استخدام المحاكاة للتحري عن حصانة اختبار الاستقطاب

د. صلاح حمزة عبد *

I- المقدمة

يعتبر التوزيع متعدد الحدود (The Multinomial Distribution) أحد أهم التوزيعات متعددة المتغيرات من النوع المتقطع ، وهو يمثل الحالة العامة لتوزيع ذي الحدين ، إذ يقال بان المتغيرات $x_1, x_2, ..., x_k$ ، التي تمثل عناصر في المتجه العشوائي $\underline{x}$ ، تخضع للتوزيع متعدد الحدود ، إذا تمثلت دالة كتلة الاحتمال المشتركة لها على النحو التالي [3] ،

$$p(\underline{x}) = p(x_1, x_2, ..., x_k) = \frac{n! \prod_{i=1}^{k} p_i^{x_i}}{\prod_{i=1}^{k} x_i!} \quad , \quad 0 < p_i < 1 \quad , \quad x_i = 0,1,...,n$$

$$i = 1,2,...,k \quad , \quad \sum_{i=1}^{k} p_i = 1 \quad , \quad \sum_{i=1}^{k} x_i = n \qquad ----(1)$$

ويشار لذلك بالشكل ، $M(\underline{x}; p_1, p_2, ..., p_k, n)$.

فتكون الدالة المولدة لعزوم هذا التوزيع بالشكل [3] ،

$$M_{\underline{x}}(t) = E\left(e^{\underline{t}'\underline{x}}\right) = \left\{\sum_{i=1}^{k} p_i e^{t_i}\right\}^n \qquad ----(2)$$

حيث أن $\underline{t} = (t_1, t_2, ..., t_k)'$.

أما متوسط وتباين أي متغير في المتجه العشوائي $\underline{x}$ فهما على التوالي $\mu_{x_i} = n p_i$ و $\sigma^2_{x_i} = n p_i q_i$ ، $(i = 1,2,...,k)$ ، إذ أن $q_i = 1 - p_i$ ، في حين إن التباين المشترك ما بين

* استاذ/ الجامعة المستنصرية/كلية الادارة والاقتصاد/قسم الاحصاء

مقبول للنشر بتاريخ 2005/10/30

أي متغيرين مثل x_i و x_j يتمثل بالشكل $\sigma_{ij} = -n\, p_i\, p_j$, $i \neq j$ ، لينتج عن ذلك إن الارتباط ما بين المتغيرين x_i و x_j ، $i \neq j$ ، هو عبارة عن ، $\rho_{ij} = -\sqrt{p_i p_j / q_i q_j}$. وبذلك فانه يمكن كتابة مصفوفة التباين والتباين المشترك للمتجه العشوائي $\underline{x}$ لتكون :

$$V = \begin{pmatrix} np_1q_1 & -np_1p_2 & \dots & -np_1p_k \\ -np_1p_2 & np_2q_2 & \dots & -np_2p_k \\ . & . & . & . \\ . & . & . & . \\ -np_1p_k & -np_2p_k & \dots & np_kq_k \end{pmatrix} \qquad ----(3)$$

ومصفوفة الارتباطات مابين كل زوج من عناصر المتجه العشوائي $\underline{x}$ لتكون ،

$$R = \begin{pmatrix} 1 & -\sqrt{p_1p_2/q_1q_2} & \dots & -\sqrt{p_1p_k/q_1q_k} \\ -\sqrt{p_1p_2/q_1q_2} & 1 & \dots & -\sqrt{p_2p_k/q_2q_k} \\ . & . & . & . \\ . & . & . & . \\ -\sqrt{p_1p_k/q_1q_k} & -\sqrt{p_2p_k/q_2q_k} & \dots & 1 \end{pmatrix} \qquad ----(4)$$

إن التوزيع متعدد الحدود هو عبارة عن توزيع شاذ (Singular distribution) ، [3] ، لكون رتبة أي من المصفوفتين V و R لا تساوي درجتها التي هي k ، بل إن رتبة أي منهما هي $k-1$.

لقد لاحظ كل من Johnson و Kotz عام 1978 ، [3] ، إن انحدار x_i على x_j ، $i \neq j$ ، هو انحدار خطي ، يمكن تمثيله على وفق المعادلة ،

$$E(x_i / x_j) = (n - x_j)\, p_i\, (1 - p_j)^{-1} \qquad ----(5)$$

كما لاحظا أيضا" إن انحدار x_i على $x_{b_1}, x_{b_2}, \dots, x_{b_r}$ ، $b_j \neq i$ ، $(j = 1,2,\dots,r)$ ، هو انحدار خطي متعدد ، يمكن تمثيله على وفق الصيغة ،

$$E(x_i / x_{b_1}, x_{b_2}, \dots, x_{b_r}) = \left(n - \sum_{j=1}^{r} x_{b_j}\right) p_i \left(1 - \sum_{j=1}^{r} p_{b_j}\right)^{-1} \qquad ----(6)$$

