Un résultat surprenant en probabilités

Michel PLATHEY, Lycée H. Fontaine, Dijon

Résumé : Étude mathématique et 3 simulations informatiques d'un jeu probabiliste dont le résultat échappe à l'intuition ou au seul bon sens. Cas où la démonstration mathématique et l'expérimentation se renforcent mutuellement.

Mots clés : Théorème des probabilités totales ; arbre de probabilité ; modélisation.

Le problème suivant est très connu, mais on ne le rencontre pas encore dans les livres d'enseignement.

Il s'agit d'un jeu télévisé qui est apparu en 1963 aux États-Unis qui s'appelait « Let's make a deal » et qui a été importé en France dans les années 2000 et a été programmé sur une chaîne privée grand public.

Cette émission a eu à l'époque un très grand succès. Mais elle en avait eu un bien plus grand encore aux États-Unis, où elle avait mis en difficulté financière la station qui l'avait lancée.

1. Rappelons la règle du jeu :

Sur le plateau télévisé, il y a 3 portes rigoureusement semblables, un meneur de jeu et un candidat joueur.
Derrière l'une des portes, il y a une voiture, objet de la convoitise du candidat et, par contre, derrière les deux autres portes, il n'y a rien, ou presque rien.

Je note 0 **la porte derrière laquelle se trouve la voiture**, 1 et 2 les deux autres portes.

Le jeu se déroule en deux temps :

Premier temps. D'abord, le candidat choisit une porte et, comme on le comprend bien, il n'est pas sûr de son choix.

Les événements :
« au cours de la première phase du jeu, la porte 0 (respectivement : 1, 2) a été choisie. »
sont désignés par :

« 0_1 », (Respectivement : « 1_1 », « 2_1 »)

Ces 3 événements forment un système complet.

Deuxième temps. Après cette première phase du jeu intervient l'animateur, qui connaît la porte derrière laquelle se trouve la voiture. L'animateur discute avec le candidat indécis et va lui dire : « Bon, vous m'êtes sympathique, je vais vous aider. Je vais vous ouvrir l'une de ces 2 portes que vous n'avez pas choisies. » Et il ouvre l'une des deux portes, derrière laquelle il n'y a rien.

Le candidat est alors autorisé à modifier son choix initial.

Puis la discussion continue jusqu'à ce que le candidat ait pris sa décision, soit de garder son choix initial, soit de le modifier, décision qui sera cette fois définitive.

« 0_2 » désigne l'événement :
« au cours de la deuxième phase du jeu, la porte 0 a été choisie. »

2. Question stratégique.

.

Le candidat doit-il rester sur son premier choix, ou pas ?

Manifestement, le candidat a bien été aidé par l'animateur, puisque, d'un choix entre 3 portes, il ne reste plus maintenant qu'un choix entre 2 portes.
Le jeu s'est simplifié.
Et, en même temps, il est devenu plus intéressant.
Le candidat va essayer d'en savoir plus encore et l'animateur du jeu va essayer d'entretenir l'incertitude du candidat et ces passes d'armes psychologiques donnent tout son sel à cette prestation télévisée.

Que doit faire le candidat ?

Il reste, dans la deuxième phase du jeu, à faire un choix entre deux portes, (porte 0 derrière laquelle se trouve la voiture et l'autre, porte 1 ou porte 2 derrière laquelle il n'y a rien, mais pour le candidat, rien ne permet de les distinguer) et on a l'impression, qu'en l'absence d'information supplémentaire, le candidat a 1 chance sur 2 de gagner la voiture. C'est évidemment mieux qu'avant la deuxième étape du jeu, où il avait 1 chance sur 3 de gagner la voiture.

On va cependant voir que cette impression est fausse, et **qu'avec une stratégie adéquate, le candidat a 2 chances sur 3 de gagner la voiture.**
C'est ce résultat, qu'il faut démontrer.
Ce résultat n'avait pas été anticipé par la firme américaine propriétaire du jeu.
Par contre certains joueurs mathématiciens l'avaient correctement calculé, ce qui a créé des difficultés financières à la firme productrice du jeu car les joueurs ont gagné plus souvent que prévu. Des débats passionnés s'en sont suivis et une grosse littérature a été consacrée à ce sujet appelé le « Monty Hall paradox ».

