$\boxed{\textbf{Épreuve ESSEC 2011 option économique Math II - Un corrigé}}$

Dans ce problème nous supposerons que le jeu de cartes est constitué d'au moins 2 cartes, soit $N \geq 2$.

1) Pour $i \in [\![2, n-1]\!]$, $\displaystyle\sum_{k=1}^{i} \Delta_k = T_1 + \sum_{k=2}^{i}(T_k - T_{k-1}) = T_i$ (somme télescopique).

   Δ_i représente le temps qu'il faut pour que la carte C_N passe de la position $N - i + 1$ à la position $N - i$, c'est-à-dire le temps qu'il faudra, à partir de l'instant T_{i-1}, pour obtenir une insertion entre les places N et $N - i + 1$ (compris).

2) $(\Delta_1 > n)$ est l'événement : pendant les n premiers instants les numéros d'insertions sont toujours entre 1 et $N - 1$.

   Par hypothèse, un numéro d'insertion entre 1 et $N - 1$ est choisi avec probabilité $\frac{N-1}{N}$ (loi uniforme) et les choix des numéros d'insertions sont supposés indépendants donc

$$\mathbf{P}(\Delta_1 > n) = \left(\frac{N-1}{N}\right)^n$$

   donc $\mathbf{P}(\Delta_1 = n) = \mathbf{P}(\Delta_1 > n-1)) - \mathbf{P}(\Delta_1 > n) = \dfrac{1}{N}\left(1 - \dfrac{1}{N}\right)^{n-1}$

   Ainsi $\boxed{\Delta_1 \text{ suit une loi géométrique de paramètre } 1/N.}$

3) (a) $(\Delta_i > n)$ correspond à l'événement : "pendant n instants successifs (à partir de T_{i-1}) les numéros d'insertions sont compris entre 1 et $N - i$ compris".

   Par hypothèse, un numéro d'insertion entre 1 et $N-i$ est choisi avec probabilité $\frac{N-i}{N}$ (loi uniforme) et les choix des numéros d'insertions sont supposés indépendants donc

$$\mathbf{P}(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n$$

   donc $\mathbf{P}(\Delta_i = n) = \mathbf{P}(\Delta_i > n-1)) - \mathbf{P}(\Delta_i > n) = \dfrac{i}{N}\left(1 - \dfrac{i}{N}\right)^{n-1}$

   On en conclut que $\boxed{\Delta_i \text{ suit une loi géométrique de paramètre } i/N.}$

   (b) Par un résultat de cours sur les lois géométriques

$$E(\Delta_i) = \frac{1}{i/N} = \frac{N}{i} \text{ et } V(\Delta_i) = \frac{1 - i/N}{(i/N)^2} = \frac{N(N-i)}{i^2}$$

4) (a) $T_2 = \Delta_1 + \Delta_2$ donc

$$(T_2 = n) = \bigcup_{k=1}^{n-1} (\Delta_1 = k) \cap (\Delta_2 = n - k)$$

$$\begin{aligned}
\mathbf{P}(T_2 = n) &= \sum_{k=1}^{n-1} \mathbf{P}\left[(\Delta_1 = k) \cap (\Delta_2 = n-k)\right] && \text{union d'événements incompatibles} \\
&= \sum_{k=1}^{n-1} \mathbf{P}(\Delta_1 = k)\mathbf{P}(\Delta_2 = n-k) && \Delta_1 \text{ et } \Delta_2 \text{ sont indépendantes}
\end{aligned}$$

(b) Calcul d'une somme géométrique de raison $\rho = \frac{1-1/N}{1-2/N} \neq 1$.

$$\begin{aligned}
\sum_{k=1}^{n-1} \left(\frac{1-1/N}{1-2/N}\right)^k &= \frac{1-1/N}{1-2/N} \sum_{k=0}^{n-2} \left(\frac{1-1/N}{1-2/N}\right)^k \\
&= \frac{1-1/N}{1-2/N} \frac{1-\left(\frac{1-1/N}{1-2/N}\right)^{n-1}}{1-\frac{1-1/N}{1-2/N}} \\
&= \frac{1-1/N}{1-2/N} \frac{1-2/N}{(1-2/N)^{n-1}} \frac{(1-2/N)^{n-1}-(1-1/N)^{n-1}}{-1/N} \\
&= N\left(1-\frac{1}{N}\right)\left[\left(\frac{1-1/N}{1-2/N}\right)^{n-1}-1\right]
\end{aligned}$$

