



Indian Statistical Institute

A Note on Blackwell Sufficiency and a Skibinsky Characterization of Distributions

Author(s): D. Basu and Carlos A. B. Pereira

Reviewed work(s):

Source: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, Vol. 45, No. 1 (Feb., 1983), pp. 99-104

Published by: [Springer](#) on behalf of the [Indian Statistical Institute](#)

Stable URL: <http://www.jstor.org/stable/25050417>

Accessed: 27/09/2012 15:37

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Springer and Indian Statistical Institute are collaborating with JSTOR to digitize, preserve and extend access to *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*.

<http://www.jstor.org>

A NOTE ON BLACKWELL SUFFICIENCY AND A SKIBINSKY CHARACTERIZATION OF DISTRIBUTIONS

By D. BASU* and CARLOS A. B. PEREIRA**

The Florida State University, Tallahassee

SUMMARY. A Skibinsky (1970) characterization of the family of hypergeometric distributions is re-examined from the point of view of sufficient experiments and a number of other distributions similarly characterized.

1. INTRODUCTION

Consider an urn containing N balls x of which are white. If a simple random sample of n ($n \leq N$) balls is drawn from the urn, then the number of white balls in the sample has the hypergeometric distribution with parameters N , n , and x [denoted by $h(N, n, x)$]. Skibinsky (1970) introduced the following characterization of $h(N, n, x)$:

“A family of $N+1$ probability distributions (indexed say by $x = 0, 1, \dots, N$), each supported on a subset of $\{0, 1, \dots, n\}$ is the hypergeometric family having population and sample size parameters N and n respectively (the remaining parameter of the x -th member being x), if and only if for each number θ , $0 < \theta < 1$, the mixture of the family with binomial (N, θ) mixing distribution is the binomial (n, θ) distribution.”

Writing $b(N, \theta)$ for the binomial distribution over $\{0, 1, \dots, N\}$ and the symbol \sim for “distributed as”, we may restate Skibinsky’s characterization as follows :

Let $X \sim b(N, \theta)$, $0 < \theta < 1$, and let $\{\tau_x : x = 0, 1, \dots, N\}$ be a family of probability distributions on $\{0, 1, \dots, n\}$, where $n \leq N$. Consider the random variable Y such that the conditional probability distribution of Y given $\{X = x\}$ is τ_x for all x (i.e., $Y | X = x \sim \tau_x$). Then $Y \sim b(n, \theta)$ for all θ in $(0, 1)$ if and only if τ_x is $h(N, n, x)$ for all x .

Skibinsky (1970) proved the above result in several interesting ways, but somehow the perspective of Blackwell sufficiency eluded him. Written for its pedagogical interest, this note is an elucidation on the notion of Blackwell sufficiency and an unification of a number of results analogous to Skibinsky’s characterization of the Hypergeometric distribution.

*Research partially supported by NSF Grant No. 79-04693.

**Research supported by CAPES and USP—Brazil.

2. BLACKWELL SUFFICIENCY

A statistical experiment related to a parameter $\theta \in \Theta$ is idealized as an observable random variable (or vector), X , associated with a sample space \mathcal{X} and a family $\{p_\theta : \theta \in \Theta\}$ of probability functions (distributions) on \mathcal{X} indexed by θ . We avoid all measurability difficulties by restricting ourselves only to discrete sample spaces. Given two spaces \mathcal{X} and \mathcal{Y} , a transition function τ , from \mathcal{X} to \mathcal{Y} , is a family

$$\tau = \{\tau_x : x \in \mathcal{X}\}$$

of probability functions, τ_x , on indexed by $x \in \mathcal{X}$. Thus, the family of Hypergeometric probability functions $\{h(N, n, x) : x = 0, 1, \dots, N\}$ is a transition function from $\{0, 1, \dots, N\}$ to $\{0, 1, \dots, n\}$.

Let X and Y be two experiments with models $(\mathcal{X}, \{p_\theta : \theta \in \Theta\})$ and $(\mathcal{Y}, \{q_\theta : \theta \in \Theta\})$ respectively.

