تقدير معلمات التوزيع فوق الهندسي السالب

د. عبدالحليم مولود الصويعي
قسم الإحصاء – كلية العلوم – الزاوية
جامعة الزاوية

مقدمة:

بفرض أن مجتمع محدود يتكون من صفتين حيث يمكن التفكير بهما كالتالي: مجموعة كرات حمراء ومجموعة كرات بيضاء موضوعين في صندوق، عليه هناك N كرة في الصندوق بحيث عدد D كرة يحملون صفة اللون الأحمر. إذا تم سحب عدد من الكرات عشوائيا، كم من الكرات يجب أن يتم سحبها حتى الحصول على r كرة حمراء حيث أن $1 \le r \le D$. هذا المتغير يتبع في تغيراته التوزيع فوق الهندسي السالب الذي عادة لايتم التطرق إليه في الكتب الإحصائية العربية ونادرا جدا ما نجده في الكتب التي تعرض التوزيعات الاحتمالية المنفصلة أيضا. أعتقد أنه هناك عدداً من الأسباب لعدم دراسة هذا التوزيع من أهمها أن توزيع فوق

الهندسي السالب قليل الظهور في التطبيقات العملية وإذا ما عرض فإنه يتم استخدام توزيع ذات الحدين السالب كتقريب له (Jones, 2013). بالرغم من ذلك فإن التوزيـــع فـــوق الهندســـي السالب يعتبر من التوزيعات الاحتمالية المنفصلة المهمة (Norman and Kotz, 1977). كذلك هناك من يطالب بإدراج هذا التوزيع من ضمن مقررات التوزيعات الاحتمالية باعتباره يقـــرب الصورة من حيث السحب بدون ارجاع. التوزيع فوق الهندسي السالب له نفس أهمية توزيع ذات الحدين السالب حيث إن التوزيع فوق الهندسي السالب يعرض الوقت المنتظر عند السحب بدون ارجاع من عينة محدودة بينما توزيع ذات الحدين السالب يعرض الوقت المنتظر عند الســـحب بالارجاع (Miller and Fridel, 2007).

في هذا البحث سيتم دراسة التوزيع فوق الهندسي السالب من حيث عرض دالة الكتلـــة الاحتمالية للتوزيع واشتقاق متوسطه وتباينه، ومن ثم سيتم التطرق إلى الهدف الأساسي من هذا البحث وهو تقدير معلمات التوزيع فوق الهندسي السالب حيث هناك حالتين يمكن دراستهما:

1) تقدير عدد العناصر التي تحمل الصفة D عندما يكون عدد العناصر الكلية N معلوم.

2) تقدير عدد العناصر الكلية N عندما يكون عدد العناصر التي تحمل الصفة D معلوم.

أخيرا، سنستخدم أسلوب المحاكاة لتقييم جودة المقدرات.

دالة الكتلة الاحتمالية للتوزيع:

إذا كان لدينا مجتمع مكون من N عنصرا منها D عنصـــرا يحمـــل صـــفة معينـــة و $N-D$ لايحملون تلك الصفة وتم اختيار مفردة بعد الأخرى وبدون ارجاع إلى حين الحصول على r عنصرا يحمل الصفة بالعينة، وأن عدد العناصر التي لاتحمل هذه الصفة بالعينة متغيـــر عشوائي x بحيث أن $x+r$ تمثل عدد العناصر الكلية للتجربة، فإنه يقال للمتغير العشـــوائي x

بأنه يتبع في تغيراته توزيعا احتماليا سمي بالتوزيع فوق الهندسي السالب وله دالة كتلة احتمالية تعطى بالعلاقة التالية:

$$P_X(x) = \frac{\binom{N-D}{x}\binom{D}{r-1}}{\binom{N}{x+r-1}} \cdot \frac{D-r+1}{N-x-r+1} \qquad , \qquad x = 0,1,2,\dots,N-D$$

يمكن تبسيط الصيغة السابقة كالاتي:

$$P_X(x) = \frac{\binom{N-D}{x}\binom{D}{r-1}}{\binom{N}{x+r-1}} \cdot \frac{D-r+1}{N-x-r+1}$$

$$= \frac{\frac{D!}{(r-1)!(D-r+1)!}\frac{(N-D)!}{x!(N-D-x)!}}{\frac{N!}{(x+r-1)!(N-x-r+1)!}} \cdot \frac{D-r+1}{N-x-r+1}$$

$$= \frac{D!}{(r-1)!(D-r)!} \cdot \frac{(N-D)!}{x!(N-D-x)!} \cdot \frac{(x+r-1)!(N-x-r)!}{N!}$$

$$= \frac{\frac{(x+r-1)!}{(r-1)!x!} \cdot \frac{(N-x-r)!}{(D-r)!(N-D-x)!}}{\frac{N!}{D!(N-D)!}}$$

$$P_X(x) = \frac{\binom{x+r-1}{r-1}\binom{N-x-r}{D-r}}{\binom{N}{D}} \;, \quad x = 0,1,2,\dots,N-D$$

عليه باستخدام النتيجة التالية يمكن اثبات أن مجموع دالة الكتلة الاحتمالية للتوزيع على جميع قيم المتغير العشوائي x يساوي الواحد الصحيح.