(c) En utilisant les lois de Δ_1 et Δ_2, on obtient

$$\begin{aligned}
\mathbf{P}(T_2 = n) &= \frac{1}{N}\frac{2}{N} \sum_{k=1}^{n-1} \left(1-\frac{1}{N}\right)^{k-1} \left(1-\frac{2}{N}\right)^{n-k-1} \\
&= \frac{2}{N^2} \frac{\left(1-\frac{2}{N}\right)^{n-1}}{1-\frac{1}{N}} \sum_{k=1}^{n-1} \left(\frac{1-1/N}{1-2/N}\right)^k \\
&= \frac{2}{N^2} \frac{\left(1-\frac{2}{N}\right)^{n-1}}{1-\frac{1}{N}} N\left(1-\frac{1}{N}\right)\left[\left(\frac{1-1/N}{1-2/N}\right)^{n-1}-1\right] \\
&= \frac{2}{N}\left[\left(1-\frac{1}{N}\right)^{n-1}-\left(1-\frac{2}{N}\right)^{n-1}\right]
\end{aligned}$$

5) Les deux probabilités demandées sont égales (il y a autant de chances que la carte placée à l'instant T_2 le soit en position $N-1$ ou N, l'autre est alors à l'autre place possible). Ainsi, les deux probabilités valent $1/2$.

6) Il y a $3! = 6$ possibilités pour le triplet (a_1, a_2, a_3).
Il y a autant de chances que la carte placée à l'instant T_3 le soit en position $N-2$, $N-1$ ou N. Et vu ce qui précède sur les cartes placées aux instants T_1 et T_2, on en déduit que les $3! = 6$ configurations possibles sont équiprobables.
Dans les deux exemples cités, les probabilités sont alors égales à $1/3! = 1/6$.

7) Comme dans les exemples précédents, à l'instant $T-1 = T_{N-1}$ la carte C_N se trouve au dessus du paquet et toutes les configurations des $N-1$ cartes situées au dessous sont équiprobables (ce qui s'établit prouve par récurrence).
À l'instant $T = T_{N-1}+1$, on glisse la carte C_N en position k en tirant le nombre k au hasard selon une loi uniforme sur $[\![1, N]\!]$. On obtient alors un jeu de cartes dans l'une des $N!$ configurations possibles, avec autant de chances pour chacune des configurations.

8) $T = \Delta_1 + \Delta_2 + \cdots + \Delta_{N-1} + 1$. Par linéarité de l'espérance :

$$\begin{aligned}
E(T) &= E(\Delta_1) + E(\Delta_2) + \cdots + E(\Delta_{N-1}) + 1 \\
&= \sum_{i=1}^{N-1} \frac{N}{i} + 1 = \sum_{i=1}^{N} \frac{N}{i} \\
&= N H_N
\end{aligned}$$

En utilisant l'indépendance des variables $(\Delta_1, \cdots, \Delta_{N-1}, 1)$, on obtient

$$\begin{aligned}
V(T) &= V(\Delta_1) + V(\Delta_2) + \cdots + V(\Delta_{N-1}) + V(1) \\
&= \sum_{i=1}^{N-1} \frac{N(N-i)}{i^2} + 0 \\
&= N^2 \sum_{i=1}^{N-1} \frac{1}{i^2} - N \sum_{i=1}^{N-1} \frac{1}{i} \\
&= N^2 \sum_{i=1}^{N} \frac{1}{i^2} - N \sum_{i=1}^{N} \frac{1}{i} && \text{car } 1 = N^2 \frac{1}{N^2} = N\frac{1}{N}
\end{aligned}$$

9) (a) Soit $k \in \mathbb{N}^*$. La fonction $f : t \mapsto 1/t$ est décroissante sur l'intervalle $[k, k+1]$. Donc $\forall t \in [k, k+1]$, $\frac{1}{k+1} \leq \frac{1}{t} \leq \frac{1}{k}$. En intégrant cette inégalité sur l'intervalle $[k, k+1]$, on obtient l'encadrement demandé.

(b) (i) $u_{n+1} - u_n = H_{n+1} - H_n - \ln(n+1) + \ln n = \frac{1}{n+1} - [\ln(n+1) - \ln(n)] = \frac{1}{n+1} - \int_n^{n+1} \frac{\mathrm{d}t}{t} \leq 0$ (en utilisant l'inégalité de gauche de l'encadrement qui précède). La suite u est donc décroissante.

