Probabilités, Statistiques, Combinatoire

Philippe Duchon

Université de Bordeaux – Licence Informatique

2017-2018

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.
- Au lieu de chercher seulement à fournir une estimation d'un paramètre, on inclut une marge d'erreur : on donne un **intervalle** (aux extrêmités définies par des variables aléatoires) dont les extrêmités sont probablement de part et d'autres du paramètre à estimer.

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.
- Au lieu de chercher seulement à fournir une estimation d'un paramètre, on inclut une marge d'erreur : on donne un **intervalle** (aux extrêmités définies par des variables aléatoires) dont les extrêmités sont probablement de part et d'autres du paramètre à estimer.
- **Remarque :** la condition $|T_n - \theta| \leq a$ peut se réécrire de manière équivalente de deux manières :

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.
- Au lieu de chercher seulement à fournir une estimation d'un paramètre, on inclut une marge d'erreur : on donne un **intervalle** (aux extrêmités définies par des variables aléatoires) dont les extrêmités sont probablement de part et d'autres du paramètre à estimer.
- **Remarque :** la condition $|T_n - \theta| \leq a$ peut se réécrire de manière équivalente de deux manières :
  - $T_n \in [\theta - a, \theta + a]$

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.
- Au lieu de chercher seulement à fournir une estimation d'un paramètre, on inclut une marge d'erreur : on donne un **intervalle** (aux extrêmités définies par des variables aléatoires) dont les extrêmités sont probablement de part et d'autres du paramètre à estimer.
- **Remarque :** la condition $|T_n - \theta| \leq a$ peut se réécrire de manière équivalente de deux manières :
  - $T_n \in [\theta - a, \theta + a]$
  - $\theta \in [T_n - a, T_n + a]$

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.
- Au lieu de chercher seulement à fournir une estimation d'un paramètre, on inclut une marge d'erreur : on donne un **intervalle** (aux extrêmités définies par des variables aléatoires) dont les extrêmités sont probablement de part et d'autres du paramètre à estimer.
- **Remarque :** la condition $|T_n - \theta| \leq a$ peut se réécrire de manière équivalente de deux manières :
  - $T_n \in [\theta - a, \theta + a]$
  - $\theta \in [T_n - a, T_n + a]$
- **Donc**, si on peut prouver que l'estimateur T_n est probablement proche de θ, on prouve que θ est probablement dans l'intervalle $[T_n - a, T_n + a]$

Notion d'intervalle de confiance

- Notion dont la **signification** est souvent mal comprise.
- Au lieu de chercher seulement à fournir une estimation d'un paramètre, on inclut une marge d'erreur : on donne un **intervalle** (aux extrêmités définies par des variables aléatoires) dont les extrêmités sont probablement de part et d'autres du paramètre à estimer.
- **Remarque :** la condition $|T_n - \theta| \leq a$ peut se réécrire de manière équivalente de deux manières :
  - $T_n \in [\theta - a, \theta + a]$
  - $\theta \in [T_n - a, T_n + a]$
- **Donc**, si on peut prouver que l'estimateur T_n est probablement proche de θ, on prouve que θ est probablement dans l'intervalle $[T_n - a, T_n + a]$
- **Définition :** pour $0 < \alpha < 1$, on appelle *intervalle de confiance au niveau de risque* α pour le paramètre θ, la donnée de deux fonctions $A_n(X_1, \ldots, X_n)$ et $B_n(X_1, \ldots, X_n)$ telle qu'on ait

$$\mathbb{P}(\theta \in [A_n, B_n]) \geq 1 - \alpha$$

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$
- On prend comme estimateur de p, $\overline{B}_n = (B_1 + \cdots + B_n)/n$

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$
- On prend comme estimateur de p, $\overline{B}_n = (B_1 + \cdots + B_n)/n$
- $n\overline{B}_n$ est binomiale (paramètres n et p, p inconnu), donc d'espérance np et de variance $np(1-p)$

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$
- On prend comme estimateur de p, $\overline{B}_n = (B_1 + \cdots + B_n)/n$
- $n\overline{B}_n$ est binomiale (paramètres n et p, p inconnu), donc d'espérance np et de variance $np(1-p)$
- donc $\overline{B}_n$ est un estimateur sans biais, de variance $p(1-p)/n \leq 1/(4n)$