إن تطبيقات التوزيع متعدد الحدود واسعة ، بل هي أوسع من تطبيقات توزيع ذي الحدين الشهير ، لكون التوزيع متعدد الحدود ، كما أسلفنا ، عبارة عن الحالة العامة لتوزيع ذي الحدين ، وذلك عند وجود أصناف متعددة من الحوادث ، بدلا" من اثنتين كما هو الحال للظواهر التي تخضع لتوزيع ذي الحدين . إن ابرز تطبيقات التوزيع متعدد الحدود تتمثل في :

1) النظرية الحركية في الفيزياء [3] ، إذ تحتل الجزيئات ، وفي بعض الدراسات الذرات ، خلايا معينة (أصناف) في حيز حالة الظاهرة ، بمعنى أن كل جزيئه تقع ضمن خلية في حيز ذي ستة أبعاد ، ثلاثة منها للموقع ، الطول والعرض والارتفاع ، وثلاثة أخرى للسرعة .

2) عند كون البيانات مأخوذة كعينة عشوائية ، يمكن تصنيفها على وفق مجاميع محددة ، على أن تكون المشاهدات مستقلة عن بعضها البعض ، فيكون تطبيق متعدد الحدود ممكنا" لجداول التوافق التي تنطبق عليها هذه المواصفات .

3) التوزيع متعدد الحدود ذي أهمية بالغة في تحليل استطلاعات الرأي [1] ، التي تسبق الانتخابات ، فلو أن k من الأحزاب السياسية تشترك في انتخابات برلمانية ، وان p_i عبارة عن نسبة الناخبين المصوتين لصالح الحزب i في استطلاع للرأي قبل الانتخابات الفعلية ، فان المتجه الذي يحتوي على متغيرات تمثل أعداد المصوتين لهذا الحزب أو ذاك ، كعناصر ، سيخضع للتوزيع متعدد الحدود .

4) عند دراسة التغيرات التي تطرأ على كثافة مجتمع من الحيوانات المهاجرة [1] ، كأحد جوانب الدراسة في علم البيئة ،فعلى سبيل المثال ، مجتمع الأسماك في بحيرة ، إذ يلاحظ تجمع الأسماك في جيوب معينة في أوقات محددة من السنة بينما تنتشر في أرجاء مختلفة في أوقاتٍ أخرى ، فلإغراض الصيد يكون من المفيد تحديد الوقت الذي تبتدئ فيه الأسماك بالحركة خارجة" من تلك الجيوب . والحالة أعلاه تنطبق أيضا" على بعض أصناف الطيور المهاجرة .

II- الاستقطاب (The Polarization)

يقال للتوزيع متعدد الحدود ، ذي k من المتغيرات بأنه مستقطب **(Polarized)** ، إذا تمركزت كتلته الاحتمالية إجمالا" في عدد من المتغيرات اقل من k . ففي سنة **1981** قدم كلٌ من **Alam** و **Mitra** **[1]** ، معيارا" لقياس الاستقطاب ، واختبارا" لتحديد فيما إذا كانت مجموعة المتغيرات $x_1, x_2, ..., x_k$ تخضع لتوزيع متعدد حدود مستقطب بقيمة معيار استقطاب معينة أم لتوزيع متعدد حدود بقيمة معيار استقطاب اكبر .

إن ما قدمه كلٌ من **Alam** و **Mitra** يستند على الترتيب الجزئي لمتجهات مثل $\underline{p}$ ، هذا الترتيب الذي يتم بناء" على علاقة البروز **[1]** ، **(Majorization relation)** التالية ، إذ يقال بأن $\underline{p}$ أكثر بروزا" من $\underline{p}^*$ إذا كان $\underline{p}^* < \underline{p}$ ، أو بمعنى آخر إذا كان $\sum_{i=1}^{j} p_{(i)} \geq \sum_{i=1}^{j} p^*_{(i)}$ ، حيث أن $p_{(i)}$ تمثل القيمة ذات الترتيب i بعد ترتيب قيم $p_1, p_2, ..., p_k$ تنازليا" .

ومما ورد أعلاه ، فأنه يقال **[1]** ، للتوزيع متعدد الحدود ذي المتجه الاحتمالي $\underline{p}$ بأنه أكثر أو أقل استقطابا" من التوزيع متعدد الحدود ذي المتجه الاحتمالي $\underline{p}^*$ إذا كان $\underline{p}^* < \underline{p}$ أو $\underline{p} < \underline{p}^*$ على التوالي . فعلى سبيل المثال يمكن على وفق الوصف أعلاه كتابة ،

$$\left(\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k}\right) < \left(\frac{1}{k-1}, \frac{1}{k-1}, ..., 0\right) < (1, 0, ..., 0)$$

لقد عرّف كلٌ من **Hardy** و **Littewood** و **polya** عام **1952** **[2]** ، صنفا"من الدوال أسموه دوال شور المحدبة **(Schur-Convex function)** ، إذ يقال بان الدالة المتماثلة $f(\underline{p})$ ، دالة شور محدبة ، إذا كان $f(\underline{p}) \geq f(\underline{p}^*)$ لكل $\underline{p}^* < \underline{p}$ ، فعلى سبيل المثال الدالة $Q = \sum_{i=1}^{k} p_i^2$ هي دالة شور محدبة .

