Concours général 2017, problème 2 : Plus de succès que d'échecs

Voici un sujet annexe, en orbite autour du problème 2 « Franck passe un examen puis un autre ».

Certaines questions présentes dans le sujet original sont reprises sans modifications. Disons que voici une mouture plus généraliste à destination de candidats au CAPES curieux et éventuellement à d'autres adeptes.

Il va de soi qu'il vaut mieux traiter le sujet authentique du Concours Général, du moins commencer par lui.

Ce sujet n'a d'autre ambition que d'apporter un éclairage (hum ?) un peu différent.

1. Le sujet

On considère une répétition de n $(n \geq 1)$ expériences identiques et indépendantes, numérotées de 1 à n, amenant chacune à l'une ou l'autre de deux issues et deux seulement, S ou $\overline{S}$, de probabilités respectives p $(0 < p < 1)$ et $q = 1 - p$.

On s'intéresse à la question suivante :

Quelle est la probabilité qu'à l'issue des n expériences, le nombre de réalisations de S soit strictement supérieur au nombre de réalisations de son contraire ?

Pour chaque entier j tel que $1 \leq j \leq n$, on note B_j la variable aléatoire qui prend la valeur 1 si l'expérience numéro j aboutit à la réalisation de S, et 0 si cette expérience aboutit à la réalisation de son contraire.

Pour chaque entier j tel que $1 \leq j \leq n$, on note $X_n = \sum_{i=1}^{n} B_i$ la variable aléatoire qui dénombre les réalisations de S observées lors des expériences numérotées de 1 à n.

X_n est une variable aléatoire qui suit de ce fait la loi binomiale $B(n,\ p)$ de paramètres n et p.

Plus généralement, pour chaque entier j tel que $1 \leq j \leq n$, on notera $X_j = \sum_{i=1}^{j} B_i$ la variable aléatoire qui dénombre les réalisations de S observées lors des expériences numérotées de 1 à j.

X_j est une variable aléatoire qui suit de ce fait la loi binomiale $B(j,\ p)$ de paramètres j et p.

Le nombre d'expériences étant égal à n, la condition $X_n > \frac{n}{2}$ est nécessaire et suffisante pour obtenir davantage de réalisations de S que de réalisations de son contraire.

On s'intéresse donc à l'évènement : $E_n = "X_n > \frac{n}{2}"$ et on note P_n sa probabilité.

Plus généralement, pour chaque entier j tel que $1 \leq j \leq n$, on considèrera l'évènement : $E_j = "X_j > \frac{j}{2}"$ et on notera P_j sa probabilité.

1. Calculer en fonction de *p* les probabilités $P_1,\ P_2,\ P_3,\ P_4,\ P_5$.

2. Suivant la valeur de *p*, classer par ordre les probabilités $P_1,\ P_3,\ P_5$

3. Prouver que, pour tout entier *k* vérifiant $1 \le k \le \frac{n}{2}$, on a $P_{2k-1} > P_{2k}$.

4. Prouver que, pour tout entier *k* vérifiant $0 \le k \le \frac{n-1}{2}$, on a $P_{2k+1} = \sum_{i=k+1}^{i=2k+1} \binom{2k+1}{i} p^i (1-p)^{2k+1-i}$

5. 1. Prouver que, pour tout entier *k* vérifiant $0 \le k \le \frac{n-3}{2}$: $P_{2k+3} - P_{2k+1} = \binom{2k+1}{k+1} p^{k+1} (1-p)^{k+1} (1-2p)$

5.2. Etudier suivant la valeur de *p* le sens de variations de la suite $\left(P_{2k+1}\right)_{0 \le k \le \frac{n-1}{2}}$

6.1. On suppose dans cette question que $p > \frac{1}{2}$. Prouver que pour tout entier *n* : $1 - P_n \le \dfrac{p(1-p)}{n\left(p-\frac{1}{2}\right)^2}$.

On pourra appliquer l'inégalité de Bienaymé-Tchebychev (à savoir $P\left(\left|X - E(X)\right| \ge \varepsilon\right) \le \frac{V(X)}{\varepsilon^2}$ pour toute variable aléatoire *X* et tout réel strictement positif ε) à X_n avec un réel strictement positif ε adéquat.

