RAPPORT SUR L'ÉPREUVE
MATHÉMATIQUES DE L'INFORMATIQUE

1. Description du problème.

Le problème portait sur l'étude d'algorithmes d'optimisation combinatoire liés à l'empilement de "paquets" dans des boîtes de capacité fixe.

La partie I avait pour objet de présenter quelques propriétés combinatoires simples liés à la "physique" du problème. La partie II avait pour objet l'étude combinatoire du placement séquentiel. Les parties III et IV étaient de nature plus analytique et visaient à établir le coût moyen du placement séquentiel décroissant ainsi que du placement optimal, sur des données uniformément réparties. La partie V consistait à établir par une chaine de réductions le caractère NP-complet du problème.

Le problème ne présuppose aucune connaissance combinatoire particulière (si ce n'est un peu de bon sens et d'aptitude au raisonnement). Les parties III et IV utilisent les notions élémentaires de probabilités et d'espérance (redéfinies dans ce cas particulier), et quelques connaissances du programme d'analyse. La partie V necessitait, pour être traitée convenablement que les candidats aient acquis la notion de probème NP-complet et de réduction polynomiale qui figure au programme de l'option.

La logique du problème est simple: la détermination de l'empilement optimal est un problème pour lequel on ne connait pas de solution efficace (NP-complétude même dans le cas de 2 boîtes), alors que sur des données uniformément réparties sur [0,1], on peut proposer des heuristiques satisfaisantes.

2. Commentaires sur les copies.

Pratiquement tous les candidats ont abordé la partie I qui constituait un test de bon sens mathématique. Les correcteurs ont été sur cette partie, et plus généralement sur les questions combinatoires du problème, particulièrement sensibles à la clarté de raisonnement des candidats. Un quart environ des copies présentent en effet des preuves fausses ou des arguments de récurrence qui ne s'appliquent pas aux questions.

Les questions de nature plus analytiques, lorsqu'elles étaient abordées, ont souvent été traitées correctement, ce qui s'explique sans doute par leur caractère relativement classique. Cependant, certains candidats fournissent des réponses (Q. 8) en contradiction flagrante avec le fait que l'espérance d'une somme est la somme des espérances ..., et manquent d'intuition probabiliste élémentaire.

La partie V a généralement été bien traitée par les candidats qui l'avaient abordée et a permis à un certain nombre de candidats de remonter efficacement leur note globale.

La répartition des notes sur 119 copies corrigées (non blanches) était la suivante:

Notes:	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40
Nb. Copies:	12	35	27	26	11	4	2	1	1

Il apparait que 25% des candidats ont obtenu une note ≤ 7, que 50% ont obtenu une note ≤ 11 et 25% une note ≥ 17. La moyenne s'établissait à 12,90 points sur un total de 40.

3. Commentaires sur les questions.

PARTIE I. GÉNÉRALITÉS. Cette partie (sauf peut-être la borne supérieure de la Q. 2) ne présente pas de difficulté particulière.

Q1. Soit $\mathcal{E}(X;c)$ l'ensemble des empilements de l'entrée $(X;c)$. Il est clair qu'on a les implications

$$\mathcal{E}(X;c)=\mathcal{E}(Y;c) \Longrightarrow \omega(X;c)=\omega(Y;c) \quad \text{et} \quad \mathcal{E}(X;c)\subseteq\mathcal{E}(Y;c) \Longrightarrow \omega(X;c)\geq\omega(Y;c). \quad (1)$$

Le principe (1) et des raisonnements découlant directement des définitions permettent de conclure.

Q2. La borne inférieure est évidente. La borne supérieure signifie que, dans l'empilement optimal, toute boîte, sauf au plus une, est au moins remplie à moitié (sinon on pourrait combiner 2 boîtes). Formellement, soient B_j le contenu de la boîte j dans un empilement optimal de coût M. On a la suite d'inégalités

$$B_1+B_2>c,\ B_2+B_3>c,\ \ldots,\ B_M+B_1>c$$

d'où résulte par sommation la majoration.

