

Quart d'heure académique du SéminDoc 06/05/2009 LIRMM – Montpellier

Estimation du nombre de citations de papillotes et de blagues Carambar

Philippe Gambette (équipes MAB/AIGco)



Introduction

- Papillotes créées en 1790
 - un billet doux pour enrober un chocolat, à l'origine
 - depuis, rébus, dessins d'humour, citations



papillotesrevillon.fr

- Carambars créés en 1954
 - mélange accidentel de caramel et cacao
 - devinettes et blagues sur l'emballage depuis 1969



<http://fr.wikipedia.org/wiki/Carambar>

Problématique

Combien de citations ou blagues différentes ?

- **pour le fabricant :**

- limiter les coûts de production → nombre fini
- satisfaire le consommateur

- **pour le consommateur :**

- frustration de retomber sur une blague déjà lue
- souci d'exhaustivité : combien en manger pour les lire toutes ?

- **pour le statisticien :**

- estimer ce nombre n d'après un échantillon

Echantillonnage

- tirer un échantillon aléatoire de k papillotes

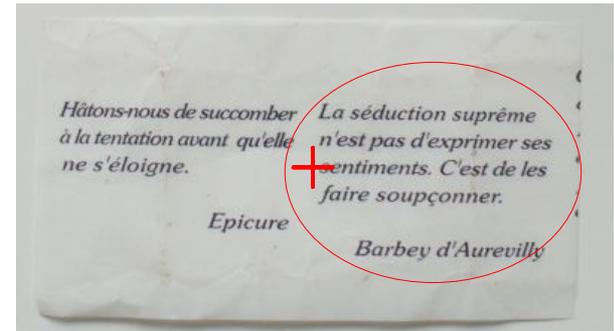
on suppose que les citations sont uniformément réparties dans les sachets

- discrétiser les données

associer une citation à chaque papillote

- identifier les doublons

associer un entier unique à chaque citation



choix de la citation la plus proche du centre du papier

Modélisation de l'échantillonnage :

tirer un mot aléatoire de k lettres, choisies parmi un alphabet de n lettres.

Modélisation du problème

Sachant qu'il y a

n papillotes différentes au total

un alphabet de n lettres

quelle est la probabilité

de tirer 40 citations

d'avoir 40 lettres

différentes, exactement,

parmi un échantillon de 52 papillotes

dans un mot de 52 lettres

?

Modélisation du problème :

trouver la valeur de n qui maximise cette probabilité

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\mathbf{P}_{d,k}(n) = \frac{\text{nombre de mots de } k \text{ lettres dont } d \text{ différentes}}{\text{nombre de mots de } k \text{ lettres}}$$

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\mathbf{P}_{d,k}(n) = \frac{\text{nombre de mots de } k \text{ lettres dont } d \text{ différentes}}{\text{nombre de mots de } k \text{ lettres}}$$

Nombre $a_{d,k}(n)$ de mots de k lettres dont d différentes :

$n=3, k=3, d=2$:

aab aba abb baa bab bba

aac aca acc caa cac cca

bbc bcb bcc cbb cbc ccb

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\mathbf{P}_{d,k}(n) = \frac{\text{nombre de mots de } k \text{ lettres dont } d \text{ différentes}}{\text{nombre de mots de } k \text{ lettres}}$$

Nombre $a_{d,k}(n)$ de mots de k lettres dont d différentes :

$n=3, k=3, d=2$:

aab aba abb baa bab bba
aac aca acc caa cac cca
bbc bcb bcc cbb cbc ccb

*on trouve les mots sur $d=2$ lettres
on en déduit les mots sur n lettres
par projection.*

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\mathbf{P}_{d,k}(n) = \frac{\text{nombre de mots de } k \text{ lettres dont } d \text{ différentes}}{\text{nombre de mots de } k \text{ lettres}}$$

Nombre $a_{d,k}(n)$ de mots de k lettres dont d différentes :

$n=3, k=3, d=2$:

aab aba abb baa bab bba
aac aca acc caa cac cca
bbc bcb bcc cbb cbc ccb

*on trouve les mots sur $d=2$ lettres
on en déduit les mots sur n lettres
par projection : $a_{d,k}(n) = a_{d,k}(k) C_n^d$*

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\mathbf{P}_{d,k}(n) = \frac{\text{nombre de mots de } k \text{ lettres dont } d \text{ différentes}}{\text{nombre de mots de } k \text{ lettres } n^k}$$

