

J.-P. MARCIANO

J. VORANGER

Sur l'optimisation de la taille d'un échantillonnage

Revue française d'informatique et de recherche opérationnelle. Série verte, tome 3, n° V2 (1969), p. 107-112

http://www.numdam.org/item?id=RO_1969__3_2_107_0

© AFCET, 1969, tous droits réservés.

L'accès aux archives de la revue « Revue française d'informatique et de recherche opérationnelle. Série verte » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

SUR L'OPTIMISATION DE LA TAILLE D'UN ECHANTILLONNAGE

par J.-P. MARCIANO (1) et J. VORANGER (2)

Résumé. — A partir de différentes classes de taux de défections de machines ou services, si l'on peut attribuer une probabilité a priori à ces taux, l'analyse Bayésienne nous permet, à partir d'un échantillonnage, d'attribuer à ces taux une probabilité a posteriori, suivant un processus de Bernoulli, voire d'en déduire un coût espéré quel que soit le taux. Face à une éventuelle décision de révision de la machine, on peut alors chercher la taille optimale de l'échantillon à étudier. Pour chaque taille n , on calcule un coût espéré et le gain par échantillonnage en tenant compte du coût de l'échantillonnage. C'est un problème type d'analyse pré a posteriori.

Nous avons voulu, à partir d'un exemple, évoquer ce problème fondamental de la Statistique Inductive qui est celui de la fixation du nombre optimum d'expériences à réaliser, ou d'observations à faire pour dégager, avec une approximation acceptable, un paramètre de comportement quelconque.

Ce problème de décision Statistique dans l'optique bayésienne a été simulé numériquement à partir d'un exemple emprunté à R. Schlaifer dans « Probability and Statistics for Business Decision ».

1. L'ANALYSE A PRIORI

Une machine ou un service fonctionne avec un taux de défection qui peut, par exemple, prendre 4 classes de valeurs t_1, t_2, t_3, t_4 , peut-être 1, 5, 15 et 25. Si l'on a une série de 500 pièces, si le coût d'ajustage par pièce est de 4 F, si l'on peut faire venir un technicien qui ramène le taux de défection à 1 % après avoir pris 60 F de frais, on peut dresser le tableau de coût suivant les décisions.

(1) Maître-Assistant de Mathématiques statistiques.

(2) Directeur du Centre d'Économétrie de l'U.E.R. de Sciences Économiques d'Aix-en-Provence.

t	A_1 (F)	A_2 (F)
0,01	20	80
0,05	100	80
0,15	300	80
0,25	500	80

A_1 est le lancement de la série sans appel au technicien,

A_2 est l'acte d'appel.

On voit que dans 3 cas sur 4 l'acte A_2 revient moins cher que l'acte A_1 . Toutefois, le fabricant n'ayant pas d'observations antérieures peut établir une distribution de « probabilités de jugement », *a priori* pour t , soit

t	PROBABILITES DE JUGEMENT
0,01	0,7
0,05	0,1
0,15	0,1
0,25	0,1

On peut alors, le coût étant fonction de t , variable aléatoire, définir un coût espéré si l'agent de décision accepte ce mode de valuation, tout au moins quand on choisit l'acte A_1 car sinon le coût d'appel au technicien est toujours 80 F quel que soit t .

		ACTE A_1		ACTE A_2
t	$p'(t)$	c	$c_t \times p_t$	
0,01	0,7	20	14	
0,05	0,1	100	10	
0,15	0,1	300	30	
0,25	0,1	500	50	
Coût quel que soit t $t(c) = 104$ F				80 F

Même avec ces probabilités *a priori* attribuées à chaque taux, l'acte A_1 semble en probabilité coûter plus cher que l'acte A_2 , il vaudra mieux faire appel au technicien.

2. L'ANALYSE A POSTERIORI

L'industriel lance la série et observe un échantillon, par exemple, les 10 premières pièces et trouve 0 pièce défectueuse.

On est en présence d'un processus de Bernoulli à 2 catégories (bonnes et mauvaises) avec $n = 10$ et $r = 0$. Suivant le taux *a priori* de pièces défectueuses, la probabilité « objective » est

$$P_0 = C_{10}^0 t_i^0 (1 - t_i)^{n-0} = P_r(r=0 / n=10 \text{ et } t_i).$$

Mais la probabilité d'avoir à la fois $t = t_i$ et $r = 0$ est

$$P_r(t = t_i \text{ ET } r=0 / n=10) = \underbrace{P_r(t = t_i)}_{\text{à priori}} \cdot \underbrace{P_r(r=0 / n=10 \text{ et } t = t_i)}_{\text{Bernoulli}}$$

D'où le tableau suivant des probabilités jointes :

t	$P(t = t_i)$ <i>a priori</i>	$P_r(r=0/n, t_i)$	P' jointe $P(t = t_i \text{ et } r = 0)$	P'' révisée
0,01	0,7	0,904	0,632 = P'_1	0,881
0,05	0,1	0,599	0,059 = P'_2	0,083
0,15	0,1	0,197	0,019 = P'_3	0,028
0,25	0,1	0,056	0,005 = P'_4	0,008

Probabilité marginale 0,718.