Q3. Prendre des X_j qui valent 0 ou $\frac{1}{2}+\epsilon$ avec un choix convenable de la quantité ϵ.

Q4. Prendre $X_1=X_2=X_3=\frac{1}{3}$ et $X_4=X_5=\cdots=\frac{2}{3}$. Pour la condition supplementaire (C3), modifier légèrement ces valeurs.

PARTIE II. LE PLACEMENT SÉQUENTIEL. La seule question délicate (et la plus interessante) était Q. 7 qui a été abordée avec succès par quelques candidats.

Q5. La preuve de correction de l'algorithme s'effectue par récurrence sur N. La borne supérieure de l'inégalité découle de ce que la somme des remplissages de 2 boîtes consécutives est toujours >1 (cf. Q. 2)

Q6. Utiliser des suites construites selon le schéma $X_1=X_3=\ldots=1-\epsilon$ et $X_2=X_4=\ldots=\eta$ pour des valeurs convenables de ϵ et η.

Q7. Soit $g(i)$ l'indice tel que $X_{g(i)}$ soit le premier élément envoyé dans la boîte i. On observe, et c'est là le point essentiel, que $g(i)\geq 2^i$ (preuve par récurrence; toute borne inférieure exponentielle pour $g(i)$ convient). Le remplissage de la boîte i est alors au moins $1-1/2^{i+1}$, d'où découle le résultat par sommation des "vides" de chaque boîte.

PARTIE III. LE PLACEMENT SÉQUENTIEL DÉCROISSANT. Contrairement au précédent, cet algorithme s'analyse en moyenne assez simplement (cf Q. 10 et suivantes) et utilise en moyenne $0,6449N$ boites pour loger N paquets (de poids total en moyenne $0,5N$). L'analyse conduit à quelques estimations probabilistes établies par le calcul intégral classique.

Q8. Un calcul direct d'intégrale montre que cette espérance vaut $N/2$. C'est aussi la somme de N variable aléatoires d'espérance $1/2$.

Q9. On calcule les premières valeurs, et l'on vérifie par récurrence que

$$\psi_k(y) = \frac{y^{k+1}}{(k+1)!}.$$

Il en découle que

$$\begin{aligned}\phi_k(y) &= \int_{1-y}^{1} \Pr[1 - X_{k+1} \leq X_1 + X_2 + \cdots + X_k \leq y]\, dX_{k+1} \\ &= \frac{ky^{k+1}}{(k+1)!}.\end{aligned}$$

La quantité $\phi_k(1) = k/(k+1)!$ mesure la probabilité que la première boîte contienne exactement $X_1, \ldots, X_k$. On obtient donc

$$\bar{\alpha}_N = \sum_{k=1}^{N} \frac{k^2}{(k+1)!},$$

et la limite lorsque $N \to \infty$ de $\bar{\alpha}_N$ vaut $e - 1 = 1.71828\ldots$.

Q10. On vérifie facilement par récurrence que si $X_i \leq Y_i$ pour tout i, alors $\gamma_{NFD}(X;1) \leq \gamma_{NFD}(Y;1)$.

(i). Borne supérieure. On considère alors le vecteur Y, tel que $X_i \leq Y_i$, défini, avec des notations évidentes, par

$$Y = (\frac{1}{1})^{w_1(X)} (\frac{1}{2})^{w_2(X)} \cdots (\frac{1}{r})^{w_r(X)}$$

pour lequel

$$\gamma_{NFD}(Y;1) \;\leq\; \sum_{l=1}^{r} \lceil \frac{w_l(X)}{l} \rceil \;\leq\; r + \sum_{l=1}^{r} \frac{w_l(X)}{l}.$$

(ii). Borne inférieure. Raisonnement du même type.

Q11. La probabilité que les n_1 premiers éléments appartiennent à I_1 etc. vaut $\iota_1^{n_1} \iota_2^{n_2} \cdots \iota_r^{n_r}$ On muliplie cette quantité par le nombre de choix possibles (le coefficient multinomial) pour obtenir $p_{\vec{n}}$.