إن التطوير البالغ الاهتمام ، وذي الأهمية الكبيرة في بحثنا هذا ، هو ذاك الناجم عن بحثين متسلسلين أعد أولهما كلٌ من **Proschan** و **Sethuraman** عام **1977** **[8]** ، بينما أشترك مع من أعدا البحث الأول في ذات العام الباحث **Nevius**
أيضا" **[5]** ، بينوا فيهما أن الشرط الضروري والكافي للدالة المتماثلة القابلة للاشتقاق $f(\underline{p})$ لتكون دالة شور محدبة ، هو أن يكون ،

$$\left(\frac{\partial f(\underline{p})}{\partial p_i} - \frac{\partial f(\underline{p})}{\partial p_j}\right)(p_i - p_j) \geq 0$$

لكل قيم $i \neq j$.

فإذا افترضنا أن $\underline{x} = (x_1, x_2, ..., x_k)'$ عبارة عن متجه عناصره عبارة عن التكرارات في عينة مسحوبة من مجتمع يخضع للتوزيع متعدد الحدود ، إذ أن ، $\sum_{i=1}^{k} x_i = n$ ، وانه يراد اختبار الفرضية $H_o : \underline{p} < \underline{p_o}$ ضد الفرضية $H_1 : \underline{p} > \underline{p_o}$ ، إذ أن $\underline{p_o}$ عبارة عن متجه معلوم القيم . لقد اختار كلٌ من Alam و Mitra عام 1981 [1] ، احصاءة الاختبار التالية ،

$$T(\underline{x}) = \sum_{i=1}^{k} x_i^2 \Big/ n = \left(\sum_{i=1}^{k} (x_i - n/k)^2 + n^2/k\right) \Big/ n \qquad ----(7)$$

التي هي عبارة عن دالة شور محدبة ، وبيّنا بأن اختبار الفرضية H_o أعلاه ضد الفرضية H_1 ، تحت مستوى المعنوية α يتم على وفق المنطقة الحرجة $\{x : T(\underline{x}) > c\}$ ، على أن c ثابت يحدد من خلال العلاقة التالية ،

$$p\left\{T(\underline{x}) > c \,\middle|\, \underline{p} = \underline{p_o}\right\} = \alpha \qquad ----(8)$$

كما وذكر هذان الباحثان بأن احصاءة الاختبار في (7) المقترحة من قبلهما تؤول إلى أقل قيمة لها والبالغة n^2/k عندما $n \to \infty$ و $p_1 = p_2 = ... = 1/k$ ، وتؤول إلى اكبر قيمة لها والبالغة n^2 عندما تتمركز كتلة احتمال التوزيع متعدد الحدود في خلية واحدة .

وبما أن $T(\underline{x})$ هي دالة شور محدبة بدلالة $\underline{x}$ ، فقد برهن كلٌ من Marshal و Olkin عام 1979 [4] ، بأن دالة قوة الاختبار ستكون هي الأخرى دالة شور محدبة ولكن بدلالة $\underline{p}$ ، وعلى هذا فإنهما قد استنتجا بان الاختبار غير متحيز .

لقد عمل Alam و Mitra عام 1981 [1] ، على إيجاد توزيع لاحصاءة اختبار الاستقطاب $T(\underline{x})$ تحت فرضية العدم H_o لغرض تحديد قيمة c بموجب المعادلة (8) ، فافترضا بأن x^+ عبارة عن أقل عدد صحيح غير سالب اكبر من أو يساوي x وان x^- هو اكبر عدد صحيح اقل من او يساوي x .

وبافتراض أن $D_k(t; p_1, p_2, ..., p_k, n) = p(T \le t)$ تمثل دالة التوزيع التجميعية للمتغير T عند قيم $k \ge 2$ ، وافتراض أن ،

$$\left.\begin{aligned} C_{jr} &= C_r^n \; p_j^r \left(1 - p_j\right)^{n-r} \\ a_j &= \frac{n}{j} - \left\{n(t - n/j)(1 - 1/j)\right\}^{1/2} \\ b_j &= \frac{n}{j} + \left\{n(t - n/j)(1 - 1/j)\right\}^{1/2} \end{aligned}\right\} \quad ----(9)$$

