Algorithmique Probabiliste

Philippe Duchon

LaBRI - ENSEIRB-Matméca - Université de Bordeaux

2014-15

Quelques points technico-calculatoires

- Majorations : inégalités probabilistes

Quelques points technico-calculatoires

- Majorations : inégalités probabilistes
- Amélioration de la probabilité de succès d'un algorithme Monte Carlo

Inégalités “de queue”

On a souvent besoin de **majorer** précisément la probabilité qu'une variable aléatoire soit “loin” d'une valeur “typique”, ou “grande”, ou “petite” : on appelle *inégalité de queue* toute inégalité de la forme

- $\mathbb{P}(X > x) \leq f(x)$

Inégalités “de queue”

On a souvent besoin de **majorer** précisément la probabilité qu’une variable aléatoire soit “loin” d’une valeur “typique”, ou “grande”, ou “petite” : on appelle *inégalité de queue* toute inégalité de la forme

- $\mathbb{P}(X > x) \leq f(x)$
- $\mathbb{P}(X < x) \leq g(x)$

Inégalité de Markov

- **Hypothèse :** X variable aléatoire **positive** (ou nulle), d'espérance finie m

Inégalité de Markov

- **Hypothèse :** X variable aléatoire **positive** (ou nulle), d'espérance finie m
- **Proposition :** on a, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda m) \leq \frac{1}{\lambda}$$

Inégalité de Markov

- **Hypothèse :** X variable aléatoire **positive** (ou nulle), d'espérance finie m
- **Proposition :** on a, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda m) \leq \frac{1}{\lambda}$$

- **Autre formulation :** pour tout $x > 0$,

$$\mathbb{P}(X \geq x) \leq \frac{m}{x}$$

Inégalité de Markov

- **Hypothèse :** X variable aléatoire **positive** (ou nulle), d'espérance finie m
- **Proposition :** on a, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda m) \leq \frac{1}{\lambda}$$

- **Autre formulation :** pour tout $x > 0$,

$$\mathbb{P}(X \geq x) \leq \frac{m}{x}$$

- **Avantage :** facile à appliquer, ne demande que peu de connaissances sur X

Inégalité de Markov

- **Hypothèse :** X variable aléatoire **positive** (ou nulle), d'espérance finie m
- **Proposition :** on a, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda m) \leq \frac{1}{\lambda}$$

- **Autre formulation :** pour tout $x > 0$,

$$\mathbb{P}(X \geq x) \leq \frac{m}{x}$$

- **Avantage :** facile à appliquer, ne demande que peu de connaissances sur X
- **Inconvénient :** souvent un peu faible

Inégalité de Tchebycheff

- **Hypothèse :** X variable aléatoire, d'espérance et variance finies m et σ^2

Inégalité de Tchebycheff

- **Hypothèse :** X variable aléatoire, d'espérance et variance finies m et σ^2
- **Proposition :** on a alors, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(|X - m| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

Inégalité de Tchebycheff

- **Hypothèse :** X variable aléatoire, d'espérance et variance finies m et σ^2
- **Proposition :** on a alors, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(|X - m| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

- **Autre formulation :**

$$\mathbb{P}(|X - m| \geq x) \leq \frac{\sigma^2}{x^2}$$

Inégalité de Tchebycheff

- **Hypothèse :** X variable aléatoire, d'espérance et variance finies m et σ^2
- **Proposition :** on a alors, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(|X - m| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

- **Autre formulation :**

$$\mathbb{P}(|X - m| \geq x) \leq \frac{\sigma^2}{x^2}$$

- **Conséquences :**

$$\begin{array}{lcl} \mathbb{P}(X \geq m + x) & \leq & \frac{\sigma^2}{x^2} \\ \mathbb{P}(X \leq m - x) & \leq & \frac{\sigma^2}{x^2} \end{array}$$

Inégalité de Tchebycheff

- **Hypothèse :** X variable aléatoire, d'espérance et variance finies m et σ^2
- **Proposition :** on a alors, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(|X - m| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

- **Autre formulation :**

$$\mathbb{P}(|X - m| \geq x) \leq \frac{\sigma^2}{x^2}$$

- **Conséquences :**

$$\begin{array}{rcl} \mathbb{P}(X \geq m + x) & \leq & \frac{\sigma^2}{x^2} \\ \mathbb{P}(X \leq m - x) & \leq & \frac{\sigma^2}{x^2} \end{array}$$

