PRÉDICTION, APPRENTISSAGE ET THÉORIE DES JEUX

... ou comment vendre sur Internet sans étude de marché préalable

Gilles STOLTZ

En collaboration avec

Gábor LUGOSI, Universitat Pompeu Fabra, Barcelone

Nicolò CESA-BIANCHI, Università di Milano

Ajustement dynamique du prix de vente (1)

Un vendeur vend n exemplaires d’un même produit à n clients.

Ajustement dynamique du prix de vente (1)

Un vendeur vend n exemplaires d'un même produit à n clients.

Ses clients viennent l'un après l'autre, et le t-ième client se voit proposer le prix $p_t \in [0, 1]$. Il a en tête un prix maximal $y_t \in [0, 1]$ qu'il est prêt à payer, mais qu'il ne dévoile pas au vendeur.

Ajustement dynamique du prix de vente (1)

Un vendeur vend n exemplaires d'un même produit à n clients.

Ses clients viennent l'un après l'autre, et le t-ième client se voit proposer le prix $p_t \in [0, 1]$. Il a en tête un prix maximal $y_t \in [0, 1]$ qu'il est prêt à payer, mais qu'il ne dévoile pas au vendeur.

Lorsque $y_t \geq p_t$, le client achète le produit et le vendeur encourt une "perte" $y_t - p_t$.

Ajustement dynamique du prix de vente (1)

Un vendeur vend n exemplaires d'un même produit à n clients.

Ses clients viennent l'un après l'autre, et le t-ième client se voit proposer le prix $p_t \in [0, 1]$. Il a en tête un prix maximal $y_t \in [0, 1]$ qu'il est prêt à payer, mais qu'il ne dévoile pas au vendeur.

Lorsque $y_t \geq p_t$, le client achète le produit et le vendeur encourt une "perte" $y_t - p_t$.

Lorsque $y_t < p_t$, le client n'achète pas et la perte du vendeur est de $c \in [0, 1]$.

Ajustement dynamique du prix de vente (1)

Un vendeur vend n exemplaires d'un même produit à n clients.

Ses clients viennent l'un après l'autre, et le t-ième client se voit proposer le prix $p_t \in [0, 1]$. Il a en tête un prix maximal $y_t \in [0, 1]$ qu'il est prêt à payer, mais qu'il ne dévoile pas au vendeur.

Lorsque $y_t \geq p_t$, le client achète le produit et le vendeur encourt une "perte" $y_t - p_t$.

Lorsque $y_t < p_t$, le client n'achète pas et la perte du vendeur est de $c \in [0, 1]$.

On définit ainsi une fonction de perte

$$\ell(p_t, y_t) = (y_t - p_t)\mathbb{I}_{y_t \geq p_t} + c\mathbb{I}_{y_t > p_t}$$

et une fonction de feedback

$$h(p_t, y_t) = \mathbb{I}_{y_t \geq p_t} \ .$$

Ajustement dynamique du prix de vente (2)

Les valeurs y_t sont arbitraires (et sont même autorisées à dépendre des actions passées du vendeur). On peut ainsi tenir compte de l'influence de notre politique tarifaire et publicitaire sur les échelles de prix des clients (voir EasyJet).

Ajustement dynamique du prix de vente (2)

Les valeurs y_t sont arbitraires (et sont même autorisées à dépendre des actions passées du vendeur). On peut ainsi tenir compte de l'influence de notre politique tarifaire et publicitaire sur les échelles de prix des clients (voir EasyJet).

Notre stratégie sera dynamique, et nous nous comparerons à tous ceux de nos concurrents qui utilisent des stratégies classiques, avec études de marché.

Ajustement dynamique du prix de vente (2)

Les valeurs y_t sont arbitraires (et sont même autorisées à dépendre des actions passées du vendeur). On peut ainsi tenir compte de l'influence de notre politique tarifaire et publicitaire sur les échelles de prix des clients (voir EasyJet).

Notre stratégie sera dynamique, et nous nous comparerons à tous ceux de nos concurrents qui utilisent des stratégies classiques, avec études de marché.

Ces dernières fournissent un prix (constant) $p \in [0, 1]$, et c'est pourquoi nous comparons notre perte cumulée à celle du meilleur prix constant :

Ajustement dynamique du prix de vente (2)

Les valeurs y_t sont arbitraires (et sont même autorisées à dépendre des actions passées du vendeur). On peut ainsi tenir compte de l'influence de notre politique tarifaire et publicitaire sur les échelles de prix des clients (voir EasyJet).