فقد حصل هذان الباحثان على صيغة لدالة التوزيع التجميعية مستندين على خاصية الإرجاع(recursively property) التي تميّز التوزيع متعدد الحدود ، وذلك بالشكل ،

$$D_2(t; p_1, 1-p_1, n) = \begin{cases} 0 & , \quad t < n/2 \\ \sum_{r=a_2^+}^{b_2^-} C_{1r} & , \quad n/2 \le t \le n \\ 1 & , \quad t > n \end{cases} \quad ----(10)$$

$$D_k(t; p_1, p_2, ..., p_k, n) = \begin{cases} 0 & , \quad t < n/k \\ \sum_{r=a_k^+}^{b_k^-} C_{kr} \; D_{k-1}\left(\frac{nt - r^2}{n-r}; \frac{p_1}{1-p_k}, ..., \frac{p_{k-1}}{1-p_k}, n-r\right) & , \quad n/k \le t \le n \\ 1 & , \quad t < n \end{cases} \quad ---(11)$$

إن علاقة الإرجاع المستخدمة للحصول على الصيغ أعلاه ، تستند [1] في الواقع على حقيقة ان التوزيع الشرطي للمتجه العشوائي $\underline{u} = (x_1, x_2, ..., x_{k-1})'$ اعتمادا" على كون x_k معلومة ، هو عبارة عن توزيع متعدد حدود ، $M\left(\underline{u}; \frac{p_1}{1-p_k}, \frac{p_2}{1-p_k}, ..., \frac{p_{k-1}}{1-p_k}, n - x_k\right)$ ، وهكذا ...

أما التوزيع التقريبي للاحصاءة $T(\underline{x})$ عند قيم n الكبيرة ، فقد توصلا له أيضا" ، بأن افترضا أن $n\Sigma = n\left(D - \underline{p}\,\underline{p}'\right)$ عبارة عن مصفوفة التباين والتباين المشترك للمتجه $\underline{x}$ ، إذ أن D عبارة

عن مصفوفة قطرية ، عناصر قطرها الرئيسي عبارة عن قيم $p_1, p_2, ..., p_k$ ، وان $\lambda_1, \lambda_2, ..., \lambda_k$ عبارة عن الجذور المميزة للمصفوفة Σ ، حيث أن $\lambda_1 = 0$ ، وان G عبارة عن مصفوفة متعامدة ، الصف i فيها ما هو إلا المتجه المميز المناظر للجذر المميز λ_i ، إذ أن $G\Sigma G' = Diag(\lambda_1, \lambda_2, ..., \lambda_k)$ ، فإذا افترضنا أن $\underline{\delta} = G\,\underline{p}$ ، فان المتجه المميز المناظر للجذر المميز λ_1 سيكون عبارة عن $\left(\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}, ..., \frac{1}{\sqrt{k}}\right)$ ، كما سيكون $\delta_1 = \frac{1}{\sqrt{k}}$ ، وعليه فان ،

$$\varphi = \underline{p}'\,\underline{p} = \underline{\delta}'\,\underline{\delta} = \sum_{i=2}^{k} \delta_i^2 + 1/k \qquad ----(12)$$

وان

$$\underline{\delta}'G\Sigma G'\,\underline{\delta} = \sum_{i=2}^{k} \lambda_i\,\delta_i^2 = \underline{p}'\Sigma\,\underline{p} = \varphi_1 - \varphi^2 \qquad ----(13)$$

حيث أن $\varphi_1 = \sum_{i=1}^{k} p_i^3$ وان $\varphi_1 \geq \varphi^2$ ، [1] .

وبما أن $\underline{x}$ تتقارب [1] من التوزيع الطبيعي متعدد المتغيرات بمتجه المتوسطات $\underline{\mu} = n\,\underline{p}$ ومصفوفة التباين والتباين المشترك $n\Sigma$ ، فان Alam و Mitra قد برهنا أن توزيع T يتقارب من توزيع ،

$$\frac{n}{k} + \sum_{i=2}^{k} \lambda_i\, z_i \qquad ----(14)$$

إذ أن z_i يخضع لتوزيع مربع كاي اللامركزي بدرجة حرية واحدة ومعلمة لامركزية $n\delta_i^2/\lambda_i$. وحيث ان $z_2, z_3, ..., z_k$ متغيرات مستقلة عن بعضها البعض [1] ، فقد افترض Alam و Mitra بان η و υ ، على التوالي عبارة عن متوسط وتباين المتغير في (14) ، فكتبا ،

$$\left.\begin{aligned} \eta &= 1 + (n-1)\varphi \\ \upsilon &= 2\left(\varphi^2 + \varphi - 2\varphi_1\right) + 4n\left(\varphi_1 - \varphi^2\right) \end{aligned}\right\} \qquad ----(15)$$

وبما أن العزم المركزي κ_r من الدرجة r لتوزيع مربع كاي اللامركزي ، $\chi^2_{m,\lambda}$ هو [3] ،

$$\kappa_r = 2^{r-1}\{m + r\lambda\}(r-1)! \qquad ----(16)$$

فان $\chi^2_{m,\lambda}$ سيتقارب من التوزيع الطبيعي عندما $\lambda \to \infty$. وعلى وفق ذلك [3] سيتقارب توزيع $(T-\eta)\upsilon^{-1/2}$ من التوزيع الطبيعي القياسي .

