelle de Θ sachant $\{R_1 = r_1, \ldots, R_t = r_t\}$ est la loi uniforme sur les permutations σ de $\mathcal{S}_N$ telles que $r_1^\sigma = r_1, \ldots, r_t^\sigma = r_t$.
En déduire que pour $f : \{1, \ldots, t\} \to \mathbb{R}$,

$$\mathbb{E}[f(\Theta_t) | R_1 = r_1, \ldots, R_t = r_t] = \frac{t!}{N!} \sum_{\sigma \in \mathcal{S}_N} \mathbf{1}_{\{r_1^\sigma = r_1, \ldots, r_t^\sigma = r_t\}} f(\sigma_t).$$

En déterminant de manière analogue la loi de Θ sachant $R_t = r_t$, montrer que

$$\mathbb{E}[f(\Theta_t) | R_t = r_t] = \frac{t}{N!} \sum_{\nu \in \mathcal{S}_N} \mathbf{1}_{\{r_t^\nu = r_t\}} f(\nu_t).$$

Remarquer qu'en permutant les $t-1$ premières valeurs d'une permutation $\sigma \in \mathcal{S}_N$ telle que $r_1^\sigma = r_1, \ldots, r_t^\sigma = r_t$, on obtient $(t-1)!$ permutations $\nu \in \mathcal{S}_N$ telles que $r_t^\nu = r_t$ et $\nu_t = \sigma_t$. Conclure que

$$\mathbb{E}[f(\Theta_t) | R_1 = r_1, \ldots, R_t = r_t] = \mathbb{E}[f(\Theta_t) | R_t = r_t].$$

Ainsi la loi conditionnelle de Θ_t sachant $R_1 = r_1, \ldots, R_t = r_t$ et la loi conditionnelle de Θ_t sachant $R_t = r_t$ sont égales.

5. En remarquant que pour $t \in \{1, \ldots, N-1\}$, $\mathbf{1}_{\{X_t \neq \Delta\}} A_t$ est fonction de $(R_1, \ldots, R_t)$, en déduire que le critère à minimiser se met aussi sous la forme
$$\mathbb{E}\left[\sum_{t=1}^{N-1} \varphi_t(X_t, A_t) + u_N(X_N)\right],$$
avec
$$u_N(x) = \begin{cases} 0 \text{ si } x = \Delta \\ \beta x - \gamma \mathbf{1}_{\{x=1\}} + \delta N \text{ sinon} \end{cases}$$
et $\varphi_t(x, a) = a f(t, x)$ pour
$$f(t, x) = \begin{cases} 0 \text{ si } x = \Delta \\ \delta t + \mathbb{E}[\beta \Theta_t - \gamma \mathbf{1}_{\{\Theta_t = 1\}} | R_t = x] \text{ sinon.} \end{cases}$$

6. Montrer que pour $r \in \{1, \ldots, t\}$,
$$\mathbb{P}(\Theta_t = s | R_t = r) = \begin{cases} 0 \text{ si } s < r \text{ ou } s > r + N - t \\ \dfrac{\binom{s-1}{r-1}\binom{N-s}{t-r}}{\binom{N}{t}} \text{ sinon.} \end{cases} \tag{3.12}$$
En déduire que
$$\sum_{k=r+1}^{r+1+N-t} \frac{\binom{k-1}{r}\binom{N+1-k}{t-r}}{\binom{N+1}{t+1}} = 1.$$
Remarquer que
$$f(t, r) = \delta t + \beta \frac{N+1}{t+1}\, r \sum_{s=r}^{r+N-t} \frac{\binom{s}{r}\binom{N-s}{t-r}}{\binom{N+1}{t+1}} - \gamma \mathbf{1}_{\{r=1\}} \frac{t}{N}$$
et conclure que $f(t, r) = \delta t + \beta \frac{N+1}{t+1} r - \gamma \mathbf{1}_{\{r=1\}} \frac{t}{N}$.

7. Montrer que pour tout $(x_1, \ldots, x_{t+1}, a_1, \ldots, a_t)$ dans $\{1\} \times \{\Delta, 1, 2\} \times \ldots \times \{\Delta, 1, \ldots, t+1\} \times \{0, 1\}^t$
$$\mathbb{P}(X_{t+1} = x_{t+1} | X_1 = x_1, A_1 = a_1, \ldots, X_t = x_t, A_t = a_t) = p_t((x_t, a_t), x_{t+1})$$
avec
$$p_t((x_t, a_t), x_{t+1}) = \begin{cases} \mathbf{1}_{\{x_{t+1} \neq \Delta\}} \mathbb{P}(R_{t+1} = x_{t+1}) \text{ si } x_t \neq \Delta \text{ et } a_t \neq 1 \\ \mathbf{1}_{\{x_{t+1} = \Delta\}} \text{ sinon.} \end{cases}$$

Dans le problème considéré ici, l'espace des actions possibles du recruteur est $\mathcal{A} = \{0, 1\}$ et le facteur d'actualisation vaut 1. La variable X_1 est à valeurs dans $\mathcal{E}_1 = \{1\}$ tandis que pour $t \in \{2, \ldots, N\}$, X_t prend ses valeurs dans $\mathcal{E}_t = \{\Delta, 1, \ldots, t\}$. Ainsi l'espace d'états dépend de t. Nous admettrons que

dans cette situation qui sort légèrement du modèle présenté au paragraphe 3.2.1, les équations d'optimalité s'écrivent pour $t \in \{1, \dots, N-1\}$,

$$\forall x_t \in \mathcal{E}_t,\ u_t(x_t) = \inf_{a \in \mathcal{A}} \sum_{x_{t+1} \in \mathcal{E}_{t+1}} p_t((x_t, a), x_{t+1}) \left[\varphi_t(x_t, a) + u_{t+1}(x_{t+1})\right].$$

8. Vérifier par récurrence descendante que pour $t \in \{2, \dots, N\}$, $u_t(\Delta) = 0$. En déduire que

$$\forall t \in \{1, \dots, N-1\},\ \forall x \in \{1, \dots, t\},\ u_t(x) = \min(f(t, x), \nu_t)$$

où $\nu_t = \mathbb{E}[u_{t+1}(R_{t+1})]$. Montrer que ν_t croît avec t.

9. Conclure à l'existence d'une suite unique

$$(r_1^*, \dots, r_{N-1}^*) \in \{0,1\} \times \{0,1,2\} \times \dots \times \{0,1,\dots,N-1\}$$

telle que la stratégie optimale du recruteur consiste à retenir le τ^*-ième candidat qui se présente où

$$\tau^* = \begin{cases} N \text{ si } \forall 1 \le t \le N-1,\ R_t > r_t^* \\ \min\{t : R_t \le r_t^*\} \text{ sinon.} \end{cases}$$

Vérifier que pour $t \in \{1, \dots, N-1\}$, si $r_t^* \le t-1$, alors $f(t, r_t^* + 1) > \nu_t$. Calculer $\mathbb{E}[u_N(R_N)] = \mathbb{E}[u_N(\Theta_N)]$ et en déduire que
- si $\gamma > N(\delta + \beta(N+1)(1/2 - 2/N))$, $r_{N-1}^* = 1$.
- si $\delta \ge \frac{\gamma}{N} + \beta(N+1)(1/2 - 1/N)$, $r_{N-1}^* = N-1$. Dans ce cas, montrer alors par récurrence que $\forall t \in \{1, \dots, N-1\},\ r_t^* = t$: comme le coût δ affecté à chaque entretien est trop important, la stratégie optimale consiste à choisir le premier candidat.