نتيجة: لأي أعداد طبيعية a, b, n, m فإن :

$$\sum_k \binom{k+b}{a}\binom{n-k}{m-a} = \binom{n+b+1}{m+1}$$

وبتطبيق النتيجة أعلاه نجد أن:

$$\sum_x \binom{x+r-1}{r-1}\binom{N-x-r}{D-r} = \binom{N}{D}$$

وبالتالي فإن

$$\sum_x P_X(x) = \frac{\binom{N}{D}}{\binom{N}{D}} = 1$$

متوسط وتباين التوزيع:

1) متوسط التوزيع:

يمكن إيجاد متوسط التوزيع كالأتي:

$$E(X) = \sum xP(x)$$

$$= \sum x \frac{\binom{N-D}{x}\binom{D}{r-1}}{\binom{N}{x+r-1}} \cdot \frac{D-r+1}{N-x-r+1}$$

$$= \sum x \frac{\frac{(N-D)!}{x!(N-D-x)!} \cdot \frac{D!}{(r-1)!(D-r-1)!}}{\frac{N!}{(x+r-1)!(N-x-r+1)!}} \cdot \frac{D-r+1}{N-x-r+1}$$

$$= \frac{D!(N-D)!}{N!} \sum \frac{(x+r-1)!(N-x-r+1)(N-x-r)!}{(x-1)!(N-D-x)!(r-1)!(D-r+1)(D-r)!} \cdot \frac{(D-r+1)}{(N-x-r+1)}$$

$$= \frac{D!(N-D)!}{N!} \sum \frac{(x+r-1)!}{(r-1)!(x-1)!} \cdot \frac{(N-x-r)!}{(D-r)!(N-D-x)!}$$

$$= \frac{D!(N-D)!}{N!} r \sum \frac{(x+r-1)!}{r!(x-1)!} \binom{N-x-r}{D-r}$$

$$= \frac{D!(N-D)!}{N!} r \sum \binom{x+r-1}{r} \binom{N-x-r}{D-r}$$

$$= \frac{D!(N-D)!}{N!} r \binom{N}{D+1}$$

$$= \frac{rD!(N-D)(N-D-1)!}{(D+1)D!(N-D-1)!}$$

$$E(X) = \frac{r(N-D)}{(D+1)}$$

2) تباين التوزيع:

يمكن إيجاد تباين التوزيع كالاتي:

$$Var(X) = E(X^2) - [E(X)]^2$$

حيث يمكن كتابة (X^2) بالصيغة

$$E(X^2) = E[X(X-1)] + E(x)$$

نوجد أولا $E[X(X-1)]$

$$E[X(X-1)] = \sum x(x-1)P(x)$$

$$= \sum x(x-1)\frac{\binom{N-D}{x}\binom{D}{r-1}}{\binom{N}{x+r-1}}\cdot\frac{D-r+1}{N-x-r+1}$$

$$= \sum x(x-1)\frac{\frac{(N-D)!}{x!(N-D-x)!}\cdot\frac{D!}{(r-1)!(D-r-1)!}}{\frac{N!}{(x+r-1)!(N-x-r+1)!}}\cdot\frac{D-r+1}{N-x-r+1}$$

$$= \frac{D!(N-D)!}{N!}\sum x(x-1)\frac{(x+r-1)!(N-x-r+1)(N-x-r)!}{x!(N-D-x)!(r-1)!(D-r+1)!}\cdot\frac{D-r+1}{N-x-r+1}$$

$$= \frac{D!(N-D)!}{N!}\sum\frac{(x+r-1)!}{(r-1)!(x-2)!}\cdot\frac{(N-x-r)!}{(D-r)!(N-D-x)!}$$

$$= \frac{D!(N-D)!}{N!}r(r+1)\sum\frac{(x+r-1)!}{(r+1)!(x-2)!}\binom{N-x-r}{D-r}$$

$$= \frac{r(r+1)D!(N-D)!}{N!}\sum\binom{x+r-1}{r+1}\binom{N-x-r}{D-r}$$

$$= \frac{r(r+1)D!(N-D)!}{N!}\binom{N}{D+2}$$

$$= \frac{r(r+1)D!(N-D)!}{N!}\cdot\frac{N!}{(D+2)!(N-D-2)!}$$

$$E[X(X-1)] = \frac{r(r+1)(N-D)(N-D-1)}{(D+2)(D+1)}$$

عليه فإن

$$E[\quad ^2] = \frac{r(r+1)(N-D)(N-D-1)}{(D+2)(D+1)} + \frac{r(N-D)}{(D+1)}$$

ومن ثم فإن

$$Var(x) = \frac{r(r+1)(N-D)(N-D-1)}{(D+2)(D+1)} + \frac{r(N-D)}{(D+1)} - \frac{r^2(N-D)^2}{(D+1)^2}$$

$$= \frac{r(N+1)(D-r+1)(N-D)}{(D+1)^2(D+2)}$$

مثال:

صندوق يحتوي على 20 مصباح كهربائي منهم 15 مصباح صالحة. ما هو احتمـــال الحصول على 3 مصابيح صالحة عند سحب المصباح السادس؟

نلاحظ هنا أن:

$$N = 20\,, \quad D = 15\,, \quad r = 3\,, \quad r + x = 6\,, \quad x = 3$$

عليه فإن:

$$P(X = 3) = \frac{\binom{6-1}{3-1}\binom{20-6}{15-3}}{\binom{20}{15}} = \frac{(10)(91)}{15504} = 0.0587$$

وإذا ما أردنا إيجاد المتوسط والتباين، فإن:

$$E(X) = \frac{3(20-15)}{15+1} = 2.81$$

$$Var(X) = \frac{3(20+1)(15-3+1)(20-15)}{(15+1)^2(15+2)} = 0.94$$

تقدير معلمات التوزيع فوق الهندسي السالب:

تعرف المجتمعات بمعلماتها التي تمثل عادة مقاييس وصفية غير معروفة القيمة حيث تهدف عملية التقدير إلى عمل استدلال عن واحد أو أكثر من هذه المعلمات بالاعتماد على المعلومات المحتواة في تجربة عشوائية مأخوذة من ذلك المجتمع. وتعتبر طريقة الإمكان الأعظم (Maximum Likelihood method) إحدى طرق التقدير المعروفة والمهمة والتي تتمتع بخصائص التقدير المرغوب بها.

عند تقدير معلمات التوزيع فوق الهندسي السالب، يجب أخذ الاعتبار الحالتين التاليتين:

1) عدد العناصر الكلية N عدد صحيح معلوم بينما عدد العناصر التي تحمل الصفة D عدد صحيح غير معلوم ونرغب في تقدير قيمته.