Que peut-on en conclure ?

6.2. On suppose que $p = 0{,}6$. Proposer un entier n_0 tel que pour $n \ge n_0$, $1 - P_n \le 0{,}001$

2. Eléments de correction

1. $P_1 = P(\text{"}X_1 = 1\text{"}) = p$; $P_2 = P(\text{"}X_2 = 2\text{"}) = p^2$;

$P_3 = P(\text{"}X_3 = 2\text{"}) + P(\text{"}X_3 = 3\text{"}) = 3p^2(1-p) + p^3 = p^2(3-2p)$

$P_4 = P(\text{"}X_4 = 3\text{"}) + P(\text{"}X_4 = 4\text{"}) = 4p^3(1-p) + p^4 = p^3(4-3p)$

$P_5 = P(\text{"}X_5 = 3\text{"}) + P(\text{"}X_5 = 4\text{"}) + P(\text{"}X_5 = 5\text{"}) = 10p^3(1-p)^2 + 5p^4(1-p) + p^5 = p^3(10-15p+6p^2)$

2. $P_3 - P_1 = p^2(3-2p) - p = p(-1+3p-2p^2) =$ $p(1-p)(2p-1)$

$P_5 - P_3 = 3p^2(2p^3 - 5p^2 + 4p - 1) = 3p^2(1-p)^2(2p-1)$ car on remarque que $(2p^3 - 5p^2 + 4p - 1)$ se factorise deux fois par $(1-p)$.

Il en résulte que :

- Si $p > \frac{1}{2}$ alors $P_5 > P_3 > P_1$
- Si $p = \frac{1}{2}$ alors $P_5 = P_3 = P_1$
- Si $p < \frac{1}{2}$ alors $P_5 < P_3 < P_1$

3. La condition $1 \le k \le \frac{n}{2}$ assure que $1 \le 2k-1 < 2k \le n$ c'est-à-dire que l'indexation des P_{2k-1} et des P_{2k} est toujours entre 1 et *n*.

$X_{2k-1} > \frac{2k-1}{2} \Leftrightarrow X_{2k-1} \ge k$ tandis que $X_{2k} > \frac{2k}{2} \Leftrightarrow X_{2k} \ge k+1$.

Or : $X_{2k} \ge k+1 \Rightarrow X_{2k-1} \ge k$ c'est-à-dire que l'évènement $\text{"}X_{2k} \ge k+1\text{"}$ est inclus dans l'évènement $\text{"}X_{2k-1} \ge k\text{"}$. L'inclusion est stricte car l'évènement $(\text{"}X_{2k-1} \ge k\text{"}) \cap (\text{"}B_{2k} = 0\text{"})$ est inclus dans $\text{"}X_{2k-1} \ge k\text{"}$ mais ne l'est pas dans $\text{"}X_{2k} \ge k+1\text{"}$. Par conséquent la probabilité de $\text{"}X_{2k} \ge k+1\text{"}$ est strictement inférieure à celle de $\text{"}X_{2k-1} \ge k\text{"}$: $P_{2k} < P_{2k-1}$

Autre méthode :

$P_{2k-1} = P(\text{"}X_{2k-1} \ge k+1\text{"}) + P(\text{"}X_{2k-1} = k\text{"}) =$ $P(\text{"}X_{2k-1} \ge k+1\text{"}) + \binom{2k-1}{k} p^k (1-p)^{k-1}$ tandis que

$P_{2k} = P(\text{"}X_{2k-1} \ge k+1\text{"}) + P(\text{"}X_{2k-1} = k\text{"} \cap \text{"}B_{2k} = 1\text{"}) = P(\text{"}X_{2k-1} \ge k+1\text{"}) + \binom{2k-1}{k} p^k (1-p)^{k-1} \times p$

Ainsi : $P_{2k-1} - P_{2k} = \binom{2k-1}{k} p^k (1-p)^k$ ce qui est strictement positif quelle que soit la valeur de *p*. On en déduit que $P_{2k-1} > P_{2k}$.

4. La condition $0 \le k \le \frac{n-1}{2}$ assure que $1 \le 2k+1 \le n$ c'est-à-dire que l'indexation des P_{2k+1} est toujours entre 1 et *n*.