Notre stratégie sera dynamique, et nous nous comparerons à tous ceux de nos concurrents qui utilisent des stratégies classiques, avec études de marché.

Ces dernières fournissent un prix (constant) $p \in [0, 1]$, et c'est pourquoi nous comparons notre perte cumulée à celle du meilleur prix constant :

$$\frac{1}{n} \sum_{t=1}^{n} \ell(p_t, y_t) - \min_{p \in [0,1]} \frac{1}{n} \sum_{t=1}^{n} \ell(p, y_t)$$

Ajustement dynamique du prix de vente (2)

Les valeurs y_t sont arbitraires (et sont même autorisées à dépendre des actions passées du vendeur). On peut ainsi tenir compte de l'influence de notre politique tarifaire et publicitaire sur les échelles de prix des clients (voir EasyJet).

Notre stratégie sera dynamique, et nous nous comparerons à tous ceux de nos concurrents qui utilisent des stratégies classiques, avec études de marché.

Ces dernières fournissent un prix (constant) $p \in [0, 1]$, et c'est pourquoi nous comparons notre perte cumulée à celle du meilleur prix constant :

$$\frac{1}{n}\sum_{t=1}^{n} \ell(p_t, y_t) - \min_{p \in [0,1]} \frac{1}{n}\sum_{t=1}^{n} \ell(p, y_t)$$

Remarque : nous cherchons des bornes indépendantes de la stratégie de l'adversaire.

Prédiction séquentielle (1)

Soit $\mathcal{X}$ un ensemble (fini) de prédictions, $\mathcal{Y}$ un ensemble (fini) d'observations.

Prédiction séquentielle (1)

Soit $\mathcal{X}$ un ensemble (fini) de prédictions, $\mathcal{Y}$ un ensemble (fini) d'observations.

On introduit deux fonctions,

1. une fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to [0, 1]$,
2. une fonction de feed-back $h : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}$,

où $\mathcal{S}$ est un ensemble fini de signaux (encodés par exemple dans $[-1, 1]$).

Prédiction séquentielle (1)

Soit $\mathcal{X}$ un ensemble (fini) de prédictions, $\mathcal{Y}$ un ensemble (fini) d'observations.

On introduit deux fonctions,

1. une fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to [0, 1]$,
2. une fonction de feed-back $h : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}$,

où $\mathcal{S}$ est un ensemble fini de signaux (encodés par exemple dans $[-1, 1]$).

Après avoir formé sa prédiction $I_t \in \mathcal{X}$, le statisticien reçoit pour information le *feed-back* $h(I_t, y_t)$.

Prédiction séquentielle (1)

Soit $\mathcal{X}$ un ensemble (fini) de prédictions, $\mathcal{Y}$ un ensemble (fini) d'observations.

On introduit deux fonctions,

1. une fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to [0,1]$,
2. une fonction de feed-back $h : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}$,

où $\mathcal{S}$ est un ensemble fini de signaux (encodés par exemple dans $[-1,1]$).

Après avoir formé sa prédiction $I_t \in \mathcal{X}$, le statisticien reçoit pour information le *feed-back* $h(I_t, y_t)$.

Son but est de former ses prédictions I_t de telle sorte que le regret

$$\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{x \in \mathcal{X}} \frac{1}{n}\sum_{t=1}^{n} \ell(x, y_t) = o(1)$$

Prédiction séquentielle (2)

A chaque tour $t = 1, 2, \ldots$,

(1) l'adversaire choisit l'observation suivante, $y_t \in \mathcal{Y}$, sans la dévoiler;

Prédiction séquentielle (2)

A chaque tour $t = 1, 2, \ldots$,

(1) l'adversaire choisit l'observation suivante, $y_t \in \mathcal{Y}$, sans la dévoiler;

(2) le statisticien choisit une probabilité $\mathbf{p}_t$ sur les actions, puis tire son action $I_t \in \mathcal{X}$ selon $\mathbf{p}_t$;

Prédiction séquentielle (2)

A chaque tour $t = 1, 2, \ldots$,

(1) l'adversaire choisit l'observation suivante, $y_t \in \mathcal{Y}$, sans la dévoiler;