2) عدد العناصر التي تحمل الصفة D عدد صحيح معلوم بينما عدد العناصر الكلية N عدد صحيح غير معلوم ونرغب في تقدير قيمته.

الحالة الأولى: تقدير عدد العناصر التي تحمل الصفة D:

لتكن التجربة العشوائية تتبع في تغيراتها التوزيع فوق الهندسي السالب بحيث عدد العناصر الكلية N معلوم وتم تحديد عناصر المجموعة التي تحمل الصفة المطلوب الحصول عليها لتتوقف التجربة r ، فإنه يكون بالإمكان الحصول على تقدير الإمكان الأعظم لعدد العناصر التي تحمل الصفة $\widehat{D} = \widehat{D}(x)$ وذلك باستخدام طريقة الإمكان الأعظم من خلال النسبة:

1) $R(D) = \frac{P_D(x)}{P_{D-1}(x)} > 1$

2) $R(D) = \frac{P_{D+1}(x)}{P_D(x)} < 1$

1) $R(D) = \frac{P_D(x)}{P_{D-1}(x)} > 1$

$$R(D) = \frac{\frac{\binom{D}{r-1}\binom{N-D}{x}}{\binom{N}{x+r-1}}}{\frac{\binom{D-1}{r-1}\left(\binom{N-D+1}{x}\right)}{\binom{N}{x+r-1}}} \cdot \frac{\frac{(D-r+1)}{(N-x-r+1)}}{\frac{(D-1-r+1)}{(N-x-r+1)}}$$

$$= \frac{\frac{D!}{(r-1)!(D-r+1)!} \cdot \frac{(N-D)!}{x!(N-D-x)!}}{\frac{(D-1)!}{(r-1)!(D-1-r+1)!} \cdot \frac{(N-D+1)!}{x!(N-D+1-x)!}} \cdot \frac{(D-r+1)}{(D-r)}$$

$$= \frac{D(D-r)!(N-D-x+1)!}{(D-r+1)!(N-D+1)(N-D-x)!} \cdot \frac{(D-r+1)}{(D-r)}$$

$$= \frac{D(N-D-x+1)}{(D-r+1)(N-D+1)} \cdot \frac{(D-r+1)}{(D-r)}$$

وحيث أن: $R(D) > 1$

عليه فإن:

$$\frac{D(N-D-x+1)}{(D-r)(N-D+1)} > 1$$

$$D(N - D - x + 1) > (D - r)(N - D + 1)$$

$$ND - D^2 - xD + D > ND - D^2 + D - rN + rD - r$$

$$-xD > -rN + rD - r$$

$$D(r + x) < r(N + 1)$$

$$D < \frac{r(N+1)}{(r+x)}$$

وبالتالي فإن قيم D التي تجعل $R(D) > 1$ و $(\quad)_D$ متزايدة مع D إذا وفقط إذا كان:

$$D < \frac{r(N + 1)}{(r + x)}$$

2) $R(D) = \frac{P_{D+1}(x)}{P_D(x)} < 1$

$$= \frac{\dfrac{\binom{D+1}{r-1}\binom{N-D-1}{x}}{\binom{N}{x+r-1}}}{\dfrac{\binom{D}{r-1}\binom{N-D}{x}}{\binom{N}{x+r-1}}} \cdot \frac{\dfrac{(D+1-r+1)}{(N-x-r+1)}}{\dfrac{(D-r+1)}{(N-x-r+1)}}$$

$$= \frac{\dfrac{(D+1)!}{(r-1)!(D+1-r+1)!} \cdot \dfrac{(N-D-1)}{x!(N-D-1-x)!}}{\dfrac{D!}{(r-1)!(D-r+1)!} \cdot \dfrac{(N-D)!}{x!(N-D-x)!}} \cdot \frac{(D-r+2)}{(D-r+1)}$$

$$= \frac{(D+1)(D-r+1)!(N-D-x)}{(N-D)(D-r-2)!} \cdot \frac{(D-r+2)}{(D-r+1)}$$

$$= \frac{(D+1)(N-D-x)}{(N-D)(D-r+1)} < 1$$

$$(D+1)(N-D-x) < (N-D)(D-r+1)$$

$$ND - D^2 - xD + N - D - x < ND - rN + N - D^2 + rD - D$$

$$-xD - x < -rN + rD$$

$$D(r+x) > r(N+1) - (r+x)$$

$$D > \frac{r(N+1)}{(r+x)} - 1$$

وبالتالي فإن قيم D التي تجعل $R(D) < 1$ و $_D(\quad)$ متناقصة مع D إذا وفقط إذا كان:

$$D > \frac{r(N+1)}{(r+x)} - 1$$

وهذا يعني أن $P_D(\quad)$ تصل إلى نهايتها العظمى عندما:

$$\widehat{D} = \left[\frac{r(N+1)}{(r+x)}\right]$$

حيث أن: $[a]$ تشير إلى أكبر عدد صحيح أقل من أو يساوي a.

الحالة الثانية: تقدير عدد العناصر الكلية N:

لتكن التجربة العشوائية تتبع في تغيراتها التوزيع فوق الهندسي السالب بحيث أن عـــدد عناصر المجموعة التي تحمل الصفة المطلوب الحصول عليها لتتوقف التجربة r، فإنه يكـــون بالإمكان الحصول على تقدير الإمكان الأعظم لعدد العناصر الكلية للمجتمع $\hat{N} = \hat{N}(x)$ وذلك باستخدام طريقة الإمكان الأعظم من خلال النسبة

$$1)\ R(N) = \frac{P_N(x)}{P_{N-1}(x)} > 1$$

$$2)\ R(N) = \frac{P_{N+1}(x)}{P_N(x)} < 1$$

$$1)\ R(N) = \frac{P_N(x)}{P_{N-1}(x)} > 1$$

$$= \frac{\dfrac{\binom{D}{r-1}\binom{N-D}{x}}{\binom{N}{x+r-1}}}{\dfrac{\binom{D}{r-1}\binom{N-1-D}{x}}{\binom{N-1}{x+r-1}}} \cdot \frac{\dfrac{(D-r+1)}{(N-x-r+1)}}{\dfrac{(D-r+1)}{(N-1-x-r+1)}}$$

$$= \frac{\dfrac{(N-D)!}{x!(N-D-x)!} \cdot \dfrac{(N-1)!}{(x+r-1)!(N-1-x-r+1)!}}{\dfrac{(N-D-1)!}{x!(N-D-x-1)!} \cdot \dfrac{N!}{(x+r-1)!(N-x-r+1)!}} \cdot \frac{(N-x-r)}{(N-x-r+1)}$$

$$= \frac{(N-D)(N-x-r+1)}{(N-D-x)(N)} \cdot \frac{(N-x-r)}{(N-x-r+1)} > 1$$

$$= \frac{(N-D)(N-x-r)}{N(N-D-x)} > 1$$

$$(N-D)(N-x-r) > N(N-D-x)$$

$$N^2 - Nx - Nr - DN + Dx + Dr > N^2 - DN - Nx$$

$$-Nr + Dx + Dr > 0$$

$$Nr < Dx + Dr$$

$$N < \frac{D(x+r)}{r}$$

عليه فإن قيم N التي تجعل النسبة $(N) > 1$ و $(\ \ )_N$ متزايدة مع N إذا وفقط إذا كان:

$$N < \frac{D(x+r)}{r}$$

$$2)\ R(N) = \frac{P_{N+1}(x)}{P_N(x)} < 1$$

$$= \frac{\frac{\binom{D}{r-1}\binom{N+1-D}{x}}{\binom{N+1}{x+r-1}}}{\frac{\binom{D}{r-1}\binom{N-D}{x}}{\binom{N}{x+r-1}}} \cdot \frac{\frac{(D-r+1)}{(N+1-x-r+1)}}{\frac{(D-r+1)}{(N-x-r+1)}}$$

$$= \frac{\frac{(N-D+1)!}{x!(N-D+1-x)!} \cdot \frac{N!}{(x+r-1)!(N-x-r+1)!}}{\frac{(N-D)!}{x!(N-D-x)!} \cdot \frac{(N+1)!}{(x+r-1)!(N+1-x-r+1)!}} \cdot \frac{(N-x-r+1)}{(N-x-r+2)}$$

$$= \frac{(N-D+1)(N-x-r+2)}{(N+1)(N-D-x+1)} \cdot \frac{(N-x-r+1)}{(N-x-r+2)}$$

$$= \frac{(N-D+1)(N-x-r+1)}{(N+1)(N-D-x+1)} < 1$$

$$(N-D+1)(N-x-r+1) < (N+1)(N-D-x+1)$$

$$N^2 - Nx - Nr + N - DN + Dx + Dr - D + N - x - r + 1$$
$$< N^2 - DN - Nx + N + N - D - x + 1$$

$$-Nr + Dx + Dr - r < 0$$

$$Nr > D(x+r) - r$$

$$N > \frac{D(x+r)}{r} - 1$$

عليه فإن قيم N التي تجعل النسبة $(N) < 1$ و $(\ \ )_N$ متناقصة مع N إذا وفقط إذا كان:

$$N > \frac{D(x+r)}{r}$$

وهذا يعني أن ($_N$) تصل إلى نهايتها العظمى عندما $\widehat{N} = \left[\frac{D(x+r)}{r}\right]$

خصائص المقدرات:

يمكن دراسة جودة المقدرات النقطية عن طريق المقارنة بين نتائج عدد من طرائق التقدير المختلفة. افترض أننا نريد تديد قيمة عددية لمعلمة المجتمع ولتكن θ . وسنرمز للمقدر بالرمز $\hat{\theta}$ فإننا نرغب أن يكون توزيع المعاينة للمقدر متمركز حول معلمة المجتمع θ. بتعبير آخر إننا نرغب في أن تكون القيمة المتوقعة لتوزيع معاينة المقدر مساوية لمعلمة المجتمع المرغوب في تقديرها. ونعبر عنها رياضيا بالصيغة $E(\hat{\theta}) = \theta$. أي أننا نقول إن $\hat{\theta}$ تقدير غير متحيز (Unbiased) للمعلمة θ . كذلك فإننا نرغب أن تكون الاختلافات بين القيم العددية للمقدرات في حالة المعاينة صغيرة. لذلك وبالإضافة إلى رغبتنا في الحصول على مقدرات غير متحيزة فإننا أيضا نرغب في أن يكون تباين هذه المقدرات أقل ما يمكن. ونعبر عنها رياضيا بالصيغة $V(\hat{\theta}) = E(\hat{\theta} - \theta)^2$. كذلك نرغب في أن يكون التوزيع الاحتمالي للمقدرات معروفا بحيث يكون التوصل إلى تقدير الفترة (Interval Estimates) لهذه المقدرات ممكنا (فاروق البشتي، 1990).

استنادا إلى هذه الصفات، هل يكون من الأفضل تحديد حجم المجتمع؟ أم تحديد عدد عناصر المجموعة التي تحمل الصفة المطلوب الحصول عليها؟ لتتوقف التجربة من خلال الحالات التالية:

1) عندما يكون عدد العناصر الكلية معلوما، والمطلوب تقدير عدد عناصر المجموعة التي تحمل الصفة.

2) عندما يكون عدد عناصر المجموعة التي تحمل الصفة معلوما، والمطلوب تقدير عـــدد العناصر الكلية.

اعتمادا على مقارنة خصائص مقدرات كل حالة.