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$
- On prend comme estimateur de p, $\overline{B}_n = (B_1 + \cdots + B_n)/n$
- $n\overline{B}_n$ est binomiale (paramètres n et p, p inconnu), donc d'espérance np et de variance $np(1-p)$
- donc $\overline{B}_n$ est un estimateur sans biais, de variance $p(1-p)/n \leq 1/(4n)$
- donc (sans connaître p) on peut affirmer que l'écart-type de $\overline{B}_n$ est au plus égal à $1/2\sqrt{n}$

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$
- On prend comme estimateur de p, $\overline{B}_n = (B_1 + \cdots + B_n)/n$
- $n\overline{B}_n$ est binomiale (paramètres n et p, p inconnu), donc d'espérance np et de variance $np(1-p)$
- donc $\overline{B}_n$ est un estimateur sans biais, de variance $p(1-p)/n \leq 1/(4n)$
- donc (sans connaître p) on peut affirmer que l'écart-type de $\overline{B}_n$ est au plus égal à $1/2\sqrt{n}$
- et pour un risque α, on peut prendre $a = \frac{1}{2\sqrt{\alpha n}}$ (Tchebycheff)

Un exemple : paramètre d'une Bernoulli

- On cherche à estimer le paramètre p d'une loi de Bernoulli
- On suppose donc qu'on a un échantillon $B_1, \ldots, B_n$
- On prend comme estimateur de p, $\overline{B}_n = (B_1 + \cdots + B_n)/n$
- $n\overline{B}_n$ est binomiale (paramètres n et p, p inconnu), donc d'espérance np et de variance $np(1-p)$
- donc $\overline{B}_n$ est un estimateur sans biais, de variance $p(1-p)/n \leq 1/(4n)$
- donc (sans connaître p) on peut affirmer que l'écart-type de $\overline{B}_n$ est au plus égal à $1/2\sqrt{n}$
- et pour un risque α, on peut prendre $a = \frac{1}{2\sqrt{\alpha n}}$ (Tchebycheff)
- **Résultat :** on a un intervalle de confiance au niveau de risque α avec $[\overline{B}_n - \frac{1}{2\sqrt{\alpha n}}, \overline{B}_n + \frac{1}{2\sqrt{\alpha n}}]$.

Exemple pratique

- Échantillon de taille 1000 : on tire 1000 Bernoulli, et on observe que 325 donnent la valeur 1 ; on veut un intervalle de confiance au niveau de risque 5% (valeur classique)
- On a donc $\overline{b}_n = 0.325$
- En appliquant l'intervalle de confiance basé sur Tchebycheff : on prend un intervalle $[\overline{b}_n - a, \overline{b}_n + a]$, en s'arrangeant pour que l'on ait $\frac{1/4}{1000a^2} = \alpha$
- On résoud : ça donne $a \simeq 0.07$.
- L'intervalle de confiance est donc $[0.255, 0.395]$ (une marge d'erreur de ± 0.07, c'est assez élevé ; et pour la diviser par 2, il faut multiplier par 4 la taille de l'échantillon).

Remarque sur l'interprétation

- On applique un intervalle de confiance, et on obtient (en appliquant les formules qu'on a définies, avec les valeurs de l'échantillon) une estimation $[a_n, b_n]$

Remarque sur l'interprétation

- On applique un intervalle de confiance, et on obtient (en appliquant les formules qu'on a définies, avec les valeurs de l'échantillon) une estimation $[a_n, b_n]$
- **Il n'est pas correct** de dire "la probabilité que le paramètre soit entre a_n et b_n est d'au moins $1 - \alpha$"

Remarque sur l'interprétation

- On applique un intervalle de confiance, et on obtient (en appliquant les formules qu'on a définies, avec les valeurs de l'échantillon) une estimation $[a_n, b_n]$
- **Il n'est pas correct** de dire "la probabilité que le paramètre soit entre a_n et b_n est d'au moins $1 - \alpha$"
- Une fois qu'on a fait les calculs, a_n et b_n ne sont plus aléatoires : ils valent certaines valeurs, et soit θ est entre les deux, soit il ne l'est pas ; ça n'a plus de probabilité.

