L. AMARAL AFONSO
P. DUARTE

Un nouveau test pour la distribution uniforme

Revue de statistique appliquée, tome 40, nº 1 (1992), p. 77-79
<http://www.numdam.org/item?id=RSA_1992__40_1_77_0>

Rev. Statistique Appliquée, 1992, **XXXX** (1), 77-79

UN NOUVEAU TEST POUR LA DISTRIBUTION UNIFORME

L. AMARAL AFONSO, P. DUARTE

Laboratório Nacional de Engenharia e Tecnologia Industrial Lisboa, Portugal

RÉSUMÉ

On propose un test pour voir si des échantillons de petite taille peuvent être considérés comme issus d'une population uniforme. Ce test est basé sur le fait que, pour une distribution en $(0,1)$, $F(x;\alpha) = \alpha x^2 + (1-\alpha)x$ est pour $|a| < 1$ une fonction de répartition qui se réduit à la loi uniforme si $\alpha = 0$. La région critique du test est de la forme $|\widehat{\alpha}| > c$, $\widehat{\alpha}$ étant l'estimateur des moindres carrés de α. Les valeurs critiques c (pour différents risques de première espèce) ont été obtenues par la méthode de Monte Carlo. On montre comment le test peut être utilisé dans le cas de la loi uniforme et d'une loi continue quelconque. Quelques remarques concernant la puissance du test pour des lois alternatives sont inclues.

Mots-clés : *Test, Loi Uniforme, Monte Carlo, Décision, Ajustement.*

ABSTRACT

A new test is presented in this paper. It is based upon the least squares parabole $F(x;\alpha) = \alpha x^2 + (1-\alpha)x$ with fixed points (0,0) and (1,1) that represents the cumulative frequencies of random number samples. When $\alpha = 0$ we have the cumulative function of the uniform distribution in $0 < x < 1$. Hence the critical region is $|\widehat{\alpha}| > c$ where $\widehat{\alpha}$ is the least square estimator of α. Confidence limits to different levels were calculated by simulation. It is shown how this test is extensible to the uniform distribution and to any continuous one as well. The power concerning some statistical laws is calculated.

Key-words : *Test, Uniform Distribution, Random Simulation, Decision, Fitting.*

Théorie. Recherche des limites de confiance

Le praticien de la statistique a souvent besoin de tester si un échantillon provient ou, au moins peut être supposé provenir, d'une population statistique uniforme.

Ce travail a pour but d'établir une méthode qui va nous rendre possible une réponse au problème suivant : supposant que nous avons un échantillon de taille assez petite, 5 à 50, nous envisageons de savoir s'il peut s'ajuster à la distribution uniforme («goodness of fit»).

On voit que l'expression da la fonction de répartition $F(x) = x$ d'une loi uniforme sur (0,1) est un cas particulier, $\alpha = 0$, de la parabole $F(x;\alpha) = \alpha x^2 + (1-\alpha)x$. L'estimation $\widehat{\alpha}$ de α obtenue par les moindres carrés est

$$\widehat{\alpha} = \left(\sum x_k^2 y_k - \sum x_k y_k - \sum x_k^3 + \sum x_k^2\right) / \left(\sum x_k^4 - 2\sum x_k^3 + \sum x_k^2\right) \quad (1)$$

où les x_k sont les valeurs ordonnées de l'échantillon de taille N et $y_k = (k-0,5)/N$ les ordonnées de la fonction cumulative. Notons que l'on peut envisager simplement $\widehat{\alpha}$ comme une statistique et donc la condition $|\alpha| < 1$ de croissance de $F(x;\alpha)$ et de localisation des paraboles dans le domaine $0 \leq x \leq 1$ et $0 \leq y \leq 1$ ne doit pas être restrictive à l'utilisation du test.

Les limites de confiance de la statistique des valeurs absolues de α ont été obtenues par simulation de Monte Carlo de la façon suivante : on a calculé (pour les tailles 5(5)50) la valeur de α d'après 50000 à 100000 échantillons. Les percentiles 95%, 97,5% et 99% sont évalués selon les moyennes de 8 répétitions par souci de rigueur. La table en annexe montre les valeurs 5(1)50 obtenues par interpolation. Les calculs ont été effectués par l'ordinateur NORSK DATA (compilateur ND-500, Fortran).

