<u>Problème</u>: la procédure « $int(10^5.(\sqrt{n} - int(\sqrt{n})))$ »sur TI.83 avec int(x) = E(x) où E(x) désigne la partie entière de x envoie des chiffres 0123456789 par blocs de 5 et chaque chiffre apparaît avec la probabilité $\frac{1}{10}$, on va dire que ces chiffres sont aléatoires, mais un échantillon de taille n peut- il faire office de table de chiffres aléatoires ?
Pour cela on dispose d'un test (dit du poker) car il ne suffit pas que chaque chiffre dans la table apparaisse avec la probabilité $\frac{1}{10}$, il faut aussi vérifier la fréquence de différents blocs.
La description de ce test est celle ci :
Les chiffres sont regroupés par blocs de 5. La probabilité d'un tel bloc est 10^{-5}. Ces blocs sont regroupés en 7 catégories dont les probabilités sont calculées. On compare alors ces probabilités avec les fréquences observées .

Le tableau suivant donne la description des 7 catégories et leurs probabilités:

n°	description	type	exemple	probabilité
1	chiffres différents	abcde	34961	0,3024
2	une paire	aabcd	29512	0,5040
3	deux paires	aabbc	44533	0,1080
4	un triplet	aaabc	60366	0,0720
5	paire triplet	aaabb	23223	0,0090
6	quadruplet	aaaab	29222	0,0045
7	quintuplet	aaaaa	55555	0,0001

Un échantillon de taille 100 est trop petit pour tirer un enseignement du test mais on est gêné par le calcul des fréquences qui se font « à la main » pour créer des échantillons de taille significativement supérieure. L'élaboration d'un programme sur TI.83 envoyant le calcul des fréquences de chacune des 7 catégories permet de faire le test sur un échantillon de taille n = 900 blocs de 5 chiffres (la TI.83 est limitée à 1000 pour la dimension de ses listes), c'est à dire :
On pose $v_n = int(10^5.(\sqrt{n} - int(\sqrt{n})))$ et on teste l'échantillon $\{v_1, v_2,\ldots\ldots,v_{900}\}$. Le programme renvoie alors les fréquences de chaque catégorie qui sont données dans le tableau suivant :

n°	description	type	exemple	fréquence
1	chiffres différents	abcde	34961	0,31
2	une paire	aabcd	29512	0,46556
3	deux paires	aabbc	44533	0,11
4	un triplet	aaabc	60366	0,07222
5	paire triplet	aaabb	23223	0,00667
6	quadruplet	aaaab	29222	0,00222
7	quintuplet	aaaaa	55555	0,03333

Ce qu'on peut dire c'est que le type 7 (aaaaa) est sur représenté et que nécessairement ce déséquilibre se répercute sur les autres types. Peut on en conclure pour autant que toute table de chiffres fournie par la suite (v_n) ne vérifie pas le test du poker? Je ne crois pas car on voit bien que dans cette suite il y a des termes « parasites » qui sont les v_n où n est un carré d'entier.

Dans l'échantillon $\{v_1, v_2, \ldots, v_n\}$ combien y a t'il de termes où l'indice est un carré d'entier?
Entre 1 et n il y a $p = E(\sqrt{n})$ carrés d'entiers car :

$$p \leq \sqrt{n} < p + 1 \Leftrightarrow p^2 \leq n < (p + 1)^2$$

Ces entiers sont donc $1^2, 2^2, 3^2, \ldots, p^2$ et la fréquence de ces v_n « parasites » est donc $\dfrac{E(\sqrt{n})}{n}$.

Observons maintenant la fréquence du type 7 dans le dernier tableau 0,03333 et calculons $\dfrac{E(\sqrt{n})}{n}$ pour n = 900. On trouve : 0,03333 !!! dans l'échantillon le type 7 n'est donc représenté que par ces termes « parasites ».

$$\frac{E(\sqrt{n})}{n} \leq \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}} < \frac{E(\sqrt{n})}{n} + \frac{1}{n}$$

On a donc :

$\dfrac{E(\sqrt{n})}{n}$ tend vers 0 et en remplaçant $\dfrac{E(\sqrt{n})}{n}$ par $\dfrac{1}{\sqrt{n}}$ on commet une erreur plus petite que $\dfrac{1}{n}$.

<u>La fréquence de ces termes « parasites » tend donc vers 0</u> mais lentement puisque, 10^{-4} étant la probabilité du type 7 :

$\dfrac{1}{\sqrt{n}} \leq 10^{-5} \Leftrightarrow 10^{10} \leq n$ (en remplaçant $\dfrac{E(\sqrt{n})}{n}$ par $\dfrac{1}{\sqrt{n}}$ on commet une erreur plus petite que 10^{-10})

Il faudrait donc un échantillon de taille supérieure ou égale à dix milliards pour que la fréquence des termes « parasites » soit négligeable devant la probabilité du type 7.

<u>Remarque</u> : si au lieu de prendre $\{v_1, v_2, \ldots, v_n\}$ comme échantillon on prend $\{v_n, v_{n+1}, \ldots, v_{n+p}\}$
On a un résultat analogue : la fréquence des termes « parasites » est une suite équivalente à $\dfrac{1}{\sqrt{p}}$

<u>Conclusion</u> :
si on décide de prendre comme générateur de nombres aléatoires de [0;1[la suite $\sqrt{n} - \text{int}(\sqrt{n})$, cette suite qui est équirépartie peut ou ne peut pas créer une table de chiffres aléatoires par la procédure définie au début, pour pencher dans l'un ou l'autre sens il faudrait faire le test du poker sur des échantillons de taille supérieure à dix milliards !