Definition (Blackwell): The experiment X is *sufficient for* (at least as informative as) the experiment Y and write $X > Y$ if there exists a transition function $\tau = \{\tau_x : x \in \mathcal{X}\}$ from \mathcal{X} to \mathcal{Y} such that

$$q_\theta(y) = \sum_x \tau_x(y) p_\theta(x) \quad \dots \quad (2.1)$$

for all $y \in \mathcal{Y}$ and $\theta \in \Theta$.

A transition function τ satisfying (2.1) is called here a Blackwell transition function. It is easy to check that the relation $>$ defines a partial order on the family of experiments related to θ .

If $T = T(X)$ is a sufficient statistic in the classical sense of Fisher (i.e., the conditional distribution of X given $\{T = t\}$ does not involve θ), then it follows at once that T is sufficient for X in the sense of Blackwell ($T > X$). Of course, X is sufficient for T in either sense.

The intuitive content of the relation $X > Y$ is as follows :

If we perform the experiment X , note its outcome x , and finally carry out a postrandomization exercise that chooses a point $y \in \mathcal{Y}$ in accordance with the probability function τ_x , then the experiment Y^* of such a choice of y is in a sense indistinguishable from the experiment Y in that both are endowed with the same model $(\mathcal{Y}, \{q_\theta : \theta \in \Theta\})$. Any decision rule related to θ that is based on the experiment Y can therefore be perfectly matched (in terms of their average performance characteristics) by a randomized rule based on X .

For two fixed integers N and n ($n \leq N$), consider now the simple case where $X \sim b(N, \theta)$ and $Y \sim n(n, \theta)$, $0 < \theta < 1$. To prove that $X > Y$ we consider an experiment $W = (W_1, \dots, W_N)$ where its components W_i , $i = 1, \dots, N$, are i.i.d Bernoulli variables with parameter θ . Note that since $X^* = W_1 + \dots + W_N$ is sufficient for W in the classical sense, $X^* > W$. On the other hand, for $Y^* = W_1 + \dots + W_n$, $W > Y^*$. Therefore, $X^* > Y^*$. Since X and X^* (Y and Y^*) are indistinguishable in their models, $X > Y$.

To conclude our version of Skibinsky's characterization we note that

$$Y^* | X^* = x \sim h(N, n, x).$$

Then a Blackwell transition function $\{\tau_x\}$ for our problem is the family of Hypergeometric probability functions. That is,

$$P_{\mathcal{R}}\{Y = y | \theta\} = \sum_x \tau_x(y) P_{\mathcal{R}}\{X = x | \theta\} \quad \dots \quad (2.2)$$

for every $y \in \mathcal{Y}$ and every $\theta \in (0, 1)$ where $\tau_x(\cdot)$ is the Hypergeometric probability function with parameter (N, n, x) . Finally, the uniqueness of $\{\tau_x\}$ as a Blackwell transition function follows from the fact that the family $\{b(N, \theta) : 0 < \theta < 1\}$ of probability distributions is complete. If $\{\tau'_x\}$ is another transition function satisfying (2.2), then

$$\sum_x [\tau_x(y) - \tau'_x(y)] P_{\mathcal{R}}\{X = x | \theta\} = 0$$

for every $y \in \mathcal{Y}$ and therefore $\tau_x(y) = \tau'_x(y)$ for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

3. FURTHER CHARACTERIZATIONS

Consider now an urn with N balls of k ($k \leq N$) different colors. Let $\mathbf{x} = (x_1, x_2, \dots, x_k)$ be the vector of frequency counts for the colors; that is, x_i ($i = 1, 2, \dots, k$) is the number of balls with the i -th color. If a simple random sample of n balls is drawn from the urn, then the sample vector of frequency counts has the multivariate Hypergeometric distribution with parameters, N, n , and \mathbf{x} . This distribution is denoted by $H(N, n, \mathbf{x})$ and its support by

$$Z_n^k = \{(x_1, \dots, x_k) : x_i \in Z, \sum x_i = N\}$$

where $Z = \{0, 1, \dots\}$.