Remarque sur l'interprétation

- On applique un intervalle de confiance, et on obtient (en appliquant les formules qu'on a définies, avec les valeurs de l'échantillon) une estimation $[a_n, b_n]$
- **Il n'est pas correct** de dire "la probabilité que le paramètre soit entre a_n et b_n est d'au moins $1 - \alpha$"
- Une fois qu'on a fait les calculs, a_n et b_n ne sont plus aléatoires : ils valent certaines valeurs, et soit θ est entre les deux, soit il ne l'est pas ; ça n'a plus de probabilité.
- L'interprétation correcte : "J'ai appliqué une procédure qui a probabilité au moins $1 - \alpha$ de donner un invervalle contenant θ, et l'intervalle obtenu était $[a_n, b_n]$"

Remarque sur l'interprétation

- On applique un intervalle de confiance, et on obtient (en appliquant les formules qu'on a définies, avec les valeurs de l'échantillon) une estimation $[a_n, b_n]$
- **Il n'est pas correct** de dire "la probabilité que le paramètre soit entre a_n et b_n est d'au moins $1 - \alpha$"
- Une fois qu'on a fait les calculs, a_n et b_n ne sont plus aléatoires : ils valent certaines valeurs, et soit θ est entre les deux, soit il ne l'est pas ; ça n'a plus de probabilité.
- L'interprétation correcte : "J'ai appliqué une procédure qui a probabilité au moins $1 - \alpha$ de donner un invervalle contenant θ, et l'intervalle obtenu était $[a_n, b_n]$"
- (Si on réalise un grand nombre de sondages qui fournissent tous un intervalle de confiance au risque 5%, il faut s'attendre à ce qu'un sur 20 se trompe)

Comptage probabiliste

- **Problème :** concevoir une structure de données permettant trois opérations :
  - `Initialiser()` : création
  - `Tick()`
  - `Valeur()` : retourne le nombre n d'appels à `Tick()` depuis le dernier appel à `Initialiser()`

Comptage probabiliste

- **Problème :** concevoir une structure de données permettant trois opérations :
  - `Initialiser()` : création
  - `Tick()`
  - `Valeur()` : retourne le nombre n d'appels à `Tick()` depuis le dernier appel à `Initialiser()`
- C'est facile : on initialise un compteur à 0, et `Tick()` incrémente ce compteur.

Comptage probabiliste

- **Problème :** concevoir une structure de données permettant trois opérations :
  - `Initialiser()` : création
  - `Tick()`
  - `Valeur()` : retourne le nombre n d'appels à `Tick()` depuis le dernier appel à `Initialiser()`
- C'est facile : on initialise un compteur à 0, et `Tick()` incrémente ce compteur.
- Nécessite une mémoire de $\lceil \log_2(n+1) \rceil$ bits.

Comptage probabiliste

- **Problème :** concevoir une structure de données permettant trois opérations :
  - `Initialiser()` : création
  - `Tick()`
  - `Valeur()` : retourne le nombre n d'appels à `Tick()` depuis le dernier appel à `Initialiser()`
- C'est facile : on initialise un compteur à 0, et `Tick()` incrémente ce compteur.
- Nécessite une mémoire de $\lceil \log_2(n+1) \rceil$ bits.
- On ne peut pas faire mieux si on exige une réponse *toujours exacte*.

Comptage probabiliste

- **Problème :** concevoir une structure de données permettant trois opérations :
  - `Initialiser()` : création
  - `Tick()`
  - `Valeur()` : retourne le nombre n d'appels à `Tick()` depuis le dernier appel à `Initialiser()`
- C'est facile : on initialise un compteur à 0, et `Tick()` incrémente ce compteur.
- Nécessite une mémoire de $\lceil \log_2(n+1) \rceil$ bits.
- On ne peut pas faire mieux si on exige une réponse *toujours exacte*.
- *Idée du comptage probabiliste :* utiliser moins de mémoire en échange du fait qu'on n'obtient qu'une *estimation* de n.

