Le problème du collectionneur de coupons

Après le tennis, du foot

stephane@gonnord.org – http://blog.psi945.fr

Vendredi 24 juin 2016

Plan

1. Combien d'étiquettes différentes, pour k achetées ?
   - Une belle idée
   - Des développements limités
2. Combien en acheter pour en avoir ℓ différentes ?
   - Un problème plus simple
   - Attente de la $\ell+1$ième
   - Encore des maths
3. IRL
   - Que disent les deux mathématiciens ?
   - On arrête les frais ?

De quoi est-il question ?

- Collection de n vignettes/coupons. Vendus de façon aléatoire.

$$n = 639 \text{ vs. } n = 680$$

- Après k achetées, combien de différentes ?
- Combien en acheter pour en avoir ℓ différentes ?
- Probabilité d'avoir rempli l'album après k achats ?

« Selon deux mathématiciens de l'Université de Genève, il faudrait acheter 899 paquets de 5 cartes pour parvenir à compléter l'album de 660 vignettes sans échange et avec la probabilité d'avoir de nombreuses images en double. »

Sommaire

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

- $X_i = \begin{cases} 1 & \text{si l'étiquette numéro } i \text{ est achetée} \\ 0 & \text{sinon} \end{cases}$

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

- $X_i = \begin{cases} 1 & \text{si l'étiquette numéro } i \text{ est achetée} \\ 0 & \text{sinon} \end{cases}$
- $X = X_1 + X_2 + \cdots + X_n$.

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

- $X_i = \begin{cases} 1 & \text{si l'étiquette numéro } i \text{ est achetée} \\ 0 & \text{sinon} \end{cases}$
- $X = X_1 + X_2 + \cdots + X_n$.
- $\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X_i) = n\mathbb{E}(X_1)$

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

- $X_i = \begin{cases} 1 & \text{si l'étiquette numéro } i \text{ est achetée} \\ 0 & \text{sinon} \end{cases}$
- $X = X_1 + X_2 + \cdots + X_n$.
- $\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X_i) = n\mathbb{E}(X_1)$
- $\mathbb{E}(X_1) = \mathbb{P}(X_1 = 1) =$

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

- $X_i = \begin{cases} 1 & \text{si l'étiquette numéro } i \text{ est achetée} \\ 0 & \text{sinon} \end{cases}$
- $X = X_1 + X_2 + \cdots + X_n$.
- $\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X_i) = n\mathbb{E}(X_1)$
- $\mathbb{E}(X_1) = \mathbb{P}(X_1 = 1) = 1 - \cdots$

Une vraie bonne idée

X : nombre d'étiquettes différentes achetées

- $X_i = \begin{cases} 1 & \text{si l'étiquette numéro } i \text{ est achetée} \\ 0 & \text{sinon} \end{cases}$
- $X = X_1 + X_2 + \cdots + X_n$.
- $\mathbb{E}(X) = \sum\limits_{k=1}^{n} \mathbb{E}(X_i) = n\mathbb{E}(X_1)$
- $\mathbb{E}(X_1) = \mathbb{P}(X_1 = 1) = 1 - \cdots$

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^k\right)$$

Sommaire

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim$$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

Et donc après *k* achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$
  - $\alpha_n \underset{n \to +\infty}{\longrightarrow} 0,$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$
  - $\alpha_n \underset{n\to+\infty}{\longrightarrow} 0$, $\alpha_n = o(1/n^{10})$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$
  - $\alpha_n \underset{n \to +\infty}{\longrightarrow} 0$, $\alpha_n = o(1/n^{10})$ et même $\alpha_n = o(1/10^n)$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1-\left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1-\mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$
  - $\alpha_n \underset{n\to+\infty}{\longrightarrow} 0$, $\alpha_n = o(1/n^{10})$ et même $\alpha_n = o(1/10^n)$

- Pour $k = 2n$: $\mathbb{E}(X) \sim \underbrace{\left(1-\mathrm{e}^{-2}\right)}_{\simeq 0.87}$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$
  - $\alpha_n \underset{n\to+\infty}{\longrightarrow} 0$, $\alpha_n = o(1/n^{10})$ et même $\alpha_n = o(1/10^n)$

- Pour $k = 2n$: $\mathbb{E}(X) \sim \underbrace{\left(1 - \mathrm{e}^{-2}\right)}_{\simeq 0.87}$
- Pour $k = n^2$: $\mathbb{E}(X) = n - o(1)$

Et donc après k achats...