وفي الواقع [1] ، فان هناك حالتان ستكونا قائمتين ، ذلك عند كون $\varphi_1 = \varphi^2$ هما :

1) $p_1 = p_2 = ... = p_k = 1/k$. وفي هذه الحالة سيكون $\lambda_2 = \lambda_3 = ... = \lambda_k = 1/k$ ، لنحصل من (13) على $\delta_i = 0$ لقيم $i = 2,3,...,k$ ، وعليه فان z_1 سيخضع لتوزيع مربع كاي اللامركزي بدرجة حرية واحدة ، وبالتالي من (14) فان توزيع $kT - n$ سيتقارب من توزيع مربع كاي ب $k-1$ من درجات الحرية .

2) $p_i\, p_j = 0$ لكل قيم $i \neq j$. وفي هذه الحالة سيكون $T = n$ باحتمال مقداره واحد .

ولغرض المقارنة بين الصيغة التقريبية والصيغة المضبوطة ، لتوزيع الاحصاءة T ، ليكن $T_\alpha = \eta + \sqrt{\upsilon}\, Z_\alpha$ ، إذ أن Z_α تمثل المئين الأكبر من الدرجة α للتوزيع الطبيعي القياسي ، فمن خلال توزيع T التقريبي فان قوة اختبار الاستقطاب التقريبية ماهي إلا ،

$$\Phi\left(\frac{\sqrt{n}(\varphi - \varphi_o)}{2(\varphi_1 - \varphi^2)} - \left(\frac{\varphi_{1o} - \varphi_o^2}{\varphi_1 - \varphi^2}\right)^{1/2} Z_\alpha\right) \qquad ----(17)$$

إذ أن Φ تمثل الدالة التجميعية للتوزيع الطبيعي القياسي ، وان φ_{1o} و φ_o تمثل قيم φ_1 و φ على التوالي عند $\underline{p} = \underline{p_o}$.

III- هدف البحث

بعد العرض المقدم في أعلاه لاختبار الاستقطاب واحصاءته ، والذي تتأتى أهميته من أهمية التوزيع متعدد الحدود ، قيد الاستخدام فيه ، وأهمية التطبيقات الواسعة جدا" التي يتميز بها هذا

التوزيع ، وفي مختلف حقول العلم ، ارتأينا دراسة حصانة هذا الاختبار فيما لو تم توليد البيانات على وفق توزيعات متقطعة أخرى غير التوزيع قيد الاهتمام ، هذه التوزيعات سيتم ذكرها في الفقرة اللاحقة . وذلك على وفق حجوم عينات مختلفة ، منها ما هو صغير ومنها ما هو متوسط ومنها ما هو كبير ، وبعدد متغيرات مختلف ، ومتجه معالم $\underline{p}$ مختلف .

IV- وصف تجربة المحاكاة

لغرض تحقيق ما نصبو إليه من هدف فقد تم بناء تجربة محاكاة ، مستعملين لاستخلاص نتائجها برنامجا" مكتوبا" من قبل الباحث بلغة فجوال بيزك ، وذلك على وفق الفروض التالية ،

1. تم افتراض ثلاث حجوم عينات ، صغيرة $s=10$ ومتوسطة $s=35$ وكبيرة $s=125$.
2. تم افتراض عدد المحاولات ، $n=20$ مرة , وافتراض $n=30$ ، مرة" أخرى .
3. تم افتراض عدد المتغيرات k مساوي مرة للثلاثة وأخرى للخمسة .
4. تم افتراض خضوع $\underline{x}=(x_1,x_2,x_3)'$ و $\underline{x}=(x_1,x_2,x_3,x_4,x_5)'$ للتوزيعات المتعددة المتغيرات المتقطعة التالية ، إضافة" للتوزيع متعدد الحدود ، لغرض مقارنة الحصانة معه ، كونه أساس بناء معيار الاختبار ،

a) التوزيع متعدد الحدود السالب (Negative Multinomial) : ويتمثل بدالة كتلة الاحتمال ذات الصيغة [6] ،

$$p(x_1,x_2,...,x_k)=\frac{\Gamma\left(\xi+\sum_{i=1}^{k}(x_i-1)\right)}{\left(\prod_{i=1}^{k}(x_i-1)!\right)\Gamma(\xi)}\,q^{-\xi}\prod_{i=1}^{k}\left(\frac{p_i}{q}\right)^{x_i-1} \qquad ----(18)$$

$$x_j\geq 1\ ,\ j=1,2,...,k\ ,\ \xi>0$$

$$q-\sum_{i=1}^{k}p_i=1$$

إذ أن ξ عبارة عن عدد صحيح موجب ، وان كل x_j يمثل عدد حالات الفشل حتى الوصول إلى ξ من حالات النجاح على اعتبار p_i يمثل احتمال النجاح .