Writing $M(N, \theta)$ for the Multinomial distribution over Z_N^k , we state the natural extension of Skibinsky's characterization. Here the parameter space is the simplex

$$\mathcal{S} = \{(p_1, \dots, p_k) : p_i \geq 0, \sum p_i = 1\}.$$

Proposition 1: Let $X \sim M(N, \theta)$, $\theta \in \mathcal{S}$, and let $\{\tau_x : x \in Z_N^k\}$ be a family of probability distributions on Z_n^k where $n \leq N$. Consider a random vector $Y|X = \bar{x} \sim \tau_x$ for all $x \in Z_N^k$. Then $Y \sim M(n, \theta)$ for all $\theta \in \mathcal{S}$ if and only if τ_x is $H(N, n, x)$ for all $x \in Z_N^k$.

The proof follows the same steps of the univariate case discussed in Section 2. Here, we consider the experiment $W = (W_1, \dots, W_N)$ where the components are i.i.d. with the common distribution being $M(1, \theta)$.

We write $X \sim \text{Poi}(\theta)$, $\theta > 0$, to indicate that X has Poisson distribution with parameter $\theta \in (0, \infty)$. Consider an additional experiment Y such that for a known number $r \in (0, 1)$, $Y \sim \text{Poi}(r\theta)$, $\theta > 0$. To prove that $X > Y$ we consider an experiment $W = (W_1, W_2)$ where its components W_1 and W_2 are independent with distributions $\text{Poi}(r\theta)$ and $\text{Poi}((1-r)\theta)$ respectively. Since $W > W_1$, $X^* = W_1 + W_2$ is sufficient for W in the classical sense, and X and X^* (Y and W_1) are indistinguishable in their models, it follows that $X > Y$.

Since $W_1|X^* = x \sim b(x, r)$ for all $x \in Z$, a Blackwell transition function $\{\tau_x\}$ is the family of Binomial probability functions. That is,

$$\frac{e^{-\theta r(\theta r)^y}}{y!} = \sum_x \tau_x(y) \frac{e^{-\theta \theta^x}}{x!}$$

for every $y \in Z$ and all $\theta \in (0, \infty)$ where $\tau_x(\cdot)$ now is the Binomial probability function with parameter (x, r) . The uniqueness of this family $\{\tau_x\}$ of Binomials as a Blackwell transition function follows from the completeness of the family $\{\text{Poi}(\theta) : \theta \in (0, \infty)\}$ of probability distributions on Z .

The above result in its extended form may be summarized as :

Proposition 2: Let $X \sim \text{Poi}(\theta)$, $\theta > 0$, and let $\{\tau_x : x \in Z\}$ be a family of probability distributions on the set $Z^k = \{(y_1, \dots, y_k) : y_i \in Z\}$. Consider a random vector $Y = (Y_1, \dots, Y_k)$ such that $Y|X = x \sim \tau_x$ for all $x \in Z$. For a known vector $r = (r_1, \dots, r_k) \in \mathcal{S}$, the components, Y_i ($i = 1, \dots, k$), of Y are mutually independent and $Y_i \sim \text{Poi}(r_i\theta)$, $\theta > 0$ if and only if τ_x is $M(x, r)$ for all $x \in Z$.

We end this section with a parallel characterization of the Dirichlet-Multinomial distribution. In Basu and Pereira (1980) we studied in details this distribution and indicated its use in statistics. We define the Dirichlet-Multinomial $DM(N; \alpha_1, \dots, \alpha_k)$ on Z_N^k as the mixture of the Multinomial family $\{M(N, \mathbf{p}); \mathbf{p} \in \mathcal{S}\}$ with \mathbf{p} distributed as Dirichlet (on \mathcal{S}) with parameter $(\alpha_1, \alpha_2, \dots, \alpha_k)$. Its probability function is given by

$$f(z_1, \dots, z_k) = \frac{N! \Gamma(\alpha)}{\Gamma(\alpha + N)} \prod_{i=1}^k \frac{\Gamma(\alpha_i + z_i)}{z_i! \Gamma(\alpha_i)}$$

for all $(z_1, \dots, z_k) \in Z_n^k$ where $\alpha = \sum_1^k \alpha_i$. When $k = 2$, in place of $(Z_1, Z_2, \sim DM(N, \alpha_1, \alpha_2))$, we write $Z_1 \sim Bb(N; \alpha_1, \alpha_2)$ to indicate that Z_1 is distributed as Beta-Binomial with parameter $(N; \alpha_1, \alpha_2)$.