- Pour $k = n$:

$$\mathbb{E}(X) = n\left(1 - \left(\frac{n-1}{n}\right)^n\right) \sim n\underbrace{\left(1 - \mathrm{e}^{-1}\right)}_{\simeq 0.63}$$

$$\mathbb{P}(n \neq) = \frac{n!}{n^n} = \alpha_n$$

  - $\alpha_{10} = 3.6\ 10^{-4}$ et $\alpha_{100} \simeq 10^{-42}$
  - $\alpha_n \underset{n\to+\infty}{\longrightarrow} 0$, $\alpha_n = o(1/n^{10})$ et même $\alpha_n = o(1/10^n)$
- Pour $k = 2n$: $\mathbb{E}(X) \sim \underbrace{\left(1 - \mathrm{e}^{-2}\right)}_{\simeq 0.87}$
- Pour $k = n^2$: $\mathbb{E}(X) = n - o(1)$
- Pour $k = n\sqrt{n}$: $\mathbb{E}(X) = n - o(1)$.

Sommaire

Un problème plus simple : attente du premier Pile

- - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.

Un problème plus simple : attente du premier Pile

- - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$

Un problème plus simple : attente du premier Pile

- - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$
- C'est une limite de somme (une *somme de série*).

Un problème plus simple : attente du premier Pile

- - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$
- C'est une limite de somme (une *somme de série*).
  Heu... ça converge ?

Un problème plus simple : attente du premier Pile

-
  - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$
- C'est une limite de somme (une *somme de série*).
  Heu... ça converge ?
- $\mathbb{P}(T = 1) = ?$

Un problème plus simple : attente du premier Pile

- - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$
- C'est une limite de somme (une *somme de série*).
  Heu... ça converge ?
- $\mathbb{P}(T = 1) = ?$ $\mathbb{P}(T = k) = ?$

Un problème plus simple : attente du premier Pile

- - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$
- C'est une limite de somme (une *somme de série*).
  Heu... ça converge ?
- $\mathbb{P}(T = 1) = ?$ $\mathbb{P}(T = k) = ?$
- $\mathbb{E}(T) = \sum_{k=1}^{+\infty} \frac{k}{2^k}$

Un problème plus simple : attente du premier Pile

-
  - FFPFF... : $T = 3$;
  - PFF.... : $T = 1$;
  - FFFFFFF... : $T = \infty$.
- $\mathbb{E}(T) = \mathbb{P}(T = 1) + 2\mathbb{P}(T = 2) + 3\mathbb{P}(T = 3) + \cdots$
- C'est une limite de somme (une *somme de série*).
  Heu... ça converge ?
- $\mathbb{P}(T = 1) = ?\ \mathbb{P}(T = k) = ?$
- $\mathbb{E}(T) = \sum_{k=1}^{+\infty} \frac{k}{2^k}$

$$\mathbb{E}(T) = 2$$

Sommaire

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell - 1)$ième.

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell - 1)$ième.
- $T_1 = 1$!
- T_2 = ?

