Tirages pondérés sans remise

Anna Ben-Hamou
En collaboration avec Yuval Peres et Justin Salez

Université Paris Diderot, LPMA

Journées MAS
29-31 août 2016

Introduction

Le cas uniforme

Hoeffding (1963)

Serfling (1974)

Le cas général

Une comparaison entre sommes issues de tirages avec et sans remise

Une inégalité de concentration

Tirages avec et sans remise

- Ensemble de n éléments, indexés de 1 à n.

Tirages avec et sans remise

- Ensemble de n éléments, indexés de 1 à n.
- Chaque $i \in \{1, \ldots, n\}$ possède une valeur d'intérêt $\nu(i) \in \mathbb{R}$ et un poids $\omega(i) > 0$, avec

$$\sum_{i=1}^{n} \omega(i) = 1 \,.$$

Tirages avec et sans remise

- Ensemble de n éléments, indexés de 1 à n.
- Chaque $i \in \{1, \ldots, n\}$ possède une valeur d'intérêt $\nu(i) \in \mathbb{R}$ et un poids $\omega(i) > 0$, avec

$$\sum_{i=1}^{n} \omega(i) = 1\,.$$

- Pour $t \leq n$, on considère des tirages pondérés sans remise $(\mathbf{I}_1, \ldots, \mathbf{I}_t)$.

Tirages avec et sans remise

- Ensemble de n éléments, indexés de 1 à n.
- Chaque $i \in \{1, \ldots, n\}$ possède une valeur d'intérêt $\nu(i) \in \mathbb{R}$ et un poids $\omega(i) > 0$, avec

$$\sum_{i=1}^{n} \omega(i) = 1\,.$$

- Pour $t \leq n$, on considère des tirages pondérés sans remise $(\mathbf{I}_1, \ldots, \mathbf{I}_t)$.
- Pour tout t-uplet $(i_1, \ldots, i_t)$ d'indices distincts,

$$\mathbb{P}\left((\mathbf{I}_1, \ldots, \mathbf{I}_t) = (i_1, \ldots, i_t)\right) = \prod_{k=1}^{t} \frac{\omega(i_k)}{1 - \omega(i_1) - \cdots - \omega(i_{k-1})}\,.$$

Somme de tirages sans remise

- Quantité d'intérêt

$$X = \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t)\,.$$

Somme de tirages sans remise

- Quantité d'intérêt

$$X = \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t)\,.$$

- Un cas particulier : les tirages biaisés par la taille, où $(\omega(i))_{i=1}^n$ et $(\nu(i))_{i=1}^n$ sont rangés dans le même ordre.

Somme de tirages sans remise

- Quantité d'intérêt

$$X = \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t)\,.$$

- Un cas particulier : les tirages biaisés par la taille, où $(\omega(i))_{i=1}^n$ et $(\nu(i))_{i=1}^n$ sont rangés dans le même ordre.
- Un autre cas particulier : le cas uniforme, où $\omega(\cdot) \equiv 1/n$.

Motivations

- En biologie statistique, *successive sampling*. Estimateur de Horvitz-Thompson pour $V = \sum_{i=1}^{n} \nu(i)$:

$$\widehat{V}_t = \sum_{i=1}^{n} \mathbb{1}_{\{i \in \{\mathbf{I}_1, \dots, \mathbf{I}_t\}\}} \frac{\nu(i)}{\mathbb{P}\left(i \in \{\mathbf{I}_1, \dots, \mathbf{I}_t\}\right)} .$$

Motivations

- En biologie statistique, *successive sampling*. Estimateur de Horvitz-Thompson pour $V = \sum_{i=1}^{n} \nu(i)$:

$$\widehat{V}_t = \sum_{i=1}^{n} \mathbb{1}_{\{i \in \{\mathbf{I}_1, \ldots, \mathbf{I}_t\}\}} \frac{\nu(i)}{\mathbb{P}\left(i \in \{\mathbf{I}_1, \ldots, \mathbf{I}_t\}\right)} .$$

- Quand $\nu = \omega$, $(\omega(\mathbf{I}_1), \ldots, \omega(\mathbf{I}_n))$ permutation des poids biaisée par la taille. Intervient dans plusieurs problèmes de combinatoire, dans la théorie des structures de partitions (Kingman 1978, Pitman 1996, Gnedin 1997), dans des modèles de stockage de données (*heaps process* ou *librairie de Tsetlin*, Donnelly 1991).