Nombre $a_{d,k}(n)$ de mots de k lettres dont d différentes :

$$a_{d,k}(n) = a_{d,k}(k) C_n^d$$

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\mathbf{P}_{d,k}(n) = \frac{a_{d,k}(k) C_n^d}{n^k}$$

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

constante par rapport à n

$$\mathbf{P}_{d,k}(n) = \frac{a_{d,k}(k) C_n^d}{n^k}$$

Calculs

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

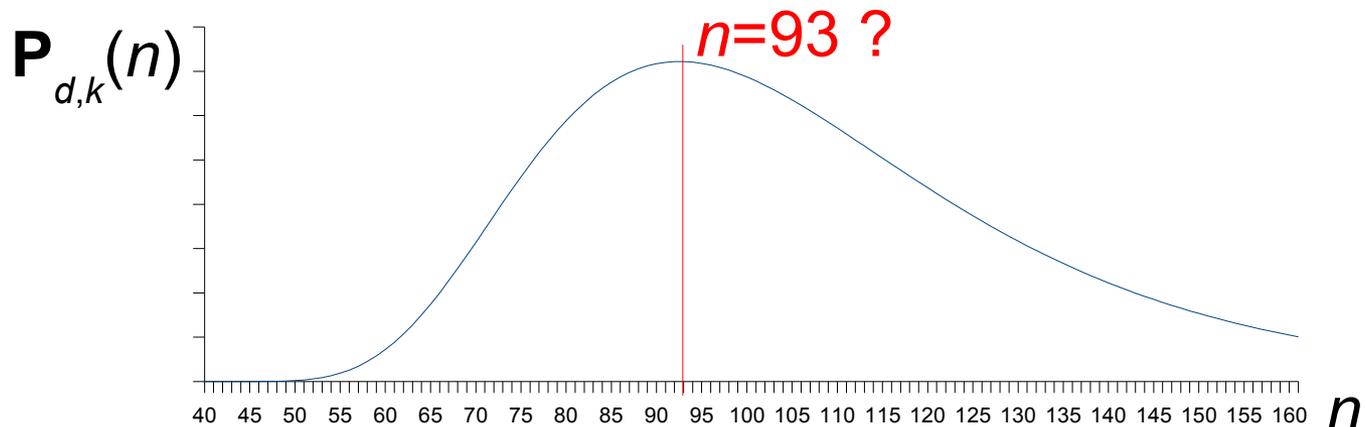
$$\operatorname{argmax}_n \mathbf{P}_{d,k}(n) = \operatorname{argmax}_n \frac{\mathbf{C}_n^d}{n^k}$$

Résultats

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

$$\operatorname{argmax}_n \mathbf{P}_{d,k}(n) = \operatorname{argmax}_n \frac{\mathbf{C}_n^d}{n^k}$$

Pour les papillotes Révillon “Festives” pour $k=52$ et $d=40$:

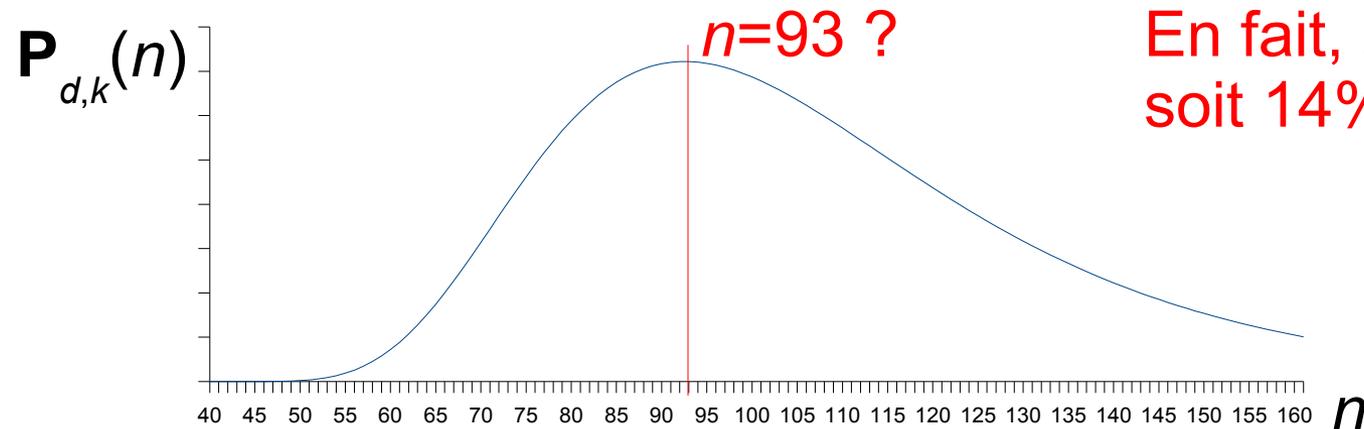


Résultats

Trouver la valeur de n qui maximise la probabilité $\mathbf{P}_{d,k}(n)$ de tirer un mot de k lettres ayant exactement d lettres différentes dans un alphabet de n lettres.