$X_{2k+1} > \dfrac{2k+1}{2} \Leftrightarrow X_{2k+1} \geq k+1$. Il y a davantage de réalisations de *S* que de son contraire si et seulement si on observe au moins $k+1$ réalisations de *S*. D'où $P_{2k+1} = \displaystyle\sum_{i=k+1}^{i=2k+1} \binom{2k+1}{i} p^i (1-p)^{2k+1-i}$

5.1. La condition $0 \leq k \leq \dfrac{n-3}{2}$ assure que $1 \leq 2k+1 < 2k+3 \leq n$ c'est-à-dire que l'indexation des P_{2k+1} et des P_{2k+3} est toujours entre 1 et *n*.

Comme on l'a vu ci-dessus : $X_{2k+1} > \dfrac{2k+1}{2} \Leftrightarrow X_{2k+1} \geq k+1$ et de même $X_{2k+3} > \dfrac{2k+3}{2} \Leftrightarrow X_{2k+3} \geq k+2$

L'évènement $"X_{2k+1} \geq k+1"$ se décompose ainsi : $"X_{2k+1} \geq k+1" = "X_{2k+1} \geq_{\text{gjulia}} k+2" \cup "X_{2k+1} = k+1"$ en deux évènements d'intersection vide, de sorte que :

$$P_{2k+1} = P\left("X_{2k+1} \geq k+2"\right) + \binom{2k+1}{k+1} p^{k+1} (1-p)^k \ .$$

L'évènement $"X_{2k+3} \geq k+2"$ se décompose ainsi :

$$"X_{2k+1} \geq k+2" \cup \left("X_{2k+1} = k+1" \cap "\left("B_{2k+2} = 1"\cup"B_{2k+3} = 1"\right)\right) \cup \left("X_{2k+1} = k" \cap "\left("B_{2k+2} = 1"\cap"B_{2k+3} = 1"\right)\right)$$

En effet, pour que $"X_{2k+3} \geq k+2"$, il faut ou bien que $k+2$ réalisations de *S* aient été observées au cours des $2k+1$ premières expériences, ou bien que $k+1$ réalisations de *S* aient été observées au cours des $2k+1$ premières expériences et qu'on en observe au moins une de plus lors des expériences numéros $2k+2$ ou $2k+3$, ou bien que *k* réalisations de *S* aient été observées au cours des $2k+1$ premières expériences et qu'on en observe une de plus lors de chacune des expériences $_{gj}$ numéros $2k+2$ et $2k+3$.

Or : $P\left("B_{2k+2} = 1" \cup "B_{2k+3} = 1"\right) = 2p - p^2$ d'après la formule donnant la probabilité d'une réunion, chaque évènement ayant pour probabilité *p* et leur intersection ayant pour probabilité p^2 puisque les deux évènements sont supposés $_{gj}$ indépendants.

Il en résulte que :

$$P_{2k+3} = P\left("X_{2k+1} \geq k+2"\right) + \binom{2k+1}{k+1} p^{k+1} (1-p)^k \times \left(2p - p^2\right) + \binom{2k+1}{k} p^k (1-p)^{k+1} \times p^2$$

Par symétrie des coefficients binomiaux : $\dbinom{2k+1}{k+1} = \dbinom{2k+1}{k}$.

$$P_{2k+3} = P\left("X_{2k+1} \geq k+2"\right) + \binom{2k+1}{k+1} p^k (1-p)^k \times \left(3p^2 - 2p^3\right)$$

$$P_{2k+3} - P_{2k+1} = \binom{2k+1}{k+1} p^k (1-p)^k \left(3p^2 - 2p^3 - p\right) = \binom{2k+1}{k+1} p^{k+1} (1-p)^{k+1} \times (2p-1)$$

Ce qui généralise les résultats obtenus dans la question **2**.

