

Exchangeable Bernoulli random variables and Bayes' postulate

Moulinath Banerjee*

*Department of Statistics
University of Michigan
Ann Arbor, Michigan 48109
e-mail: moulib@umich.edu*

and

Thomas Richardson†

*Department of Statistics
University of Washington
Seattle, WA 98195
e-mail: thomasr@uw.edu*

Abstract: We discuss the implications of Bayes' postulate in the setting of exchangeable Bernoulli random variables. Bayes' postulate, here, stipulates a uniform distribution on the total number of successes in any number of trials. For an infinite sequence of exchangeable Bernoulli variables the conditions of Bayes' postulate are equivalent to a uniform (prior) distribution on the underlying mixing variable which necessarily exists by De Finetti's representation theorem. We show that in the presence of exchangeability, the conditions of Bayes' postulate are implied by a considerably weaker assumption which only specifies the probability of n successes in n trials, for every n . The equivalence of the Bayes' postulate and the weak assumption holds for both finite and infinite sequences. We also explore characterizations of the joint distribution of finitely many exchangeable Bernoulli variables in terms of probability statements similar to the weak assumption. Finally, we consider extensions of Bayes' postulate in the framework of exchangeable multinomial trials.

AMS 2000 subject classifications: Primary 62A01; secondary 62C10.

Keywords and phrases: Bernoulli random variables, Bayes' postulate, exchangeability, multinomial trials, weak assumption.

Received July 2013.

1. Introduction

Bayes' postulate, ascribed to the Reverend Thomas Bayes, was published posthumously by Richard Price, exactly 250 years ago, in the Philosophical Transactions of the Royal Society of London (1753), under the name 'An Essay towards

*Supported by U.S. NSF Grant DMS-1007751, U.S. NSA grant H98230-11-1-0166 and a Sokol Faculty Award, University of Michigan

†Supported by the U.S. National Science Foundation grant CNS-0855230 and U.S. National Institutes of Health grant R01 AI032475.

solving a problem in the doctrine of chances'.¹ Among other things, it deals with the problem of eliciting posterior probabilities for the chance of a success in a single Bernoulli trial in the wake of observed evidence based on a finite number of identical trials and prescribes *equi-distribution of ignorance* when assigning probabilities to events about which no prior stance can be taken. An excellent discussion of this material is available in the work of Stigler (1986), Chapter 3, but see also Molina (1931).

We formulate Bayes' postulate following the discussion on Bayes' scholium in (Stigler, 1986, Ch. 3). Consider a game between two rivals to be played n times where we '*absolutely know nothing*' of the probability that player 1 will win any single match. We define random variables X_1, X_2, \dots, X_n where X_i equals 1 if the i 'th match is won by player 1, and 0 otherwise. Let $S_n = X_1 + X_2 + \dots + X_n$ denote the total number of matches won by player 1. It is not assumed that the outcomes of these matches are independent. Ruling out temporal trends, the random variables may however be assumed to be identically distributed. A natural assumption is to take the joint distribution of the X_i 's to be exchangeable: i.e. if π is any permutation of the integers 1 through n , the joint distribution of $(X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)})$ is the same as the joint distribution of (X_1, X_2, \dots, X_n) .

As explained by Stigler, what Bayes postulates in this scenario of *absolutely knowing nothing* is that the distribution of S_n should be taken to follow a discrete uniform distribution: i.e. $P(S_n = m) = 1/(n + 1)$ for $m = 0, 1, \dots, n$. A standard situation in which the outcomes are exchangeable is what is today the classical Bayesian situation where the unknown chance Θ that player 1 will win any match is considered to be a random variable, and the vector of outcomes X_1, X_2, \dots, X_n considered to be i.i.d. Bernoulli(θ), conditional on a realized value θ of Θ – indeed, this is essentially the situation considered in Bayes' Billiard Table Problem (Stigler, 1986, pp. 124–125). Stigler (1982) notes that several commentators on Bayes' postulate have unfortunately *misinterpreted* it as specifying a uniform (prior) distribution on Θ as a way of summarizing *perfect ignorance* and argued that this notion of *perfect ignorance* about Θ is not invariant to reparametrizations, since monotone (non-linear) transformations of a uniform random variable are not necessarily uniform. According to Schrieber (1987), the source of this (flawed) argument appears to be Fisher's address (interestingly enough!) to the Royal Society in the 1920s. In Schrieber's opinion, it is Fisher's criticism of Bayes' idea that eventually led to 'the emergence of different schools and is the main root of subjective Bayesianism!' *But*, Bayes' postulate *simply assigns a uniform distribution to the outcome variable S_n and does not* correspond to a unique specification of the distribution of Θ ; see the appendix, also Edwards (1978). Observe that, Bayes' postulate, correctly interpreted, is indeed invariant to monotone transformations owing to the discreteness of the random variable S_n .

¹See, for example, http://www.socsci.uci.edu/~bskyrms/bio/readings/bayes_essay.pdf for a copy of Bayes' essay.