- **Avantage :** souvent plus puissante que Markov

Inégalité de Tchebycheff

- **Hypothèse :** X variable aléatoire, d'espérance et variance finies m et σ^2
- **Proposition :** on a alors, pour tout nombre réel $\lambda > 0$,

$$\mathbb{P}(|X - m| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

- **Autre formulation :**

$$\mathbb{P}(|X - m| \geq x) \leq \frac{\sigma^2}{x^2}$$

- **Conséquences :**

$$\begin{array}{rcl} \mathbb{P}(X \geq m + x) & \leq & \frac{\sigma^2}{x^2} \\ \mathbb{P}(X \leq m - x) & \leq & \frac{\sigma^2}{x^2} \end{array}$$

- **Avantage :** souvent plus puissante que Markov
- **Inconvénient :** nécessite plus de connaissance de la loi de X

Inégalités exponentielles (Chernoff)

- **Hypothèse :** $X = \sum_i X_i$, où les X_i sont des variables de Bernoulli **indépendantes**, de paramètres respectifs p_i ; $m = \sum_i p_i$ espérance de X

Inégalités exponentielles (Chernoff)

- **Hypothèse :** $X = \sum_i X_i$, où les X_i sont des variables de Bernoulli **indépendantes**, de paramètres respectifs p_i ; $m = \sum_i p_i$ espérance de X
- **Proposition :** pour tout réel $\delta > 0$,

$$\mathbb{P}(X \geq (1+\delta)m) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{m}$$

Inégalités exponentielles (Chernoff)

- **Hypothèse :** $X = \sum_i X_i$, où les X_i sont des variables de Bernoulli **indépendantes**, de paramètres respectifs p_i ; $m = \sum_i p_i$ espérance de X
- **Proposition :** pour tout réel $\delta > 0$,

$$\mathbb{P}(X \geq (1+\delta)m) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{m}$$

- **Avantage :** inégalité très puissante, ne dépend pas des p_i individuels

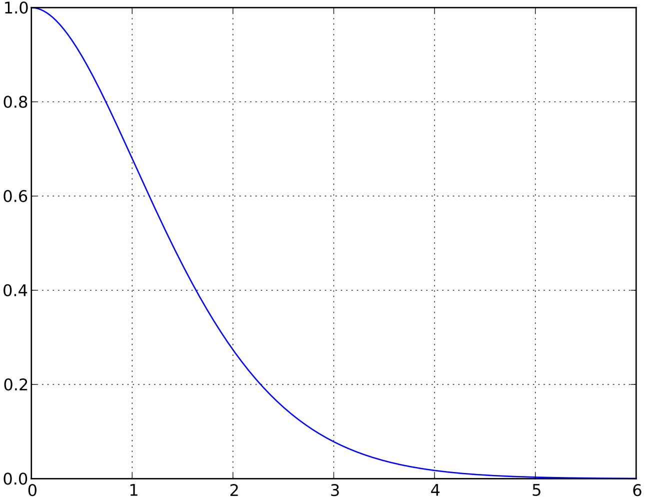
Inégalités exponentielles (Chernoff)

- **Hypothèse :** $X = \sum_i X_i$, où les X_i sont des variables de Bernoulli **indépendantes**, de paramètres respectifs p_i ; $m = \sum_i p_i$ espérance de X
- **Proposition :** pour tout réel $\delta > 0$,

$$\mathbb{P}(X \geq (1+\delta)m) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^m$$

- **Avantage :** inégalité très puissante, ne dépend pas des p_i individuels
- **Inconvénients :** l'hypothèse d'indépendance des X_i est très forte ; formule (et preuve !) un peu complexes ; pas de version symétrique

La fonction $x \mapsto e^x/(1+x)^{1+x}$



Diminution de la probabilité d'erreur MC

- On a vu qu'on pouvait facilement faire diminuer **rapidement** (exponentiellement) la probabilité d'erreur d'un algorithme **1MC** par répétitions

Diminution de la probabilité d'erreur MC

- On a vu qu'on pouvait facilement faire diminuer **rapidement** (exponentiellement) la probabilité d'erreur d'un algorithme **1MC** par répétitions
- Pour un algorithme **2MC**, ou pour un algorithme **MC** pour un problème qui n'est pas décisionnel, c'est plus compliqué.