Principe du comptage probabiliste

- Lors d'un `Tick()`, on n'incrémente le compteur que de manière aléatoire : *lorsque le compteur vaut k, on l'incrémente avec probabilité* $1/2^k$ (avec probabilité $1 - 1/2^k$, il reste inchangé)

Principe du comptage probabiliste

- Lors d'un `Tick()`, on n'incrémente le compteur que de manière aléatoire : *lorsque le compteur vaut k, on l'incrémente avec probabilité* $1/2^k$ (avec probabilité $1 - 1/2^k$, il reste inchangé)
- À la louche : passer d'un compteur valant k à $k + 1$, devrait prendre environ 2^k essais (en moyenne)

Principe du comptage probabiliste

- Lors d'un `Tick()`, on n'incrémente le compteur que de manière aléatoire : *lorsque le compteur vaut k, on l'incrémente avec probabilité* $1/2^k$ (avec probabilité $1 - 1/2^k$, il reste inchangé)
- À la louche : passer d'un compteur valant k à $k + 1$, devrait prendre environ 2^k essais (en moyenne)
- . . .donc passer d'un compteur 0 à un compteur k devrait prendre environ $1 + 2 + \cdots + 2^{k-1} = 2^k - 1$ essais (en moyenne)

Principe du comptage probabiliste

- Lors d'un `Tick()`, on n'incrémente le compteur que de manière aléatoire : *lorsque le compteur vaut k, on l'incrémente avec probabilité* $1/2^k$ (avec probabilité $1 - 1/2^k$, il reste inchangé)
- À la louche : passer d'un compteur valant k à $k + 1$, devrait prendre environ 2^k essais (en moyenne)
- . . .donc passer d'un compteur 0 à un compteur k devrait prendre environ $1 + 2 + \cdots + 2^{k-1} = 2^k - 1$ essais (en moyenne)
- "donc" on peut espérer que, si le compteur vaut k, le nombre de `Tick()` devrait être proche de 2^k

Principe du comptage probabiliste

- Lors d'un `Tick()`, on n'incrémente le compteur que de manière aléatoire : *lorsque le compteur vaut k, on l'incrémente avec probabilité* $1/2^k$ (avec probabilité $1 - 1/2^k$, il reste inchangé)
- À la louche : passer d'un compteur valant k à $k+1$, devrait prendre environ 2^k essais (en moyenne)
- . . .donc passer d'un compteur 0 à un compteur k devrait prendre environ $1 + 2 + \cdots + 2^{k-1} = 2^k - 1$ essais (en moyenne)
- "donc" on peut espérer que, si le compteur vaut k, le nombre de `Tick()` devrait être proche de 2^k
- On va faire le calcul de manière précise.

Preuve du comptage probabiliste

- On pose X_n la valeur du compteur après n appels à `Tick()`

Preuve du comptage probabiliste

- On pose X_n la valeur du compteur après n appels à `Tick()`
- On a donc $\mathbb{P}(X_{n+1} = k+1 | X_n = k) = 1/2^k$ et $\mathbb{P}(X_{n+1} = k | X_n = k) = 1 - 1/2^k$

Preuve du comptage probabiliste

- On pose X_n la valeur du compteur après n appels à `Tick()`
- On a donc $\mathbb{P}(X_{n+1} = k+1 | X_n = k) = 1/2^k$ et $\mathbb{P}(X_{n+1} = k | X_n = k) = 1 - 1/2^k$
- **Proposition :** pour tout $n \geq 0$, $\mathbb{E}(2^{X_n}) = n + 1$.

Preuve du comptage probabiliste

- On pose X_n la valeur du compteur après n appels à `Tick()`
- On a donc $\mathbb{P}(X_{n+1} = k+1 | X_n = k) = 1/2^k$ et $\mathbb{P}(X_{n+1} = k | X_n = k) = 1 - 1/2^k$
- **Proposition :** pour tout $n \geq 0$, $\mathbb{E}(2^{X_n}) = n+1$.
- (donc, $2^{X_n} - 1$ est un estimateur sans biais de n)

Preuve du comptage probabiliste

- On pose X_n la valeur du compteur après n appels à `Tick()`
- On a donc $\mathbb{P}(X_{n+1} = k+1 | X_n = k) = 1/2^k$ et $\mathbb{P}(X_{n+1} = k | X_n = k) = 1 - 1/2^k$
- **Proposition :** pour tout $n \geq 0$, $\mathbb{E}(2^{X_n}) = n+1$.
- (donc, $2^{X_n} - 1$ est un estimateur sans biais de n)
- **Preuve :** on montre la propriété par récurrence sur n, avec comme hypothèse de récurrence H_n : $\mathbb{E}(2^{X_n}) = n+1$.