خصائص مقدر عدد عناصر المجموعة التي تحمل الصفة:

أولا: التوقع الرياضي:

بما أن: $\widehat{D}_{ML} = \left[\frac{r(N+1)}{(x+r)}\right]$

فإن: $E\left(\widehat{D}_{ML}\right) = E\left[\frac{r(N+1)}{x+r}\right]$

$$= r(N+1)E\left[\frac{1}{x+r}\right]$$

وحيث أن: $E\left[\frac{1}{x+r}\right]$ غير موجود فإننا نعدل المقدر بحيث يكون $\widehat{D} = \left[\frac{r(N+1)}{(x+r-1)}\right]$

وحيث أن:

$$E\left[\frac{1}{x+r-1}\right] = \sum_x \frac{1}{x+r-1} p(x)$$

$$= \sum \frac{1}{x+r-1} \cdot \frac{\binom{D}{r-1}\binom{N-D}{x}}{\binom{N}{x+r-1}} \cdot \frac{(D-r+1)}{(N-x-r+1)}$$

$$= \sum \frac{1}{(x+r-1)} \cdot \frac{D!}{(r-1)!(D-r+1)!} \cdot \frac{(N-D)!}{x!(N-D-x)!} \cdot \frac{(x+r-1)!(N-x-r+1)!}{N!} \cdot \frac{(D-r+1)}{(N-x-r+1)}$$

$$= \sum \frac{D!(N-D)!}{(r-1)!N!} \cdot \frac{(x+r-2)!}{(r-2)!x!} \cdot \frac{(N-x-r)!}{(D-r)!(N-D-x)!}$$

$$= \frac{1}{(r-1)\binom{N}{D}} \sum \binom{x+r-2}{r-2}\binom{N-x-r}{D-r}$$

$$= \frac{1}{(r-1)\binom{N}{D}} \binom{N-1}{D-1} = \frac{D!(N-D)!}{(r-1)N!} \cdot \frac{(N-1)!}{(D-1)!(N-D)!}$$

$$E\left[\frac{1}{x+r-1}\right]=\frac{D}{(r-1)N}$$

عليه فإن:

$$E\left(\widehat{D}_{MLE}\right)=r(N+1)E\left[\frac{1}{x+r-1}\right]$$
$$=\left[r(N+1)\frac{D}{(r-1)D}\right]$$

من الواضح أن $\widehat{D}_{MLE}$ مقدر متحيز إلى D ، لذلك نلجأ إلى تعديل المقدر بحيث يكون:

$$\widehat{D}_U=\left[\frac{(r-1)N}{(x+r-1)}\right]$$

وهو مقدر غير متحيز إلى D حيث أن:

$$E\left(\widehat{D}_U\right)=\left[(r-1)N\frac{D}{(r-1)N}\right]=D$$

ثانيا التباين:

$$V\left(\widehat{D}_U\right)=V\left(\frac{(r-1)N}{(x+r-1)}\right)$$
$$=D(r-1)^2N^2V\left(\frac{1}{x+r-1}\right)$$

وهو مقدر لايمكن التعبير عنه بصيغة مختصرة، ولكن Zhang and Johnson(2011)

وجدوا تقدير لتباين $\widehat{D}$ باستخدام سلسلة تايلور (Taylor Series Method) وهي كالاتي:

$$V\left(\widehat{D}_U\right)\cong\frac{(r-1)^2N^2(D+1)^2\,r(N-D)(N+1)(D-r+1)}{(D+2)(rN-D+r-1)^4}$$

وإذا كانت D غير معلومة، يمكن التعويض عنها بـــ $\widehat{D}$ فيصبح التباين كالاتي:

$$V\left(\widehat{D}_U\right)\cong\frac{(r-1)^2N^2(\widehat{D}+1)^2\,r(N-\widehat{D})(N+1)(\widehat{D}-r+1)}{(\widehat{D}+2)(rN-\widehat{D}+r-1)^4}$$

خصائص مقدر عدد العناصر الكلية N:

أولاً- التوقع الرياضي:

بما أن $\hat{N}_{ML} = \left[D + \frac{Dx}{r}\right]$

فإن:

$$E\left[\hat{N}_{ML}\right] = E\left[D + \frac{Dx}{r}\right]$$

$$= D + \frac{D}{r}E[x]$$

$$= D + \frac{D}{r} \cdot \frac{r(N-D)}{(D+1)}$$

$$= \frac{Dr(D+1)+rD(N-D)}{r(D+1)}$$

$$= \frac{rD^2+rD+rDN-rD^2}{r(D+1)}$$

$$= \frac{DN+D}{(D+1)}$$

$$E\left[\hat{N}_{MLE}\right] = \frac{ND+D+N-N}{(D+1)}$$

$$= \frac{N(D+1)-(N-D)}{(D+1)}$$

$$= \left[N - \frac{(N-D)}{(D+1)}\right]$$

وهو تقدير متحيز إلى N ، لذلك نلجأ إلى تعديل المقدر بحيث يكون:

$$\hat{N}_U = \left[D + \frac{(D+1)x}{r}\right]$$

وهو مقدر غير متحيز إلى N حيث أن:

$$E\left[\hat{N}_U\right] = \left[D + \frac{(D+1)}{r} \cdot \frac{r(N-D)}{(D+1)}\right] = N$$

ثانياً- التباين:

$$V(\hat{N}_U) = V\left(D + \frac{(D+1)x}{r}\right) = \frac{(D+1)^2}{r^2} V(x)$$

$$= \frac{(D+1)^2}{r^2} \cdot \frac{r(N+1)(D-r+1)(N-D)}{(D+1)^2(D+2)}$$

$$V(\hat{N}_U) = \frac{(N+1)(D-r+1)(N-D)}{r(D+2)}$$

وإذا كانت Nغير معلومة، يمكن التعويض عنها بــ $\hat{N}$ فيصبح التباين كالاتي:

$$V(\hat{N}_U) = \frac{(\hat{N}+1)(D-r+1)(\hat{N}-D)}{r(D+2)}$$

استخدام أسلوب المحاكاة للمقارنة بين مقدر الإمكان الأعظم والمقدر المعدل:

في هذا البند سنستخدم أسلوب المحاكاة لتقييم المقدرات التي تم الحصول عليها نظريــا. الغرض الأساسي من المحاكاة هو المقارنة بين المقدرات التي تم الحصـــول عليهـــا باســـتخدام طريقة الإمكان الأعظم والمقدرات التي تم الحصول عليها بعد التعديل، وذلـــك فـــي الحـــالتين التاليتين:

الحالة الأولى: تقدير عدد العناصر التي تحمل الصفة D عندما يكون عدد العناصر الكليـــة N معلوم.