5.2. Cette différence est du signe de $2p-1$.

- Si $p > \dfrac{1}{2}$ alors la suite $\left(P_{2k+1}\right)_{0 \leq k \leq \frac{n-1}{2}}$ est croissante
- Si $p = \dfrac{1}{2}$ alors la suite $\left(P_{2k+1}\right)_{0 \leq k \leq \frac{n-1}{2}}$ est stationnaire
- Si $p < \dfrac{1}{2}$ alors la suite $\left(P_{2k+1}\right)_{0 \leq k \leq \frac{n-1}{2}}$ est décroissante

6.1. La variable aléatoire X_n suit la loi binomiale de paramètres n et p. Son espérance est $n\,p$ et sa variance est $n\,p(1-p)$

Or, $n\,p-\dfrac{n}{2}=n\left(p-\dfrac{1}{2}\right)$. L'évènement "$X_n\le\dfrac{n}{2}$", c'est-à-dire l'évènement contraire de celui que l'on considère, est inclus dans l'évènement "$\left|X_n-n\,p\right|\ge n\left(p-\dfrac{1}{2}\right)$"

D'après l'inégalité de Tchebychev, pour toute variable aléatoire X et tout réel strictement positif ε: $P\left(\left|X-E(X)\ge\varepsilon\right|\right)\le\dfrac{V(X)}{\varepsilon^2}$.

En appliquant cette égalité avec $\varepsilon=n\left(p-\dfrac{1}{2}\right)$, on va majorer la probabilité de "$\left|X_n-n\,p\right|\ge n\left(p-\dfrac{1}{2}\right)$" et *a fortiori* la probabilité de "$X_n\le\dfrac{n}{2}$". On va majorer $1-P_n$.

On obtient : $P\left(\left|X_n-np\right|\ge n\left(p-\dfrac{1}{2}\right)\right)\le\dfrac{np(1-p)}{n^2\left(p-\dfrac{1}{2}\right)^2}=\dfrac{p(1-p)}{n\left(p-\dfrac{1}{2}\right)^2}$ et par suite : $1-P_n\le\dfrac{p(1-p)}{n\left(p-\dfrac{1}{2}\right)^2}$

On en conclut que la suite $(1-P_n)_{n\in N^*}$ est majorée par une suite qui converge vers zéro. En augmentant indéfiniment le nombre d'expériences, il devient quasi certain que le nombre de réalisations de S va dépasser le nombre de réalisations de son contraire.

6.2. On obtient $1-P_n\le\dfrac{24}{n}$. On peut proposer $n_0=24000$ …

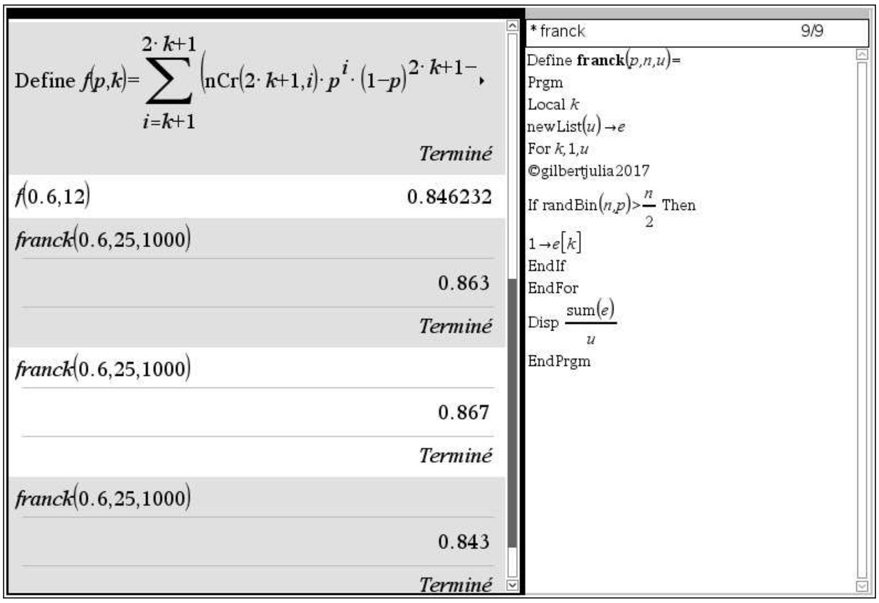
On constatera que l'inégalité de Tchebychev n'est pas très performante, elle est purement théorique. Elle a juste le mérite d'exister.