La probabilité marginale sera la probabilité d'avoir $r = 0$ quel que soit t

$$P_r[r = 0 \text{ ET } (t = t_1 \text{ ou } t_2 \text{ ou } t_3 \text{ ou } t_4)] = P'_1 + P'_2 + P'_3 + P'_4.$$

D'après le principe des probabilités composées

$$P'_r(t = t_i \text{ ET } r = 0) = P(r = 0) \times P''(t = t_i / r = 0)$$

ou

$$P''_r(t = t_i / r=0) = \frac{P'(t = t_i \text{ ET } r = 0)}{P(r = 0)} = \frac{P_r \text{ jointe}}{P_r \text{ marginale}}$$

Ce n'est autre que le théorème de Bayes qui nous permet ainsi, après le résultat de l'échantillonnage de trouver une probabilité révisée pour chaque taux, afin éventuellement de réviser notre décision suivant le nouveau tableau des coûts espérés :

t	$P''(t)$	ACTE A_1		ACTE A_2
		C	$P_t C_t$	
0,01	0,881	20	17,6	
0,05	0,083	100	8,3	
0,15	0,028	300	8,4	
0,25	0,008	500	4	
Coût espéré			38,8	80

Grâce à l'échantillonnage, l'action dont le coût semble le moindre apparaît maintenant être l'acte A_1 . Le résultat aurait été évidemment tout autre si à l'échantillonnage on avait trouvé un plus grand nombre de pièces défectueuses. Du gain espéré grâce à l'échantillonnage doit cependant être déduit le coût de l'échantillonnage lui-même.

3. ANALYSE PRE A POSTERIORI

Nous allons chercher une taille optimale de l'échantillon pour avoir un gain réel espéré maximum. Le calcul nous permet de dresser, pour une taille d'échantillon donnée $n = 10$, par exemple, et avant que l'échantillonnage ne commence, le tableau suivant :

r	P MARGINALE $P_r(r = a/n = 10/\sqrt{t})$	COÛT ESPÉRÉ DE L'ACTE A_1	COÛT DE A_2	COÛT ESPÉRÉ MINIMUM	$P_t^0(C_{\min})_t$
0	0,718	38,3	80	38,3	27,4
1	0,149	162	80	80	11,9
2	0,066	350	80	80	5,3
3	0,039	422	80	80	3,1
4	0,019	448	80	80	1,5
5	0,007	474	80	80	0,6
6	0,002				
7	ϵ				
8	ϵ				
9	ϵ				
10	ϵ				

On définira un coût espéré pour n donné, quel que soit r , appelé coût terminal espéré

$$E(C_{\min}) = \sum P_i(C_{\min})_i = 50 = E_n$$

On a ainsi avant l'échantillonnage, n étant fixe mais quel que soit le nombre de pièces défectueuses, un coût terminal espéré et en complément un gain terminal espéré grâce à l'échantillonnage

$$G_n = 80 - E_n$$

On peut d'autre part calculer le coût de l'échantillonnage proprement dit qui se monte, dans l'exemple traité, à 5 F quelle que soit la taille, auquel s'ajoute 0,6 F par pièce

$$Z_n = 5 + 0,6n.$$

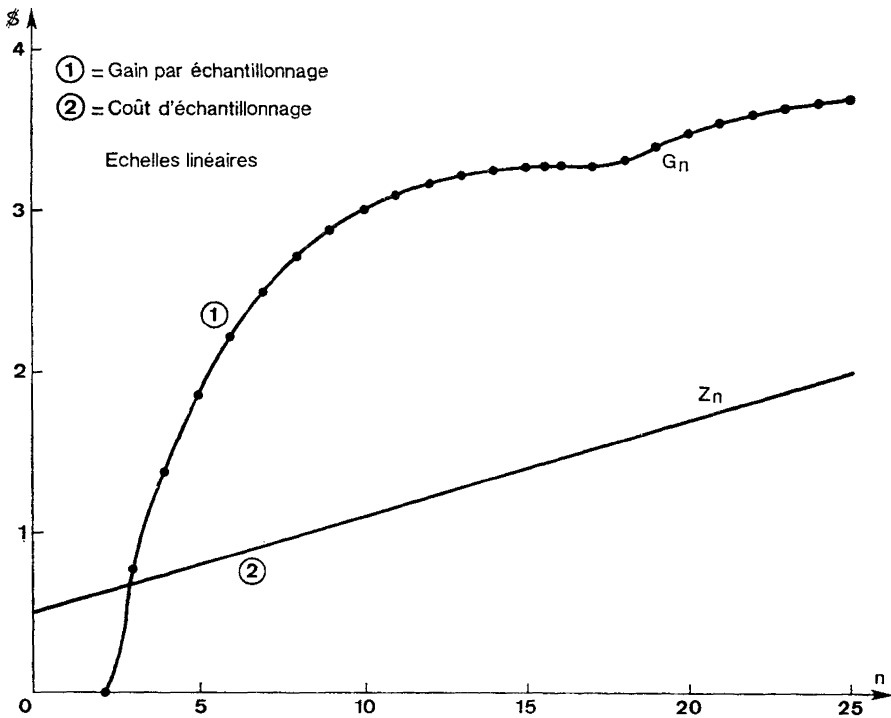
L'échantillonnage optimal sera réalisé comme nous le montrent le tableau et la courbe s'y rapportant, pour la plus grande différence $G_n - Z_n$ atteinte ici pour $n = 12$ (fig. 1 et 2).