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n + 1$.

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n + 1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n + 1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n + 1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.
- On écrit la formule pour $\mathbb{E}(2^{X_{n+1}})$:
  $\mathbb{E}(2^{X_{n+1}}) = \sum_{k \geq 0} 2^k \mathbb{P}(X_{n+1} = k)$

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n+1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.
- On écrit la formule pour $\mathbb{E}(2^{X_{n+1}})$:
  $\mathbb{E}(2^{X_{n+1}}) = \sum_{k\geq 0} 2^k \mathbb{P}(X_{n+1} = k)$
- On applique la formule des probabilités totales sur les événements $E_{n,j} = \{X_n = j\}$ (n fixe, j variable de 0 à n) :

  $$\mathbb{P}(X_{n+1} = k) = (1 - 1/2^k)\mathbb{P}(X_n = k) + 1/2^{k-1}\mathbb{P}(X_n = k-1).$$

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n+1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.
- On écrit la formule pour $\mathbb{E}(2^{X_{n+1}})$:
  $\mathbb{E}(2^{X_{n+1}}) = \sum_{k\geq 0} 2^k \mathbb{P}(X_{n+1} = k)$
- On applique la formule des probabilités totales sur les événements $E_{n,j} = \{X_n = j\}$ (n fixe, j variable de 0 à n) :

  $$\mathbb{P}(X_{n+1} = k) = (1 - 1/2^k)\mathbb{P}(X_n = k) + 1/2^{k-1}\mathbb{P}(X_n = k-1).$$

- On multiplie par 2^k, puis on somme sur tous les k :

  $$\mathbb{E}(2^{X_{n+1}}) = \sum_k \Big(2^k \mathbb{P}(X_n = k) - \mathbb{P}(X_n = k) + 2\mathbb{P}(X_n = k-1)\Big)$$

Preuve par récurrence

- **Hypothèse de récurrence H_n :** $\mathbb{E}(2^{X_n}) = n+1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.
- On écrit la formule pour $\mathbb{E}(2^{X_{n+1}})$:
  $\mathbb{E}(2^{X_{n+1}}) = \sum_{k\geq 0} 2^k \mathbb{P}(X_{n+1} = k)$
- On applique la formule des probabilités totales sur les événements $E_{n,j} = \{X_n = j\}$ (n fixe, j variable de 0 à n) :

  $$\mathbb{P}(X_{n+1} = k) = (1 - 1/2^k)\mathbb{P}(X_n = k) + 1/2^{k-1}\mathbb{P}(X_n = k-1).$$

- On multiplie par 2^k, puis on somme sur tous les k :

  $$\mathbb{E}(2^{X_{n+1}}) = \sum_k \Big(2^k \mathbb{P}(X_n = k) - \mathbb{P}(X_n = k) + 2\mathbb{P}(X_n = k-1)\Big)$$

- Soit $\mathbb{E}(2^{X_{n+1}}) = \mathbb{E}(2^{X_n}) - 1 + 2 = \mathbb{E}(2^{X_n}) + 1$

Preuve par récurrence

- **Hypothèse de récurrence** H_n **:** $\mathbb{E}(2^{X_n}) = n+1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.
- On écrit la formule pour $\mathbb{E}(2^{X_{n+1}})$:
  $\mathbb{E}(2^{X_{n+1}}) = \sum_{k\geq 0} 2^k \mathbb{P}(X_{n+1} = k)$
- On applique la formule des probabilités totales sur les événements $E_{n,j} = \{X_n = j\}$ (n fixe, j variable de 0 à n) :

  $$\mathbb{P}(X_{n+1} = k) = (1 - 1/2^k)\mathbb{P}(X_n = k) + 1/2^{k-1}\mathbb{P}(X_n = k-1).$$

