Rapport sur la génération aléatoire des performances des actions sur les critères

Raymond BISDORFF

5 mars 2007

Introduction

Dans cette note je présente une étude approfondie de la génération aléatoire des performances d'un ensemble d'actions sur une famille de critères.

La première section sera consacrée à la distribution des différences. La deuxième est consacrée à la loi Gaussienne tronquée. La vérification empirique de mes générateurs de performances aléatoires a révélée une erreur pour la loi triangulaire. Aussi je développe dans une troisième section une formulation mathématique précise et rigoureuse de la loi triangulaire afin de mieux la comprendre et par la même de dégager les formules correctes.

1 Distribution théorique des différences de performances

Sur chaque critère nous utilisons une échelle de mesure dont nous notons m la valeur minimale et M la valeur maximale. Comme cas de référence et sans perte de généralité nous fixons m à 0 et M à 100.

Il faut remarquer que la différence $X - X$ d'une variable aléatoire X avec elle-même suit une loi de probabilité de moyenne $E(X) - E(X) = 0$ et de variance $V(X - X) = 2V(X)$. Si σ est l'écart-type de la variable aléatoire X, l'écart-type de la distribution des différences de performances sera donc toujours égal à $\sqrt{2}\sigma$, et ceci quelque soit la loi X envisagée.

Pour une génération selon une loi uniforme sur une échelle de 0 à 100 nous obtenons ainsi des différences de performance aléatoires de moyenne 0 et d'écart-type 40.82, c'est-à-dire $\sqrt{2\frac{(100-0)^2}{12}}$. Il faut noter que la distribution des différences pour la loi uniforme est parfaitement triangulaire de mode 0 et de répartisseur 0.5 avec un empan de densité non nulle de -100 à 100. Cette allure triangulaire de la distribution des différences est importante car elle caractérisera également la distribution des différences de performances que l'on obtiendra avec une loi Gaussienne tronquée associée à un écart-type élevé.

2 La loi de Gauss tronquée

Nous envisageons pour le moment une génération selon une loi de Gauss tronquée de moyenne 30, 50 ou 70 et d'écart-type 20, 25, ou 30. Avec l'aide du logiciel Mathematica, j'ai calculé les moyennes et écart-types théoriques de la loi de Gauss tronquée afin de vérifier si notre dispositif expérimental donne bien des distributions de différences de performance conformes avec la théorie.

Les résultats rassemblés dans le tableau 1 ci-dessous ont été obtenus avec des échantillons de 500 évaluation (24500 différences) observées à partir des performances aléatoires de 50 actions évaluées sur une famille de 10 critères avec une loi de probabilité donnée. $\mathcal{N}$ désignera la loi de Gauss alors que $\mathcal{N}_t$ désignera la loi de Gauss tronquée.

Chaque ligne du tableau indique d'abord la loi aléatoire de génération des performances considérée, suivi de l'écart-type théorique et empirique de la distribution des différences de performance correspondante. J'indique ensuite l'empan des différences qui supportent une densité non nulle. Ensuite j'indique la loi aléatoire avec laquelle nous faisons la différence de χ^2. Nous opérons pour cela avec un regroupement des 24500 différences en 30 classes, ce qui nous donne un degré de liberté de 29. Ainsi la valeur seuil du test de χ^2 à 5% de probabilité correspond à une distance de 17.7 (resp. 52.3 pour 99.5%) . La dernière colonne montre la valeur de χ^2 obtenue avec notre échantillon.

