Exercices sur les algorithmes probabilités
Nom du rédacteur : Johanne Cohen
`https://www.lri.fr/~jcohen/pages/enseignement.html`

Algorithmes probabilistes : ranger n balles pour m boîtes

Le problème concerne le placement aléatoire de n balles dans m boîtes. Chaque balle est placée dans l'une des boîtes de façon uniforme. Une fois que toutes les balles sont dans les boîtes, nous regardons le nombre de balles dans chaque boîte ; nous appelons ce nombre la charge de la boite. Nous nous intéressons au problème suivant : quelle est la charge maximale d'un seul bac ?

Notons par X_i la variable aléatoire désignant le nombre de balles contenue dans la $i$ème boîte.

Question 1. Quelle devrait être la valeur de n pour que chaque boîte contient au moins une balle en moyenne ($\forall i \; E[X_i] \geq 1$) ?

C'est une reformulation du problème de collection de coupons. Ici

— les balles sont les boites de céréales

— les boites sont les vignettes.

Question 2. Calculer $E[X_i]$?

Remarquons

$$E[X_1] = E[X_2] = \cdots = E[X_m]$$

Comme nous avons n balles,

$$\sum_{i=1}^{n} E[X_i] = n$$

Donc

$$E[X_i] = \frac{n}{m}$$

Question 3. Calculer $Pr[X_i = k]$?

La variable X_i correspond à une loi binomiale qui modélise la fréquence du nombre de succès obtenus lors de la répétition de n expériences aléatoires identiques et indépendantes sachant que la probabilité de succès vaut p.

$$Pr[X_i = k] = \left(\begin{array}{c} n \\ k \end{array} \right) (\frac{1}{n})^k (1 - \frac{1}{n})^{n-k}$$

Question 4. Soit S un sous-ensemble de $\{1, 2, \cdots, n\}$. Notons par $\mathcal{A}_S$ l'évènement que la balle s soit dans la boite i pour tout $s \in S$.

Donner la probabilité de $Pr(\mathcal{A}_S)$?

Comme les tirages sont indépendants, nous nous focalisons uniquement pour les balles de S. L'évènement que "la balle s soit dans la boite i" se réalise avec la probabilité $\frac{1}{m}$.

$$Pr(\mathcal{A}_S) = p^{|S|} = \left(\frac{1}{m}\right)^{|S|}$$

Question 5. Montrer $Pr[X_i \geq k] \leq \left(\begin{array}{c} n \\ k \end{array}\right)\left(\frac{1}{m}\right)^k$? *utiliser la question précédente et l'inégalité de Boole.*

L'évènement "$X_i \geq k$" se réalise si il existe un sous ensemble $S \subseteq \{1, \ldots n\}$ tel que l'évènement se réalise $\mathcal{A}_S$ et tel que $|S| = k$.

$$\begin{aligned}
\Pr[X_i \geq k] = & \quad \Pr[\exists S \subseteq \{1, \ldots n\} \textit{ tel que } \mathcal{A}_S \textit{ et } |S| = k] && (1)\\
\leq & \quad \Pr[\cup_{S \subseteq \{1,\ldots n\} \textit{ et } |S|=k} \mathcal{A}_S] && (2)\\
\leq & \quad \cup_{S \subseteq \{1,\ldots n\} \textit{ et } |S|=k} Pr(\mathcal{A}_S) && (3)\\
\leq & \quad \left(\begin{array}{c} n \\ k \end{array}\right)\left(\frac{1}{m}\right)^k && (4)
\end{aligned}$$

Question 6. Montrer que $Pr[X_i \geq k] \leq \left(\frac{ne}{mk}\right)^k$ sachant que $(\frac{n}{k})^k \leq \left(\begin{array}{c} n \\ k \end{array}\right) \leq (\frac{en}{k})^k$?

$$Pr[X_i \geq k] \leq \left(\begin{array}{c} n \\ k \end{array}\right) \cdot \frac{1}{m^k} \leq \left(\frac{ne}{k}\right)^k \cdot \frac{1}{m^k}$$

En supposant qu'il existe le même nombre de boîtes et de balles : $n = m$

Question 7. En déduire $Pr(X_i \geq k) \leq (\frac{e}{k})^k$ sachant que $(\frac{n}{k})^k \leq \left(\begin{array}{c} n \\ k \end{array}\right) \leq (\frac{en}{k})^k$?