Or, la clef de la solution est à la portée d'un lycéen de première ou de terminale.

Et cette solution montre bien la puissance explicative des arbres de probabilité.

3. Modélisation de ce jeu.

Plaçons-nous dans le cas où le candidat, dans la deuxième phase du jeu, reste sur sa position avec la probabilité p et donc change d'avis avec la probabilité $1-p$.

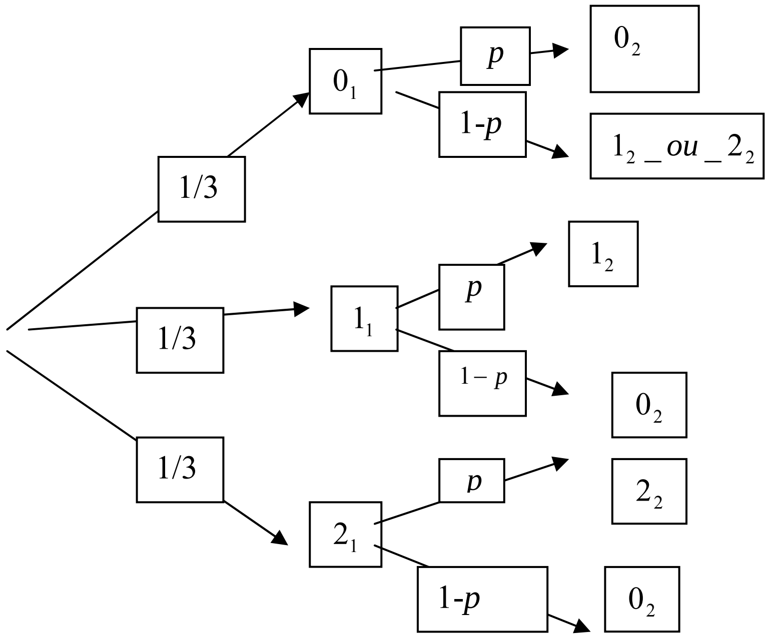
On modélise le choix du candidat au cours de la deuxième phase du jeu par un jeu de pile ou face avec :

- « si « pile », le candidat garde la position qu'il avait prise lors de la première phase du jeu,
- si « face », il change son choix initial ».

où :

$$p("pile") = p \,;\, 0 \le p \le 1.$$

Dans la première phase du jeu, la probabilité de choisir une porte particulière est $\frac{1}{3}$ et dans la deuxième phase du jeu, la probabilité, pour le candidat, de garder sa position est p, la probabilité de choisir l'autre porte restée fermée est $1-p$.
On peut alors représenter le jeu complet par l'arbre de probabilité suivant :



On obtient alors :

$$p(0_1 \cap 0_2) = \frac{1}{3} * p \;\; ; \;\; p(1_1 \cap 0_2) = \frac{1}{3} * (1-p) \;\; ; \;\; p(2_1 \cap 0_2) = \frac{1}{3} * (1-p) .$$

Donc :

$$p(0_2) = p(0_1 \cap 0_2) + p(1_1 \cap 0_2) + p(2_1 \cap 0_2) = \frac{1}{3} * (p + 1 - p + 1 - p) = \frac{1}{3} * (2-p) .$$

On constate alors que $p(0_2)$, probabilité de gagner la voiture, est une fonction affine f décroissante de $p \in [0;1]$, (p étant la probabilité pour le candidat de garder le choix initial) telle que :

$$f(0) = \frac{1}{3} * (2-0) = \frac{2}{3};$$

$$f(\frac{1}{2}) = \frac{1}{3} * (2 - \frac{1}{2}) = \frac{1}{2};$$

$$f(1) = \frac{1}{3} * (2-1) = \frac{1}{3}.$$

En conclusion :