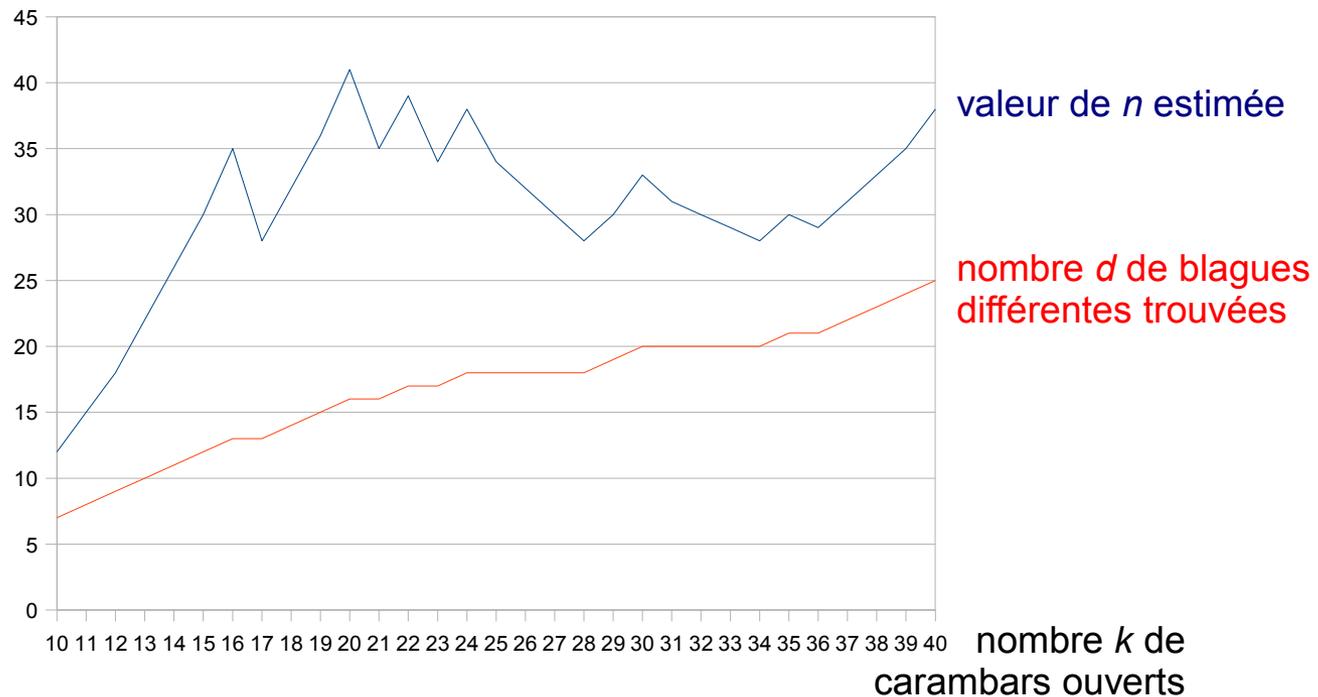
$$\operatorname{argmax}_n \mathbf{P}_{d,k}(n) = \operatorname{argmax}_n \frac{C_n^d}{n^k}$$

Pour les papillotes Révillon “Festives” pour $k=52$ et $d=40$:



Résultats

Evolution du nombre de blagues Carambar “Caramel” estimé en fonction de la taille du tirage :



Perspectives

- étude de la précision de la méthode par simulations
- formule directe pour la valeur de n estimée
- utilisations d'autres caractéristiques du tirage pour une évaluation plus précise :
 - nombre de citations présentes deux fois
 - distribution des nombres d'apparition de citations
 - taille la plus longue d'une séquence de blagues consécutives
- estimation plus précise du nombre de blagues Carambar
- estimation du nombre de surprises Kinder



ebay.fr