ولغرض التوفيق مع التوزيع متعدد الحدود تماما" ، وحتى تكون المقارنة صائبة ، فقد تم بتر التوزيع أعلاه عند n من المحاولات بقدر عدد تلك التي ستؤخذ للتوزيع متعدد الحدود ، كما وأُخذت $\xi = 3$ كتنفيذ فعلي عند إجراء تجربة المحاكاة .

b) التوزيع الهندسي متعدد المتغيرات ، وهو على نفس الصيغة في (18) أعلاه ولكن بقيمة ξ مساوية للواحد . إن نفس ما ذُكر في التوزيع متعدد الحدود السالب ، الذي يكون التوزيع الهندسي متعدد المتغيرات ، حالة خاصة منه ، سينطبق هنا أيضا" .

c) التوزيع الهندسي الزائدي متعدد المتغيرات (Multivariate Hyper-Geometric) ، ويتمثل بدالة كتلة الاحتمال ذات الصيغة [9] ،

$$p(x_1, x_2, ..., x_k) = \frac{\prod_{j=1}^{k} C_{x_j}^{M_j}}{C_n^M} \qquad ----(19)$$

$$\sum_{j=1}^{k} x_j = n \ , \ j = 1,2,...,k \ , \ 0 \le x_j \le M_j$$

إذ لو افترضنا أن هناك مجتمع مؤلف من M من المفردات ، وفي هذا المجتمع k من الأصناف ، M_1 من مفردات المجتمع تقع في الصنف الأول ، و M_2 من مفردات المجتمع تقع في الصنف الثاني ، وهكذا .. فان M_k من مفردات المجتمع تقع في الصنف k ، فيكون بالطبع $\sum_{j=1}^{k} M_j = M$. فإذا سحبنا عينة عشوائية بحجم n من هذا المجتمع فان التوزيع المشترك للمتغيرات العشوائية $x_1, x_2, ..., x_k$ ، التي تمثل عدد المفردات المأخوذة ضمن العينة من كل صنف من أصناف المجتمع على التوالي ، سيكون هو المذكور في الصيغة (19) .

لقد تم افتراض أن $M_j = n$ ، كما تم بتر التوزيع عند $x_j = 1$ عند إجراء المحاكاة لغرض التوافق مع خصائص التوزيع متعدد الحدود ، لأجل أن تكون المقارنة معه صائبة .

d) توزيع المتسلسلة اللوغارتمية متعدد المتغيرات (Multivariate Logarithmic series) ، بالمعلمات $\theta_1, \theta_2, ..., \theta_k$ ، ويتمثل بدالة كتلة الاحتمال ذات الصيغة [7] ،

$$p(x_1, x_2, \dots, x_k) = \frac{(n-1)! \prod_{j=1}^{k} \theta_j^{x_j}}{\left(\prod_{j=1}^{k} (x_j)!\right)\left\{-Log\left(1 - \sum_{j=1}^{k} \theta_j\right)\right\}} \qquad ----(20)$$

$$\sum_{j=1}^{k} x_j = n > 0 \ , \ 0 \le x_j \ , \ 0 < \sum_{j=1}^{k} \theta_j < 1$$

لقد تم بتر هذا التوزيع من الجانبين ، عند القيمة واحد من الحد الأدنى وعند القيمة n من الحد الأعلى ، عند إجراء المحاكاة وذلك لغرض التوافق مع خصائص التوزيع متعدد الحدود أساس إنشاء الاختبار ومعياره ، لأجل أن تكون المقارنة معه صائبة .

5. **تم اعتماد مستويات المعنوية الشائعة لإغراض الاختبار 0.01 و 0.05 و 0.10 .**
6. **تم استخدام حجم مكرر مقداره 1000 لكل من التوافيق المذكورة في أعلاه .**
7. **تم الاستناد على مستوى المعنوية التجريبي وقوة الاختبار التجريبية للحكم على مدى حصانة الاختبار ، اذ أن مستوى المعنوية التجريبي ، يتمثل بنسبة عدد مرات رفض الفرضية وهي واقعا" صحيحة ، وان قوة الاختبار التجريبية ، تتمثل بنسبة عدد مرات قبول الفرضية وهي واقعا" خطأ ، مطروحة" من الواحد .**

V- استعراض النتائج

بعد تنفيذ تجربة المحاكاة المشار إليها في الفقرة (IV) السابقة ، تم الحصول على النتائج الواردة في الجدول رقم (1) أدناه :

جدول رقم (1) : يمثل نتائج مستوى المعنوية التجريبي $\hat{\alpha}$ وقوة الاختبار التجريبية $1-\hat{\beta}$ ، لعدد متغيرات k مختلفة ، وحجوم عينات s مختلفة ، وقيم n مختلفة ، ومستويات معنوية فعلية α مختلفة ، وتوزيعات مختلفة .

