BORNES INFÉRIEURES EN PRÉDICTION AVEC AVIS D'EXPERTS

Gilles STOLTZ, Université Paris-Sud

www.math.u-psud.fr/∼stoltz

Directeur de thèse : Gábor LUGOSI, Universitat Pompeu Fabra, Barcelone

www.econ.upf.es/∼lugosi

Plan de l’exposé

1. Prédire une suite individuelle avec avis d’experts
2. Extension au cas où le nombre d’observations est limité
3. Mesure de la qualité d’une stratégie de prédiction par le regret externe
4. Indication de bornes supérieures sur le regret
5. Obtention de bornes inférieures, résolution du problème minimax

Prédire une suite individuelle (1)

On se donne un ensemble (quelconque) d'observations, $\mathcal{Y}$

On suppose que les observations $y_1, y_2, \ldots, y_n, \ldots$ sont faites séquentiellement.

Ces observations ne sont pas la réalisation d'un quelconque processus stochastique sous-jacent (c'est-à-dire que l'on ne cherche pas à modéliser le problème). Elles sont individualisées.

Notre mission : juste avant le temps t, après avoir vu $y_1^{t-1} = (y_1, y_2, \ldots, y_{t-1})$, prédire ce que sera y_t

Notre prédiction I_t est un élément d'un ensemble $\mathcal{X}$, fixé.

Exemple : le temps qu'il fera demain, $\mathcal{Y} = \{0, 1\}$ et $\mathcal{X} = [0, 1]$

Prédire une suite individuelle (2)

Pour réaliser notre tâche, nous avons accès à N experts, qui juste avant le temps t, nous donnent leur avis $f_{j,t} = f_{j,t}(y_1^{t-1}) \in \mathcal{X}$, où $j = 1, \ldots, N$

Pour prédire y_t, nous disposons donc de y_1^{t-1}, ainsi que des $f_{j,s}$, $s \leqslant t$

On introduit une fonction de perte (connue) $\ell : \mathcal{X} \times \mathcal{Y} \to [0,1]$

Notre perte instantanée au temps t est ainsi $\ell(I_t, y_t)$

Les experts subissent quant à eux les pertes instantanées respectives $\ell(f_{j,t}, y_t)$

Plan de l’exposé

1. Prédire une suite individuelle avec avis d’experts
2. Extension au cas où le nombre d’observations est limité
3. Mesure de la qualité d’une stratégie de prédiction par le regret externe
4. Indication de bornes supérieures sur le regret
5. Obtention de bornes inférieures, résolution du problème minimax

Nombre limité d'observations (1)

Dans le cadre précédent, au temps t, après avoir formé notre prédiction, nous avions accès à la vraie valeur y_t ; nous étions en situation dite d'information complète.

Imaginons cependant que voir y_t soit coûteux et que nous n'ayons qu'un budget limité de $m(n)$ observations au temps n

2 On détermine donc à chaque temps s, si on demande à voir l'observation ($Z_s = 1$) ou non ($Z_s = 0$).

On n'a ainsi accès, au temps t, qu'aux observations y_s telles que $s \leqslant t - 1$ et $Z_s = 1$. C'est une situation d'information incomplète, introduite par Helmbold et Panizza, '97.

Nombre limité d’observations (2)

Jeu répété entre le Statisticien et la Nature (omnisciente)

Paramètres : observations $\mathcal{Y}$, prédictions $\mathcal{X}$, perte ℓ, N experts, horizon n, budget de m observations

A chaque instant $t = 1, \ldots, n$,

(1) la Nature choisit $y_t \in \mathcal{Y}$ sans le révéler,

(2) le Statisticien forme une prédiction I_t (grâce aux conseils f_j des experts et en fonctions de leurs performances passées),

(3) les pertes $\ell(f_{j,t}, y_t)$ et $\ell(I_t, y_t)$ sont calculées (mais non révélées),

(4) si moins de m observations ont été faites, le Statisticien peut demander à voir y_t (et ainsi connaître les pertes du (3)), sans quoi il reste dans l’ignorance de y_t

Plan de l'exposé

1. Prédire une suite individuelle avec avis d'experts
2. Extension au cas où le nombre d'observations est limité
3. Mesure de la qualité d'une stratégie de prédiction par le regret externe
4. Indication de bornes supérieures sur le regret
5. Obtention de bornes inférieures, résolution du problème minimax

Définition du regret externe

Sauf dans le cas où $\mathcal{X}$ et ℓ sont convexes, notre prédiction I_t sera randomisée. A chaque temps t, il faudra donc donner la distribution de I_t en fonction des observations passées et des conseils présents, et la simulation de I_t se fera avec une randomisation auxiliaire.