Diminution de la probabilité d'erreur MC

- On a vu qu'on pouvait facilement faire diminuer **rapidement** (exponentiellement) la probabilité d'erreur d'un algorithme **1MC** par répétitions
- Pour un algorithme **2MC**, ou pour un algorithme **MC** pour un problème qui n'est pas décisionnel, c'est plus compliqué.
- **Idée :** chercher une stratégie de la forme : exécuter k fois l'algorithme, retourner la réponse la plus fréquemment obtenue

Diminution de la probabilité d'erreur MC

- On a vu qu'on pouvait facilement faire diminuer **rapidement** (exponentiellement) la probabilité d'erreur d'un algorithme **1MC** par répétitions
- Pour un algorithme **2MC**, ou pour un algorithme **MC** pour un problème qui n'est pas décisionnel, c'est plus compliqué.
- **Idée :** chercher une stratégie de la forme : exécuter k fois l'algorithme, retourner la réponse la plus fréquemment obtenue
- **Problème :** il faut calibrer k

Diminution de la proba d'erreur **MC**

- On suppose qu'on a un algorithme $\mathcal{A}$, **MC**, de probabilité d'erreur $p < 1/2$

Diminution de la proba d'erreur **MC**

- On suppose qu'on a un algorithme $\mathcal{A}$, **MC**, de probabilité d'erreur $p < 1/2$
- Donc pour toute donnée x, $\mathbb{P}(\mathcal{A}(x) \neq f(x)) = p(x) \leq p$.

Diminution de la proba d'erreur **MC**

- On suppose qu'on a un algorithme $\mathcal{A}$, **MC**, de probabilité d'erreur $p < 1/2$
- Donc pour toute donnée x, $\mathbb{P}(\mathcal{A}(x) \neq f(x)) = p(x) \leq p$.
- Par conséquent, si on répète k fois l'algorithme $\mathcal{A}$ sur x, le nombre R de réponses incorrectes suit la **loi binomiale** de paramètres k et $p(x)$

Diminution de la proba d'erreur **MC**

- On suppose qu'on a un algorithme $\mathcal{A}$, **MC**, de probabilité d'erreur $p < 1/2$
- Donc pour toute donnée x, $\mathbb{P}(\mathcal{A}(x) \neq f(x)) = p(x) \leq p$.
- Par conséquent, si on répète k fois l'algorithme $\mathcal{A}$ sur x, le nombre R de réponses incorrectes suit la **loi binomiale** de paramètres k et $p(x)$
- Donc l'algorithme de répétition et majorité, donne $f(x)$ dès que $R < k/2$: on veut donc majorer la probabilité que R soit supérieur à $k/2$.

Tentative 1 : Markov

- R suit la loi binomiale $(k, p(x))$, d'espérance $k.p(x)$

Tentative 1 : Markov

- R suit la loi binomiale $(k, p(x))$, d'espérance $k.p(x)$
- donc $k/2$ représente $1/(2p(x))$ fois l'espérance

Tentative 1 : Markov

- R suit la loi binomiale $(k, p(x))$, d'espérance $k.p(x)$
- donc $k/2$ représente $1/(2p(x))$ fois l'espérance
- donc on peut majorer la probabilité d'erreur par $2p(x)$

Tentative 1 : Markov

- R suit la loi binomiale $(k, p(x))$, d'espérance $k.p(x)$
- donc $k/2$ représente $1/(2p(x))$ fois l'espérance
- donc on peut majorer la probabilité d'erreur par $2p(x)$
- (ça ne dépend pas de k ! on ne peut rien conclure !)

Tentative 2 : Tchebycheff

- La binomiale est d'espérance $m = k.p(x)$, et de variance $\sigma^2 = k.p(x).(1 - p(x)) \leq k/4$

Tentative 2 : Tchebycheff

- La binomiale est d'espérance $m = k.p(x)$, et de variance $\sigma^2 = k.p(x).(1 - p(x)) \leq k/4$
- Donc $k/2 = m + (1/2 - p(x)).k \geq m + \sigma.(1 - 2p)\sqrt{k}$