(2) le statisticien choisit une probabilité $\mathbf{p}_t$ sur les actions, puis tire son action $I_t \in \mathcal{X}$ selon $\mathbf{p}_t$;

(3) le statisticien et chaque action x encourent des pertes respectives $\ell(I_t, y_t)$ et $\ell(x, y_t)$, mais ces valeurs demeurent secrètes;

Prédiction séquentielle (2)

A chaque tour $t = 1, 2, \ldots$,

(1) l'adversaire choisit l'observation suivante, $y_t \in \mathcal{Y}$, sans la dévoiler;

(2) le statisticien choisit une probabilité $\mathbf{p}_t$ sur les actions, puis tire son action $I_t \in \mathcal{X}$ selon $\mathbf{p}_t$;

(3) le statisticien et chaque action x encourent des pertes respectives $\ell(I_t, y_t)$ et $\ell(x, y_t)$, mais ces valeurs demeurent secrètes;

(4) le feed-back $h(I_t, y_t)$ est dévoilé au statisticien.

Prédiction séquentielle (2)

A chaque tour $t = 1, 2, \ldots$,

(1) l'adversaire choisit l'observation suivante, $y_t \in \mathcal{Y}$, sans la dévoiler;

(2) le statisticien choisit une probabilité $\mathbf{p}_t$ sur les actions, puis tire son action $I_t \in \mathcal{X}$ selon $\mathbf{p}_t$;

(3) le statisticien et chaque action x encourent des pertes respectives $\ell(I_t, y_t)$ et $\ell(x, y_t)$, mais ces valeurs demeurent secrètes;

(4) le feed-back $h(I_t, y_t)$ est dévoilé au statisticien.

La randomisation auxiliaire est rendue nécessaire par le cadre "worst case bound" en les stratégies de l'adversaire – qui équivaut à supposer que notre adversaire est le diable et qu'il sait lire dans nos pensées...

Exemples

Problème des bandits : La seule information qui soit retournée au statisticien est sa propre perte (ou son propre gain) : $h = \ell$.

Exemples

Problème des bandits : La seule information qui soit retournée au statisticien est sa propre perte (ou son propre gain) : $h = \ell$.

Dégustation de pommes : $\mathcal{X} = \mathcal{Y} = \{0, 1\}$,

$$\mathbf{L} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} b & c \\ a & a \end{bmatrix}$$

Le statisticien ne reçoit de nouvelles informations que lorsqu'il choisit la première action.

Exemples

Problème des bandits : La seule information qui soit retournée au statisticien est sa propre perte (ou son propre gain) : $h = \ell$.

Dégustation de pommes : $\mathcal{X} = \mathcal{Y} = \{0, 1\}$,

$$\mathbf{L} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} b & c \\ a & a \end{bmatrix}$$

Le statisticien ne reçoit de nouvelles informations que lorsqu'il choisit la première action.

Désormais, $\mathcal{X}$ et $\mathcal{Y}$ sont finis, et on considère les matrices $\mathbf{L} = [\ell(x, y)]_{(x,y)}$ et $\mathbf{H} = [h(x, y)]_{(x,y)}$ en lieu et place des fonctions ℓ et h.

Une stratégie générale (1)

Paramètres : $\mathbf{L}$ et $\mathbf{H}$ sont connues.

Une stratégie générale (1)

Paramètres : $\mathbf{L}$ et $\mathbf{H}$ sont connues.

Initialisation : $\widetilde{L}_{1,0} = \cdots = \widetilde{L}_{N,0} = 0$.

A chaque tour de jeu $t = 1, 2, \ldots$,

Une stratégie générale (1)

Paramètres : $\mathbf{L}$ et $\mathbf{H}$ sont connues.

Initialisation : $\widetilde{L}_{1,0} = \cdots = \widetilde{L}_{N,0} = 0$.

A chaque tour de jeu $t = 1, 2, \ldots$,

(1) $\eta_t \sim t^{-2/3}$ et $\gamma_t \sim t^{-1/3}$;

Une stratégie générale (1)

Paramètres : $\mathbf{L}$ et $\mathbf{H}$ sont connues.

Initialisation : $\widetilde{L}_{1,0} = \cdots = \widetilde{L}_{N,0} = 0$.