في الحالة الأولى، عملية المحاكاة تمت بناء على الخطوات التالية:

1) توليد عينة عشوائية من التوزيع فوق الهندسي السالب حيث تم استخدام حجم المجتمـــع $N = 100$ بمعلمة D تأخذ أخذ القـــيم $D = \{90,70,50,30\}$ والتـــي تمثـــل عـــدد العناصر التي تحمل الصفة وبالتالي $N - D = 100 - D$ تمثل عدد العناصر التـــي لاتحمل الصفة.

2) لكل قيمة من قيم D و r تأخذ القيم من 3 إلى 25 $(r = 3,4,\dots,25)$ ، تـــم توليـــد 10000 عينة عشوائية.

3) لكل عينة، تم حساب مقدر الإمكان الأعظم $\widehat{D}_{MLE}$ والمقدر المعدل $\widehat{D}_{\mathrm{U}}$ ومن ثم حساب قيمة متوسط هذه المقدرات. $\overline{\widehat{\mathrm{D}}} = \frac{\sum \widehat{\mathrm{D}}_{\mathrm{i}}}{10000}$

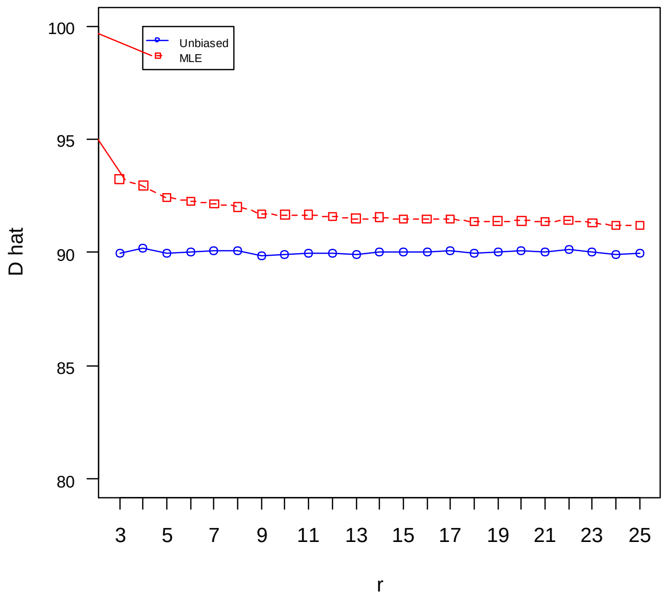
4) النتائج تم عرضها في الأشكال (1)،(2)،(3)،(4)

يتم تقييم جودة المقدر بناء على الفرق المتوقع بين المقدرات والقيمة الحقيقيـــة لـــــ D. المقدر الذي له أقل فرق هو المقدر المفضل.

بالنظر إلى الأشكال (1)، (2)، (3)، (4) نلاحظ أنه مهما كانت قيمة:

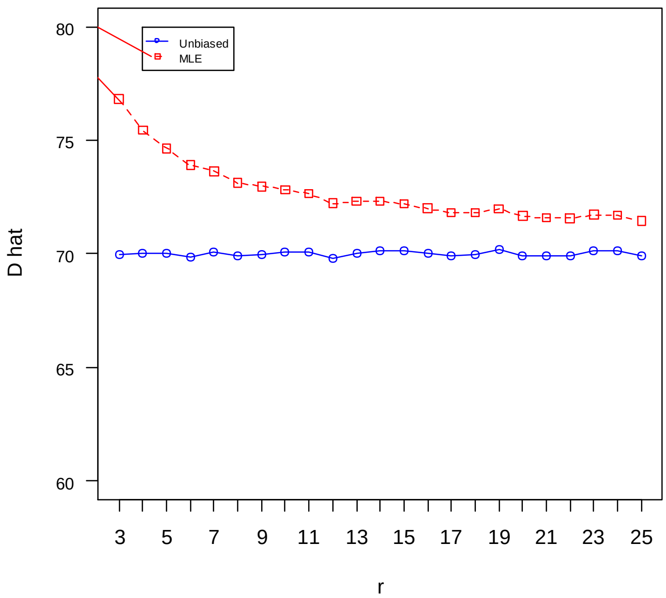
$$D = \{90,70,50,30\}$$

فإن المقدر غير المتحيز $\widehat{D}_{\mathrm{U}}$ هو المقدر الأفضل مقارنة بمقدر الإمكان الأعظـــم $\widehat{D}_{MLE}$ لأن معظم المقدرات قريبة جدا من القيم الحقيقية لــ D مع ملاحظة أن مقدر الإمكـــان $\widehat{D}_{MLE}$ يقترب من القيمة الحقيقية لــ D كلما زادت قيمة r. بشكل عام فإن مقدر الإمكان $\widehat{D}_{MLE}$ يعطي قيمة أكبر من القيمة الحقيقية لــ D (Over Estimate).