Montrer que pour $2 \le t \le N$, $\mathbb{P}(\tau^* \ge t) = \prod_{k=1}^{t-1} \left(1 - \frac{r_k^*}{k}\right)$. En déduire la loi de τ^* et vérifier que

$$\mathbb{E}[\tau^*] = \sum_{t=1}^{N} \mathbb{P}(\tau^* \ge t) = 1 + \sum_{t=2}^{N} \prod_{k=1}^{t-1} \left(1 - \frac{r_k^*}{k}\right).$$

En utilisant le résultat de la question 4, montrer que la probabilité $\mathbb{P}(\Theta_{\tau^*} = 1)$ d'obtenir le meilleur candidat est égale à

$$\sum_{t=1}^{N} \mathbf{1}_{\{r_t^* \ge 1\}} \mathbb{P}(\Theta_t = 1 | R_t = 1) \mathbb{P}(R_1 > r_1^*, \dots, R_{t-1} > r_{t-1}^*, R_t = 1).$$

En déduire que

$$\mathbb{P}(\Theta_{\tau^*} = 1) = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1}_{\{r_t^* \ge 1\}} \prod_{k=1}^{t-1} \left(1 - \frac{r_k^*}{k}\right).$$

où on adopte la convention $r_N^* = N$.

10. On s'intéresse à la monotonie de la suite $(r_1^*, \ldots, r_{N-1}^*)$.
   – Montrer que si $\delta = 0$, alors pour $t \in \{1, \ldots, N-1\}$ et $r \in \{1, \ldots, t\}$, $f(t+1, r) \leq f(t, r)$. En déduire que dans ce cas la suite $(r_1^*, \ldots, r_{N-1}^*)$ est croissante.
   – Soit $t \in \{1, \ldots, N-2\}$. Montrer que

$$\nu_t = \frac{t+1-r_{t+1}^*}{t+1}\nu_{t+1} + \delta r_{t+1}^* + \beta \frac{N+1}{(t+1)(t+2)} \frac{r_{t+1}^*(r_{t+1}^*+1)}{2} - \frac{\gamma}{N}\mathbf{1}_{\{r_{t+1}^* \geq 1\}}.$$

   Supposons que $r_{t+1}^* \leq t-1$ sans quoi nécessairement $r_t^* \leq r_{t+1}^*$. Vérifier en utilisant l'expression de ν_t ci-dessus que

$$f(t, r_{t+1}^*+1) - \nu_t + \frac{r_{t+1}^* - (t+1)}{t+1}(f(t+1, r_{t+1}^*+1) - \nu_{t+1}) = \beta \frac{(N+1)(r_{t+1}^*+1)(r_{t+1}^*+2)}{2(t+1)(t+2)} + \frac{\gamma}{N} - \delta.$$

   Conclure que si $\delta \leq \frac{(N+1)\beta}{N(N-1)} + \frac{\gamma}{N}$, la suite $(r_1^*, \ldots, r_{N-1}^*)$ est toujours croissante.

11. On se place dans le cas particulier $(\beta, \gamma, \delta) = (0, 1, 0)$ où le recruteur souhaite maximiser la probabilité de retenir le candidat le meilleur. Comme nous allons le voir, il est alors possible d'expliciter la stratégie optimale. Calculer ν_{N-1} et montrer la relation de récurrence

$$\forall t \in \{1, \ldots, N-2\},\ \nu_t = \begin{cases} \left(-\frac{1}{N} + \frac{t}{t+1}\nu_{t+1}\right) \text{ si } \nu_{t+1} \geq -\frac{t+1}{N} \\ \nu_{t+1} \text{ sinon.} \end{cases}$$

   En déduire que pour $t^*(N) = \min\{t \geq 1 :\ \frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{N-1} \leq 1\}$,

$$\forall t \in \{t^*(N)-1, \ldots, N-1\},\ \nu_t = -\frac{t}{N}\left(\frac{1}{t} + \frac{1}{t+1} + \ldots + \frac{1}{N-1}\right)$$
$$\text{et } \forall t \in \{1, \ldots, t^*(N)-2\},\ \nu_t = \nu_{t^*(N)-1}.$$

   Conclure que

$$r_t^* = \begin{cases} 0 \text{ si } t \leq t^*(N)-1 \\ 1 \text{ si } t^*(N) \leq t \leq N-1. \end{cases}$$

   Ainsi la stratégie optimale du recruteur consiste à observer les $t^*(N)-1$ premiers candidats qui se présentent puis à choisir ensuite tout candidat meilleur que ces $t^*(N)-1$ premiers. Si le meilleur des candidats figure dans les $t^*(N)-1$ premiers, il choisit le dernier candidat qu'il reçoit. Montrer que $\lim_{N\to+\infty} t^*(N)/N = \lim_{N\to+\infty} \mathbb{P}(\Theta_{\tau^*} = 1) = 1/e$.

12. Pour différentes valeurs de (β, γ, δ) dont $(1, 0, 0)$ et $(0, 1, 0)$
    - Déterminer la stratégie optimale en résolvant numériquement les équations d'optimalité. La suite $(r_1^*, \ldots, r_{N-1}^*)$ est-elle toujours croissante ?
    - Illustrer son caractère optimal en la comparant par la méthode de Monte-Carlo à d'autres stratégies.

♦

3.3 Résolution du problème dynamique de gestion de stock

D'après l'exemple 3.2.1, lorsque le produit se présente sous forme d'unités, le problème dynamique de gestion de stock décrit au paragraphe 3.1.2 entre dans le cadre du paragraphe précédent consacré au contrôle de chaînes de Markov. L'espace d'état est $\mathcal{E} = \mathbb{Z}$ pour le stock système X_t et l'ensemble d'actions $\mathcal{A} = \mathbb{N}$ pour la quantité Q_t de produit commandée. En outre, pour tout (t, x, q, y) dans $\{0, \ldots, N-1\} \times \mathbb{Z} \times \mathbb{N} \times \mathbb{Z}$,

$$\begin{cases} p_t((x,q),y) = \mathbb{P}(D_1 = x + q - y), \\ \varphi_t(x,q,y) = c_F \mathbf{1}_{\{q>0\}} + cq + c_S y^+ + c_M y^-. \end{cases}$$

Les équations d'optimalité (3.10) s'écrivent : pour $t \in \{0, \ldots, N-1\}$, $x \in \mathbb{Z}$

$$\begin{aligned} u_t(x) &= \inf_{q\in\mathbb{N}} \sum_{y\in\mathbb{Z}} p_t((x,q),y)\left[\varphi_t(x,q,y) + \alpha u_{t+1}(y)\right] \\ &= \inf_{q\in\mathbb{N}} \sum_{y\in\mathbb{Z}} \mathbb{P}(D_1 = x+q-y)\left[c_F \mathbf{1}_{\{q>0\}} + cq + c_S y^+ + c_M y^- + \alpha u_{t+1}(y)\right] \\ &= \inf_{q\in\mathbb{N}} \sum_{z\in\mathbb{Z}} \mathbb{P}(D_1 = z)[c_F \mathbf{1}_{\{q>0\}} + cq + c_S(x+q-z)^+ + c_M(x+q-z)^- \\ &\qquad + \alpha u_{t+1}(x+q-z)] \text{ en posant } z = x+q-y, \\ &= \inf_{q\in\mathbb{N}} \mathbb{E}[c_F \mathbf{1}_{\{q>0\}} + cq + c_S(x+q-D_1)^+ + c_M(x+q-D_1)^- \\ &\qquad + \alpha u_{t+1}(x+q-D_1)]. \end{aligned}$$