- **La plus mauvaise des stratégies de jeu consiste à garder son choix initial quoi qu'il arrive, ce qui revient à ne pas tenir compte de l'information donnée par l'ouverture de la porte vide et la probabilité de gain est alors : $\frac{1}{3}$.**
- **Le procédé consistant à jouer à pile ou face entre les 2 tours du jeu, n'exploite pas non plus l'information à son maximum.**
- **La méthode qui maximise la probabilité de gain est de changer d'avis systématiquement et la probabilité de gain est alors : $\frac{2}{3}$.**

4. Simulation informatique du jeu.

On peut aussi faire une simulation de ce jeu.

Voici ce que j'ai obtenu, en répétant 50 fois l'expérience dans un tableur Excel :

Toutes les colonnes sont donc de taille 50.

La colonne 1 (A) contient les expressions P1, P2, …, P50 pour partie n°1, partie n°2, …, partie n°50, à partir de la ligne 1.

La colonne 2 (B) contient les numéros des portes choisies par le candidat au premier tour :

[=ENT (ALEA ()*3)],

portes choisies de façon équiprobable.

La première phase du jeu est alors exécutée.

On passe à la deuxième étape.

La colonne 3 (C) contient un nombre au hasard :

[=ALEA ()]

compris entre 0 et 1 (loi continue uniforme sur $[0;1]$) qui servira à la construction de la colonne 4, qui désigne le choix au hasard de la porte ouverte par l'animateur.

La colonne 4 (D) contient le numéro de la porte ouverte par l'animateur.

[=SI (B1<>0 ; 3-B1;SI (C1<0,5 ; 1 ; 2))].

(Lorsque le candidat a choisi la mauvaise porte au départ, 1 ou 2, l'animateur ouvre l'autre porte vide : 3-1=2 ou 3-2=1 et lorsque le candidat a choisi la bonne porte au départ, l'animateur choisit d'ouvrir l'une des 2 autres portes vides avec équiprobabilité.)

La colonne 5 (E) contient le numéro de la porte qui sera choisie par le candidat dans la deuxième phase de jeu dans le cas où il change son choix initial :

[=SI (B1<>0 ; 0 ; 3-D1]

(s'il n'avait pas choisi 0 dans la première phase, alors il choisit 0 dans le deuxième temps du jeu et s'il avait choisi 0 dans la première phase, alors il choisit l'autre porte non ouverte par l'animateur dans le deuxième temps du jeu.

La colonne 6 (F) contient un nombre au hasard :

[=ALEA ()]

compris entre 0 et 1 (loi continue uniforme sur $[0;1]$) qui servira à la construction de la colonne 7, qui désigne le choix au hasard de la porte ouverte par le candidat dans la deuxième phase du jeu.

La colonne 7 (G) contient le numéro de la porte ouverte par le candidat dans la deuxième phase du jeu :

[SI (E1<$J$1 ; B1 ; E1].

(avec la probabilité p, le candidat garde son choix, et avec la probabilité $1-p$, il change son choix initial.
Auparavant, en J1, on aura placé la probabilité p qui décide si le candidat reste ou non sur sa position.

On répète 50 fois ces instructions.

Ensuite :

On calcule en J4, le nombre de 0 (nombre de succès) obtenus dans la colonne G :
[=NB.SI($G$1 :$G$50 ;0]
On calcule en I7, la fréquence des succès obtenus en colonne F :
[=$J$4/50].
On répète des échantillons de 50 expériences de ce type en appuyant sur la touche F9 de la première ligne du clavier.

En J10, on calcule la fréquence théorique des succès, obtenue par la théorie précédente :
[=1/3*(2-$J$1)]
En I13, on calcule la différence entre la fréquence obtenue et la fréquence théorique : [=$J$10-$J$1].
On peut ainsi se convaincre expérimentalement que la probabilité de gain au jeu énoncé est bien de $\frac{2}{3}$ lorsqu'on change systématiquement d'avis entre les deux tours.
La connaissance par les candidats de cette règle tue évidemment le jeu puisqu'elle enlèvera tout suspense dans la phase intermédiaire de discussion entre les deux tours.