(ii) Pour $n \in \mathbb{N}^*$, sommons les inégalités de droite de l'encadrement de la question 9.a pour $k = 1..n$. On obtient alors

$$\underbrace{\int_1^{n+1} \frac{\mathrm{d}t}{t}}_{=\ln(n+1)} \leq \underbrace{\sum_{k=1}^{n} \frac{1}{k}}_{=H_n}$$

Si $n = 1$, $H_1 = 1 \leq \ln(1) + 1$. Si $n \geq 2$, on somme les inégalités de gauche pour $k = 1$ à $n - 1$ et on obtient : $(H_n - 1) \leq \int_1^n \frac{\mathrm{d}t}{t} = \ln(n)$.

(c) $\forall n \in \mathbb{N}^*, \quad 0 \leq \ln(n+1) - \ln(n) \leq u_n \leq 1$.
Décroissante et minorée par 0, la suite (u_n) est donc convergente. L'encadrement permet de dire que sa limite γ appartient au segment $[0, 1]$.

10) (a) En utilisant les questions 8 et 9.b.ii on sait que :
$N \ln(N+1) \leq E(T) = NH_N \leq N \ln(N) + N$. Donc

$$\frac{\ln(N+1)}{\ln(N)} \leq \frac{E(T)}{N \ln(N)} \leq 1 + \frac{1}{\ln(N)}$$

Or $\ln(N+1) = \ln(N) + \ln\left(1 + \frac{1}{N}\right)$, donc $\frac{\ln(N+1)}{\ln(N)} = 1 + \frac{\ln\left(1 + \frac{1}{N}\right)}{\ln(N)} \longrightarrow 1$

Ainsi, par encadrement, la suite $\left(\frac{E(T)}{N\ln(N)}\right)_N$ tend vers 1, c'est-à-dire $E(T) \sim N \ln N$.
De plus $E(T) - N\ln(N) = Nu_N = N(\gamma + o(1)) = N\gamma + o(N)$.
N.B. Remarquons que ce développement asymptotique permet de retrouver directement l'équivalent $E(T) \sim N \ln(N)$ car $N = o(N \ln(N))$.

(b) Pour tout entier $n \geq 1$, posons $S_n = \sum_{k=1}^{n} 1/k^2$. Cette suite $(S_n)_n$ est convergente (série de Riemann). Notons α sa limite. On sait que $(S_n)_n$ est croissante donc

$$\forall n \geq 1, \quad \alpha \geq S_n \geq S_1 > 0$$

Or $\frac{V(T)}{N^2} = S_N - \frac{H_N}{N}$ et $\frac{H_N}{N} \sim \frac{\ln(N)}{N} \longrightarrow 0$, donc la suite $(V(T)/N^2)_N$ est convergente vers α.
On a bien

$$V(T) \sim \alpha N^2$$

De plus,

$$\forall N \in \mathbb{N}^*, \ \frac{V(T)}{N^2} = S_N - \frac{H_N}{N} \leq \alpha - \frac{H_N}{N} \leq \alpha$$

11) (a) •Pour $\omega \in \Omega$, $T(\omega) - N\ln(N) = T - E(T) + Nu_N$. En utilisant l'inégalité triangulaire, il vient

$$|T(\omega) - N\ln(N)| \leq |T(\omega) - E(T)| + |Nu_N| \leq |T(\omega) - E(T)| + N$$

car avec 9.b.ii, $0 \leq Nu_N \leq N$.

•Supposons que pour un $\omega \in \Omega$, on ait $|T(\omega) - N\ln(N)| \geq cN$, alors grâce à l'inégalité précédente, on en déduit : $|T(\omega) - E(T)| + N \geq cN$, donc $|T(\omega) - E(T)| \geq (c-1)N$. On a donc l'inclusion des événements :

$$\Big(|T - N\ln(N)| \geq cN\Big) \subset \Big(|T - E(T)| \geq N(c-1)\Big)$$

(b) On en déduit $\mathbf{P}\Big(|T - N\ln(N)| \geq cN\Big) \leq \mathbf{P}\Big(|T - E(T)| \geq N(c-1)\Big)$ et en utilisant l'inégalité de Bienaymé-Chebychev à la variable aléatoire T, on a la majoration

$$\mathbf{P}\Big(|T - E(T)| \geq N(c-1)\Big) \leq \frac{V(T)}{(N(c-1))^2}$$

En utilisant enfin la question 10b, on peut conclure

$$\mathbf{P}\Big(|T - N\ln(N)| \geq cN\Big) \leq \mathbf{P}\Big(|T - E(T)| \geq N(c-1)\Big) \leq \frac{\alpha}{(c-1)^2}$$