A chaque tour de jeu $t = 1, 2, \ldots$,

(1) $\eta_t \sim t^{-2/3}$ et $\gamma_t \sim t^{-1/3}$;

(2) tirer au hasard une action I_t dans $\mathcal{X}$, selon la probabilité $\mathbf{p}_t$ définie par

$$p_{x,t} = (1-\gamma_t)\frac{e^{-\eta_t \widetilde{L}_{x,t-1}}}{\sum_{z\in\mathcal{X}} e^{-\eta_t \widetilde{L}_{z,t-1}}} + \frac{\gamma_t}{N} \ ;$$

Une stratégie générale (1)

Paramètres : $\mathbf{L}$ et $\mathbf{H}$ sont connues.

Initialisation : $\widetilde{L}_{1,0} = \cdots = \widetilde{L}_{N,0} = 0$.

A chaque tour de jeu $t = 1, 2, \ldots$,

(1) $\eta_t \sim t^{-2/3}$ et $\gamma_t \sim t^{-1/3}$;

(2) tirer au hasard une action I_t dans $\mathcal{X}$, selon la probabilité $\mathbf{p}_t$ définie par

$$p_{x,t} = (1 - \gamma_t) \frac{e^{-\eta_t \widetilde{L}_{x,t-1}}}{\sum_{z \in \mathcal{X}} e^{-\eta_t \widetilde{L}_{z,t-1}}} + \frac{\gamma_t}{N} \ ;$$

(3) soit $\widetilde{L}_{x,t} = \widetilde{L}_{x,t-1} + \widetilde{\ell}(x, y_t)$ pour tout $x \in \mathcal{X}$.

Une stratégie générale (1)

Paramètres : $\mathbf{L}$ et $\mathbf{H}$ sont connues.

Initialisation : $\widetilde{L}_{1,0} = \cdots = \widetilde{L}_{N,0} = 0$.

A chaque tour de jeu $t = 1, 2, \ldots$,

(1) $\eta_t \sim t^{-2/3}$ et $\gamma_t \sim t^{-1/3}$;

(2) tirer au hasard une action I_t dans $\mathcal{X}$, selon la probabilité $\mathbf{p}_t$ définie par

$$p_{x,t} = (1-\gamma_t) \frac{e^{-\eta_t \widetilde{L}_{x,t-1}}}{\sum_{z \in \mathcal{X}} e^{-\eta_t \widetilde{L}_{z,t-1}}} + \frac{\gamma_t}{N} \; ;$$

(3) soit $\widetilde{L}_{x,t} = \widetilde{L}_{x,t-1} + \widetilde{\ell}(x, y_t)$ pour tout $x \in \mathcal{X}$.

Il ne reste plus qu'à voir comment définir ces estimateurs $\widetilde{\ell}(x, y_t)$ des pertes $\ell(x, y_t)$...

Une stratégie générale (2)

Inspirée de Piccolboni & Schindelhauer ’01.

Hypothèse essentielle : il existe un encodage de $\mathbf{H}$ tel qu’il existe une matrice $\mathbf{K} = [k(x,z)]$ vérifiant $\mathbf{L} = \mathbf{K} \cdot \mathbf{H}$.

Une stratégie générale (2)

Inspirée de Piccolboni & Schindelhauer ’01.

Hypothèse essentielle : il existe un encodage de $\mathbf{H}$ tel qu’il existe une matrice $\mathbf{K} = [k(x,z)]$ vérifiant $\mathbf{L} = \mathbf{K} \cdot \mathbf{H}$.

Ainsi,

$$\ell(x,y) = \sum_{z \in \mathcal{X}} k(x,z) h(z,y)$$

Une stratégie générale (2)

Inspirée de Piccolboni & Schindelhauer '01.

Hypothèse essentielle : il existe un encodage de $\mathbf{H}$ tel qu'il existe une matrice $\mathbf{K} = [k(x,z)]$ vérifiant $\mathbf{L} = \mathbf{K} \cdot \mathbf{H}$.

Ainsi,

$$\ell(x,y) = \sum_{z \in \mathcal{X}} k(x,z) h(z,y)$$

Les pertes sont alors estimées par

$$\widetilde{\ell}(x, y_t) = \frac{k(x, I_t) h(I_t, y_t)}{p_{I_t,t}}$$

Une stratégie générale (2)

Inspirée de Piccolboni & Schindelhauer '01.

Hypothèse essentielle : il existe un encodage de $\mathbf{H}$ tel qu'il existe une matrice $\mathbf{K} = [k(x,z)]$ vérifiant $\mathbf{L} = \mathbf{K} \cdot \mathbf{H}$.