Je donne ci-après trois simulations de 50 parties, une première série avec $p=1$: le candidat reste sur sa position ; une autre avec $p=\frac{1}{2}$: le candidat choisit au hasard avec équiprobabilité s'il garde sa position ou change d'avis ; et une autre avec $p=0$: le candidat change sa position de façon systématique.

Les colonnes utilisées sont les colonnes A à J.

La colonne H (mince) est vide.

La colonne I (large) contient des commentaires.

La colonne J contient les données ainsi que les résultats du problème.

Les intitulés des colonnes ont été mis en ligne 51.

Mon collègue Jean-Marie Thomassin propose quelques questions supplémentaires :

- Comment peut-on rééquilibrer le jeu ?
- Que se passe-t-il avec 4 portes, 5 portes, n portes ?

- Et si on ajoute une porte avec un gain négatif ou une mise initiale ?

- Il semble qu'une mésaventure analogue ait eu lieu dans le passé avec un jeu appelé « le tapis vert » mais je n'en connais pas le détail.

Il est clair que les tentatives de résolution de ces diverses questions peuvent commencer par des investigations informatiques suivies ensuite par des démonstrations des propriétés constatées.

Ce n'est pas l'ordre que j'ai choisi dans l'exposition du « Monty Hall paradox » mais il était peut-être préférable.

Je remercie Monsieur Lafond et Monsieur Thomassin pour leurs critiques constructives.

P1	0	0,734	2	1	0,107	0	
P2	2	0,647	1	0	0,721	2	
P3	2	0,638	1	0	0,942	2	
P4	0	0,276	1	2	0,125	0	
P5	2	0,999	1	0	0,588	2	
P6	0	0,733	2	1	0,511	0	
P7	0	0,526	2	1	0,723	0	
P8	1	0,714	2	0	0,229	1	
P9	2	0,154	1	0	0,608	2	
P10	1	0,067	2	0	0,217	1	
P11	1	0,160	2	0	0,788	1	
P12	2	0,261	1	0	0,027	2	
P13	1	0,543	2	0	0,918	1	
P14	0	0,299	1	2	0,516	0	
P15	1	0,858	2	0	0,852	1	
P16	2	0,458	1	0	0,550	2	
P17	0	0,254	1	2	0,751	0	
P18	1	0,471	2	0	0,026	1	
P19	0	0,821	2	1	0,016	0	
P20	0	0,956	2	1	0,142	0	
P21	2	0,710	1	0	0,722	2	
P22	0	0,661	2	1	0,078	0	
P23	2	0,503	1	0	0,986	2	
P24	0	0,198	1	2	0,162	0	
P25	0	0,592	2	1	0,305	0	
P26	1	0,370	2	0	0,463	1	
P27	0	0,737	2	1	0,013	0	
P28	1	0,456	2	0	0,623	1	
P29	0	0,115	1	2	0,171	0	
P30	2	0,635	1	0	0,226	2	
P31	2	0,546	1	0	0,150	2	
P32	2	0,356	1	0	0,620	2	
P33	0	0,289	1	2	0,179	0	
P34	1	0,714	2	0	0,919	1	
P35	1	0,732	2	0	0,261	1	
P36	0	0,590	2	1	0,753	0	
P37	0	0,899	2	1	0,070	0	
P38	1	0,342	2	0	0,776	1	
P39	0	0,912	2	1	0,162	0	
P40	1	0,431	2	0	0,558	1	
P41	2	0,626	1	0	0,713	2	
P42	2	0,531	1	0	0,710	2	
P43	1	0,983	2	0	0,317	1	
P44	2	0,232	1	0	0,056	2	
P45	2	0,065	1	0	0,158	2	
P46	0	0,305	1	2	0,753	0	
P47	1	0,525	2	0	0,002	1	
P48	0	0,248	1	2	0,554	0	
P49	1	0,174	2	0	0,386	1	
P50	1	0,474	2	0	0,439	1	

Probabilité que le candidat garde son choix de la première phase de jeu :	**1**
Nombre de succès :	**19**
Fréquence des succès :	**0,38**
Probabilité théorique de succès :	**0,333333**
Différence entre fréquence observée et probabilité de succès :	**0,046667**

Intitulé.