Consider a sequence of Bernoulli trials with probability of success $\theta \in (0, 1)$. If $X + \alpha$ is the number of trials needed to obtain a fixed number α of success, then X is said to be a Negative Binomial experiment with parameter $(\alpha; \theta)$ and we write $X \sim nb(\alpha; \theta)$, $0 < \theta < 1$. Its probability function is

$$g(x|\theta) = \frac{\Gamma(\alpha+x)}{\Gamma(\alpha)x!} \theta^\alpha (1-\theta)^x \quad \dots \quad (3.1)$$

for every $x \in Z$ and all $\theta \in (0, 1)$. Note that

$$\sum_{x=0}^{\infty} \frac{\Gamma(\alpha+x)}{\Gamma(\alpha)x!} (1-\theta)^x = \theta^{-\alpha} \text{ for every } \alpha \in (0, \infty) \text{ and all } \theta \in (0, 1).$$

Then the following results hold not only for $\alpha \in Z$ but in general for any $\alpha \in (0, \infty)$. In this case, we still write $X \sim nb(\alpha; \theta)$, $0 < \theta < 1$, to indicate that the family of probability functions associated with the experiment X is (3.1). It is easy to check that this family is complete.

For $\alpha \geq \alpha_1 > 0$, let X and Y be two experiments such that $X \sim nb(\alpha; \theta)$ and $Y \sim nb(\alpha_1; \theta)$, $0 < \theta < 1$. To prove that $X > Y$, consider the experiment $W = (W_1, W_2)$ where now W_1 and W_2 are independent with distributions $nb(\alpha_1; \theta)$ and $nb(\alpha - \alpha_1; \theta)$ respectively. Following our previous chain of arguments, one can easily check that (i) $W_1 + W_2 \sim X$, (ii) $W_1 + W_2 > W > W_1$, (iii) $X > Y$, (iv) $W_1 | W_1 + W_2 = x \sim Bb(x; \alpha_1, \alpha - \alpha_1)$, and (v) the family $\{Bb(x; \alpha_1, \alpha - \alpha_1) : x \in Z\}$ of probability functions is the unique Blackwell transition function, and thus arrive at a Skibinsky type characterization of the Beta-Binomial.

The following is a summary of an extended version of the above results.

Proposition 3 : Let $X \sim nb(\alpha; \theta)$, $0 < \theta < 1$, and let $\{\tau_x : x \in Z\}$ be a family of probability distributions on the set Z^k . Consider a random vector $Y = (Y_1, \dots, Y_k)$ such that $Y|X = x \sim \tau_x$ for all $x \in Z$. For a fixed vector $(\alpha_1, \dots, \alpha_k)$ where $0 < \alpha_i < \infty$, $i = 1, 2, \dots, k$, and $\alpha = \sum_1^k \alpha_i$, the components Y_i ($i = 1, \dots, k$) of Y are mutually independent with $Y_i \sim nb(\alpha_i, \theta)$, $0 < \theta < 1$, if and only if τ_x is $DM(x; \alpha_1, \dots, \alpha_k)$ for all $x \in Z$.

REFERENCES

- BASU, D. and PEREIRA, C. A. B. (1982): On the Bayesian analysis of categorical data: The problem of nonresponse. *Journal of Statistical Planning and Inference*, **6**, 345-362.
- BLACKWELL, D. and GIRSHICK, M. A. (1954): *Theory of Games and Statistical Experiments*, John Wiley, N. Y.
- LEHMANN, E. L. (1959): *Testing Statistical Hypothesis*, John Wiley, N. Y.
- SKIBINSKY, M. (1970): A characterization of hypergeometric distributions. *JASA*, **65**, 926-929.

Paper received : March, 1981.