Nous allons étudier ces équations dans le cadre continu où le stock système est réel, les demandes sont des variables aléatoires positives non nécessairement entières et les quantités commandées sont des réels positifs, car leur résolution est plus simple que dans le cadre discret où nous les avons établies. Elles s'écrivent alors : pour $t \in \{0, \ldots, N-1\}$,

$$\begin{aligned} u_t(x) = \inf_{q\geq 0} \mathbb{E}[&c_F \mathbf{1}_{\{q>0\}} + cq + c_S(x+q-D_1)^+ \\ &+ c_M(x+q-D_1)^- + \alpha u_{t+1}(x+q-D_1)]. \end{aligned} \tag{3.13}$$

3.3.1 Gestion sans coût fixe d'approvisionnement

En plus de la nullité du coût fixe ($c_F = 0$), nous supposerons dans ce paragraphe que la fonction de coût terminale est $u_N(x) = -cx$, ce qui revient à associer une valeur unitaire c au stock système terminal X_N. Cette hypothèse simplificatrice qui va nous permettre d'obtenir une stratégie optimale stationnaire (i.e. telle que la règle de décision à l'instant t ne dépend pas de t) est discutable : si $X_N \leq 0$, il y a $-X_N$ manquants et il est naturel de leur associer le coût $-cX_N$ puisque c'est le prix à payer au fournisseur pour obtenir la quantité $-X_N$; en revanche affecter le coût $-cX_N$ et donc la valeur cX_N au stock physique résiduel $X_N \geq 0$ est moins naturel.
Commençons par déterminer la quantité de produit commandée optimale à l'instant $N-1$. Comme $u_N(x) = -cx$,

$$\begin{aligned} u_{N-1}(x) &= \alpha c\mathbb{E}[D_1] - cx + \inf_{q\geq 0}\Big((1-\alpha)c(x+q) + c_S\mathbb{E}[(x+q-D_1)^+] \\ &\qquad + c_M\mathbb{E}[(D_1-x-q)^+]\Big) \\ &= \alpha c\mathbb{E}[D_1] - cx + \inf_{q\geq 0} g_\alpha(x+q) \end{aligned} \tag{3.14}$$

où la fonction $g_\alpha(y) = (1-\alpha)cy + c_S\mathbb{E}[(y-D_1)^+] + c_M\mathbb{E}[(D_1-y)^+]$ est définie comme la fonction de coût moyen g du modèle étudié au paragraphe 3.1.1 (voir équation (3.1)) à ceci près que le coût unitaire c est remplacé par $\boxed{(1-\alpha)c}$. On se place dans le cas intéressant où $c_M > (1-\alpha)c > -c_S$. D'après l'analyse menée au paragraphe 3.1.1, la fonction g_α est continue, décroissante sur $]-\infty, S_\alpha]$ et croissante sur $[S_\alpha, +\infty[$ avec

$$S_\alpha = \inf\{z \geq 0 : F(z) \geq (c_M - (1-\alpha)c)/(c_M + c_S)\} \in \mathbb{R}_+,$$

où F est la fonction de répartition commune des demandes. La décision optimale à l'instant $N-1$ consiste donc à commander la quantité $(S_\alpha - x)^+$ (i.e. $(S_\alpha - x)$ si $x \leq S_\alpha$ et rien du tout sinon) si le stock système est x. Nous allons vérifier que cela reste vrai à tout instant t dans $\{0, \ldots, N-2\}$.

Théorème 3.3.1. *On suppose $c_F = 0$, $c_M > (1-\alpha)c > -c_S$ et $u_N(x) = -cx$. La stratégie avec stock objectif*

$$S_\alpha = \inf\{z \geq 0 : F(z) \geq (c_M - (1-\alpha)c)/(c_M + c_S)\},$$

qui consiste pour tout $t \in \{0, \ldots, N-1\}$ à commander la quantité $(S_\alpha - x)^+$ lorsque le stock système vaut x est optimale.

La figure 3.2 illustre l'optimalité de la stratégie de stock objectif $S_\alpha = 55$ dans le cas particulier où les demandes sont distribuées suivant la loi de Poisson de paramètre 50, $N = 10$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $c_F = 0$, $u_N(x) = -cx$ et le stock initial est $X_0 = 20$. Pour $z \in \{40, 41, \ldots, 70\}$ les

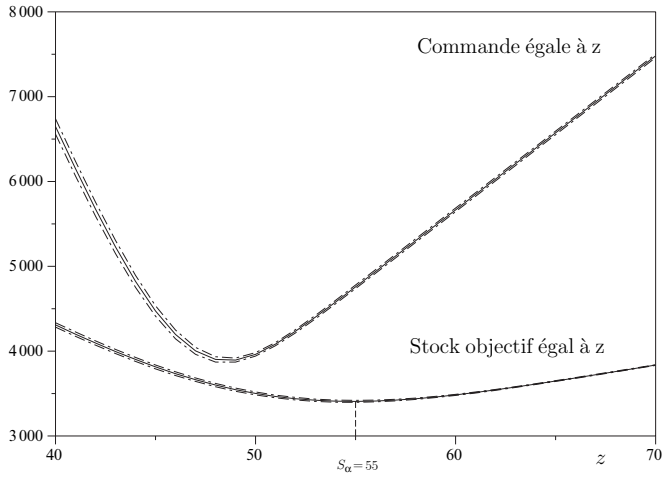

Fig. 3.2. Comparaison des coûts entre la stratégie de stock objectif z et la stratégie de commande constante égale à z ($N = 10$, stock initial $X_0 = 20$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $c_F = 0$, $u_N(x) = -cx$, demandes distribuées suivant la loi de Poisson de paramètre 50)

coûts moyens (3.8) associés à la stratégie qui consiste à commander z à chaque instant d'une part et à la stratégie de stock objectif z d'autre part ont été évalués en effectuant la moyenne empirique des coûts sur 1 000 réalisations indépendantes $(D_1^i, \ldots, D_{10}^i)_{1 \le i \le 1\,000}$ des demandes. Les mêmes réalisations de ces variables ont été utilisées pour chacune des stratégies. Plus précisément, le coût moyen est approché par

$$\frac{1}{1\,000} \sum_{i=1}^{1\,000} \left[\sum_{t=0}^{9} (0.9)^t \left(10 Q_t^i + 5 (X_{t+1}^i)^+ + 20 (X_{t+1}^i)^- \right) - 10 (0.9)^{10} X_N^i \right]$$

où pour $i \in \{1, \ldots, 1\,000\}$, $X_0^i = 20$ et pour $t \in \{0, \ldots, 9\}$, $Q_t^i = z$ et $X_{t+1}^i = X_t^i + z - D_{t+1}^i$ dans la stratégie qui consiste à commander z à chaque instant et $Q_t^i = (z - X_t^i)^+$ et $X_{t+1}^i = \max(X_t^i, z) - D_{t+1}^i$ dans la stratégie de stock objectif z. Les courbes en pointillés représentent les bornes des intervalles de confiance à 95 % obtenus pour chacun des coûts moyens. On observe bien que parmi les stratégies considérées, le coût moyen minimal est obtenu pour la stratégie de stock objectif $S_\alpha = 55$.

Démonstration. Comme la quantité q optimale dans (3.14) est $(S_\alpha - x)^+$, on a

$$u_{N-1}(x) = \begin{cases} \alpha c \mathbb{E}[D_1] - cx + g_\alpha(S_\alpha) \ \text{ si } \ x \le S_\alpha \\ \alpha c \mathbb{E}[D_1] - cx + g_\alpha(x) \ \text{ sinon.} \end{cases}$$

Ainsi $u_{N-1}(x) = K_{N-1} - cx + w_{N-1}(x)$ où $K_{N-1} = \alpha c \mathbb{E}[D_1] + g_\alpha(S_\alpha)$ est une constante et $w_{N-1}(x) = \mathbf{1}_{\{x \ge S_\alpha\}}(g_\alpha(x) - g_\alpha(S_\alpha))$ est une fonction croissante et nulle sur $]-\infty, S_\alpha]$.

