Supplementary Material for 'A Finite Population Likelihood Ratio Test of the Sharp Null Hypothesis for Compliers'

Wen Wei Loh
Department of Statistics
University of Washington
wloh@u.washington.edu

Thomas S. Richardson
Department of Statistics
University of Washington
thomasr@u.washington.edu

A MAXIMIZING THE HYPERGEOMETRIC PROBABILITY IN A 2×2 TABLE

We review some existing results from [1] and [2] below. Recall the following 2×2 table where the row totals $(k, N-k)$ and the counts in one row $(b, (N-k)-b)$ are fixed.

Table 1: 2×2 Table With Unknown Column Totals

	Red	Green	Row
Not drawn	x	$k-x$	k
Drawn	b	$(N-k)-b$	$N-k$
Column	$b+x$	$N-(b+x)$	N

Our interest is in maximizing the hypergeometric probability corresponding to Table 1 with respect to x:

$$\Pr(x\,|\,(k,b,N)) = \binom{b+x}{x}\binom{N-(b+x)}{k-x}\Big/\binom{N}{k}.$$

Theorem 1. *In a 2×2 table where the row totals $(k, N-k)$ and the counts in one row $(b, (N-k)-b)$ are fixed, the most likely value of $x \in [0,k]$ under the randomization assumption is:*

$$\hat{x} = \arg\max_{x\in[0,k]} \left\{ x < (k+1)\frac{b}{N-k} \right\} = \left\lfloor\left\lfloor (k+1)\frac{b}{N-k} \right\rfloor\right\rfloor,$$

where the 'basement' function $\lfloor\lfloor a \rfloor\rfloor$ is defined as:

$$\lfloor\lfloor a \rfloor\rfloor = \max\{0, \lceil a \rceil - 1\}.$$

Equivalently,

$$b+\hat{x} = \begin{cases} \left\lfloor b\frac{N}{N-k} \right\rfloor & \text{if } b\frac{N+1}{N-k} \leq \left\lceil b\frac{N}{N-k} \right\rceil, \\ \left\lceil b\frac{N}{N-k} \right\rceil & \text{otherwise.} \end{cases}$$

Proof. We are interested in the following $k+1$ binomial coefficient products to find the most likely value of x:

$$\left\{ \binom{b+x}{x}\binom{N-(b+x)}{k-x}, \quad x\in[0,k] \right\}.$$

The first terms in each product are an increasing sequence in x, the second terms are a decreasing sequence in x for all values of $x \in \{0,\ldots,k-1\}$.

$$\begin{aligned} \binom{b+x}{x} &= \binom{b+x+1}{x+1} - \binom{b+x}{x+1} \\ &< \binom{b+x+1}{x+1}; \\ \binom{N-(b+x)}{k-x} &= \binom{N-(b+x+1)}{k-x} + \binom{N-(b+x+1)}{k-(x+1)} \\ &> \binom{N-(b+x+1)}{k-(x+1)}. \end{aligned}$$

The relationship between consecutive products due to a unit increase in x is thus:

$$\begin{aligned} & \binom{b+x+1}{x+1}\binom{N-(b+x+1)}{k-(x+1)} \\ &= \binom{b+x}{x}\binom{N-(b+x)}{k-x} \times \frac{b+x+1}{x+1} \times \frac{k-x}{N-(b+x)} \\ &> \binom{b+x}{x}\binom{N-(b+x)}{k-x}; \\ \iff\quad & \frac{b+x+1}{x+1} > \frac{N-(b+x)}{k-x} \\ \iff\quad & \frac{k-x}{x+1} > \frac{N-(b+x)}{b+x+1} \\ \iff\quad & \frac{k-x+(x+1)}{x+1} > \frac{N-(b+x)+(b+x+1)}{b+x+1} \\ \iff\quad & \frac{b+x+1}{x+1} > \frac{N+1}{k+1} \\ \iff\quad & x+1 < (k+1)\frac{b}{N-k}. \end{aligned} \tag{1}$$

Since (k,b,N) are fixed, the sequence of binomial coefficient products will increase with x until it exceeds the critical value $(k+1)\frac{b}{N-k}$ and the inequality no longer holds.