- On multiplie par 2^k, puis on somme sur tous les k :

  $$\mathbb{E}(2^{X_{n+1}}) = \sum_k \Big(2^k \mathbb{P}(X_n = k) - \mathbb{P}(X_n = k) + 2\mathbb{P}(X_n = k-1)\Big)$$

- Soit $\mathbb{E}(2^{X_{n+1}}) = \mathbb{E}(2^{X_n}) - 1 + 2 = \mathbb{E}(2^{X_n}) + 1$
- Par H_n, $\mathbb{E}(2^{X_n}) = n+1$ donc $\mathbb{E}(2^{X_{n+1}}) = n+2$, soit H_{n+1}.

Preuve par récurrence

- **Hypothèse de récurrence** H_n **:** $\mathbb{E}(2^{X_n}) = n+1$.
- H_0 est vraie : $X_0 = 0$ (de manière certaine) donc $\mathbb{E}(2^{X_0}) = 1$.
- Soit $n \geq 0$ tel que H_n soit vraie ; voyons si on peut en déduire H_{n+1}.
- On écrit la formule pour $\mathbb{E}(2^{X_{n+1}})$:
  $\mathbb{E}(2^{X_{n+1}}) = \sum_{k\geq 0} 2^k \mathbb{P}(X_{n+1} = k)$
- On applique la formule des probabilités totales sur les événements $E_{n,j} = \{X_n = j\}$ (n fixe, j variable de 0 à n) :

  $$\mathbb{P}(X_{n+1} = k) = (1 - 1/2^k)\mathbb{P}(X_n = k) + 1/2^{k-1}\mathbb{P}(X_n = k-1).$$

- On multiplie par 2^k, puis on somme sur tous les k :

  $$\mathbb{E}(2^{X_{n+1}}) = \sum_k \Big(2^k \mathbb{P}(X_n = k) - \mathbb{P}(X_n = k) + 2\mathbb{P}(X_n = k-1)\Big)$$

- Soit $\mathbb{E}(2^{X_{n+1}}) = \mathbb{E}(2^{X_n}) - 1 + 2 = \mathbb{E}(2^{X_n}) + 1$
- Par H_n, $\mathbb{E}(2^{X_n}) = n+1$ donc $\mathbb{E}(2^{X_{n+1}}) = n+2$, soit H_{n+1}.
- Donc, par récurrence sur n, H_n est vraie pour tout n et on a bien $\mathbb{E}(2^{X_n}) = n+1$.

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.
- Mais pour avoir une idée de sa qualité (convergence ?), il faudrait connaître sa variance.

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.
- Mais pour avoir une idée de sa qualité (convergence ?), il faudrait connaître sa variance.
- $\mathbf{Var}(2^{X_n}) = \mathbb{E}(2^{2X_n}) - (\mathbb{E}(2^{X_n}))^2$, donc il faudrait calculer $\mathbb{E}(4^{X_n})$.

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.
- Mais pour avoir une idée de sa qualité (convergence ?), il faudrait connaître sa variance.
- $\mathbf{Var}(2^{X_n}) = \mathbb{E}(2^{2X_n}) - (\mathbb{E}(2^{X_n}))^2$, donc il faudrait calculer $\mathbb{E}(4^{X_n})$.
- On fait le même type de calcul : $\mathbb{E}(4^{X_{n+1}}) = \sum_k 4^k \mathbb{P}(X_{n+1} = k)$.