Ainsi,

$$\ell(x,y) = \sum_{z \in \mathcal{X}} k(x,z) h(z,y)$$

Les pertes sont alors estimées par

$$\widetilde{\ell}(x,y_t) = \frac{k(x,I_t)h(I_t,y_t)}{p_{I_t,t}}$$

En effet, la loi de I_t est $\mathbf{p}_t$, et

$$\mathbb{E}\left[\widetilde{\ell}(x,y_t) \,|\, I_1,\ldots,I_{t-1}\right] = \sum_{z \in \mathcal{X}} p_{z,t} \frac{k(x,z)h(z,y_t)}{p_{z,t}} = \ell(x,y_t)$$

Borne supérieure sur le regret

Théorème : Pour toute stratégie de l'adversaire, pour tout $n \geq 1$,

Borne supérieure sur le regret

Théorème : Pour toute stratégie de l'adversaire, pour tout $n \geq 1$, pour tout $\delta \in [0,1]$, et avec probabilité au moins $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,y_t) - \min_{x\in\mathcal{X}}\frac{1}{n}\sum_{t=1}^{n}\ell(x,y_t) \leqslant C n^{-1/3} N^{2/3}\sqrt{\ln(N/\delta)}\ ,$$

où C dépend de $\mathbf{K}$.

Borne supérieure sur le regret

Théorème : Pour toute stratégie de l'adversaire, pour tout $n \geq 1$, pour tout $\delta \in [0,1]$, et avec probabilité au moins $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{x \in \mathcal{X}} \frac{1}{n}\sum_{t=1}^{n} \ell(x, y_t) \leqslant C n^{-1/3} N^{2/3} \sqrt{\ln(N/\delta)}\ ,$$

où C dépend de $\mathbf{K}$.

Application : Pour l'ajustement dynamique du prix, on divise le temps en les périodes $[2^{r-1}, 2^r - 1]$, $r = 1, 2, \ldots$, de longueur $n_r = 2^{r-1}$. On relance l'algorithme précédent au début de chaque période r,

Borne supérieure sur le regret

Théorème : Pour toute stratégie de l'adversaire, pour tout $n \geq 1$, pour tout $\delta \in [0,1]$, et avec probabilité au moins $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,y_t) - \min_{x\in\mathcal{X}}\frac{1}{n}\sum_{t=1}^{n}\ell(x,y_t) \leqslant Cn^{-1/3}N^{2/3}\sqrt{\ln(N/\delta)}\ ,$$

où C dépend de $\mathbf{K}$.

Application : Pour l'ajustement dynamique du prix, on divise le temps en les périodes $[2^{r-1}, 2^r - 1]$, $r = 1, 2, \ldots$, de longueur $n_r = 2^{r-1}$. On relance l'algorithme précédent au début de chaque période r, en discrétisant $[0,1]$ par $\{0, 1/N_r, 2/N_r, \ldots, 1\}$, où $N_r \sim (2^r)^{1/5}$.

Borne supérieure sur le regret

Théorème : Pour toute stratégie de l'adversaire, pour tout $n \geq 1$, pour tout $\delta \in [0,1]$, et avec probabilité au moins $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{x \in \mathcal{X}} \frac{1}{n}\sum_{t=1}^{n} \ell(x, y_t) \leqslant C n^{-1/3} N^{2/3} \sqrt{\ln(N/\delta)} \,,$$

où C dépend de $\mathbf{K}$.

Application : Pour l'ajustement dynamique du prix, on divise le temps en les périodes $[2^{r-1}, 2^r - 1]$, $r = 1, 2, \ldots$, de longueur $n_r = 2^{r-1}$. On relance l'algorithme précédent au début de chaque période r, en discrétisant $[0,1]$ par $\{0, 1/N_r, 2/N_r, \ldots, 1\}$, où $N_r \sim (2^r)^{1/5}$.

L'effet de la discrétisation est de l'ordre de $\sum_r 1/N_r \sim n^{-1/5}$, et la borne proposée par le théorème est de l'ordre de $\sum_r n_r^{-1/3} N_r^{2/3} \sim n^{-1/5}$.