50 parties de "Let's make a deal"

en B1:=ENT(ALEA()*3
en C1:=ALEA()
en D1:=SI(B1<>0;3-B1;SI(C1<0,5;1;2))
en E1:=SI(B1<>0;0;3-D1)
en F1:=ALEA()
en G1:=SI(F1<$J$1;B1;E1)
où
en J1 se trouve la probabilité de garder son choix entre les 2 tours.

On répète 50 fois les instructions précédentes

puis:

en J4:=NB.SI($J$1:$J$50;0)
en J7:=$J$4/50
en J10:=(1/3)*(2-$J$1)
en J13:=$J$7-$J$10

A	B	C	D	E	F	G	H	I	J
P1	**0**	**0,275**	**1**	**2**	**0,045**	**0**		**Probabilité que le candidat garde son**	**0,5**
P2	2	0,045	1	0	0,326	2		**choix de la première phase de jeu :**	

P3	2	0,271	1	0	0,274	2			
P4	2	0,333	1	0	0,734	0		**Nombre de succès :**	**20**
P5	1	0,219	2	0	0,268	1			
P6	1	0,174	2	0	0,208	1			
P7	2	0,519	1	0	0,168	2		**Fréquence des succès :**	**0,4**
P8	0	0,907	2	1	0,320	0			
P9	0	0,831	2	1	0,565	1			
P10	1	0,965	2	0	0,442	1		**Probabilité théorique de succès :**	**0,5**
P11	0	0,250	1	2	0,606	2			
P12	0	0,103	1	2	0,898	2			
P13	0	0,440	1	2	0,468	0		**Différence entre fréquence observée**	**-0,1**
P14	1	0,083	2	0	0,260	1		**et probabilité de succès :**	
P15	0	0,199	1	2	0,641	2			
P16	0	0,205	1	2	0,366	0		**Intitulé.**	
P17	1	0,226	2	0	0,415	1			
P18	2	0,318	1	0	0,562	0		**50 parties de "Let's make a deal"**	
P19	2	0,243	1	0	0,417	2			
P20	1	0,002	2	0	0,620	0			
P21	2	0,396	1	0	0,060	2			
P22	0	0,529	2	1	0,232	0			
P23	0	0,794	2	1	0,747	1		en E1:=SI(B1<>0;0;3-D1)	
P24	1	0,409	2	0	0,189	1		en F1:=ALEA()	
P25	0	0,768	2	1	0,686	1		en G1:=SI(F1<$J$1;B1;E1)	
P26	2	0,538	1	0	0,903	0		où	
P27	1	0,095	2	0	0,894	0		en J1 se trouve la probabilité de garder	
P28	0	0,651	2	1	0,194	0		son choix entre les 2 tours.	
P29	0	0,164	1	2	0,360	0			
P30	1	0,741	2	0	0,279	1		On répète 50 fois	
P31	0	0,166	1	2	0,543	2		en B1:=ENT(ALEA()*3	
P32	2	0,977	1	0	0,653	0		en C1:=ALEA()	
P33	2	0,922	1	0	0,525	0		en D1:=SI(B1<>0;3-B1;SI(C1<0,5;1;2))	
P34	2	0,365	1	0	0,442	2			
P35	2	0,118	1	0	0,814	0		en J4:=NB.SI($J$1:$J$50;0)	
P36	0	0,924	2	1	0,572	1		en J7:=$J$4/50	
P37	2	0,605	1	0	0,922	0		en J10:=(1/3)*(2-$J$1)	
P38	0	0,652	2	1	0,941	1		en J13:=$J$7-$J$10	
P39	1	0,976	2	0	0,840	0			
P40	0	0,567	2	1	0,917	1			
P41	2	0,961	1	0	0,050	2			
P42	0	0,735	2	1	0,573	1			
P43	2	0,254	1	0	0,403	2			
P44	1	0,962	2	0	0,221	1			
P45	0	0,543	2	1	0,349	0			
P46	2	0,398	1	0	0,464	2			
P47	1	0,847	2	0	0,034	1			
P48	0	0,796	2	1	0,810	1			
P49	0	0,141	1	2	0,232	0			
P50	2	0,404	1	0	0,845	0			
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**