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.
- Mais pour avoir une idée de sa qualité (convergence ?), il faudrait connaître sa variance.
- $\mathbf{Var}(2^{X_n}) = \mathbb{E}(2^{2X_n}) - (\mathbb{E}(2^{X_n}))^2$, donc il faudrait calculer $\mathbb{E}(4^{X_n})$.
- On fait le même type de calcul : $\mathbb{E}(4^{X_{n+1}}) = \sum_k 4^k \mathbb{P}(X_{n+1} = k)$.
- Résultat : $\mathbb{E}(4^{X_{n+1}}) = \mathbb{E}(4^{X_n}) + 3(n+1)$

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.
- Mais pour avoir une idée de sa qualité (convergence ?), il faudrait connaître sa variance.
- $\mathbf{Var}(2^{X_n}) = \mathbb{E}(2^{2X_n}) - (\mathbb{E}(2^{X_n}))^2$, donc il faudrait calculer $\mathbb{E}(4^{X_n})$.
- On fait le même type de calcul : $\mathbb{E}(4^{X_{n+1}}) = \sum_k 4^k \mathbb{P}(X_{n+1} = k)$.
- Résultat : $\mathbb{E}(4^{X_{n+1}}) = \mathbb{E}(4^{X_n}) + 3(n+1)$
- . . .dont on déduit (avec $\mathbb{E}(4^{X_0}) = 1$) $\mathbb{E}(4^{X_n}) = 1 + 3n(n+1)/2$

Et après ?

- On a prouvé que $2^{X_n} - 1$ est un estimateur sans biais de n.
- Mais pour avoir une idée de sa qualité (convergence ?), il faudrait connaître sa variance.
- $\mathbf{Var}(2^{X_n}) = \mathbb{E}(2^{2X_n}) - (\mathbb{E}(2^{X_n}))^2$, donc il faudrait calculer $\mathbb{E}(4^{X_n})$.
- On fait le même type de calcul : $\mathbb{E}(4^{X_{n+1}}) = \sum_k 4^k \mathbb{P}(X_{n+1} = k)$.
- Résultat : $\mathbb{E}(4^{X_{n+1}}) = \mathbb{E}(4^{X_n}) + 3(n+1)$
- . . .dont on déduit (avec $\mathbb{E}(4^{X_0}) = 1$) $\mathbb{E}(4^{X_n}) = 1 + 3n(n+1)/2$
- **Résultat :** l'écart-type de 2^{X_n} est d'environ $1/\sqrt{2}n \simeq 0.7n$, donc du même ordre de grandeur que l'espérance : l'estimateur n'est sans doute pas convergent.

Amélioration

- **Idée simple (pas forcément utilisable en statistiques classiques) :** pour diminuer la variance de l'estimateur, on peut en prendre k versions indépendantes, et prendre leur moyenne empirique : ça divise la variance par k

Amélioration

- **Idée simple (pas forcément utilisable en statistiques classiques) :** pour diminuer la variance de l'estimateur, on peut en prendre k versions indépendantes, et prendre leur moyenne empirique : ça divise la variance par k
- (En statistiques classiques, ça veut souvent dire multiplier par k la taille de l'échantillon)

Amélioration

- **Idée simple (pas forcément utilisable en statistiques classiques) :** pour diminuer la variance de l'estimateur, on peut en prendre k versions indépendantes, et prendre leur moyenne empirique : ça divise la variance par k
- (En statistiques classiques, ça veut souvent dire multiplier par k la taille de l'échantillon)
- Par exemple, avec $k = 7$, l'écart-type est d'environ $0.1n$: $5\sigma < n/2$; (Tchebycheff) le nouvel estimateur tombe entre $n/2$ et $3n/2$ avec proba. au moins $1 - 1/25$, soit 96%.

Amélioration

- **Idée simple (pas forcément utilisable en statistiques classiques) :** pour diminuer la variance de l'estimateur, on peut en prendre k versions indépendantes, et prendre leur moyenne empirique : ça divise la variance par k
- (En statistiques classiques, ça veut souvent dire multiplier par k la taille de l'échantillon)
- Par exemple, avec $k = 7$, l'écart-type est d'environ $0.1n$: $5\sigma < n/2$; (Tchebycheff) le nouvel estimateur tombe entre $n/2$ et $3n/2$ avec proba. au moins $1 - 1/25$, soit 96%.
- Mémoire utilisée : on maintient k compteurs indépendants, qui seront d'ordre de grandeur $\log_2(n)$, soit $k \log_2 \log_2(n)$ bits.