Techniques de preuve

Par des techniques de concentration de la mesure (inégalité de Bernstein pour les accroissements de martingales), on prouve que

$$\left(\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{n} \ell(x, y_t) \right) = \left(\sum_{t=1}^{n} \widetilde{\ell}(\mathbf{p}_t, y_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{n} \widetilde{\ell}(x, y_t) \right) + O_P(n^{2/3})$$

Techniques de preuve

Par des techniques de concentration de la mesure (inégalité de Bernstein pour les accroissements de martingales), on prouve que

$$\left(\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{n} \ell(x, y_t) \right) = \left(\sum_{t=1}^{n} \widetilde{\ell}(\mathbf{p}_t, y_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{n} \widetilde{\ell}(x, y_t) \right) + O_P(n^{2/3})$$

Ensuite, on remarque que $\eta_t \, \widetilde{\ell}(x, y_t) \leqslant 1$ par construction, et ainsi, en utilisant

$$e^x \leqslant 1 + x + x^2 \quad \text{pour } x \leqslant 1 \, ,$$

Techniques de preuve

Par des techniques de concentration de la mesure (inégalité de Bernstein pour les accroissements de martingales), on prouve que

$$\left(\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{n} \ell(x, y_t) \right) = \left(\sum_{t=1}^{n} \widetilde{\ell}(\mathbf{p}_t, y_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{n} \widetilde{\ell}(x, y_t) \right) + O_P(n^{2/3})$$

Ensuite, on remarque que $\eta_t \, \widetilde{\ell}(x, y_t) \leqslant 1$ par construction, et ainsi, en utilisant

$$e^x \leqslant 1 + x + x^2 \quad \text{pour } x \leqslant 1 \ ,$$

le second membre est essentiellement borné par

$$\frac{\ln N}{\eta_{n+1}} + \sum_{t=1}^{n} \eta_t \sum_{x \in \mathcal{X}} p_{x,t} \mathbb{E}\left[\widetilde{\ell}(x, y_t)^2 \,|\, I_1, \ldots, I_{t-1} \right] + O_P(n^{2/3})$$

Optimalité de la stratégie (1)

Considérons le problème de prédiction avec un nombre limité d'observations (il existe bien $\mathbf{K}$ tel que $\mathbf{L} = \mathbf{K} \cdot \mathbf{H}$) :

$$\mathbf{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix}$$

Optimalité de la stratégie (1)

Considérons le problème de prédiction avec un nombre limité d'observations (il existe bien $\mathbf{K}$ tel que $\mathbf{L} = \mathbf{K} \cdot \mathbf{H}$) :

$$\mathbf{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix}$$

Théorème : Pour tout $n \geqslant 8$ et pour toute stratégie (randomisée), il existe une suite d'observations $y_1, \ldots, y_n$ telle que

$$\mathbb{E}_A \left[\frac{1}{n} \sum_{t=1}^{n} \ell(I_t, y_t) \right] - \min_{i=1,\ldots,N} \frac{1}{n} \sum_{t=1}^{n} \ell(i, y_t) \geqslant \frac{n^{-1/3}}{5} ,$$

où $\mathbb{E}_A$ désigne l'espérance par rapport à la randomisation auxiliaire éventuellement utilisée par le statisticien.

Optimalité de la stratégie (2)

Idée de la preuve : on construit une probabilité sur les observations possibles, ces dernières sont considérées comme tirées aléatoirement par l'adversaire, et sont notées $Y_1, Y_2, \ldots$. Alors,

$$\sup_{y_1^n} \left(\mathbb{E}_{\mathbb{A}} \left[\sum_{t=1}^{n} \ell(I_t, y_t) \right] - \min_{j=1,2,3} \sum_{t=1}^{n} \ell(j, y_t) \right) \geqslant \mathbb{E} \left[\mathbb{E}_{\mathbb{A}} \left[\sum_{t=1}^{n} \ell(I_t, Y_t) \right] - \min_{j=1,2,3} \sum_{t=1}^{n} \ell(j, Y_t) \right]$$

Optimalité de la stratégie (2)

Idée de la preuve : on construit une probabilité sur les observations possibles, ces dernières sont considérées comme tirées aléatoirement par l'adversaire, et sont notées $Y_1, Y_2, \ldots$. Alors,

$$\sup_{y_1^n} \left(\mathbb{E}_{\mathbb{A}} \left[\sum_{t=1}^n \ell(I_t, y_t) \right] - \min_{j=1,2,3} \sum_{t=1}^n \ell(j, y_t) \right)$$
$$\geqslant \mathbb{E} \left[\mathbb{E}_{\mathbb{A}} \left[\sum_{t=1}^n \ell(I_t, Y_t) \right] - \min_{j=1,2,3} \sum_{t=1}^n \ell(j, Y_t) \right]$$

L'inégalité de Pinsker permet de minorer le membre de droite par

$$m_1 + n\varepsilon \left(\frac{1}{2} - \varepsilon \sqrt{3 m_1} \right) ,$$

pour tout $0 \leqslant \varepsilon \leqslant 1/\sqrt{6}$, où m_1 désigne le nombre moyen de fois où l'action 1 est jouée.