TAB. 1 – Différences de performance générées selon la loi de Gauss tronquée

$X \sim$	$\sigma^{\text{th}}_{(X-X)}$	$\sigma^{\text{empirique}}_{(X-X)}$	empan	$X - X \approx$	χ^2
$\mathcal{N}_t(50, 10)$	14.14	14.15	-45,+45	$\mathcal{N}(0, 14.15)$	2.3
$\mathcal{N}_t(50, 20)$	**27.00**	**27.10**	**-90,+90**	$\mathcal{N}(0, 27.00)$	**20.2**
$\mathcal{N}_t(50, 25)$	31.10	30.58	-100,+100	$\mathcal{N}(0, 30.58)$	80.0
$\mathcal{N}_t(50, 30)$	33.76	32.52	-100,+100	$\mathcal{N}(0, 32.52)$	136.1
$\mathcal{N}_t(50, 40)$	36,71	35.59	-100,+100	$\mathcal{N}(0, 35.59)$	193.3
$\mathcal{N}_t(50, 50)$	38.15	38.05	-100,+100	$\mathcal{N}(0, 38.05)$	320.2
$\mathcal{N}_t(50, 1000)$	40.81	40.07	-100,+100	$\mathcal{N}(0, 40.07)$	2087.0
$\mathcal{U}(0, 100)$	40.82	41.00	-100,+100	$\mathcal{T}(0, 0.5)$	0
$\mathcal{N}_t(70, 20)$	24.81	24.60	-90,+90	$\mathcal{N}(0, 24.60)$	24.6
$\mathcal{N}_t(30, 20)$	24.81	25.05	-80,+80	$\mathcal{N}(0, 25.05)$	4.7
$\mathcal{N}_t(70, 30)$	31.98	31.60	-100,+100	$\mathcal{N}(0, 31.60$	376.3

D'emblée on remarquera que les écart-types empiriques des différences obtenues avec notre dispositif expérimental sont parfaitement en règle avec les écart-types théoriques.

Ensuite on doit noter que $\mathcal{N}_t(50, 20)$ nous donne encore une distribution des différences de performance à peu près Gaussienne de moyenne nulle et d'écart-type 27. Par contre, dès que la troncature commence à montrer son effet statistique, on note que les lois $\mathcal{N}_t(50, 25 - 50)$ produisent des distributions de différences de performance qui s'écartent de plus en plus d'une loi Gaussienne pour se rapprocher de la distribtion triangulaire observée avec la génération uniforme des performances (voir section précédente). Pour un écart-type forcé de 1000, on obtient ainsi une distribution très proche de la triangulaire, ce qui produit une valeur de χ^2 de 2087.

Le déplacement du mode vers 70 ou 30, ne produit pas de résultats contrastés étant donné que la distribution symmétrique des différence de performance ne garde pas de trace de cette position du mode des performances. Par contre, l'effet troncature est amplifié. Ainsi, pour $\mathcal{N}_t(70, 30)$, nous obtenons déjà un χ^2 loin au-dessus de la valeur test, comparable à $\mathcal{N}_t(70, 50)$.

Passons maintenant à l'étude de loi triangulaire.

3 La loi triangulaire

Une variable aléatoire X suivant une loi de probabilité triangulaire de mode x_m et de répartisseur r sera notée $X \sim \mathcal{T}(x_m, r)$. Comme cas de référence nous retenons $X \sim \mathcal{T}(50, 0.5)$

La densité de la loi triangulaire est donnée par deux triangles qui dessinent avec leur surfaces respectives la probabilité r de m à x_m, respectivement $1 - r$ de x_m à M.

Pour générer des évaluations suivant une loi $X \sim \mathcal{T}(x_m, r)$, nous allons utiliser un générateur de nombres aléatoires u compris entre 0 et 1 que nous injectons dans l'inverse de la fonction de répartition de $\mathcal{T}(x_m, r)$.

Pour $0 \leq u < r$ nous avons ainsi :

$$X \;=\; m + \sqrt{\frac{u}{r}}\,(x_m - m), \tag{1}$$

et, pour $r \leq u \leq 1$, nous avons de même :

$$X \;=\; M - \sqrt{\frac{(1-u)}{(1-r)}}\,(M - x_m). \tag{2}$$

En effet, pour $m \leq X < x_m$, la fonction de répartition de la loi triangulaire de mode x_m et de répartisseur r est donnée par l'intégrale définie :

$$u = \int_m^X (ax+b)dx \;=\; \frac{r(X-m)^2}{(x_m-m)^2} \tag{3}$$

où les paramètres a et b sont déterminés par les deux conditions suivantes :

$$am + b = 0; \quad ax_m + b \;=\; \frac{2r}{x_m - m}. \tag{4}$$

Et, pour $x_m \leq X < M$, la fonction de répartition de la même loi triangulaire est donnée par l'intégrale définie :

$$u = r + \int_{x_m}^X (ax+b)dx = r + \frac{(1-r)(2M - X - x_m)(X - x_m)}{(M - x_m)^2}. \tag{5}$$