Notre perte cumulée au temps n est

$$\widehat{L}_n = \widehat{L}_n(y_1^n) = \sum_{t=1}^{n} \ell(I_t, y_t)$$

tandis que celle du meilleur expert est $\min_{j=1,\ldots,N} L_{j,n}$, où

$$L_{j,n} = L_{j,n}(y_1^n) = \sum_{t=1}^{n} \ell(f_{j,t}, y_t)$$

Nous nous intéresserons au regret externe,

$$\widehat{R}_n = \widehat{R}_n(y_1^n) = \widehat{L}_n - \min_{j=1,\ldots,N} L_{j,n}$$

Plan de l’exposé

1. Prédire une suite individuelle avec avis d’experts
2. Extension au cas où le nombre d’observations est limité
3. Mesure de la qualité d’une stratégie de prédiction par le regret externe
4. Indication de bornes supérieures sur le regret
5. Obtention de bornes inférieures, résolution du problème minimax

Bornes supérieures sur le regret externe (1)

Considérons l'algorithme suivant :

Paramètres : Deux nombres réels, $\eta > 0$ et $0 \leqslant \varepsilon \leqslant 1$

Initialisation : $\mathbf{w}_1 = (1, \ldots, 1)$

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

(1) choisir un expert $J_t \in \{1, \ldots, N\}$ selon la distribution

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} \qquad i = 1, \ldots, N \ ,$$

et prédire comme lui, $I_t = f_{J_t,t}$

(2) tirer une variable de Bernoulli Z_t de paramètre ε,

(3) si $Z_t = 1$, demander l'observation y_t et calculer

$$w_{i,t+1} = w_{i,t}\, e^{-\eta\, \ell(f_{i,t}, y_t)/\varepsilon} \qquad \text{pour tout } i = 1, \ldots, N \ ,$$

et sinon, $\mathbf{w}_{t+1} = \mathbf{w}_t$

Bornes supérieures sur le regret externe (2)

Théorème. Fixons $\delta \in]0,1[$. Pour un choix convenable de ε et η (dépendant de n, m et δ), pour toute suite d'observations $y_1, y_2, \ldots, y_n$, et avec probabilité $1-\delta$ par rapport à la randomisation auxiliaire, le Statisticien ne demande effectivement pas plus de m observations et

$$\widehat{R}_n = \widehat{L}_n - \min_{j=1,\ldots,N} L_{j,n} \leqslant C_1 \frac{n}{\sqrt{m}} \sqrt{\ln \frac{N}{\delta}} \, ,$$

où C_1 est une constante universelle.

Remarque : Si l'on ne s'intéresse qu'au regret moyen, on a une borne supérieure de la forme

$$\sup_{y_1^n \in \mathcal{Y}^n} \mathbb{E}\left[\widehat{L}_n - \min_{j=1,\ldots,N} L_{j,n} \right] \leqslant C_2 \frac{n}{\sqrt{m}} \sqrt{\ln N}$$

Plan de l'exposé

1. Prédire une suite individuelle avec avis d'experts
2. Extension au cas où le nombre d'observations est limité
3. Mesure de la qualité d'une stratégie de prédiction par le regret externe
4. Indication de bornes supérieures sur le regret
5. Obtention de bornes inférieures, résolution du problème minimax

Problème minimax (1)

On va maintenant montrer que les bornes précédentes ne peuvent pas être améliorées en général.

Pour ce faire, nous prouverons pour tout N et tout $\gamma \ln N \leqslant m \leqslant n$, qu'il existe $\mathcal{Y}$, $\mathcal{X}$, ℓ et des experts tels que

$$\inf_{\text{algo.}} \sup_{y_1^n \in \mathcal{Y}^n} \mathbb{E}\left[\widehat{L}_n - \min_{1 \leqslant j \leqslant N} L_{j,n}\right] \geqslant C_3 \frac{n}{\sqrt{m}} \sqrt{\ln N} \,,$$

où γ et C_3 sont des constantes universelles

Problème minimax (2)

On prend par exemple $\mathcal{X} = \{1/N, 2/N, \ldots, 1\}$, $\mathcal{Y} = [0,1]$, $f_{j,t} \equiv j/N$.