4. ALGORITHME DE CALCUL

Il suffit d'introduire les taux de défections t_i , les probabilités de jugement correspondantes p_i , et les coûts c_i . On a alors l'algorithme suivant :

1. FAIRE $N = 1,100$
2. FAIRE $r = 0, n$
3. FAIRE $i = 1,4$
4. CALCULER $PJ\emptyset(i) = C_n^r p_i t_i^r (1 - t_i)^{n-r}$
5. CALCULER $PMA(r) = \sum PJ\emptyset(i)$
6. ALLER EN 3 JUSQU'A FIN DE BOUCLE
7. CALCULER $PREV(i) = PJ\emptyset(i) / PMA(R) \forall_i$
8. CALCULER $C(R) = \sum PREV(i) \cdot C(i)$
9. SI $[C(r) - 8]$ NEGATIF, ALLER EN 10, SINON EN 11
10. FAIRE $c(r) = 8$
11. ALLER EN 2 JUSQU'A FIN DE BOUCLE
12. $CT(N) = \sum PMA(r) \cdot C(r)$
13. CALCULER $G(N) = 8 - CT(N)$
14. CALCULER $Z(N) = an + b$
15. CALCULER $D(N) = G(N) - Z(N)$
16. IMPRIMER $G(N), Z(N), D(N)$
17. ALLER EN 1 JUSQU'EN FIN DE BOUCLE
18. GRAPHES DE $G(N)$ ET $Z(N)$
19. FIN

Comme t_i et $1 - t_i$ peuvent être petits avec des puissances élevées, on aura intérêt à passer par les logarithmes pour l'instruction de calcul 4.



Valeurs de : $E_n, G_n, Z_n, G_n - Z_n$, en fonction de N . N varie de 3 à 25 par pas de 1
 26 à 50 par pas de 2
 50 à 100 par pas de 5

RECAPITULATION DES RESULTATS									
N	E_n	G_n	Z_n	$G_n - Z_n$	N	E_n	G_n	Z_n	$G_n - Z_n$
3	7.2191	.7808	.6800	.1008	26	4.2438	3.7561	2.0600	1.6961
4	6.6057	1.3942	.7400	.6542	28	4.2031	3.7968	2.1800	1.6168
5	6.1334	1.8665	.8000	1.0665	30	4.1764	3.8235	2.3000	1.5235
6	5.7700	2.2299	.8600	1.3699	32	4.1604	3.8395	2.4200	1.4195
7	5.4908	2.5091	.9200	1.5891	34	4.1525	3.8474	2.5400	1.3074
8	5.2771	2.7228	.9799	1.7428	36	4.1510	3.8489	2.6600	1.1889
9	5.1141	2.8858	1.0400	1.8458	38	4.1104	3.8895	2.7800	1.1095
10	4.9900	3.0090	1.1000	1.9090	40	4.0780	3.9219	2.9000	1.0219
11	4.8938	3.1011	1.1600	1.9411	42	4.0525	3.9474	3.0200	.9274
12	4.8312	3.1687	1.2200	1.9487	44	4.0328	3.9671	3.1400	.8271
13	4.7828	3.2171	1.2800	1.9371	46	4.0176	3.9823	3.2600	.7223
14	4.7497	3.2502	1.3400	1.9102	48	4.0061	3.9938	3.3800	.6138
15	4.7287	3.2712	1.4000	1.8712	50	3.9977	4.0022	3.5000	.5022
16	4.7173	3.2826	1.4600	1.8226	55	3.9867	4.0132	3.8000	.2132
17	4.7136	3.2863	1.5200	1.7663	60	3.9754	4.0245	4.1000	-.0754
18	4.6711	3.3288	1.5800	1.7488	65	3.9536	4.0463	4.4000	-.3536
19	4.5820	3.4179	1.6400	1.7779	70	3.9394	4.0605	4.7000	-.6394
20	4.5063	3.4936	1.7000	1.7936	75	3.9303	4.0696	5.0000	-.9303
21	4.4421	3.5578	1.7600	1.7978	80	3.9248	4.0751	5.3000	-1.2248
22	4.3878	3.6121	1.8200	1.7921	85	3.9223	4.0776	5.6000	-1.5223
23	4.3419	3.6580	1.8799	1.7780	90	3.9182	4.0817	5.9000	-1.8182
24	4.3033	3.6966	1.9400	1.7566	95	3.9096	4.0903	6.2000	-2.1096
25	4.2709	3.7290	2.0000	1.7290	100	3.9025	4.0974	6.5000	-2.4025