En reprenant la question précédente

$$\begin{aligned}
Pr[X_i \geq k] \leq & \quad \left(\begin{array}{c} n \\ k \end{array}\right)\left(\frac{1}{m}\right)^k && (5)\\
Pr[X_i \geq k] \leq & \quad \left(\begin{array}{c} n \\ k \end{array}\right)\left(\frac{1}{n}\right)^k && (6)\\
Pr[X_i \geq k] \leq & \quad \left(\frac{en}{k}\right)^k\left(\frac{1}{n}\right)^k && (7)\\
Pr[X_i \geq k] \leq & \quad \left(\frac{e}{k}\right)^k && (8)
\end{aligned}$$

Question 8. En posant $k = \frac{3 \ln n}{\ln \ln n}$, montrer que $\Pr(X_i \geq k) \leq \frac{1}{n^2}$

$$
\begin{aligned}
\Pr[X_i \geq k] \leq & \qquad (\frac{e}{k})^k && (9)\\
\leq & \qquad (\frac{e \ln\ln n}{3\ln n})^{\frac{3\ln n}{\ln\ln n}} && (10)\\
\leq & \qquad exp(\frac{3\ln n}{\ln\ln n}(\ln\ln\ln n - lnn \ln n)) && (11)\\
\leq & \qquad exp(-3\ln n + \frac{3\ln n(\ln\ln\ln n)}{\ln\ln n}) && (12)\\
\leq & \qquad exp(-2\ln n) && (13)\\
\leq & \qquad \frac{1}{n^2} && (14)
\end{aligned}
$$

Notons par X, la variable aléatoire désignant la charge maximum des boîtes

$$X = maxX_i$$

Question 9. Montrer que $Pr(X \geq k) \leq n(\frac{e}{k})^k$.

L'évènement "$X \geq k$" se réalise si il existe une boite i telle que "$X_i \geq k$" se réalise.

$$Pr(X \geq k) = Pr(\exists i | X_i \geq k) \leq Pr(\cup_{i=1}^{n}(X_i \geq k))$$

En utilisant la relation Union Bound

$$
\begin{aligned}
\Pr(X \geq k) &\leq \sum_{i=1}^{n} \Pr(X_i \geq k))\\
&\leq \sum_{i=1}^{n} \left(\frac{e}{k}\right)^k\\
&\leq n\left(\frac{e}{k}\right)^k
\end{aligned}
$$

Question 10. En posant $k = \frac{3\ln n}{\ln\ln n}$, montrer que $Pr[X \geq k] \leq \frac{1}{n}$.

Tout d'abord, nous allons nous concentrer sur le fait que $Pr[X \geq k] \leq n\left(\frac{e}{k}\right)^k$

$$
\begin{aligned}
Pr[X \geq k] = & \qquad Pr[\textit{il existe un sac ayant au moins } k \textit{ balles}] && (15)\\
\leq & \qquad n\frac{1}{n^2} && (16)\\
\leq & \qquad \frac{1}{n} && (17)
\end{aligned}
$$

Question 11. En posant $k = \frac{3\ln n}{\ln\ln n}$, montrer que $Pr(X < k)) \geq 1 - \frac{1}{n}$

De la question précédente on peut déduire que

$$Pr[\textit{tous les sacs ont au plus } k \textit{ balles}] \geq 1 - \frac{1}{n}$$

Question 12. Soit k fixé. Montrer que $E[X] \leq kPr(X < k) + nPr(X \geq k)$.

Comme $X \leq n$,

$$\begin{aligned} E[X] = & \sum_{i=1}^{n} iPr(X = i) && (18) \\ = & \sum_{i=1}^{k-1} iPr(X = i) + \sum_{i=k}^{n} iPr(X = i) && (19) \end{aligned}$$

Observons que

1. $\sum_{i=1}^{k-1} iPr(X = i) \leq k \sum_{i=1}^{k} Pr(X = i) = kPr(X < k)$.
2. $\sum_{i=k}^{n} iPr(X = i) \leq n \sum_{i=k}^{n} Pr(X \geq k)$.

$$E[X] \leq kPr(X < k) + nPr(X \geq k)$$

Question 13. Lorsque $m = n$, quelle est la charge moyenne maximale sur l'ensemble des boîtes ?

Maintenant il faut instancier pour $k = \frac{3 \ln n}{\ln \ln n}$.

$$\begin{aligned} E[X] \leq & \; kPr(X < k) + nPr(X \geq k) && (20) \\ \leq & \; k + n\frac{1}{n} && (21) \\ \leq & \; \frac{3 \ln n}{\ln \ln n} + 1 && (22) \end{aligned}$$