Nous allons vérifier que cette écriture se généralise aux instants antérieurs en démontrant par récurrence descendante que pour tout t dans $\{0,\dots,N-1\}$, la décision optimale à l'instant t consiste à commander $(S_\alpha - x)^+$ si le stock système est x et que $u_t(x) = K_t - cx + w_t(x)$ où K_t est une constante et w_t est une fonction croissante et nulle sur $]-\infty, S_\alpha]$.

Supposons que l'hypothèse de récurrence est vérifiée au rang $t+1$. En insérant l'égalité $u_{t+1}(x) = K_{t+1} - cx + w_{t+1}(x)$ dans l'équation (3.13) avec $c_F = 0$, on obtient

$$\begin{aligned} u_t(x) = \inf_{q\geq 0} \mathbb{E}\Big[& cq + c_S(x+q-D_1)^+ + c_M(x+q-D_1)^- \\ & + \alpha(K_{t+1} - c(x+q-D_1) + w_{t+1}(x+q-D_1))\Big] \\ = \alpha(K_{t+1} + c\mathbb{E}[D_1]) - cx + & \inf_{q\geq 0} \left(g_\alpha(x+q) + \alpha\mathbb{E}[w_{t+1}(x+q-D_1)]\right). \end{aligned}$$

Pour analyser la minimisation du dernier terme, on remarque que

- la fonction $g_\alpha(y)$ est croissante sur $[S_\alpha, +\infty[$ et atteint son minimum pour $y = S_\alpha$;
- la fonction $y \to \mathbb{E}[w_{t+1}(y - D_1)]$ est croissante par croissance de w_{t+1} ;
- cette fonction est nulle pour $y \leq S_\alpha$ puisque comme $D_1 \geq 0$, $y - D_1$ est alors dans l'ensemble $]-\infty, S_\alpha]$ où w_{t+1} s'annule.

On en déduit que l'infimum est atteint pour $q = (S_\alpha - x)^+$ et que $u_t(x) = K_t - cx + w_t(x)$ avec

$$\begin{cases} K_t = \alpha(K_{t+1} + c\mathbb{E}[D_1]) + g_\alpha(S_\alpha) \\ w_t(x) = \mathbf{1}_{\{x\geq S_\alpha\}}(g_\alpha(x) - g_\alpha(S_\alpha) + \alpha\mathbb{E}[w_{t+1}(x - D_1)]). \end{cases}$$

La fonction w_t est clairement croissante et nulle sur $]-\infty, S_\alpha]$. □

Remarque 3.3.2. Notons que si le produit se présente sous forme d'unités, les demandes des clients sont entières et on peut vérifier comme dans la démonstration de la proposition 3.1.1 que le stock objectif S_α est entier. À l'instant t, le stock système X_t est entier et il est possible pour le gestionnaire de commander la quantité $(S_\alpha - X_t)^+$ qui est un entier positif. Cette stratégie est optimale car elle l'est pour le modèle où le produit se présente sous forme continue qui offre plus d'opportunités en termes de quantité commandée. ◊

3.3.2 Gestion avec coût fixe

Nous supposons maintenant qu'en plus du coût unitaire c, toute commande supporte un coût fixe $c_F \geq 0$ et nous nous plaçons dans le cas intéressant où $c_M > (1-\alpha)c > -c_S$. Les équations d'optimalité (3.13) s'écrivent alors : pour $t \in \{0,\dots,N-1\}$,

$$u_t(x) = -cx + \inf_{q\geq 0}\left(c_F\mathbf{1}_{\{q>0\}} + g(x+q) + \alpha\mathbb{E}[u_{t+1}(x+q-D_1)]\right)$$
$$= -cx + \min\left(f_t(x), c_F + \inf_{q>0} f_t(x+q)\right)$$

où la fonction $g(y) = cy + c_S\mathbb{E}[(y-D_1)^+] + c_M\mathbb{E}[(D_1-y)^+]$ est la fonction de coût du modèle à une période de temps du paragraphe 3.1.1 et $f_t(y) = g(y) + \alpha\mathbb{E}[u_{t+1}(y-D_1)]$.

Afin de préciser les hypothèses faites sur la fonction de coût terminale u_N, nous introduisons la notion de C-convexité qui a été utilisée pour la première fois par Scarf [9] en 1960 :

Définition 3.3.3. *Pour $C \geq 0$, une fonction $u : \mathbb{R} \to \mathbb{R}$ est dite C-convexe si*

$$\forall x \leq y,\ \forall \beta \in [0,1],\ u(\beta x + (1-\beta)y) \leq \beta u(x) + (1-\beta)(u(y) + C).$$

La notion de convexité usuelle correspond à la 0-convexité et pour $0 \leq C' \leq C$, toute fonction C'-convexe est C-convexe.

Exemple 3.3.4. Si $C \geq 0$ et f est une fonction convexe continue sur $\mathbb{R}$ qui atteint son minimum f^* en S, alors la fonction u définie par

$$u(x) = \begin{cases} f^* + C \text{ si } x \leq s \\ f(x) \text{ si } x > s \end{cases} \qquad \text{où } s = \sup\{z \leq S : f(z) \geq f^* + C\}$$

avec la convention $\sup\emptyset = -\infty$, est C-convexe d'après le lemme 3.3.9 énoncé plus loin. La figure 3.3 illustre cette construction. ◊

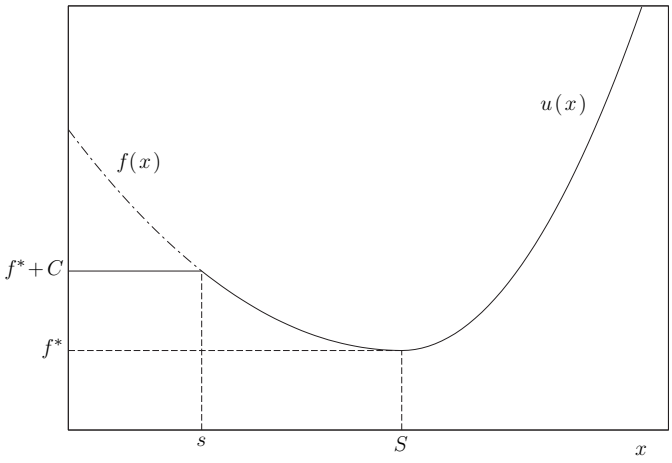

Fig. 3.3. En trait plein : exemple de fonction C-convexe u obtenue par la construction de l'exemple 3.3.4 à partir d'une fonction f convexe continue atteignant son minimum f^*en S ($s = \sup\{z \leq S : f(z) \geq f^* + C\}$)

Remarque 3.3.5. Dans le cas où $c_M \geq -c_S$, la fonction g définie par (3.1) est convexe d'après la remarque 3.1.3. Elle atteint son minimum en S d'après la proposition 3.1.1. Et le stock seuil s est défini comme $\sup\{z \leq S : g(z) \geq g(S) + c_F\}$ d'après la proposition 3.1.5. D'après l'exemple 3.3.4, la fonction $(g(S)+c_F)\mathbf{1}_{\{x\leq s\}}+g(x)\mathbf{1}_{\{x>s\}}$ est c_F-convexe. Et la fonction de coût minimale $u(x) = -cx+(g(S)+c_F)\mathbf{1}_{\{x\leq s\}}+g(x)\mathbf{1}_{\{x>s\}}$ obtenue dans la remarque 3.1.6 pour le modèle à une période de temps avec coût fixe $c_F > 0$ est également c_F-convexe. Il n'est donc pas étonnant que la notion de c_F-convexité intervienne dans la résolution des équations d'optimalité pour le modèle à plusieurs périodes de temps. ◊