P1	1	0,588	2	0	0,111	0		**Probabilité que le candidat garde son**	**0**
P2	2	0,690	1	0	0,745	0		**choix de la première phase de jeu :**	
P3	0	0,365	1	2	0,069	2			
P4	1	0,304	2	0	0,781	0		**Nombre de succès :**	**35**
P5	1	0,252	2	0	0,006	0			
P6	0	0,194	1	2	0,860	2			

A	B	C	D	E	F	G	H
P7	1	0,434	2	0	0,385	0	
P8	1	0,188	2	0	0,131	0	
P9	2	0,865	1	0	0,510	0	
P10	0	0,285	1	2	0,743	2	
P11	0	0,319	1	2	0,737	2	
P12	2	0,041	1	0	0,971	0	
P13	2	0,283	1	0	0,840	0	
P14	1	0,139	2	0	0,731	0	
P15	1	0,461	2	0	0,493	0	
P16	0	0,476	1	2	0,261	2	
P17	2	0,858	1	0	0,211	0	
P18	2	0,953	1	0	0,135	0	
P19	0	0,861	2	1	0,899	1	
P20	0	0,760	2	1	0,743	1	
P21	1	0,073	2	0	0,772	0	
P22	1	0,786	2	0	0,292	0	
P23	2	0,187	1	0	0,607	0	
P24	0	0,187	1	2	0,847	2	
P25	1	0,481	2	0	0,688	0	
P26	2	0,500	1	0	0,016	0	
P27	0	0,085	1	2	0,005	2	
P28	1	0,601	2	0	0,838	0	
P29	2	0,157	1	0	0,125	0	
P30	1	0,908	2	0	0,419	0	
P31	2	0,319	1	0	0,357	0	
P32	1	0,732	2	0	0,024	0	
P33	1	0,037	2	0	0,792	0	
P34	0	0,979	2	1	0,606	1	
P35	2	0,830	1	0	0,600	0	
P36	1	0,271	2	0	0,639	0	
P37	1	0,894	2	0	0,960	0	
P38	2	0,449	1	0	0,980	0	
P39	1	0,695	2	0	0,136	0	
P40	2	0,136	1	0	0,307	0	
P41	2	0,529	1	0	0,476	0	
P42	1	0,476	2	0	0,483	0	
P43	1	0,791	2	0	0,796	0	
P44	0	0,581	2	1	0,235	1	
P45	0	0,802	2	1	0,754	1	
P46	1	0,135	2	0	0,400	0	
P47	0	0,916	2	1	0,616	1	
P48	2	0,355	1	0	0,176	0	
P49	0	0,436	1	2	0,570	2	
P50	0	0,847	2	1	0,811	1	

I	J
Fréquence des succès :	**0,7**
Probabilité de succès :	**0,6667**
Différence entre fréquence observée et probabilité de succès :	**0,0333**

Intitulé.

50 parties de "Let's make a deal"

en B1:=ENT(ALEA()*3
en C1:=ALEA()
en D1:=SI(B1<>0;3-B1;SI(C1<0,5;1;2))
en E1:=SI(B1<>0;0;3-D1)
en F1:=ALEA()
en G1:=SI(F1<$J$1;B1;E1)
où
en J1 se trouve la probabilité de garder son choix entre les 2 tours.

On répète 50 fois
les instructions précédentes

puis:

en J4:=NB.SI($J$1:$J$50;0)
en J7:=$J$4/50
en J10:=(1/3)*(2-$J$1)
en J13:=$J$7-$J$10