Les paramètres a et b sont déterminés cette fois par les deux conditions suivantes :

$$aM + b = 0; \quad ax_m + b = \frac{2(1-r)}{M - x_m}. \tag{6}$$

La moyenne de la loi triangulaire $\mathcal{T}(x_m, r)$ est donnée par la formule :

$$E(X) = \frac{mr + M(1-r) + 2x_m}{3}, \tag{7}$$

et la variance en est donnée par la formule un peu moins sympathique suivante :

$$V(X) = \frac{1}{18}\Big(m^2(3-2r)r - x_m(2mr+x_m) - 2M(1-r)(2mr+x_m) + M^2(1+r-2r^2)\Big) \tag{8}$$

Dans le cas de notre échelle de mesure de référence($m = 0$ et $M = 100$), la formule se simplifie cependant pour donner :

$$V(X) = \frac{1}{18}\Big(100^2(1+r-r^2) - 2\cdot 100(1+r)x_m + x_m^2\Big) \tag{9}$$

L'analyse empirique par inspection des histogrammes obtenus à partir des distributions d'évaluations aléatoires générées selon les formules (1) et (2) confirme leur caractère triangulaire déterminé par la position du mode x_m et par la grandeur de chaque triangle en fonction du choix du répartisseur r. Notre dispositif expérimental fonctionne par conséquent correctement.

TAB. 2 – Différences de performance générées selon la loi triangulaire

$X \sim$	$\sigma^{\text{th}}_{(X-X)}$	$\sigma^{\text{empirique}}_{(X-X)}$	empan	$X - X \approx$	χ^2
$\mathcal{T}(50, 0.5)$	**28.87**	**28.93**	**-85,+85**	$\mathcal{N}(0, 28.93)$	**570.4**
$\mathcal{T}(30, 0.5)$	29.63	29.90	-100,+100	$\mathcal{N}(0, 29.83)$	67.6
$\mathcal{T}(70, 0.5)$	29.63	29.83	-100,+100	$\mathcal{N}(0, 29.83)$	129.6
$\mathcal{T}(50, 0.33)$	27.74	27.34	-90,+90	$\mathcal{N}(0, 27.34)$	106.9
$\mathcal{T}(50, 0.66)$	27.86	28.50	-90,+90	$\mathcal{N}(0, 28.50)$	155.7
$\mathcal{T}(30, 0.33)$	29.82	29.60	-85,+85	$\mathcal{N}(0, 29.60)$	567.2
$\mathcal{T}(30, 0.66)$	27.38	27.44	-90,+90	$\mathcal{N}(0, 27.44)$	474.6
$\mathcal{T}(70, 0.33)$	27.17	27.77	-90,+90	$\mathcal{N}(0, 27.77)$	527.4
$\mathcal{T}(70, 0.66)$	29.87	29.24	-85,+85	$\mathcal{N}(0, 29.24)$	497.4

Les résultats statistiques, rassemblés dans le tableau 2 ci-dessus ont été obtenus avec des échantillons de 2500 évaluations (122500 différences) observées à partir des performances aléatoires de 50 actions évaluées sur une famille de 50 critères.

Il faut d'emblée remarquer que la loi triangulaire produit une distribution des différences de performance sensiblement différente d'une distribution Gaussienne (valeurs de χ^2 supérieures à la valeur test à 99.5%). Le changement de position du mode vers la valeur 30 ou 70 avec un répartisseur inchangé $r = 0.5$, augmente l'empan des différences obtenues et ramène leure distribution plus près d'une distribution Gaussienne.

Par contre, on doit remarquer que la valeur du χ^2 la plus élevée (> 500) est chaque fois observée avec les couples ($x_m = 50$,$r = 0.5$), ($x_m = 30$,$r = 0.33$) et ($x_m = 70$,$r = 0.66$). L'histogramme de la distribution des différences de performance pour ces deux derniers cas donne la même forme triangulaire cassée et rabaissée de manière symétrique aux points 0.33 et 0.66.