Motivations

- En biologie statistique, *successive sampling*. Estimateur de Horvitz-Thompson pour $V = \sum_{i=1}^{n} \nu(i)$:

$$\widehat{V}_t = \sum_{i=1}^{n} \mathbb{1}_{\{i \in \{\mathbf{I}_1, \dots, \mathbf{I}_t\}\}} \frac{\nu(i)}{\mathbb{P}\left(i \in \{\mathbf{I}_1, \dots, \mathbf{I}_t\}\right)} .$$

- Quand $\nu = \omega$, $(\omega(\mathbf{I}_1), \dots, \omega(\mathbf{I}_n))$ permutation des poids biaisée par la taille. Intervient dans plusieurs problèmes de combinatoire, dans la théorie des structures de partitions (Kingman 1978, Pitman 1996, Gnedin 1997), dans des modèles de stockage de données (*heaps process* ou *librairie de Tsetlin*, Donnelly 1991).
- Modèle de configuration pour les graphes aléatoires.

- Hajek (1960) : dans le cas uniforme, TCL pour $X = \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t)$.

- Hajek (1960) : dans le cas uniforme, TCL pour $X = \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t)$.
- Holst(1973) et Gordon (1983) : TCL avec poids variables.

- Hajek (1960) : dans le cas uniforme, TCL pour $X = \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t)$.
- Holst(1973) et Gordon (1983) : TCL avec poids variables.
- Rosen (1972) : approximation pour les probabilités d'inclusion $\mathbb{P}\left(i \in \{\mathbf{I}_1, \ldots, \mathbf{I}_t\}\right)$ et pour $\mathbb{E}X$.

Concentration de X : le cas uniforme

$$\begin{aligned} X &= \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t) \\ Y &= \nu(\mathbf{J}_1) + \cdots + \nu(\mathbf{J}_t) \end{aligned}$$

où

$$\mathbb{P}\left((\mathbf{J}_1, \ldots, \mathbf{J}_t) = (j_1, \ldots, j_t)\right) = \prod_{k=1}^{t} \omega(j_k) \cdot$$

Concentration de X : le cas uniforme

$$\begin{aligned} X &= \nu(\mathbf{I}_1) + \cdots + \nu(\mathbf{I}_t) \\ Y &= \nu(\mathbf{J}_1) + \cdots + \nu(\mathbf{J}_t) \end{aligned}$$

où

$$\mathbb{P}\left((\mathbf{J}_1, \ldots, \mathbf{J}_t) = (j_1, \ldots, j_t)\right) = \prod_{k=1}^{t} \omega(j_k) \cdot$$

Théorème (Hoeffding (1963))

Dans le cas uniforme, X est majoré par Y dans l'ordre convexe : pour toute fonction convexe $f\colon \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}f(X) \;\leq\; \mathbb{E}f(Y)\,.$$

- Jolie preuve récente de Luh et Pippenger (2011).

  $X \prec_{\mathrm{oc}} Y$ ssi il existe un couplage $(X^\star, Y^\star)$ de X et Y tel que $\mathbb{E}\left[Y^\star | X^\star\right] = X^\star$.

- Jolie preuve récente de Luh et Pippenger (2011).

  $$X \prec_{\mathrm{oc}} Y \text{ ssi il existe un couplage } (X^\star, Y^\star) \text{ de } X \text{ et } Y \text{ tel que } \mathbb{E}\left[\, Y^\star | X^\star \right] = X^\star.$$

- Tirer $\mathbf{J}_1, \mathbf{J}_2$... Si T_k = temps d'arrivée du $k^{\text{ème}}$ symbole distinct, poser $\mathbf{I}_k = \mathbf{J}_{T_k}$. Alors $\mathbb{E}\left[\, Y | X \right] = X$.