Le résultat que nous allons démontrer est une généralisation de celui obtenu dans le paragraphe 3.1.1 pour le modèle à une période avec coût fixe :

Théorème 3.3.6. *On suppose que $c_M > (1-\alpha)c > -c_S$ et que $u_N(x)$ est une fonction continue, c_F-convexe, minorée par $K_N - cx$ où $K_N \in \mathbb{R}$ et vérifiant $|u_N(x)| \leq \eta_N + \gamma_N|x|$ où $\eta_N, \gamma_N \geq 0$. Alors il existe une stratégie optimale dont la règle de décision à chaque instant $t \in \{0,\ldots,N-1\}$ est du type (s_t, S_t) i.e. consiste à commander $\mathbf{1}_{\{x\leq s_t\}}(S_t - x)^+$ si le stock système vaut x.*

Remarque 3.3.7. La fonction de coût terminale $u_N(x) = -cx$ introduite au paragraphe précédent est continue, vérifie l'hypothèse de minoration pour $K_N = 0$ et celle de domination pour $\eta_N = 0$ et $\gamma_N = c$. Enfin, elle est linéaire donc convexe et à fortiori c_F-convexe. Ainsi elle vérifie les hypothèses du théorème. La fonction de coût $u_N(x) = cx^- + c_D x^+$ où $c_D \geq 0$ s'interprète comme le coût supporté par le gestionnaire pour se débarrasser d'une unité de stock résiduel à l'instant terminal N vérifie également les hypothèses. ◊

Démonstration. Le principe de la démonstration est le suivant :
- nous allons vérifier par récurrence descendante que $\forall t \in \{0,\ldots,N\}$ la fonction $u_t(x)$ est continue, c_F-convexe, minorée par $K_t - cx$ où $K_t \in \mathbb{R}$,
- en vérifiant que pour $t \in \{0,\ldots,N-1\}$, l'hypothèse de récurrence au rang $t+1$ implique l'hypothèse de récurrence au rang t, nous montrerons qu'il existe un couple (s_t, S_t) tel que la règle de décision (s_t, S_t) est optimale à l'instant t.

L'hypothèse de récurrence est clairement vérifiée au rang N. Supposons-la vérifiée au rang $t+1$ avec $t \in \{0,\ldots,N-1\}$. Il est facile d'en déduire que $y \to \alpha\mathbb{E}[u_{t+1}(y - D_1)]$ est αc_F-convexe et donc c_F-convexe ($\alpha \in [0,1]$). La continuité de cette fonction se déduit de celle de u_{t+1} en utilisant le théorème de convergence dominée et une majoration technique du type $|u_{t+1}(y)| \leq \eta_{t+1} + \gamma_{t+1}|y|$ avec $\eta_{t+1}, \gamma_{t+1} \geq 0$ que nous ne démontrerons pas ici (mais qui s'obtient également par récurrence descendante sur t). Comme la fonction g est continue et convexe d'après la remarque 3.1.3, la fonction

$$\boxed{f_t(y) = g(y) + \alpha\mathbb{E}[u_{t+1}(y - D_1)]}$$

est c_F-convexe continue.

Par hypothèse de récurrence, $u_{t+1}(y-D_1) \geq K_{t+1} - cy + cD_1$. Donc $f_t(y)$ est minoré par

$$(1-\alpha)cy + c_S\mathbb{E}[(y-D_1)^+] + c_M\mathbb{E}[(D_1-y)^+] + \alpha(K_{t+1} + c\mathbb{E}[D_1])$$
$$= ((1-\alpha)c + c_S)y + (c_S + c_M)\mathbb{E}[(D_1-y)^+] + \alpha K_{t+1} + (\alpha c - c_S)\mathbb{E}[D_1].$$

En remarquant que pour $y \leq 0$, le second membre est égal à $((1-\alpha)c - c_M)y + \alpha K_{t+1} + (\alpha c + c_M)\mathbb{E}[D_1]$, on déduit de l'inégalité $c_M > (1-\alpha)c > -c_S$ que $\lim_{|y|\to+\infty} f_t(y) = +\infty$.
D'où l'existence de (s_t, S_t) avec

$$\boxed{f_t(S_t) = \inf_{y\in\mathbb{R}} f_t(y) \ \text{ et } \ s_t = \sup\{z \leq S_t,\ f_t(z) \geq c_F + \inf_y f_t(y)\}.}$$

Posons $f_t^* = \inf_y f_t(y)$. Par continuité de f_t, $f_t(s_t) = c_F + f_t^*$. À l'instant t, comme

$$u_t(x) = -cx + \min\left(f_t(x), c_F + \inf_{q>0} f_t(x+q)\right),$$

le gestionnaire doit choisir la commande $q \geq 0$ qui minimise $c_F\mathbf{1}_{\{q>0\}} + f_t(x+q)$. Montrons que la règle de décision (s_t, S_t) est optimale. Pour cela, on distingue trois situations

cas 1 : $x \leq s_t$. Lorsque $c_F = 0$, $s_t = S_t$ et commander $S_t - x$ est optimal. On suppose donc $c_F > 0$. Alors $s_t \in [x, S_t[$ i.e. $s_t = \beta x + (1-\beta)S_t$ pour $\beta \in]0,1]$ et par c_F-convexité de f_t,

$$f_t(s_t) \leq \beta f_t(x) + (1-\beta)(f_t^* + c_F).$$

Comme $f_t(s_t) = f_t^* + c_F$, on en déduit que

$$f_t(x) \geq c_F + f_t^* = c_F + \inf_{q>0} f_t(x+q)$$

et il est optimal de commander $S_t - x$.

cas 2 : $s_t \leq x < S_t$. Par définition de s_t, $f_t(x) \leq c_F + f_t^* = c_F + \inf_{q>0} f_t(x+q)$ et il est optimal de ne rien commander.

cas 3 : $x \geq S_t$. Pour $q \geq 0$, comme $x \in [S_t, x+q]$, il existe $\beta \in [0,1]$ tel que $x = \beta S_t + (1-\beta)(x+q)$. Par c_F-convexité de f_t, puis en utilisant $f_t^* = \inf_y f_t(y)$,

$$f_t(x) \leq \beta f_t^* + (1-\beta)(f_t(x+q) + c_F) \leq f_t(x+q) + (1-\beta)c_F \leq f_t(x+q) + c_F.$$

On en déduit que $f_t(x) \leq c_F + \inf_{q>0} f_t(x+q)$ et qu'il est optimal de ne rien commander.

En conséquence,

$$\boxed{u_t(x) = \begin{cases} -cx + f_t^* + c_F \ \text{ si } \ x \leq s_t \\ -cx + f_t(x) \ \text{ si } \ x > s_t \end{cases},}$$

et cette fonction est continue et minorée par $K_t - cx$ pour $K_t = f_t^*$. Le lemme suivant assure que la fonction $\mathbf{1}_{\{x \le s_t\}}(f_t^* + c_F) + \mathbf{1}_{\{x > s_t\}} f_t(x)$ est c_F-convexe. On en déduit facilement que u_t qui s'obtient en ajoutant à cette fonction la fonction linéaire $-cx$ est c_F-convexe, ce qui achève la démonstration. □

Remarque 3.3.8. Lorsque les demandes des clients $D_1, \ldots, D_N$ ne sont pas identiquement distribuées mais restent indépendantes et intégrables, on peut vérifier que les équations d'optimalité s'écrivent

$$u_t(x) = -cx + \inf_{q \ge 0} \left(c_F \mathbf{1}_{\{q>0\}} + g_t(x+q) + \alpha \mathbb{E}[u_{t+1}(x+q-D_{t+1})] \right),$$

où $g_t(y) = cy + c_S \mathbb{E}\left[(y - D_{t+1})^+\right] + c_M \mathbb{E}\left[(D_{t+1} - y)^+\right]$. La démonstration précédente permet toujours de conclure à l'existence d'une stratégie optimale de la forme (s_t, S_t). ◊