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell - 1)$ième.
- $T_1 = 1$!
- T_2 = ? $1 \leq T_1$; T_2 non bornée.
- $\mathbb{P}(T_2 = 1) = ?$ $\mathbb{P}(T_2 = 2) = ?$ $\mathbb{P}(T_2 = k) = ?$

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell - 1)$ième.
- $T_1 = 1$!
- T_2 = ? $1 \leq T_1$; T_2 non bornée.
- $\mathbb{P}(T_2 = 1) = ?$ $\mathbb{P}(T_2 = 2) = ?$ $\mathbb{P}(T_2 = k) = ?$

$$\mathbb{E}(T_2) = \mathbb{P}(T_2 = 1) + 2\mathbb{P}(T_2 = 2) + \cdots =$$

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell-1)$ième.
- $T_1 = 1$!
- T_2 = ? $1 \leq T_1$; T_2 non bornée.
- $\mathbb{P}(T_2 = 1) = ?$ $\mathbb{P}(T_2 = 2) = ?$ $\mathbb{P}(T_2 = k) = ?$

$$\mathbb{E}(T_2) = \mathbb{P}(T_2 = 1) + 2\mathbb{P}(T_2 = 2) + \cdots = \frac{n}{n-1}.$$

- $\mathbb{P}(T_3 = k) = ?$ $\mathbb{E}(T_3) = ?$

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell-1)$ième.
- $T_1 = 1$!
- T_2 = ? $1 \leq T_1$; T_2 non bornée.
- $\mathbb{P}(T_2 = 1) = ?$ $\mathbb{P}(T_2 = 2) = ?$ $\mathbb{P}(T_2 = k) = ?$

$$\mathbb{E}(T_2) = \mathbb{P}(T_2 = 1) + 2\mathbb{P}(T_2 = 2) + \cdots = \frac{n}{n-1}.$$

- $\mathbb{P}(T_3 = k) = ?$ $\mathbb{E}(T_3) = ?\mathbb{E}(T_k) = ?$

Retour aux vignettes

- T_ℓ : temps d'attente de la ℓième vignette, après la $(\ell-1)$ième.
- $T_1 = 1$!
- T_2 = ? $1 \leq T_1$; T_2 non bornée.
- $\mathbb{P}(T_2 = 1) = ?$ $\mathbb{P}(T_2 = 2) = ?$ $\mathbb{P}(T_2 = k) = ?$

$$\mathbb{E}(T_2) = \mathbb{P}(T_2 = 1) + 2\mathbb{P}(T_2 = 2) + \cdots = \frac{n}{n-1}.$$

- $\mathbb{P}(T_3 = k) = ?$ $\mathbb{E}(T_3) = ?\mathbb{E}(T_k) = ?$

$$\mathbb{E}(T_1 + \cdots + T_\ell) = n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-(\ell-1)}\right)$$

Sommaire

Quelques valeurs particulières

- $k = n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_n) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)$$

Quelques valeurs particulières

- $k = n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_n) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \sim n \ln n$$

(comparaison somme/intégrale).

Quelques valeurs particulières

- $k = n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_n) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \sim n \ln n$$

  (comparaison somme/intégrale).

- $k = n/2$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_{n/2}) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n/2}\right) \sim n \ln 2$$

- $k = \lambda n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_{\lambda n}) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{\lambda n}\right) \sim n \ln \frac{1}{1-\lambda}.$$

Quelques valeurs particulières

- $k = n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_n) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \sim n \ln n$$

(comparaison somme/intégrale).

- $k = n/2$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_{n/2}) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n/2}\right) \sim n \ln 2$$

- $k = \lambda n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_{\lambda n}) = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{\lambda n}\right) \sim n \ln \frac{1}{1-\lambda}.$$

- $k = (1 - \mathrm{e}^{-1})n$:

$$\mathbb{E}(T_1 + T_2 + \cdots + T_k) \sim n$$

Sommaire

Que disent les deux mathématiciens ?

Selon deux mathématiciens de l'Université de Genève, il faudrait acheter 899 paquets de 5 cartes pour parvenir à compléter l'album de 660 vignettes sans échange et avec la probabilité d'avoir de nombreuses images en double. »

$$n = 660$$

- $\mathbb{E}(T_n) \simeq n \ln n \simeq 4285...$

Que disent les deux mathématiciens ?