- Jolie preuve récente de Luh et Pippenger (2011).

  $X \prec_{\mathrm{oc}} Y$ ssi il existe un couplage $(X^\star, Y^\star)$ de X et Y tel que $\mathbb{E}\left[Y^\star | X^\star\right] = X^\star$.

- Tirer $\mathbf{J}_1, \mathbf{J}_2 \dots$ Si $T_k =$ temps d'arrivée du $k^{\text{ème}}$ symbole distinct, poser $\mathbf{I}_k = \mathbf{J}_{T_k}$. Alors $\mathbb{E}\left[Y | X\right] = X$.
- Conséquence : $\mathbb{E}\, \mathrm{e}^{\lambda X} \leq \mathbb{E}\, \mathrm{e}^{\lambda Y}$ et $\mathbb{E} X = \mathbb{E} Y$. Par exemple

$$\mathbb{P}\left(X - \mathbb{E} X > x\right) \leq \exp\left(-\frac{2x^2}{t\Delta^2}\right),$$

où $\Delta = \max_{1 \leq i \leq n} \nu(i) - \min_{1 \leq i \leq n} \nu(i)$.

$$\operatorname{Var} Y = t\sigma^2 \qquad \operatorname{Var} X = \frac{n-t}{n-1} t\sigma^2 \,.$$

- Quand $\frac{t}{n} \to 1$, la variance de X est bien plus petite que celle de Y.

$$\operatorname{Var} Y = t\sigma^2 \qquad \operatorname{Var} X = \frac{n-t}{n-1} t\sigma^2 \,.$$

- Quand $\frac{t}{n} \to 1$, la variance de X est bien plus petite que celle de Y.
- Serfling (1974), amélioré par Bardenet et Maillard (2015)

$$\mathbb{P}\left(X - \mathbb{E}X > x\right) \leq \exp\left(-\frac{2x^2}{\frac{n-t}{n-1} t\Delta^2}\right) \,.$$

$$\operatorname{Var} Y = t\sigma^2 \qquad \operatorname{Var} X = \frac{n-t}{n-1} t\sigma^2 \,.$$

- Quand $\frac{t}{n} \to 1$, la variance de X est bien plus petite que celle de Y.
- Serfling (1974), amélioré par Bardenet et Maillard (2015)

$$\mathbb{P}\left(X - \mathbb{E}X > x\right) \leq \exp\left(-\frac{2x^2}{\frac{n-t}{n-1} t\Delta^2}\right) \,.$$

- Inégalité de « Bernstein-Serfling » obtenue par Bardenet et Maillard (2015).

Le cas des poids non-uniformes

- Le cas des tirages biaisés par la taille : pour tout $i, j \in \{1, \ldots, n\}$,

$$\omega(i) \geq \omega(j) \Longleftrightarrow \nu(i) \geq \nu(j)\,. \tag{1}$$

Le cas des poids non-uniformes

- Le cas des tirages biaisés par la taille : pour tout $i, j \in \{1, \dots, n\}$,

$$\omega(i) \geq \omega(j) \iff \nu(i) \geq \nu(j) \,. \tag{1}$$

Théorème (B., Peres, Salez, 2016)

Sous la condition (1), *X est majoré par Y dans l'ordre convexe croissant : pour toute fonction convexe croissante $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\left[f\left(X\right)\right] \;\leq\; \mathbb{E}\left[f\left(Y\right)\right] .$$

- Preuve inspirée de celle de Luh et Pippenger (2011).

  $X \prec_{\text{occ}} Y$ ssi il existe un couplage $(X^\star, Y^\star)$ de X et Y tel que $\mathbb{E}\left[\, Y^\star | X^\star \right] \geq X^\star$.