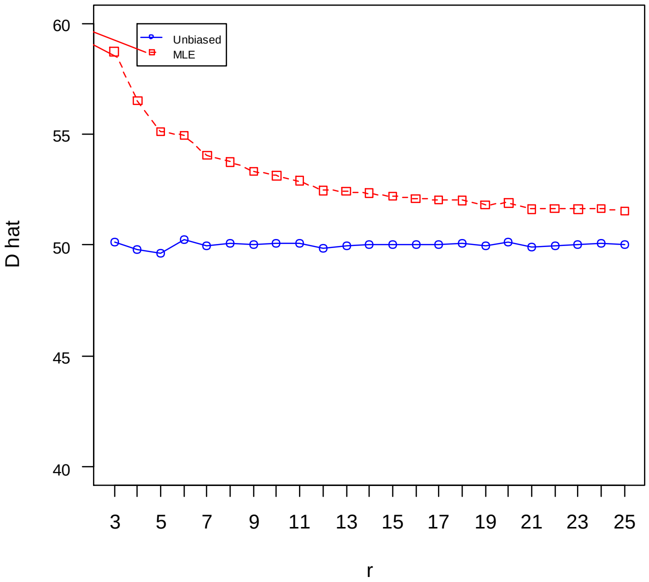
شكل (1) متوسط قيم مقدرات D

(عدد مرات التكرار = 10000 ، N=100 , D=90)



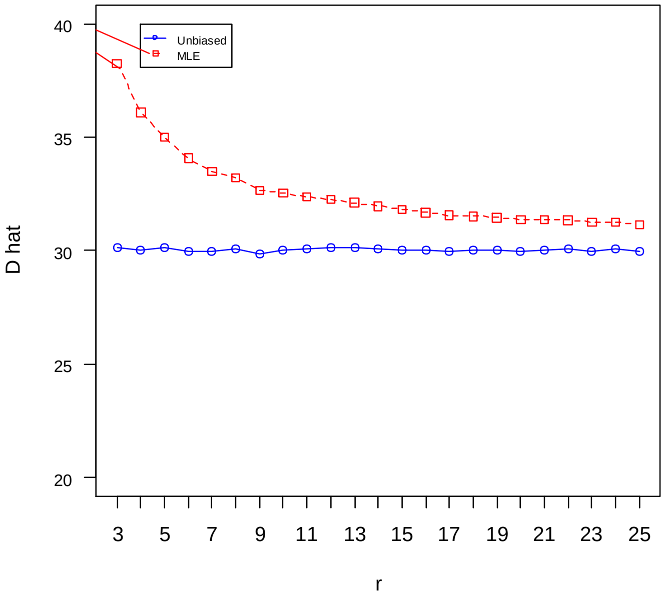
شكل (2) متوسط قيم مقدرات D

(عدد مرات التكرار = 10000 ، N=100 , D=70)



شكل (2) متوسط قيم مقدرات D

(عدد مرات التكرار = 10000 ، N=100 , D=50)



شكل (4) متوسط قيم مقدرات D

(عدد مرات التكرار = 10000 ، N=100 , D=30)

الحالة الثانية: تقدير عدد العناصر الكلية N عندما يكون عدد العناصر التي تحمل الصــفة D معلوم.

في الحالة الثانية، عملية المحاكاة تمت بناء على الخطوات التالية:

1) توليد عينة عشوائية من التوزيع فوق الهندسي السالب حيث تم استخدام عدد العناصــر التي تحمل الصفة $D = 50$ بمعلمة N تأخذ أخــذ القــيم $N = \{120,100,80,60\}$ والتي تمثل عدد العناصر الكلية.

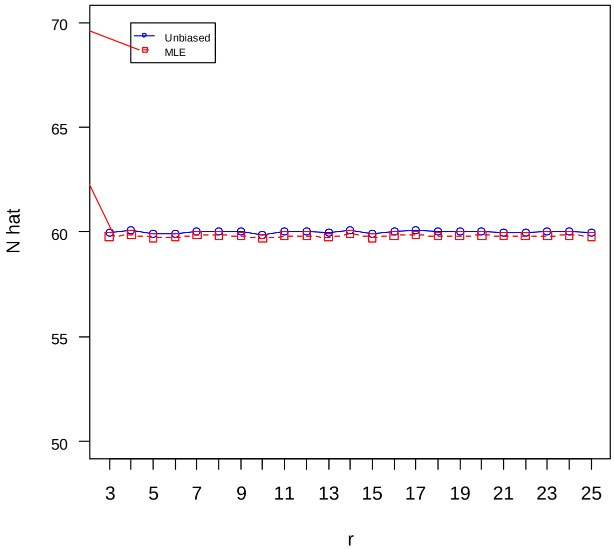
2) لكل قيمة من قيم N و r تأخذ القيم من 3 إلى 25 $(r = 3,4,\ldots,25)$ ، تــم توليــد 10000 عينة عشوائية.

3) لكل عينة، تم حساب مقدر الإمكان الأعظم $\widehat{N}_{MLE}$ والمقدر المعدل $\widehat{N}_{\mathrm{U}}$ ومن ثم حساب قيمة متوسط هذه المقدرات. $\overline{\widehat{\mathrm{N}}} = \frac{\sum \widehat{\mathrm{N}}_{\mathrm{i}}}{10000}$

4) النتائج تم عرضها في الأشكال (5)،(6)،(7)،(8)

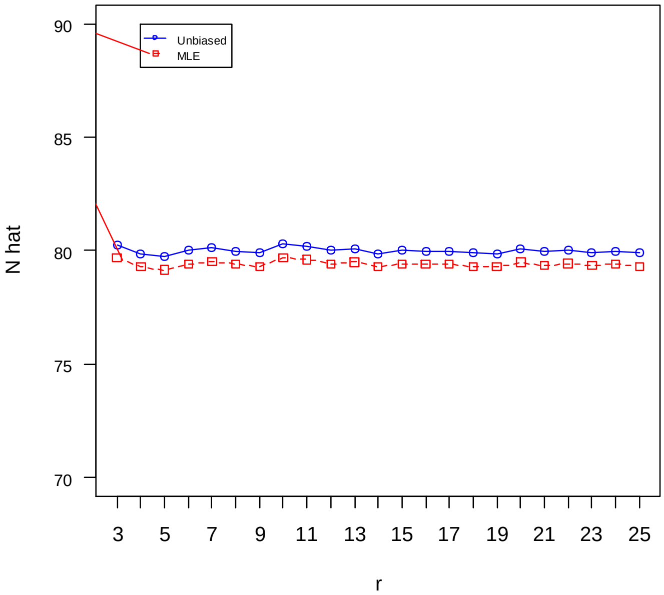
يتم تقييم جودة المقدر بناء على الفرق المتوقع بين المقدرات والقيمة الحقيقيــة لـــــ N. المقدر الذي له أقل فرق هو المقدر المفضل.