	α		$k=3$						$k=5$					
			$s=10$		$s=35$		$s=125$		$s=10$		$s=35$		$s=125$	
			n=20	n=30	n=20	n=30	n=20	n=30	n=20	n=30	n=20	n=30	n=20	n=30
متعدد الحدود	0.01	$\hat{\alpha}$	0.0084	0.0092	0.0088	0.0094	0.0092	0.0098	0.0085	0.0094	0.0090	0.0096	0.0093	0.0101
		$1-\hat{\beta}$	0.7162	0.7231	0.7222	0.7658	0.7610	0.8083	0.7214	0.7333	0.7303	0.7668	0.7989	0.8465
	0.05	$\hat{\alpha}$	0.0432	0.0449	0.0441	0.0462	0.0453	0.0472	0.0436	0.0452	0.0450	0.0466	0.0459	0.0485
		$1-\hat{\beta}$	0.8065	0.8149	0.8094	0.8222	0.8144	0.8337	0.8118	0.8372	0.8294	0.8441	0.8366	0.8764
	0.1	$\hat{\alpha}$	0.0833	0.0924	0.0876	0.0965	0.0925	0.0967	0.0877	0.0944	0.0929	0.0963	0.0938	0.0973

		$1-\hat{\beta}$	0.8818	0.9202	0.9087	0.9432	0.9237	0.9543	0.9109	0.9431	0.9199	0.9448	0.9348	0.9765
متعدد الحدود السالب	0.01	$\hat{\alpha}$	0.0077	0.0085	0.0081	0.0087	0.0085	0.0091	0.0078	0.0087	0.0083	0.0089	0.0086	0.0094
		$1-\hat{\beta}$	0.6155	0.6224	0.6215	0.6651	0.6603	0.7076	0.6207	0.6326	0.6296	0.6661	0.6982	0.7458
	0.05	$\hat{\alpha}$	0.0425	0.0442	0.0434	0.0455	0.0446	0.0465	0.0429	0.0445	0.0443	0.0459	0.0452	0.0478
		$1-\hat{\beta}$	0.7058	0.7142	0.7087	0.7215	0.7137	0.7330	0.7111	0.7365	0.7287	0.7434	0.7359	0.7757
	0.10	$\hat{\alpha}$	0.0826	0.0917	0.0869	0.0958	0.0918	0.0960	0.0870	0.0936	0.0922	0.0956	0.0931	0.0968
		$1-\hat{\beta}$	0.7811	0.8195	0.8080	0.8425	0.8230	0.8536	0.8102	0.8424	0.8192	0.8441	0.8341	0.8759
هندسي متعدد متغيرات	0.01	$\hat{\alpha}$	0.0074	0.0081	0.0078	0.0083	0.0081	0.0087	0.0075	0.0083	0.0079	0.0085	0.0082	0.0090
		$1-\hat{\beta}$	0.5890	0.5956	0.5948	0.6365	0.6319	0.6772	0.5940	0.6054	0.6025	0.6375	0.6682	0.7137
	0.05	$\hat{\alpha}$	0.0407	0.0423	0.0415	0.0435	0.0446	0.0445	0.0411	0.0426	0.0424	0.0439	0.0432	0.0457
		$1-\hat{\beta}$	0.6755	0.6835	0.6782	0.6905	0.6830	0.7015	0.6805	0.7048	0.6974	0.7114	0.7043	0.7423
	0.10	$\hat{\alpha}$	0.0790	0.0878	0.0832	0.0917	0.0879	0.0919	0.0833	0.0896	0.0882	0.0915	0.0891	0.0926
		$1-\hat{\beta}$	0.7475	0.7843	0.7733	0.8063	0.7876	0.8169	0.7754	0.8062	0.7840	0.8078	0.7982	0.8382
هندسي زائدي متعدد	0.01	$\hat{\alpha}$	0.0068	0.0075	0.0072	0.0076	0.0075	0.0079	0.0069	0.0076	0.0073	0.0078	0.0075	0.0082
		$1-\hat{\beta}$	0.5399	0.5459	0.5484	0.5834	0.5792	0.6207	0.5444	0.5548	0.5522	0.5843	0.6124	0.6541
	0.05	$\hat{\alpha}$	0.0373	0.0387	0.0380	0.0399	0.0409	0.0408	0.0376	0.0390	0.0388	0.0402	0.0396	0.0419
		$1-\hat{\beta}$	0.6191	0.6265	0.6215	0.6329	0.6260	0.6430	0.6237	0.6459	0.6391	0.6520	0.6455	0.6803
	0.10	$\hat{\alpha}$	0.0724	0.0805	0.0763	0.0840	0.0806	0.0842	0.0764	0.0822	0.0808	0.0838	0.0817	0.0849
		$1-\hat{\beta}$	0.6851	0.7189	0.7087	0.7390	0.7219	0.7512	0.7107	0.7389	0.7186	0.7404	0.7315	0.7682
متسلسلة لوغارتمية متعدد	0.01	$\hat{\alpha}$	0.0073	0.0080	0.0077	0.0082	0.0080	0.0085	0.0074	0.0082	0.0078	0.0084	0.0081	0.0088
		$1-\hat{\beta}$	0.5796	0.5861	0.5888	0.6263	0.6218	0.6664	0.5845	0.5957	0.5929	0.6273	0.6575	0.7023
	0.05	$\hat{\alpha}$	0.0400	0.0416	0.0408	0.0428	0.0439	0.0438	0.0404	0.0419	0.0417	0.0432	0.0425	0.0450
		$1-\hat{\beta}$	0.6647	0.6726	0.6673	0.6795	0.6721	0.6903	0.6696	0.6935	0.6862	0.7000	0.6930	0.7304
	0.10	$\hat{\alpha}$	0.0777	0.0864	0.0819	0.0902	0.0865	0.0904	0.0820	0.0882	0.0868	0.0900	0.0877	0.0911
		$1-\hat{\beta}$	0.7355	0.7718	0.7609	0.7934	0.7750	0.8065	0.7630	0.7933	0.7715	0.7949	0.7854	0.8248