- Preuve inspirée de celle de Luh et Pippenger (2011).

  $X \prec_{\mathrm{occ}} Y$ ssi il existe un couplage $(X^\star, Y^\star)$ de X et Y tel que $\mathbb{E}\left[Y^\star | X^\star \right] \geq X^\star$.

- Tirer $\mathbf{J}_1, \mathbf{J}_2$... Si $T_k =$ temps d'arrivée du $k^{\text{ème}}$ symbole distinct, poser $\mathbf{I}_k = \mathbf{J}_{T_k}$. Alors $\mathbb{E}\left[Y | X \right] \geq X$.

- Preuve inspirée de celle de Luh et Pippenger (2011).

  $X \prec_{\mathrm{occ}} Y$ ssi il existe un couplage $(X^\star, Y^\star)$ de X et Y tel que $\mathbb{E}\,[\,Y^\star | X^\star] \geq X^\star$.
- Tirer $\mathbf{J}_1, \mathbf{J}_2$... Si T_k = temps d'arrivée du $k^{\text{ème}}$ symbole distinct, poser $\mathbf{I}_k = \mathbf{J}_{T_k}$. Alors $\mathbb{E}\,[\,Y|X] \geq X$.
- Conséquence : $\mathbb{E}\,\mathrm{e}^{\lambda X} \leq \mathbb{E}\,\mathrm{e}^{\lambda Y}$ pour $\lambda > 0$. Mais on peut avoir $\mathbb{E}X \neq \mathbb{E}\,Y$. Par exemple

$$\mathbb{P}\,(X - \mathbb{E}\,Y > x) \leq \exp\left(-\frac{2x^2}{t\Delta^2}\right) .$$

Une inégalité de concentration

- Poids $(\omega(i))_{i=1}^{n}$ arbitraires.

Une inégalité de concentration

- Poids $(\omega(i))_{i=1}^{n}$ arbitraires.
- On pose $\alpha = \dfrac{\min_{1\leq i\leq N}\omega(i)}{\max_{1\leq i\leq N}\omega(i)}\cdot$

Théorème (B., Peres, Salez, 2016)

Supposons $\alpha < 1$. Pour tout $t > 0$,

$$\mathbb{P}\left(X - \mathbb{E}X > x\right) \leq \exp\left(-\frac{x^2}{2v}\right) ,$$

avec

$$v \;=\; \min\left(4\Delta^2 t\,,\, \frac{1+4\alpha}{\alpha(1-\alpha)}\Delta^2 n\left(\frac{n-t}{n}\right)^{\alpha}\right)$$

- Preuve fondée sur le même couplage, qui permet d'exprimer X comme une fonction des variables I.I.D. $\mathbf{J}_1, \ldots, \mathbf{J}_t$:

$$X = \sum_{i=1}^{+\infty} \nu(\mathbf{J}_i) \mathbb{1}_{\{\mathbf{J}_i \notin \{\mathbf{J}_1, \ldots, \mathbf{J}_{i-1}\}\}} \mathbb{1}_{\{T_n \geq i\}} .$$

- Preuve fondée sur le même couplage, qui permet d'exprimer X comme une fonction des variables I.I.D. $\mathbf{J}_1, \dots, \mathbf{J}_t$:

$$X = \sum_{i=1}^{+\infty} \nu(\mathbf{J}_i) \mathbb{1}_{\{\mathbf{J}_i \notin \{\mathbf{J}_1, \dots, \mathbf{J}_{i-1}\}\}} \mathbb{1}_{\{T_n \geq i\}} .$$

- Méthode entropique : si $\mathrm{Ent}\left[e^{\lambda X}\right] \leq \frac{\lambda^2 v}{2} \mathbb{E}\left[e^{\lambda X}\right]$, alors $\mathbb{P}\left(X - \mathbb{E}X > x\right) \leq \exp\left(-\frac{x^2}{2v}\right)$.