The most likely value of x is then:

$$\hat{x} = \arg\max_{x\in[0,k]} \left\{ x < (k{+}1)\frac{b}{N-k} \right\}$$
$$= \begin{cases} \left\lfloor (k+1)\frac{b}{N-k} \right\rfloor & \text{if } (k+1)\frac{b}{N-k} > \left\lfloor (k+1)\frac{b}{N-k} \right\rfloor, \\ (k+1)\frac{b}{N-k} - 1 & \text{if } (k+1)\frac{b}{N-k} = \left\lfloor (k+1)\frac{b}{N-k} \right\rfloor. \end{cases}$$

This is the same result shown in [2]. Johnson and Kotz [1, page 146] give an equivalent result in terms of the column total $(b + x)$.

If we define the 'basement' function as:

$$\lfloor\lfloor x \rfloor\rfloor = \max\{0, \lceil x \rceil - 1\},$$

then $\hat{x} = \left\lfloor \left\lfloor (k+1)\frac{b}{N-k} \right\rfloor \right\rfloor$.

However, we may rewrite the critical value $(k+1)\frac{b}{N-k}$ as follows:

$$(k+1)\frac{b}{N-k} = k\frac{b}{N-k} + \frac{b}{N-k}$$
$$= \left\lceil k\frac{b}{N-k} \right\rceil + \frac{b}{N-k} - \left(\left\lceil k\frac{b}{N-k} \right\rceil - k\frac{b}{N-k} \right).$$

Since $0 < \frac{b}{N-k} < 1$ and $0 \le \left\lceil k\frac{b}{N-k} \right\rceil - k\frac{b}{N-k} < 1$, we see that:

$$-1 < (k+1)\frac{b}{N-k} - \left\lceil k\frac{b}{N-k} \right\rceil < 1.$$

There are now two cases to consider:

1. If $1 > (k+1)\frac{b}{N-k} - \left\lceil k\frac{b}{N-k} \right\rceil > 0$, then

$$(k+1)\frac{b}{N-k} > \left\lceil k\frac{b}{N-k} \right\rceil,$$
$$\Rightarrow \left\lceil k\frac{b}{N-k} \right\rceil = \arg\max_{x\in[0,k]} \left\{ x < (k{+}1)\frac{b}{N-k} \right\}.$$

   When $k\frac{b}{N-k}$ is a positive integer, such as in a balanced table where $k = N - k$, then $k\frac{b}{N-k} = \left\lceil k\frac{b}{N-k} \right\rceil$. Since $\frac{b}{N-k} > 0$,

$$\Rightarrow (k+1)\frac{b}{N-k} > k\frac{b}{N-k} = \left\lceil k\frac{b}{N-k} \right\rceil.$$

2. If $-1 < (k+1)\frac{b}{N-k} - \left\lceil k\frac{b}{N-k} \right\rceil \le 0$, then

$$\left\lceil k\frac{b}{N-k} \right\rceil - 1 < (k+1)\frac{b}{N-k} \le \left\lceil k\frac{b}{N-k} \right\rceil$$
$$\Rightarrow \arg\max_{x\in[0,k]} \left\{ x < (k{+}1)\frac{b}{N-k} \right\}$$
$$= \left\lceil k\frac{b}{N-k} \right\rceil - 1$$
$$= \left\lfloor k\frac{b}{N-k} \right\rfloor.$$

This gives us the following result:

$$\hat{x} = \arg\max_{x\in[0,k]} \left\{ x < (k{+}1)\frac{b}{N-k} \right\}$$
$$= \begin{cases} \left\lfloor k\frac{b}{N-k} \right\rfloor & \text{if } (k+1)\frac{b}{N-k} \le \left\lceil k\frac{b}{N-k} \right\rceil, \\ \left\lceil k\frac{b}{N-k} \right\rceil & \text{otherwise.} \end{cases}$$

□

References

[1] N L Johnson and S Kotz. *Discrete distributions*. Houghton Mifflin, Boston, 1969.

[2] H Zhang. A note about maximum likelihood estimator in hypergeometric distribution. *Comunicaciones En Estadística*, 2(2):169–174, 2009.