Par encadrement, si N est fixé : $\lim\limits_{c \to +\infty} \mathbf{P}\Big(|T - N\ln N| \geq cN\Big) = 0$

12) Soit $\varepsilon > 0$. En utilisant l'inégalité de la question 11.b avec $c = \varepsilon\ln(N)$ (c peut être supposé > 1 pour N assez grand), on obtient

$$\mathbf{P}\Big(|T - N\ln(N)| \geq \varepsilon N\ln(N)\Big) \leq \frac{\alpha}{(\varepsilon\ln(N) - 1)^2} \longrightarrow_{N \to +\infty} 0$$

Par encadrement :

$$\lim_{N \to +\infty} \mathbf{P}\Big(|T - N\ln(N)| \geq \varepsilon N\ln(N)\Big) = 0$$

13)
```
PROGRAM ESSEC2011;
TYPE Paquet=ARRAY[1..32] OF INTEGER;
VAR Jeu :Paquet;
    S,k :INTEGER
```
a)
```
PROCEDURE Init( VAR Jeu :Paquet);
VAR k :INTEGER;
BEGIN
   FOR k :=1 TO 32 DO Jeu[k] :=k
END;
```

b)
```
PROCEDURE Insertion(VAR Jeu :Paquet);
VAR i,k,cartedessus :INTEGER;
BEGIN
   k :=1+RANDOM(32);
   cartedessus :=Jeu[1];
   IF k>1 THEN FOR i :=1 TO k-1 DO Jeu[i] := Jeu[i+1];
   Jeu[k] :=cartedessus
END;
```
c) Comme écrite dans le sujet, la fonction T est une simulation de la variable $T - 1 = T_{N-1}$. C'est un peu dommage, cela aurait été plus pertinent d'écrire une simulation de la variable T...

d)

```
BEGIN  { programme principal }
   S :=0;
   FOR k :=1 TO 100 DO
      S :=S+T(Jeu);
   WRITE('Moyenne =',S/100)
END.
```

14) (a) Nous supposons d'abord que $n \geq N$.
Si $n \geq T(\omega)$: à l'instant n, toutes les configurations du paquet sont équiprobables, donc : $\mathbf{P}_{(T\leq n)}(E_n) = \pi(A)$.
Ainsi $\mathbf{P}\left(E_n \cap (T \leq n)\right) = \mathbf{P}_{(T\leq n)}(E_n)\mathbf{P}(T \leq n) = \pi(A)\mathbf{P}(T \leq n)$.
Dans le cas où $n < N$, l'événement $(T \leq n)$ est impossible et

$$\mathbf{P}\left(E_n \cap (T \leq n)\right) = 0 = \pi(A)\mathbf{P}(T \leq n)$$

(cependant dans ce dernier cas, le calcul de la probabilité conditionnelle n'est pas bien défini.)

(b) $(E_n \cap (T > n)) \subset (T > n)$ donc $\mathbf{P}(E_n \cap (T > n)) \leq \mathbf{P}(T > n)$.

(c)

$$\begin{aligned}\mu_n(A) = \mathbf{P}(E_n) &= \mathbf{P}(E_n \cap (T \leq n)) + \mathbf{P}(E_n \cap (T > n))\\ &\leq \pi(A)\mathbf{P}(T \leq n) + \mathbf{P}(T > n)\\ &\leq \pi(A) + \mathbf{P}(T > n)\end{aligned}$$

15) (a) μ_n et π étant des probabilités, on a $\mu_n(\overline{A}) = 1 - \mu_n(A)$ et $\pi(\overline{A}) = 1 - \pi(A)$ donc

$$\mu_n(\overline{A}) - \pi(\overline{A}) = \pi(A) - \mu_n(A)$$

(b) L'inégalité de la question 14c donne pour tout $A \subset \mathcal{S}_N$: $\mu_n(A) - \pi(A) \leq \mathbf{P}(T > n)$. En utilisant cette inégalité avec la partie $\overline{A}$ au lieu de A, on a aussi $\mu_n(\overline{A}) - \pi(\overline{A}) = \pi(A) - \mu_n(A) \leq \mathbf{P}(T > n)$ soit finalement

$$|\mu_n(A) - \pi(A)| \leq \mathbf{P}(T > n)$$

16) L'inégalité précédente vaut pour toute partie A de $\mathcal{S}_N$, donc en particulier pour une partie réalisant le maximum ; ainsi $\boxed{d(\mu_n, \pi) \leq \mathbf{P}(T > n)}$ et bien sûr $\underline{0 \leq d(\mu_n, \pi)}$.
D'autre part, $T = \Delta_1 + \Delta_2 + \cdots + \Delta_{N-1} + 1$ est une somme de variables aléatoires finies presque-sûrement (lois géométriques), donc la variable T est elle-même finie presque-sûrement et $\lim_{N\to+\infty} \mathbf{P}(T > n) = 0$.
Par encadrement : $\boxed{\lim\limits_{n\to+\infty} d(\mu_n, \pi) = 0}$.