Selon deux mathématiciens de l'Université de Genève, il faudrait acheter 899 paquets de 5 cartes pour parvenir à compléter l'album de 660 vignettes sans échange et avec la probabilité d'avoir de nombreuses images en double. »

$$n = 660$$

- $\mathbb{E}(T_n) \simeq n \ln n \simeq 4285...$ et $5 \times 899 = 4495$.

Que disent les deux mathématiciens ?

Selon deux mathématiciens de l'Université de Genève, il faudrait acheter 899 paquets de 5 cartes pour parvenir à compléter l'album de 660 vignettes sans échange et avec la probabilité d'avoir de nombreuses images en double. »

$$n = 660$$

- $\mathbb{E}(T_n) \simeq n \ln n \simeq 4285...$ et $5 \times 899 = 4495$.
- $\mathbb{E}(\neq \text{après } 4285 \text{ achats}) \simeq 659.01$;
  $\mathbb{E}(\neq \text{après } 4495 \text{ achats}) \simeq 659.28$

Que disent les deux mathématiciens ?

Selon deux mathématiciens de l'Université de Genève, il faudrait acheter 899 paquets de 5 cartes pour parvenir à compléter l'album de 660 vignettes sans échange et avec la probabilité d'avoir de nombreuses images en double. »

$$n = 660$$

- $\mathbb{E}(T_n) \simeq n \ln n \simeq 4285...$ et $5 \times 899 = 4495$.
- $\mathbb{E}(\neq \text{ après } 4285 \text{ achats}) \simeq 659.01$;
  $\mathbb{E}(\neq \text{ après } 4495 \text{ achats}) \simeq 659.28$
- $\mathbb{P}(\text{rempli après } k) = \mathbb{P}(\text{surjection}) = \cdots$

Que disent les deux mathématiciens ?

Selon deux mathématiciens de l'Université de Genève, il faudrait acheter 899 paquets de 5 cartes pour parvenir à compléter l'album de 660 vignettes sans échange et avec la probabilité d'avoir de nombreuses images en double. »

$$n = 660$$

- $\mathbb{E}(T_n) \simeq n \ln n \simeq 4285...$ et $5 \times 899 = 4495$.
- $\mathbb{E}(\neq \text{ après } 4285 \text{ achats}) \simeq 659.01$;
  $\mathbb{E}(\neq \text{ après } 4495 \text{ achats}) \simeq 659.28$
- $\mathbb{P}(\text{rempli après } k) = \mathbb{P}(\text{surjection}) = \cdots$

$$\frac{1}{n^k} \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} i^k$$

Sommaire

On arrête les frais ?

- Prix p_1 (hasard) ou p_2 (étiquette choisie).

On arrête les frais ?

- Prix p_1 (hasard) ou p_2 (étiquette choisie).
- $\mathbb{P}$(nouvelle) : $\frac{n-k}{n}$ vs. 1.

On arrête les frais ?

- Prix p_1 (hasard) ou p_2 (étiquette choisie).
- $\mathbb{P}(\text{nouvelle})$: $\frac{n-k}{n}$ vs. 1.
- But : maximiser $\dfrac{\mathbb{P}(\text{nouvelle})}{p}$.

On arrête les frais ?

- Prix p_1 (hasard) ou p_2 (étiquette choisie).
- $\mathbb{P}(\text{nouvelle})$: $\frac{n-k}{n}$ vs. 1.
- But : maximiser $\dfrac{\mathbb{P}(\text{nouvelle})}{p}$.

$$\frac{n-k}{np_1} \leq \frac{1}{p_2} \iff \frac{k}{n} \geq 1 - \frac{p_1}{p_2}$$

- $(0.1, 0.2)$: à la moitié ; $(0.1, 0.3)$: aux deux-tiers.

C'est fini...

Merci de votre attention !

C'est fini...

Merci de votre attention !

Et bonnes vacances \o/