3. Simulation

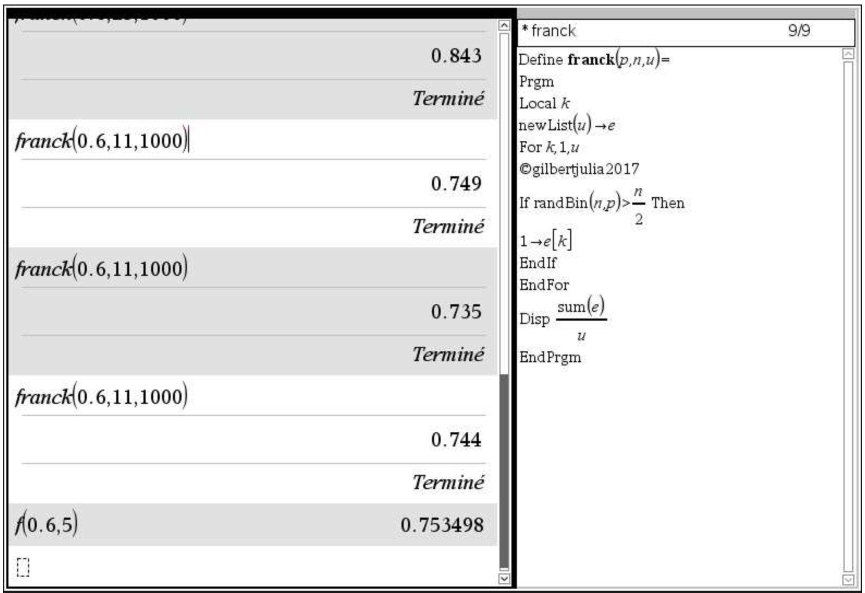
Le programme **franck** simule une série de m expériences dans lesquelles Franck répond à n questions.

Il est lancé plusieurs fois avec $n=25$ (c'est-à-dire en supposant que Franck répond à toutes les questions sauf à la dernière) et en supposant que $p=0{,}6$.

On peut comparer à la probabilité théorique.

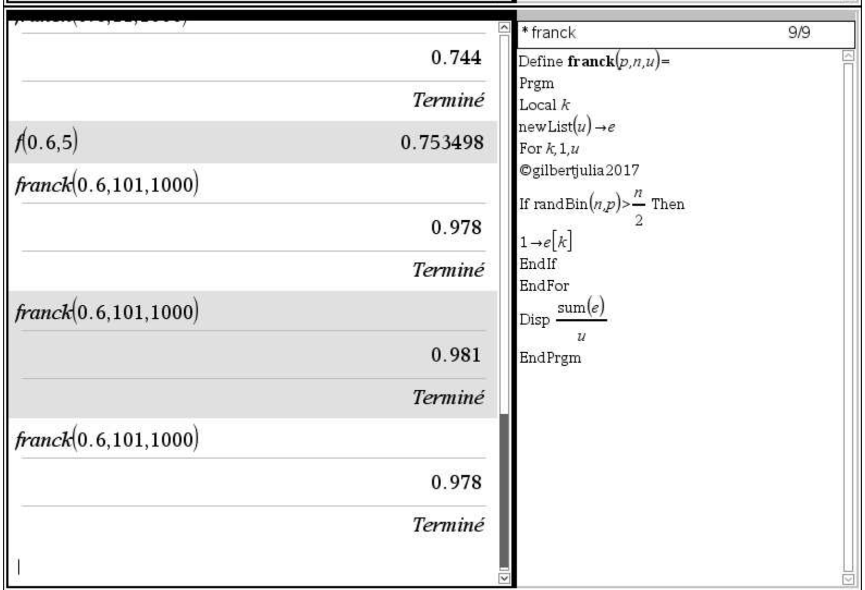


Si Franck décide de répondre à 11 des questions dont il ne connaît pas la réponse puis de s'arrêter, c'est mieux que 0,6 mais moins bien que la stratégie précédente.



si le QCM comportait 101 questions, la fréquence des expériences réussies est de l'ordre de 0,98. C'est pas mal, mais non parfait.

Avec 101 questions et une probabilité égale à 0,6 de répondre correctement, Franck n'aurait quand même pas son examen « presque à tous les coups ».



Avec 201, c'est presque gagné. Encore un petit effort …

On est tout de même fort loin des 24000 prévues par l'inégalité façon « Je n'y vais pas avec le dos de la cuiller » de Tchebychev.