Optimalité de la stratégie (3)

Une modification simple d'un argument introduit par Piccolboni & Schindelhauer '01 conduit à l'alternative suivante.

Optimalité de la stratégie (3)

Une modification simple d'un argument introduit par Piccolboni & Schindelhauer '01 conduit à l'alternative suivante.

Théorème : Tout problème de prédiction $(\mathbf{H}, \mathbf{L})$

- soit se ramène à un problème $(\mathbf{H}', \mathbf{L}')$ pas plus difficile, avec $\mathbf{L}' = \mathbf{K}'\mathbf{L}'$, de sorte que le regret est $O(n^{2/3})$ avec la stratégie générale,
- soit est tel qu'aucune stratégie n'encourt un regret uniformément en $o(n)$.

Optimalité de la stratégie (3)

Une modification simple d'un argument introduit par Piccolboni & Schindelhauer '01 conduit à l'alternative suivante.

Théorème : Tout problème de prédiction $(\mathbf{H}, \mathbf{L})$

- soit se ramène à un problème $(\mathbf{H}', \mathbf{L}')$ pas plus difficile, avec $\mathbf{L}' = \mathbf{K}'\mathbf{L}'$, de sorte que le regret est $O(n^{2/3})$ avec la stratégie générale,
- soit est tel qu'aucune stratégie n'encourt un regret uniformément en $o(n)$.

Donc, en général, notre algorithme et la vitesse $n^{2/3}$, quand elle peut être atteinte, sont optimaux.

Optimalité de la stratégie (4)

Cependant, dans certains cas particuliers où la quantité d’information dans les feed-back est plus importante, on peut atteindre des vitesses plus rapides.

Optimalité de la stratégie (4)

Cependant, dans certains cas particuliers où la quantité d'information dans les feed-back est plus importante, on peut atteindre des vitesses plus rapides.

Information complète : Lorsque $h(x, y) = y$, il n'y a pas besoin d'estimer les pertes $\ell(x, y_t)$, et la stratégie dite de pondération exponentielle atteint une vitesse (optimale) de $\sqrt{n}$ (Hannan '56, Blackwell '57).

Optimalité de la stratégie (4)

Cependant, dans certains cas particuliers où la quantité d'information dans les feed-back est plus importante, on peut atteindre des vitesses plus rapides.

Information complète : Lorsque $h(x, y) = y$, il n'y a pas besoin d'estimer les pertes $\ell(x, y_t)$, et la stratégie dite de pondération exponentielle atteint une vitesse (optimale) de $\sqrt{n}$ (Hannan '56, Blackwell '57).

Problème des bandits : Lorsque $h = \ell$, une variante de la stratégie générale atteint une vitesse (optimale) de $\sqrt{n}$ pour la convergence du regret (Auer, Cesa-Bianchi, Freund et Schapire '02).

Optimalité de la stratégie (4)

Cependant, dans certains cas particuliers où la quantité d'information dans les feed-back est plus importante, on peut atteindre des vitesses plus rapides.

Information complète : Lorsque $h(x,y) = y$, il n'y a pas besoin d'estimer les pertes $\ell(x, y_t)$, et la stratégie dite de pondération exponentielle atteint une vitesse (optimale) de $\sqrt{n}$ (Hannan '56, Blackwell '57).

Problème des bandits : Lorsque $h = \ell$, une variante de la stratégie générale atteint une vitesse (optimale) de $\sqrt{n}$ pour la convergence du regret (Auer, Cesa-Bianchi, Freund et Schapire '02).

Problèmes ouverts : Caractériser la classe des problèmes $(\mathbf{H}, \mathbf{L})$ qui permettent l'obtention de vitesses plus rapides que $n^{2/3}$. Des vitesses optimales autres que $\sqrt{n}$ ou $n^{2/3}$ peuvent-elles être atteintes ?