Mode d'emploi du test

On a affaire à un l'échantillon de valeurs réelles de taille N', $\{x'_k \varepsilon R, k = 1, \ldots, N'\}$. Nous allons considérer deux cas : celui de l'ajustement à la loi uniforme (rectangulaire) et celui dont la loi à tester est continue arbitraire.

test de la loi uniforme

a) on range les valeurs par ordre croissant. Notons-les $x_{(k)}$,

b) pour assimiler nos valeurs à des nombres aléatoires uniformes sur (0,1) effectuons la transformation $u_k = (x_{(k)} - x_{(1)}) \;/\; (x_{(N')} - x_{(1)})$. Rejetons les valeurs extrêmes 0 et 1 : on obtient un échantillon de taille $N'' = N' - 2$ où $x_k = u_{k+1},\ k = 1, 2, \ldots, N'-2$,

c) calculer $\widehat{\alpha}$ selon la formule (1) avec $y_k = (k-0,5)/N''$,

d) consulter la table des limites de confiance : si la valeur absolue de $\widehat{\alpha}$ est plus petite que celle correspondante à la taille, on peut accepter au niveau de confiance préalablement choisi que l'échantillon provient d'une population uniforme. Au contraire si elle est plus grande on refuse cette hypothèse.

test d'une loi quelconque de fonction de répartition $F(x)$

On range les x'_k par ordre croissant : notons-les $x'_{(k)}$. On pose $x_k = F(x'_{(k)})$. On obtient $\widehat{\alpha}$ (à partir des $x_{(k)}$) comme ci-dessus et on consulte la table. Les conclusions suivent comme auparavant.

Il faut bien remarquer que l'équation va transformer l'échantillon avec la loi de densité $f(x) = dF/dx$, en un autre ayant la loi uniforme. Comme les paramètres de la loi doivent être présents on a toujours le problème de les estimer à partir de l'échantillon lui-même. S'il y a des tests spécifiques pour la loi qu'on va tester, en principe ils sont préférables. Cependant seule une étude approfondie du pouvoir du test de l'uniformité en comparaison avec ceux-ci peut indiquer le meilleur choix.

Remarques sur la puissance

Pour évaluer la puissance du test α pour quelques autres lois, on s'est basé sur l'intervalle de confiance $(0, v)$ au niveau de 95% tel que prob $(|\alpha| < v) = 0,95$. Pour des tailles entre 10 et 30 on a obtenu les puissances suivantes : pour la loi normale entre 5,8 et 12,5%, pour celle du χ^2 à 2 degrés de liberté (exponentielle) elles se situent entre 50,0 et 98,6%, et à 6 degrés de liberté entre 20,5 et 73,3%.

Il faut comparer ces résultats avec ceux issus de l'utilisation du test classique du χ^2, aussi au niveau de 95%, (on a fait 3 classes et on n'a pas accompli la correction de continuité) : on a obtenu de 3,3 à 4,4% pour la loi uniforme (ce qui correspond en fait à l'estimation du risque de première espèce à savoir 5%, puisque l'hypothèse nulle correspond à la loi uniforme, et ce qui montre que le test du χ^2 est un test approché, correct seulement avec des tailles infinies), et de 6,8 à 42,4% pour la loi normale. Dans le cas des lois du χ^2 à 2 et 6 degrés de liberté on a obtenu respectivement des puissances de 41,6 à 96,0% et de 18,0 à 69,2%.

En conclusion : le test alpha est préférable si la distribution est asymétrique, au contraire si elle est normale ou symétrique le test du χ^2 est plus performant.

ANNEXE
Valeurs critiques du test alpha

N	95%	97.5	99	N	95%	97.5	99	N	95%	97.5	99
5	1.703	2.049	2.489	20	0.783	0.909	1.074	35	0.588	0.678	0.794
6	1.544	1.844	2.224	21	0.764	0.887	1.047	36	0.581	0.670	0.785
7	1.412	1.675	2.007	22	0.746	0.866	1.021	37	0.575	0.662	0.777
8	1.303	1.536	1.831	23	0.729	0.846	0.996	38	0.568	0.654	0.768
9	1.211	1.421	1.688	24	0.712	0.826	0.972	39	0.561	0.647	0.759
10	1.135	1.327	1.571	25	0.697	0.808	0.949	40	0.555	0.639	0.750
11	1.072	1.249	1.477	26	0.682	0.791	0.927	41	0.548	0.630	0.740
12	1.019	1.185	1.400	27	0.668	0.774	0.906	42	0.540	0.622	0.730
13	0.974	1.131	1.336	28	0.655	0.758	0.887	43	0.532	0.613	0.719
14	0.936	1.086	1.283	29	0.643	0.744	0.869	44	0.524	0.603	0.707
15	0.903	1.047	1.238	30	0.632	0.731	0.853	45	0.516	0.594	0.695
16	0.874	1.013	1.199	31	0.622	0.718	0.839	46	0.510	0.589	0.691
17	0.848	0.984	1.163	32	0.612	0.707	0.826	47	0.504	0.583	0.687
18	0.825	0.957	1.131	33	0.604	0.696	0.814	48	0.498	0.577	0.684
19	0.803	0.932	1.102	34	0.596	0.687	0.804	49	0.492	0.571	0.680
								50	0.486	0.566	0.676