Lemme 3.3.9. *Soit $C \ge 0$ et $f : \mathbb{R} \to \mathbb{R}$ une fonction C-convexe continue qui atteint son minimum f^* en S. On suppose que l'ensemble $\{z \le S,\ f(z) \ge f^* + C\}$ est non vide et on note s sa borne supérieure. Alors la fonction $u(x) = \mathbf{1}_{\{x \le s\}}(f^* + C) + \mathbf{1}_{\{x > s\}} f(x)$ est C-convexe.*

Démonstration. La fonction u est C-convexe sur chacun des intervalles $]-\infty, s]$ et $[s, +\infty[$ car constante sur le premier et égale à f sur le second. Pour conclure, il suffit donc de montrer l'inégalité de C-convexité (voir définition 3.3.3) pour $x \le s \le y$. Pour ce faire, on distingue deux cas :

cas 1 : $\beta \in [0,1]$ est t.q. $\beta x + (1-\beta)y \le s$. Alors,

$$u(\beta x + (1-\beta)y) = f^* + C = \beta(f^* + C) + (1-\beta)(f^* + C),$$

et comme $u(x) = f^* + C$ et $u(y) = f(y) \ge f^*$, l'inégalité de C-convexité est vérifiée.

cas 2 : $\beta x + (1-\beta)y \ge s$. Alors pour $\tilde{\beta} = \beta(y-x)/(y-s) \in [\beta, 1]$, $\beta x + (1-\beta)y = \tilde{\beta} s + (1-\tilde{\beta})y$. En utilisant la C-convexité de f puis l'égalité $f(s) = f^* + C$, on a

$$\begin{aligned} u(\beta x + (1-\beta)y) &= f(\tilde{\beta} s + (1-\tilde{\beta})y) \le \tilde{\beta} f(s) + (1-\tilde{\beta})(f(y) + C) \\ &= \beta(f^* + C) + (1-\beta)(f(y) + C) + (\beta - \tilde{\beta})(f(y) - f^*) \\ &= \beta u(x) + (1-\beta)(u(y) + C) + (\beta - \tilde{\beta})(f(y) - f^*) \\ &\le \beta u(x) + (1-\beta)(u(y) + C), \end{aligned}$$

puisque $(\beta - \tilde{\beta})(f(y) - f^*) \le 0$.

□

Dans le cas particulier où $N = 10$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $u_N(x) = -cx$ et les demandes sont distribuées suivant la loi de Poisson de paramètre 50, les Tableaux 3.1 et 3.2 fournissent respectivement les valeurs de la suite $(S_t,\ 0 \le t \le 9)$ des stocks objectifs et de

Tableau 3.1. Suite des stocks objectifs en fonction de c_F dans le cas $N = 10$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $u_N(x) = -cx$ et demandes distribuées suivant la loi de Poisson de paramètre 50

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
$c_F = 0$	55	55	55	55	55	55	55	55	55	55
$c_F = 100$	55	55	55	55	55	55	55	55	55	55
$c_F = 200$	55	55	55	55	55	55	55	55	55	55
$c_F = 300$	55	55	55	55	55	55	55	55	55	55
$c_F = 400$	100	100	100	100	100	100	100	100	100	55
$c_F = 500$	100	100	100	100	100	100	100	100	100	55
$c_F = 600$	100	100	100	100	100	100	100	101	100	55
$c_F = 700$	100	100	100	100	100	101	100	139	100	55

Tableau 3.2. Suite des stocks seuils en fonction de c_F dans le cas $N = 10$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $u_N(x) = -cx$ et demandes distribuées suivant la loi de Poisson de paramètre 50

	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
$c_F = 0$	55	55	55	55	55	55	55	55	55	55
$c_F = 100$	42	42	42	42	42	42	42	42	42	42
$c_F = 200$	36	36	36	36	36	36	36	36	36	36
$c_F = 300$	31	31	31	31	31	31	31	31	31	31
$c_F = 400$	28	27	29	27	29	26	29	26	30	26
$c_F = 500$	27	23	27	22	28	22	28	21	29	20
$c_F = 600$	25	19	25	19	26	18	27	17	28	15
$c_F = 700$	21	17	22	17	22	16	23	15	28	10

la suite $(s_t,\ 0 \leq t \leq 9)$ des stocks seuils pour c_F parcourant l'ensemble $\{0, 100, 200, 300, 400, 500, 600, 700\}$.

Ces suites ont été calculées en effectuant par récurrence descendante sur t l'ensemble des étapes données par les formules encadrées dans la démonstration du théorème 3.3.6.

Plus précisément, lors de l'implémentation informatique, il faut borner l'ensemble des valeurs du stock système pour lesquelles on effectue les calculs. À cet effet, on se donne deux entiers relatifs $x_{\min} < x_{\max}$ et on note $\mathcal{D} = \{x_{\min}, x_{\min} + 1, \ldots, x_{\max}\}$. On commence par calculer $(g(x),\ x \in \mathcal{D})$ en utilisant par exemple la formule (3.5). Puis partant de $(u_N(x) = -cx,\ x \in \mathcal{D})$, on effectue par récurrence descendante sur $t \in \{N-1, \ldots, 0\}$ l'ensemble des étapes suivantes :

- calcul de $f_t(x) = g(x) + \alpha \sum_{k=0}^{K} u_{t+1}(\max(x-k, x_{\min}))\mathbb{P}(D_1 = k)$ pour $x \in \mathcal{D}$ où K est fixé de façon à ce que $\mathbb{P}(D_1 \leq K)$ soit suffisamment proche de 1 (pour D_1 distribuée suivant la loi de Poisson de paramètre 50, le choix $K = 80$ assure $\mathbb{P}(D_1 \leq K) \simeq 1 - 3.4 \times 10^{-5}$),
- détermination de $S_t \in \mathcal{D}$ tel que $f_t(S_t) = \min_{x \in \mathcal{D}} f_t(x)$,

- calcul de $s_t = \max\{x \in \{x_{\min}, \dots, S_t\} : f_t(x) \geq f_t(S_t) + c_F\}$ avec la convention $\max \emptyset = x_{\min} - 1$,
- calcul de $u_t(x) = -cx + \mathbf{1}_{\{x \leq s_t\}}(f_t(S_t) + c_F) + \mathbf{1}_{\{x > s_t\}} f_t(x)$ pour $x \in \mathcal{D}$.

Pour mesurer les effets des réductions de domaine effectuées à la fois pour les valeurs du stock système (choix de $x_{\min}$ et de $x_{\max}$) et pour les valeurs des demandes (choix de K), il convient de regarder si diminuer $x_{\min}$ en augmentant simultanément $x_{\max}$ et K modifie les valeurs des stocks objectifs et des stocks seuils calculées par l'algorithme. Sur notre exemple, nous avons obtenu les résultats qui figurent dans les tableaux 3.1 à la fois pour $(x_{\min}, x_{\max}, K) = (-700, 300, 80)$ et pour $(x_{\min}, x_{\max}, K) = (-3\,000, 1\,000, 100)$.
Ces résultats appellent quelques commentaires. Tout d'abord, en absence de coût fixe ($c_F = 0$), on a $s_t = S_t = 55$ pour tout t ce qui signifie que la stratégie optimale consiste à chaque instant à commander $(55 - x)^+$ lorsque le stock système est égal à x. On retrouve bien la stratégie optimale donnée par le théorème 3.3.1 puisque pour le choix des paramètres considéré, $S_\alpha = 55$ (voir Fig. 3.2). Ensuite, il est normal que la valeur de S_9 ne dépende pas du coût fixe c_F puisque S_9 s'obtient comme réalisant le minimum de f_9, fonction qui s'écrit à partir de u_{10} et de g qui ne dépendent pas de c_F. Enfin, lorsque l'on augmente le niveau du coût fixe c_F, on distingue deux régimes :

- dans un premier temps ($c_F \leq 300$) le stock objectif S reste égal à 55 tandis que le stock seuil s, constant en temps, diminue. La quantité commandée minimale (lors d'une commande effective) qui est égale à la différence $S - s$ augmente donc pour compenser l'augmentation du coût fixe.
- dans un second temps ($c_F \geq 400$), le stock objectif passe à $100 = 2\mathbb{E}[D_1]$ sauf au dernier instant tandis que la moyenne temporelle du stock seuil continue à diminuer. On peut interpréter le doublement approximatif du stock objectif de la façon suivante : lorsque le gestionnaire décide de passer une commande auprès de son fournisseur, il approvisionne suffisamment de stock pour faire face aux demandes des clients sur deux périodes et non plus sur une seule période comme pour des valeurs plus faibles du coût fixe.