بالنظر إلى الشكل (5) نلاحظ أنه عندما كانت قيمة $N = 60$، فإن المقدر غير المتحيز $\widehat{N}_{\mathrm{U}}$ ومقدر الإمكان $\widehat{N}_{MLE}$ قريبين جدا من القيمة الحقيقية لـــ N . لاتوجد أفضلية لمقدر علـــى آخر. بينما عندما N تأخذ القيم $N = \{120,100,80\}$ فإن المقدر المعدل $\widehat{N}_{\mathrm{U}}$ هـــو المقـــدر الأفضل مقارنة بمقدر الإمكان الأعظم $\widehat{N}_{MLE}$ لأن معظم المقدرات قريبة جدا من القيم الحقيقية لـــ N ، أنظر الأشكال (6)،(7)،(8). مع ملاحظة أن مقدر الإمكان $\widehat{N}_{MLE}$ يعطي قيمة أقل من القيمة الحقيقية لـــ N (Under Estimate) ويكون هذا واضح خصوصا عنـــدما $N = 120$ أنظر الشكل (8).



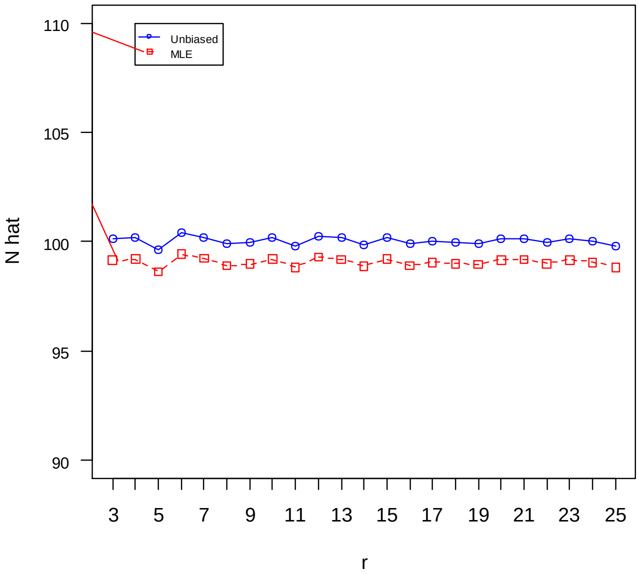
شكل (5) متوسط قيم مقدرات N

(عدد مرات التكرار = 10000 ، N=60 , D=50)



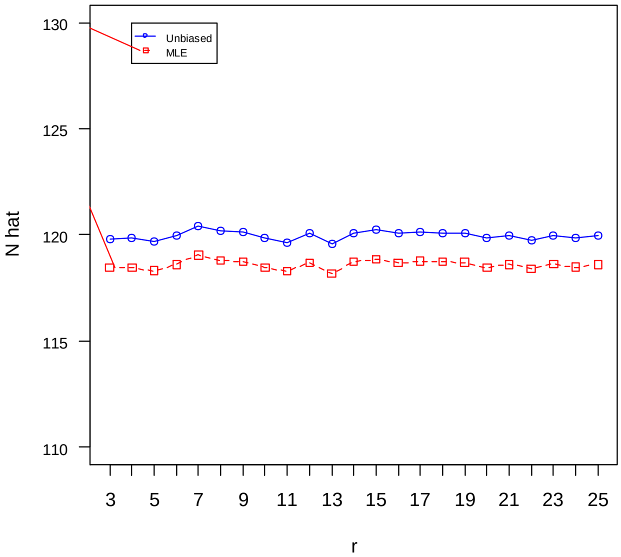
شكل (6) متوسط قيم مقدرات N

(عدد مرات التكرار = 10000 ، N=80 , D=50)



شكل (7) متوسط قيم مقدرات N

(عدد مرات التكرار = 10000 ، N=100 , D=50)



شكل (8) متوسط قيم مقدرات N

(عدد مرات التكرار = 10000 ، N=120 , D=50)

في الخلاصة يمكن القول أن:

1) المقدر غير المتحيز $\widehat{D}$ يكون أقرب بانتظام للقيمة الحقيقية لـــ D مقارنة إلى مقـــدر الإمكان الأعظم $\widehat{D}_{MLE}$. عليه نوصي باستخدام $\widehat{D}_U$ عند تقدير عدد العناصر التي تحمل الصفة.

2) المقدر غير المتحيز $\widehat{N}$ يكون أقرب بانتظام للقيمة الحقيقية لـــ N مقارنة إلى مقـــدر الإمكان الأعظم $\widehat{N}_{MLE}$. عليه نوصي باستخدام $\widehat{N}_U$ عند تقدير عدد العناصر الكلية.

<u>المراجع:</u>

1) *Johnson, Norman L., and Samuel Kotz. Urn Models and Their Application. New York: John Wiley & Sons, 1977.*
2) *Jones, Steven Norman, "A Gaming Application of the Negative Hypergeometric Distribution" (2013). UNLV Theses/Dissertations/ Professional Papers/Capstones. Paper 1846.*
3) *Lei Zhang, L. and Johnson, W. (2011). Approximate Confidence Intervals for a Parameter of the Negative Hypergeometric Distribution. JSM2014, ASA (1753-1767).*
4) *Miller, Gregory K., and Stephanie L. Fridel. \A Forgotten Discrete Distribution? Reviving the Negative Hypergeometric Model." The American Statistician 61 (November 2007): 347-350.*
5) *فاروق البشتي، نظرية التقدير، معهد الاتحاد العربي – بيروت (1990).*