Q12. Par la formule multinomiale,

$$Q(t_1, t_2, \ldots, t_r) = (\iota_1 t_1 + \iota_2 i_2 + \cdots + \iota_r t_r)^N.$$

Par dérivation par rapport aux t_j, puis en faisant $t_j = 1$, on obtient l'espérance de $W_r(X)$:

$$N\left(\frac{1}{r^2} + \sum_{l=1}^{r-1} \frac{1}{l^2(l+1)}\right),$$

et celle de $V_r(X)$:

$$N\sum_{l=1}^{r-1}\frac{1}{l^2(l+1)}.$$

Q13. De Q. 12 et de l'encadrement de Q. 10, on trouve

$$\frac{\bar{\gamma}_{NFD}(N)}{N}\to\frac{\pi^2}{6}-1\approx 0,6449.$$

PARTIE IV. COÛT DU PLACEMENT OPTIMAL. L'algorithme de placement complémentaire est un algorithme "théorique" (cf Q. 22) dont l'introduction permet d'établir le coût moyen du placement optimal. L'analyse se ramène à l'évaluation d'irrégularités de distribution, traitées par séries génératrices (Q. 15) et analyse complexe. On constate (Q. 18) qu'en moyenne, le placement PC et le placement optimal utilisent $\sim N/2$ boîtes.

Q14. Utiliser l'égalité $\gamma_{PC}=N-\sum_{j=1}^{s}\lambda_j$ et l'inégalité $\lambda_j\geq\frac{N}{r}-\Delta_r$.

Q15. D'après Q. 11,

$$p'_{\vec{n}}=\frac{N!}{n_1!n_2!\cdots n_r!}\left(\frac{1}{r}\right)^N$$

et par sommation,

$$\begin{aligned}q_{N,b}&=\sum_{n_i>b}\left(\frac{N!}{n_1!n_2!\cdots n_r!}\left(\frac{1}{r}\right)^N\right)\\&=\frac{N!}{r^N}[z^N]\left(e^z-e_b(z)\right)^r\end{aligned}$$

où $[z^N]f(z)$ représente le coefficient de z^N dans le développement de Taylor de $f(z)$ à l'origine.

Q16. L'expression intégrale de $q_{N,b}$ s'établit par le théorème de Cauchy à partir de Q. 15. La majoration résulte de bornes triviales en observant que l'integrande est à coefficients positifs.

Q17. En notant que $(1-t)^r\geq 1-rt$ pour $0\leq t\leq 1$ et $r\geq 0$, on obtient

$$1-q_{N,b}\leq\frac{N!e^N}{N^N}rt\qquad\text{avec}\qquad t=\frac{e^{-N/r}}{b!}(b+1)(\frac{N}{r})^b,$$

et le calcul se conclut par application de la formule de Stirling et de développements asymptotiques élémentaires de $\log x$ et $\exp x$.

Q18. On utilise les valeurs de b et r dans Q. 17. Soit X tel que $\forall l,\ \varsigma_l(X)>b$. On a alors $\Delta_r(X)=O(\sqrt{N})$ et d'après Q. 14:

$$\gamma_{PC}(X;1)\leq\frac{N}{2}+O(N^{5/6}).$$

Cet évènement se produit avec probabilité $q_{N,b}$. Dans le cas contraire, et avec probabilité $(1-q_{N,b})$, on majore $\gamma_{PC}(X)$ par N. Donc, l'espérance de $\gamma_{PC}(X;1)$ est majorée par

$$\bar{\gamma}_{PC}(N)\ \leq\ q_{N,b}(\frac{N}{2}+O(N^{5/6}))+(1-q_{N,b})N\ \leq\ \frac{N}{2}+O(N^{5/6}).$$

Comme $\bar{\pi}(N)=N/2$, on en déduit

$$\lim_{N\to\infty}\frac{\bar{\gamma}_{PC}(N)}{N}=\frac{1}{2}$$

ainsi que $\bar{\omega}(N)\sim N/2$.