17) S_1 est la variable aléatoire certaine égale à 1.

18) Soit $k \in [\![2, N]\!]$, $S_k(\Omega) \subset \mathbb{N}^*$. La variable S_k est le temps d'attente d'un premier succès lors de la répétition d'épreuves aléatoires de Bernoulli identiques et indépendantes, la probabilité p_k d'un succès étant égale à la probabilité d'obtenir un des $N - (k - 1)$ timbres encore non reçus parmi les N possibles, soit $p_k = \frac{N-k+1}{N}$. Ainsi, $\boxed{S_k \text{ suit une loi géométrique de paramètre } \frac{N-k+1}{N}.}$ Pour $n \in \mathbb{N}^*$, $\mathbf{P}(S_k = n) = \dfrac{N-(k-1)}{N}\left(\dfrac{k-1}{N}\right)^{n-1}$

19) La loi de S_k est celle de la variable $\Delta_{N-(k-1)}$ (en convenant que Δ_N est la variable certaine égale à 1). En ré-ordonnant l'ordre des termes, on peut écrire $S = S_N + S_{N-1} + \cdots + S_2 + S_1$ et comme les variables S_k sont indépendantes, <u>S suit bien la même loi que la variable</u> : $\Delta_1 + \Delta_2 + \cdots + \Delta_{N-1} + \Delta_N = \underline{T}$

20) Soit $m \in \mathbb{N}^*$.

(a) $(S > m)$ correspond à l'événement "le jour m, la collection des timbres reçus n'est pas encore complète" ou de manière équivalente "le jour m, il existe au moins un des N timbres qui n'a pas été obtenu". Donc $\boxed{(S > m) = B_1^m \cup B_2^m \cup \ldots \cup B_N^m}$.

(b) Soit $j \in [\![1, N]\!]$. $\mathbf{P}(B_j^m) = \left(\frac{N-1}{N}\right)^m = \left(1 - \frac{1}{N}\right)^m$

(c) $\mathbf{P}(S > m) = \mathbf{P}(\bigcup_{j=1}^{N} B_j^m) \leq \sum_{j=1}^{N} \mathbf{P}(B_j^m) = N\left(1 - \dfrac{1}{N}\right)^m$.

21) (a) On établit l'inégalité : $\forall\, x \in]-1, +\infty[,\ \ln(1+x) \leq x$, par exemple grâce à une étude des variations de la fonction $x \mapsto \ln(1+x) - x$ sur $]-1, +\infty[$.

(b) $\ln\left(1 - \frac{1}{N}\right) \leq -\frac{1}{N}$ car $-1/N > -1$ et comme exp est croissante : $e^{m\ln\left(1-\frac{1}{N}\right)} \leq e^{-m/N}$. Donc :

$$\mathbf{P}(T > m) = \mathbf{P}(S > m) = N\left(1 - \frac{1}{N}\right)^m = Ne^{m\ln\left(1-\frac{1}{N}\right)} \leq Ne^{-m/N}$$

22) (a) Soit $c > 0$ et $n \geq N\ln N + cN$ on a : $d(\mu_n, \pi) \leq \mathbf{P}(T > n) \leq Ne^{-n/N}$.
Or $-n/N \leq -\ln(N) - c$ et comme exp est croissante : $e^{-n/N} \leq e^{-\ln(N)-c} = \frac{1}{N}e^{-c}$. Ainsi

$$\boxed{d(\mu_n, \pi) \leq e^{-c}}$$

(b) *Application numérique.* Avec $N = 32$ on cherche n de sorte que $d(\mu_n, \pi) \leq 0.2$. Vu l'inégalité précédente, il suffit de choisir c de sorte que $e^{-c} \leq 0.2$ soit $c \geq \ln 5$. Donc $n \geq N\ln(N) + \ln(5)N = 32\ln(160) \approx 162.4$.
Il faudra donc au moins <u>163 battages par insertions pour considérer le paquet bien mélangé.</u>