Amélioration

- **Idée simple (pas forcément utilisable en statistiques classiques) :** pour diminuer la variance de l'estimateur, on peut en prendre k versions indépendantes, et prendre leur moyenne empirique : ça divise la variance par k
- (En statistiques classiques, ça veut souvent dire multiplier par k la taille de l'échantillon)
- Par exemple, avec $k = 7$, l'écart-type est d'environ $0.1n$: $5\sigma < n/2$; (Tchebycheff) le nouvel estimateur tombe entre $n/2$ et $3n/2$ avec proba. au moins $1 - 1/25$, soit 96%.
- Mémoire utilisée : on maintient k compteurs indépendants, qui seront d'ordre de grandeur $\log_2(n)$, soit $k \log_2 \log_2(n)$ bits.
- Pour que ce soit intéressant par rapport à la méthode naïve, il faut que $k \log_2(\log_2(n))$ soit petit par rapport à $\log_2(n)$: pour $k = 10$ ça correspond à de très grandes valeurs de n ($n > 5.10^{17}$ pour $k = 10$).

Amélioration

- **Idée simple (pas forcément utilisable en statistiques classiques) :** pour diminuer la variance de l'estimateur, on peut en prendre k versions indépendantes, et prendre leur moyenne empirique : ça divise la variance par k
- (En statistiques classiques, ça veut souvent dire multiplier par k la taille de l'échantillon)
- Par exemple, avec $k = 7$, l'écart-type est d'environ $0.1n$: $5\sigma < n/2$; (Tchebycheff) le nouvel estimateur tombe entre $n/2$ et $3n/2$ avec proba. au moins $1 - 1/25$, soit 96%.
- Mémoire utilisée : on maintient k compteurs indépendants, qui seront d'ordre de grandeur $\log_2(n)$, soit $k \log_2 \log_2(n)$ bits.
- Pour que ce soit intéressant par rapport à la méthode naïve, il faut que $k \log_2(\log_2(n))$ soit petit par rapport à $\log_2(n)$: pour $k = 10$ ça correspond à de très grandes valeurs de n ($n > 5.10^{17}$ pour $k = 10$).
- (On peut être plus efficace : au lieu de la *moyenne empirique* de k estimateurs, on prend leur *médiane empirique* ; mais les calculs sont un peu moins simples)

Variantes du comptage probabiliste

- L'idée originale du comptage probabiliste remonte à **Morris (1978)**

Variantes du comptage probabiliste

- L'idée originale du comptage probabiliste remonte à **Morris (1978)**
- Analyse détaillée (plus complexe que celle présentée ici !), **Flajolet-Martin (1985)**

Variantes du comptage probabiliste

- L'idée originale du comptage probabiliste remonte à **Morris (1978)**
- Analyse détaillée (plus complexe que celle présentée ici !), **Flajolet-Martin (1985)**
- Variante : "loglog counting" (**Durand-Flajolet, 2003**) pour compter les valeurs *distinctes* dans une longue liste (en un seul passage de la liste, et avec très peu de mémoire)

Variantes du comptage probabiliste

- L'idée originale du comptage probabiliste remonte à **Morris (1978)**
- Analyse détaillée (plus complexe que celle présentée ici !), **Flajolet-Martin (1985)**
- Variante : "loglog counting" (**Durand-Flajolet, 2003**) pour compter les valeurs *distinctes* dans une longue liste (en un seul passage de la liste, et avec très peu de mémoire)
- L'algorithme LogLogCounting et ses variantes, sont typiquement utilisés dans le domaine du traitement de "données massives" (BigData)

Variantes du comptage probabiliste

- L'idée originale du comptage probabiliste remonte à **Morris (1978)**
- Analyse détaillée (plus complexe que celle présentée ici !), **Flajolet-Martin (1985)**
- Variante : "loglog counting" (**Durand-Flajolet, 2003**) pour compter les valeurs *distinctes* dans une longue liste (en un seul passage de la liste, et avec très peu de mémoire)
- L'algorithme LogLogCounting et ses variantes, sont typiquement utilisés dans le domaine du traitement de "données massives" (BigData)
- Exemple : en **une passe** sur le texte de l'œuvre de Shakespeare, avec 128 octets de mémoire, donne une estimation du nombre de mots distincts de 30897 (valeur exacte : 28239 ; +9.4%)