- Preuve fondée sur le même couplage, qui permet d'exprimer X comme une fonction des variables I.I.D. $\mathbf{J}_1, \ldots, \mathbf{J}_t$:

$$X = \sum_{i=1}^{+\infty} \nu(\mathbf{J}_i) \mathbb{1}_{\{\mathbf{J}_i \notin \{\mathbf{J}_1, \ldots, \mathbf{J}_{i-1}\}\}} \mathbb{1}_{\{T_n \geq i\}} .$$

- Méthode entropique : si $\mathrm{Ent}\left[\mathrm{e}^{\lambda X}\right] \leq \frac{\lambda^2 v}{2} \mathbb{E}\left[\mathrm{e}^{\lambda X}\right]$, alors $\mathbb{P}\left(X - \mathbb{E}X > x\right) \leq \exp\left(-\frac{x^2}{2v}\right)$.
- On montre

$$\mathrm{Ent}\left[\mathrm{e}^{\lambda X}\right] \leq \sum_{i=1}^{+\infty} \mathbb{E}\left[\lambda^2 \mathrm{e}^{\lambda X} (X - X^i)_+^2\right] .$$

- Preuve fondée sur le même couplage, qui permet d'exprimer X comme une fonction des variables I.I.D. $\mathbf{J}_1, \ldots, \mathbf{J}_t$:

$$X = \sum_{i=1}^{+\infty} \nu(\mathbf{J}_i) \mathbb{1}_{\{\mathbf{J}_i \notin \{\mathbf{J}_1, \ldots, \mathbf{J}_{i-1}\}\}} \mathbb{1}_{\{T_n \geq i\}} \,.$$

- Méthode entropique : si $\mathrm{Ent}\left[\mathrm{e}^{\lambda X}\right] \leq \frac{\lambda^2 v}{2} \mathbb{E}\left[\mathrm{e}^{\lambda X}\right]$, alors $\mathbb{P}\left(X - \mathbb{E}X > x\right) \leq \exp\left(-\frac{x^2}{2v}\right)$.
- On montre

$$\mathrm{Ent}\left[\mathrm{e}^{\lambda X}\right] \leq \sum_{i=1}^{+\infty} \mathbb{E}\left[\lambda^2 \,\mathrm{e}^{\lambda X} (X - X^i)_+^2\right] \,.$$

- Il faut établir

$$V := \mathbb{E}\left[\sum_{i=1}^{+\infty} (X - X^i)_+^2 \middle| \mathbf{I}_1, \ldots, \mathbf{I}_n\right] \leq \frac{v}{2} \,.$$

Conclusion

- En étudiant X par l'intermédiaire de $(\mathbf{J}_1, \dots, \mathbf{J}_t)$, on paye un coût dans le facteur variance. Ex : dans le cas des poids uniformes, on paye un facteur $\log t$.

Conclusion

- En étudiant X par l'intermédiaire de $(\mathbf{J}_1, \ldots, \mathbf{J}_t)$, on paye un coût dans le facteur variance. Ex : dans le cas des poids uniformes, on paye un facteur $\log t$.
- Association négative de $(\nu(\mathbf{I}_1), \ldots, \nu(\mathbf{I}_n))$ dans le cas biaisé par la taille ?

Conclusion

- En étudiant X par l'intermédiaire de $(\mathbf{J}_1, \ldots, \mathbf{J}_t)$, on paye un coût dans le facteur variance. Ex : dans le cas des poids uniformes, on paye un facteur $\log t$.
- Association négative de $(\nu(\mathbf{I}_1), \ldots, \nu(\mathbf{I}_n))$ dans le cas biaisé par la taille ?
- Alexander (1989) : les événements $\{i \in \{\mathbf{I}_1, \ldots, \mathbf{I}_t\}\}$ et $\{j \in \{\mathbf{I}_1, \ldots, \mathbf{I}_t\}\}$ peuvent être positivement corrélés.