On construit soigneusement ℓ pour qu'il existe N probabilités $\mathbb{P}_j$ sur $(Y_1, Y_2, \ldots, Y_n, \ldots,) \in \mathcal{Y}^{\mathbb{N}}$ telles que sous $\mathbb{P}_j$,

- les variables $\ell(k/N, Y_t)$ soient toutes indépendantes (quand k et t varient),
- les $\ell(j/N, Y_t)$ suivent toutes, quelque soit t, une loi de Bernoulli de paramètre $1/2 - \varepsilon$,
- les $\ell(k/N, Y_t)$, $k \neq j$, suivent une loi de Bernoulli de paramètre $1/2$

Problème minimax (3)

Tout d'abord, on écrit que le cas le pire est pire que le cas moyen,

$$\sup_{y_1^n \in \mathcal{Y}^n} \mathbb{E}\left[\widehat{L}_n - \min_{j=1,\ldots,N} L_{j,n}\right] \geqslant \max_{j=1,\ldots,N} \mathbb{E}_j \left[\mathbb{E}\left[\widehat{L}_n - L_{j,n}\right]\right]$$

Ensuite, en supposant que l'algorithme que l'on considère est déterministe, on remarque que

$$\mathbb{E}_j \widehat{L}_n = \frac{n}{2} - \varepsilon \sum_{t=1}^{n} \mathbb{P}_j \left[I_t = j\right] \quad \text{et} \quad \mathbb{E}_j L_{j,n} = \frac{n}{2} - n\,\varepsilon$$

De sorte que la borne inférieure est plus grande que

$$n\,\varepsilon \max_{j=1,\ldots,N} \left(1 - \frac{1}{n} \sum_{t=1}^{n} \mathbb{P}_j \left[I_t = j\right]\right)$$

Problème minimax (4)

Mais, par lemme de Fano généralisé,

$$\min_{j=1,\dots,N} \frac{1}{n} \sum_{t=1}^{n} \mathbb{P}_j\left[I_t = j\right] \leqslant \max\left(\frac{\bar{K}}{\ln(N-1)}, \frac{e}{e+1}\right)$$

où

$$\bar{K} = \frac{1}{n(N-1)} \sum_{i=2}^{N} \sum_{t=1}^{n} \mathcal{K}\left(\mathbb{P}_i^{I_t}, \mathbb{P}_1^{I_t}\right)$$

Nous allons montrer que $\mathcal{K}(\mathbb{P}_i^{I_t}, \mathbb{P}_1^{I_t}) \leqslant 5\, m\, \varepsilon^2$ et la preuve sera conclue, puisque la borne inférieure sera au moins

$$n\, \varepsilon \left(1 - \max\left(\frac{e}{e+1}, \frac{5\, m\, \varepsilon^2}{\ln N}\right)\right)$$

et que le choix $\varepsilon \sim \sqrt{(\ln N)/m}$ convient

Problème minimax (5)

Soient $1 \leqslant T_1 < T_2 \ldots < T_m \leqslant n$ les temps où l'on demande à voir les observations : ces temps $T_j = T_j(Y_1^n)$ sont aléatoires. Soit
$K_t = \max\{k \,|\, T_k \leqslant t-1\}$

Au temps t, I_t est fonction de $Y_{T_1}, Y_{T_2}, \ldots, Y_{T_{K_t}}$, c'est-à-dire de $Z = (Y_{T_1}, Y_{T_2}, \ldots, Y_{T_m})$, puisque K_t lui-même est fonction de Z

Ainsi, $\mathcal{K}(\mathbb{P}_i^{I_t}, \mathbb{P}_1^{I_t}) \leqslant \mathcal{K}(\mathbb{P}_i^{Z}, \mathbb{P}_1^{Z})$ pour tout t

Mais on prouve que Z est formé de m composantes indépendantes, de loi commune sous $\mathbb{P}_j$ un N-vecteur de variables de Bernoulli, de paramètres $1/2$ pour toutes les composantes, sauf la j-ième, de paramètre $1/2 - \varepsilon$

Un calcul explicite montre alors que $\mathcal{K}(\mathbb{P}_i^{Z}, \mathbb{P}_1^{Z}) \leqslant 5\, m\, \varepsilon^2$ pour $\varepsilon \leqslant 1/10$