3.3.3 Délai de livraison

Nous introduisons la généralisation très utile en pratique qui consiste à supposer que les commandes effectuées par le gestionnaire auprès de son fournisseur lui sont livrées avec un délai de $d \in \mathbb{N}$ périodes de temps : plus précisément la quantité Q_t commandée en $t \in \{0, \dots, N-1\}$ est livrée sur la période $[t+d, t+d+1]$. Jusqu'à présent, nous nous sommes intéressé au cas $d = 0$. Nous supposons également que pour $t \in \{0, \dots, N+d-1\}$, les clients formulent la demande D_{t+1} sur la période $[t, t+1]$ avec $D_1, \dots, D_{N+d}$ des variables aléatoires positives intégrables indépendantes et de fonction de répartition commune F.

Pour $t \in \{0, \ldots, N+d\}$, on note Y_t la quantité égale au stock physique moins les manquants à l'instant t. Et pour $t \in \{0, \ldots, N\}$, on définit le **stock système** X_t en t comme la somme de Y_t et de l'**en-commande** c'est-à-dire la quantité commandée $Q_{t-d} + Q_{t-d+1} + \ldots + Q_{t-1}$ qui se trouve en attente de livraison :

$$X_t = Y_t + Q_{t-d} + Q_{t-d+1} + \ldots + Q_{t-1}. \tag{3.15}$$

Pour $t \in \{0, \ldots, N-1\}$, entre t et $t+1$, l'en-commande varie de $Q_t - Q_{t-d}$ (quantité commandée en t moins quantité livrée sur la période) tandis que Y_t varie de $Q_{t-d} - D_{t+1}$ (quantité livrée moins demandes des clients sur la période) si bien que

$$X_{t+1} = X_t + (Q_t - Q_{t-d}) + (Q_{t-d} - D_{t+1}) = X_t + Q_t - D_{t+1}.$$

Ainsi l'équation (3.7) donnant l'évolution du stock système en absence de délai de livraison est préservée.
Le stock physique et les manquants en $t+d+1$ sont donnés par Y^+_{t+d+1} et Y^-_{t+d+1} si bien que le coût sur la période $[t+d, t+d+1]$ induit par la commande de Q_t en $t \in \{0, \ldots, N-1\}$ est donné par

$$c_F \mathbf{1}_{\{Q_t>0\}} + cQ_t + c_S Y^+_{t+d+1} + c_M Y^-_{t+d+1}.$$

En ajoutant un coût final donné par $u_{N+d}(Y_{N+d})$, on obtient que l'espérance du coût total avec facteur d'actualisation α est donnée par

$$\mathbb{E}\left[\sum_{t=0}^{N-1} \alpha^t \left(c_F \mathbf{1}_{\{Q_t>0\}} + cQ_t + c_S Y^+_{t+d+1} + c_M Y^-_{t+d+1}\right) + \alpha^N u_{N+d}(Y_{N+d})\right].$$

Pour $t \in \{0, \ldots, N-1\}$, la quantité stock physique moins manquants en $t+d+1$ s'obtient en ajoutant à cette quantité en t les commandes $Q_{t-d}, \ldots, Q_t$ qui sont livrées par le fournisseur sur la période $[t, t+d+1]$ et en y retranchant les demandes $D_{t+1}, \ldots, D_{t+d+1}$ qui sont formulées par les clients sur la même période :

$$\begin{aligned} Y_{t+d+1} &= Y_t + Q_{t-d} + \ldots + Q_{t-1} + Q_t - D_{t+1} - \ldots - D_{t+d+1} \\ &= X_t + Q_t - D_{t+1} - \ldots - D_{t+d+1}, \text{ d'après (3.15).} \end{aligned}$$

Comme X_t et Q_t ne dépendent que des demandes $D_1, D_2, \ldots, D_t$ jusqu'à l'instant t qui sont indépendantes des demandes ultérieures $D_{t+1}, \ldots, D_{N+d}$, le couple (X_t, Q_t) est indépendant du vecteur aléatoire $(D_{t+1}, \ldots, D_{t+d+1})$.

D'après la proposition A.1.21, on en déduit que

$$\mathbb{E}\left[c_F\mathbf{1}_{\{Q_t>0\}} + cQ_t + c_S Y_{t+d+1}^+ + c_M Y_{t+d+1}^-\right] = \mathbb{E}[\varphi(X_t, Q_t)],$$

$$\text{où } \varphi(x,q) = \mathbb{E}\Big[c_F\mathbf{1}_{\{q>0\}} + cq + c_S(x+q-D_{t+1}-\ldots-D_{t+d+1})^+ + c_M(x+q-D_{t+1}-\ldots-D_{t+d+1})^-\Big].$$

De manière analogue, $Y_{N+d} = X_N - D_{N+1} - \ldots - D_{N+d}$ où la variable aléatoire X_N est indépendante du vecteur aléatoire $(D_{N+1}, \ldots, D_{N+d})$, ce qui entraîne que $\mathbb{E}[u_{N+d}(Y_{N+d})] = \mathbb{E}[u_N(X_N)]$ où

$$u_N(x) = \mathbb{E}[u_{N+d}(x - D_{N+1} - \ldots - D_{N+d})].$$

On en déduit que l'espérance du coût total actualisé s'écrit

$$\mathbb{E}\left[\sum_{t=0}^{N-1} \alpha^t \varphi(X_t, Q_t) + \alpha^N u_N(X_N)\right].$$

Comme les demandes sont identiquement distribuées, les équations d'optimalité s'écrivent : pour $t \in \{0, \ldots, N-1\}$,

$$\begin{aligned} u_t(x) &= \inf_{q\geq 0}\Big(\varphi(x,q) + \alpha\mathbb{E}[u_{t+1}(x+q-D_1)]\Big) \\ &= \inf_{q\geq 0}\mathbb{E}\Big(c_F\mathbf{1}_{\{q>0\}} + cq + c_S(x+q-D_1-\ldots-D_{d+1})^+ \\ &\qquad + c_M(x+q-D_1-\ldots-D_{d+1})^- + \alpha u_{t+1}(x+q-D_1)\Big). \end{aligned}$$