Tentative 2 : Tchebycheff

- La binomiale est d'espérance $m = k.p(x)$, et de variance $\sigma^2 = k.p(x).(1 - p(x)) \leq k/4$
- Donc $k/2 = m + (1/2 - p(x)).k \geq m + \sigma.(1 - 2p)\sqrt{k}$
- L'inégalité de Tchebycheff donne par conséquent,

$$\mathbb{P}(R \geq k/2) \leq \frac{1}{k.(1 - 2p)^2}$$

Tentative 2 : Tchebycheff

- La binomiale est d'espérance $m = k.p(x)$, et de variance $\sigma^2 = k.p(x).(1 - p(x)) \leq k/4$
- Donc $k/2 = m + (1/2 - p(x)).k \geq m + \sigma.(1 - 2p)\sqrt{k}$
- L'inégalité de Tchebycheff donne par conséquent,

$$\mathbb{P}(R \geq k/2) \leq \frac{1}{k.(1-2p)^2}$$

- On ramène la probabilité d'erreur en dessous de ϵ en prenant $k \geq \frac{1}{\epsilon(1-2p)^2}$

Tentative 2 : Tchebycheff

- La binomiale est d'espérance $m = k.p(x)$, et de variance $\sigma^2 = k.p(x).(1 - p(x)) \leq k/4$
- Donc $k/2 = m + (1/2 - p(x)).k \geq m + \sigma.(1 - 2p)\sqrt{k}$
- L'inégalité de Tchebycheff donne par conséquent,

$$\mathbb{P}(R \geq k/2) \leq \frac{1}{k.(1-2p)^2}$$

- On ramène la probabilité d'erreur en dessous de ϵ en prenant $k \geq \frac{1}{\epsilon(1-2p)^2}$
- Numériquement : avec $p = 1/3$, $\epsilon = 10^{-3}$: $k = 9000$

Tentative 3 : Chernoff

- Une binomiale (k, p) est représentable comme somme de k Bernoulli (de paramètre p) indépendantes ; $m = p.k$

Tentative 3 : Chernoff

- Une binomiale (k, p) est représentable comme somme de k Bernoulli (de paramètre p) indépendantes ; $m = p.k$
- Donc $k/2 = m/(2p)$: $\delta = 1/(2p) - 1 = (1-2p)/(2p)$.

Tentative 3 : Chernoff

- Une binomiale (k, p) est représentable comme somme de k Bernoulli (de paramètre p) indépendantes ; $m = p.k$
- Donc $k/2 = m/(2p)$: $\delta = 1/(2p) - 1 = (1-2p)/(2p)$.
- Chernoff donne une majoration de la probabilité d'erreur :

$$p' \leq \cdots \leq \left(e^{-1}(2pe)^{1/(2p)}\right)^{kp} = (e^{-p}\sqrt{2pe})^k$$

Tentative 3 : Chernoff

- Une binomiale (k, p) est représentable comme somme de k Bernoulli (de paramètre p) indépendantes ; $m = p.k$
- Donc $k/2 = m/(2p)$: $\delta = 1/(2p) - 1 = (1-2p)/(2p)$.
- Chernoff donne une majoration de la probabilité d'erreur :

$$p' \leq \cdots \leq \left(e^{-1}(2pe)^{1/(2p)}\right)^{kp} = (e^{-p}\sqrt{2pe})^k$$

- Numériquement : $p = 1/3$, $p' \leq 0.965^k$

Tentative 3 : Chernoff

- Une binomiale (k, p) est représentable comme somme de k Bernoulli (de paramètre p) indépendantes ; $m = p.k$
- Donc $k/2 = m/(2p)$: $\delta = 1/(2p) - 1 = (1-2p)/(2p)$.
- Chernoff donne une majoration de la probabilité d'erreur :

$$p' \leq \cdots \leq \left(e^{-1}(2pe)^{1/(2p)}\right)^{kp} = (e^{-p}\sqrt{2pe})^k$$

- Numériquement : $p = 1/3$, $p' \leq 0.965^k$
- La proba d'erreur passe en dessous de 10^{-3} pour $k \geq 192$

Tentative 3 : Chernoff

- Une binomiale (k, p) est représentable comme somme de k Bernoulli (de paramètre p) indépendantes ; $m = p.k$
- Donc $k/2 = m/(2p)$: $\delta = 1/(2p) - 1 = (1-2p)/(2p)$.
- Chernoff donne une majoration de la probabilité d'erreur :

$$p' \leq \cdots \leq \left(e^{-1}(2pe)^{1/(2p)}\right)^{kp} = (e^{-p}\sqrt{2pe})^k$$

- Numériquement : $p = 1/3$, $p' \leq 0.965^k$
- La proba d'erreur passe en dessous de 10^{-3} pour $k \geq 192$
- La décroissance est comme pour un **1MC** : pour $\epsilon = 10^{-6}$, $k = 384$ suffit, etc.