ومن هذه النتائج يمكن القول بالحصانة النسبية لاختبار الاستقطاب ، على أن أجود النتائج هي تلك العائدة لحالة التوزيع متعدد الحدود لمتجهات البيانات ، وهذا متوقع كون التوزيع المذكور هو أساس بناء معيار الاختبار ، على أننا ينبغي أن نأخذ بنظر الاعتبار ما يلي ،

1) تدرجت حصانة الاختبار نسبيا" على وفق التوزيعات المفترضة لمتجهات البيانات ، من أفضلها إلى اقلها ، وكما يلي ، التوزيع متعدد الحدود السالب ، ومن ثم التوزيع الهندسي متعدد المتغيرات ومن ثم توزيع المتسلسلة اللوغارتمية متعدد المتغيرات ومن ثم التوزيع الهندسي الزائدي متعدد المتغيرات .
2) هناك علاقة طردية ما بين حصانة الاختبار و حجم العينة s ، إذ تزداد الحصانة بزيادة حجم العينة ، لأن الأقيام التجريبية لمستوى المعنوية تقترب من نظائرها الفعلية ، كما أن أقيام قوة الاختبار التجريبية تأخذ بالزيادة بزيادة حجم العينة .
3) تزداد حصانة الاختبار بزيادة عدد المحاولات n ، حيث تزداد أقيام قوة الاختبار التجريبية بزيادة عدد المحاولات ، كما أن الأقيام التجريبية لمستوى المعنوية تأخذ بالاقتراب من نظائرها الفعلية.
4) لعدد متغيرات k ، تأثير طردي هو الآخر على الحصانة ، إذ تزداد الحصانة بزيادة عدد متغيرات، لأن الأقيام التجريبية لمستوى المعنوية تأخذ بالاقتراب من نظائرها الفعلية بزيادة عدد متغيرات ، كما أن أقيام قوة الاختبار التجريبية تأخذ بالزيادة هي الأخرى بزيادة عدد متغيرات .

المصادر

1) Alam , K. and Mitra , A. (1981) " Polarization test for the multinomial distribution " , JASA , Vol. 76 , No. 373 , march , pp. 107-109 .
2) Hardy , G. & Littlewood , J. and Polya , G. (1952) "Inequalities " , Cambridge university press .
3) Johnson , N. and Kotz , S. (1976) " Distributions in statistics : Discrete Distributions " , wiley series , USA .
4) Marshal , A. and Olkin , I. (1979) " Inequalities : Theory of majorization and its applications " , Academic press, New York ,USA .
5) Nevius , S. & Proschan , F. and Sethuraman , J. (1977) " Schur Functions in statistics : II , stochastic majorization " , Annals of statistics , 5 , pp. 263-273 .

6) **Neyman , J. (1963) " Certain chance mechanisms involving discrete distributions " proceedings of the international symposium on discrete distributions , USA .**
7) **Patil ,G. and Bildikan , S. (1976) " Multivariate logarithmic series distribution as a probability model in population and community ecology and some of its statistical properties", JASA , 62 , pp. 655-674 .**
8) **Proschan , F. and Sethuraman , J. (1977) " Schur Functions in statistics : I , The preservation theorem " , Annals of statistics , 5 , pp. 256-262 .**
9) **Steyn , H. (1955) " On discrete multivariate probability functions of hypergeometric type " , Annals of mathematical statistics , 8 , pp. 621-633 .**

……

……

……