Par rapport aux équations d'optimalité (3.13), les termes $(x+q-D_1)^+$ et $(x+q-D_1)^-$ sont respectivement remplacés ici par $(x+q-D_1-\ldots-D_{d+1})^+$ et $(x+q-D_1-\ldots-D_{d+1})^-$, ce qui traduit le fait que les premières conséquences en termes de coût de la décision de commander la quantité q ont lieu $d+1$ périodes plus tard i.e. lorsque les clients ont exprimé leurs demandes sur $d+1$ périodes. Le terme $\mathbb{E}[u_{t+1}(x+q-D_1)]$ est inchangé car la dynamique du stock système l'est.
Lorsque la fonction $u_{N+d}(x)$ est continue, c_F-convexe, minorée par $K-cx$ où $K \in \mathbb{R}$ et vérifie $|u_{N+d}(x)| \leq \eta + \gamma|x|$ (avec $\eta, \gamma \geq 0$) il est facile de vérifier que u_N vérifie les mêmes propriétés. Lorsque $u_{N+d}(x) = -cx$, $u_N(x) = -cx + cd\,\mathbb{E}[D_1]$ où la constante $cd\,\mathbb{E}[D_1]$ ne change rien à la stratégie optimale. En reprenant l'approche et les démonstrations des deux paragraphes précédents, on obtient :

Théorème 3.3.10. *On suppose $c_M > (1-\alpha)c > -c_S$.*
Si $u_{N+d}(x)$ est continue, minorée par $K-cx$ où $K \in \mathbb{R}$, c_F-convexe et vérifie

$|u_{N+d}(x)| \leq \eta + \gamma|x|$ avec $\eta, \gamma \geq 0$, alors il existe une stratégie optimale dont la règle de décision à l'instant $t \in \{0, \ldots, N-1\}$ est de la forme (s_t, S_t) i.e. consiste à commander $\mathbf{1}_{\{x \leq s_t\}}(S_t - x)^+$ si le stock système à l'instant t est x. Dans le cas particulier où $c_F = 0$ et $u_{N+d}(x) = -cx$, la stratégie avec stock objectif $S_{\alpha,d}$ où

$$\begin{cases} S_{\alpha,d} = \inf\{z \geq 0 : \ F_{d+1}(z) \geq (c_M - (1-\alpha)c)/(c_M + c_S)\} \\ \text{avec } F_{d+1}(z) = \mathbb{P}(D_1 + \ldots D_{d+1} \leq z) \end{cases}$$

qui consiste à commander $(S_{\alpha,d} - x)^+$ lorsque le stock système vaut x est optimale.

La figure 3.4 illustre l'optimalité de la stratégie de stock objectif $S_{\alpha,1} = 107$ dans le cas particulier où le délai de livraison est $d = 1$, les demandes sont distribuées suivant la loi de Poisson de paramètre 50, $N = 10$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $c_F = 0$, $f(x) = -cx$ et le stock initial est $X_0 = 50$. Pour $z \in \{90, 91, \ldots, 120\}$ le coût moyen associé à la stratégie de stock objectif z a été évalué en effectuant la moyenne empirique des coûts sur 1 000 réalisations des demandes $(D_1, \ldots, D_{11})$. Les mêmes réalisations de ces variables ont été utilisées pour chacune des stratégies. Les courbes en pointillés représentent les bornes des intervalles de confiance à 95 % obtenus pour chacun des coûts moyens.

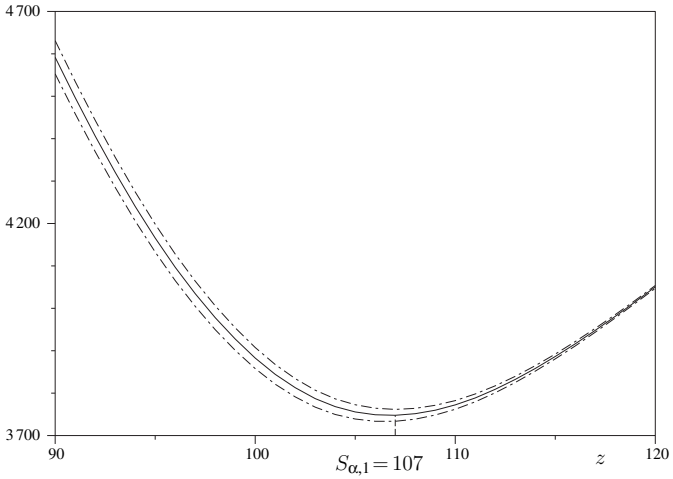

Fig. 3.4. Coût associé à la stratégie de stock objectif z pour un délai de livraison $d = 1$ ($N = 10$, stock initial $X_0 = 50$, $\alpha = 0.9$, $c = 10$, $c_M = 20$, $c_S = 5$, $c_F = 0$, $u_{N+d}(x) = -cx$, demandes distribuées suivant la loi de Poisson de paramètre 50)

Remarque 3.3.11. Dans le cas où les demandes sont distribuées suivant la loi de Poisson de paramètre μ (resp. la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$), $D_1 + \ldots + D_{d+1}$

suit la loi de Poisson de paramètre $(d+1)\mu$ (resp. la loi normale $\mathcal{N}((d+1)\mu, (d+1)\sigma^2)$) et F_{d+1} est la fonction de répartition de cette loi. $\Diamond$

3.4 Conclusion

Dans ce chapitre nous avons mis en évidence l'intérêt des règles de décision de type (s, S) pour la gestion dynamique du stock d'un produit (pièce de rechange par exemple). En effet, nous avons montré l'optimalité d'une stratégie composée de règles de décision markoviennes de ce type à chaque période parmi toutes les stratégies composées de règles de décision fonctions de tout le passé. Ce résultat reste valable pour des modèles plus généraux que celui que nous avons étudié : fonctions de coût de surplus et de coût de manquants convexes (elles sont supposées linéaires ici), demandes indépendantes mais non identiquement distribuées, fonctions de coût dépendant du temps, coûts fixes dépendant du temps sous réserve que $\alpha c_F(t+1) \leq c_F(t)$, etc. Nous renvoyons à la présentation de Porteus [7] ou au livre de Liu et Esogbue [6] pour la description de ces modèles généraux. L'optimalité d'une stratégie (s, S) reste également valable pour le problème de gestion de stock à horizon temporel infini, auquel est consacré le Chap. 13 du livre de Whittle [11].
D'un point de vue pratique, on doit identifier la loi de la demande. À cet effet, il est par exemple possible d'effectuer un traitement statistique des demandes passées. On peut ensuite déterminer les (s_t, S_t) en résolvant par récurrence descendante les équations d'optimalité.

Références

1. R. Bellman. *Dynamic programming*. Princeton University Press, Princeton, N.J., 1957.
2. D.P. Bertsekas. *Dynamic programming and stochastic control*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. Mathematics in Science and Engineering, 125.
3. D.P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. 1 et 2*. Athena Scientific, 1995.
4. D.P. Bertsekas et S.E. Shreve. *Stochastic optimal control*, volume 139 de *Mathematics in Science and Engineering*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1978. The discrete time case.
5. D. Lamberton et B. Lapeyre. *Introduction au calcul stochastique appliqué à la finance*. Ellipses Édition Marketing, Paris, seconde édition, 1997.
6. B. Liu et A.O. Esogbue. *Decision criteria and optimal inventory processes*. International Series in Operations Research & Management Science, 20. Kluwer Academic Publishers, Boston, MA, 1999.
7. E.L. Porteus. Stochastic inventory theory. In *Stochastic models*, volume 2 de *Handbooks Oper. Res. Management Sci.*, pages 605–652. North-Holland, Amsterdam, 1990.

8. M.L. Puterman. *Markov decision processes : discrete stochastic dynamic programming*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1994.
9. H. Scarf. The optimality of (s, S) policies in the dynamic inventory problem. In S.U. Press, editor, *Proceeding of the 1959 Stanford Symposium on Mathematical Methods in the Social Sciences*, pages 196–202, 1960.
10. D.J. White. *Markov decision processes*. John Wiley & Sons Ltd., Chichester, 1993.
11. P. Whittle. *Optimization over time. Vol. I*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1982. Dynamic programming and stochastic control.