Variants and extensions of Bayes' postulate have been proposed by a number of authors. Schriber (1987) wrote about an extended version of Bayes' postulate and its potential effect on statistical methods: the extended postulate requires the statement of two prior uniform distributions and provides a unique parameter representation (leaving no freedom for nonlinear parameter transformations that arise in Fisher's argument) and unique posterior statements which are useful for both small and large sample sizes. Coolen (1998A,B) also revisited the postulate, formulated a revised version based on Hill's assumption $A_{(n)}$ (Hill, 1968), and used it for statistical inferences about future trials given past trials via the notion of *imprecise probabilities*. Hill's assumption, $A_{(n)}$, asserts that given past observations, O_1, \dots, O_n of a continuously valued quantity, a future observation, O_{n+1} , is equally likely to fall in any of the open intervals between successive order statistics of the past observations: i.e.,

$$P(O_{(i)} < O_{n+1} < O_{(i+1)} \mid O_1, \dots, O_n) = \frac{1}{n+1}.$$

Hill (1968) goes on to show that if ties have probability zero, for no n can we find a countably additive exchangeable distribution on the space of observations such that the conditional probabilities conform to $A_{(n)}$ for almost all O_1, \dots, O_n ; however, it is possible to exploit $A_{(n)}$ to make valid statistical inferences if one is willing to work with finitely additive distributions as pursued in a later paper Hill (1993). Very recently, an interesting extension of Bayes' postulate to *trinomial trials* was achieved by Diniz and Polpo (2012), who also discussed the philosophical implications of Bayes' postulate and its misinterpretation, but see also Good (1979). Their work appears in connection with the material in Section 5.

In this paper, we revisit the implications of Bayes' postulate for a sequence of exchangeable Bernoulli trials. For an infinite sequence of Bernoulli trials, the conditions of Bayes' postulate, which stipulates a discrete uniform distribution on S_n for all n , imply a uniform prior distribution on the probability of success, a fact independently (and differently) established by Murray (1930) and de Finetti (1930). We are able to show that under the exchangeability assumption, the conditions of Bayes' postulate are implied by a considerably weaker assumption on the distribution of S_n which only specifies its distribution at a *single point*. We show that this is true not only for an infinite sequence of exchangeable Bernoulli variables but also for a *finite exchangeable sequence* in which situation a prior need not exist. In the process, we also develop a characterization of the distribution of a vector of exchangeable Bernoulli random variables in terms of a vector of monotonically decreasing probabilities. This characterization is closely related to prior results of de Finetti (1964) and also Wood (1992). Finally, we provide a generalization of Bayes' postulate to multinomial trials.

2. Bayes' postulate and the uniform prior

Consider the classical Bayesian situation as considered in the last paragraph of the previous section. In this situation, we have:

$$\begin{aligned} P(S_n = m) &= E[P(S_n = m) \mid \Theta] \\ &= \int_0^1 \binom{n}{m} \theta^m (1 - \theta)^{n-m} dG(\theta), \text{ for } m = 0, 1, \dots, n, \end{aligned}$$

where G is the (prior) distribution on Θ . If G is the uniform $(0, 1)$ distribution, it is easy to check using properties of beta integrals that $P(S_n = m) = 1/(n+1)$ for all $0 \leq m \leq n$, so that the conditions of Bayes' postulate are satisfied. However, the converse is not true. Indeed, take any integer $N > 0$ and assume that

$$P(S_n = m) = \frac{1}{n+1} \quad \forall 0 \leq m \leq n, \quad \forall 1 \leq n \leq N. \quad (2.1)$$

Proposition. For any finite N , the above set of conditions is not sufficient to guarantee that G is the Uniform $(0, 1)$ distribution.

This proposition is established in the appendix where for every fixed N we show that there exists a non-uniform G_N that satisfies all of the above conditions. Note that Bayes' postulate therefore does not imply uniformity of the underlying Θ , a point raised in the previous section. However, if N is allowed to go to infinity, the conditions of Bayes' postulate do imply that Θ has a Uniform $(0, 1)$ distribution.

Theorem 2.1. Consider an infinite sequence of Bernoulli random variables X_1, X_2, \dots , that are conditionally i.i.d. Bernoulli (θ) , given $\Theta = \theta$. Suppose that (2.1) is satisfied for all positive integers N . Then Θ has a Uniform $(0, 1)$ distribution.

Proof. Since $P(S_n = n) = \int_0^1 \theta^n dG(\theta) = 1/(n+1)$ equals the n 'th moment of G for each $n \geq 1$ and since a compactly supported distribution is uniquely determined by its moments, it follows that G is the Uniform $(0, 1)$ distribution (as $EU^n = (n+1)^{-1}$, for each $n \geq 1$, if U is Uniform $(0, 1)$). \square

The Weak Assumption. For an exchangeable Bernoulli sequence X_1, X_2, \dots, X_n , Bayes' postulate would stipulate a uniform distribution on S_n . Suppose we now make the considerably weaker assumption that $P(S_n = n) = 1/(n+1)$. Thus, we only specify the distribution of S_n at a single point, as opposed to the entire distribution.

The following then holds.

Theorem 2.2. Consider an infinite sequence of exchangeable Bernoulli random variables X_1, X_2, \dots . Then there exists a random variable Θ such that the X_i 's are conditionally i.i.d. given $\Theta = \theta$ and provided that the weak assumption is satisfied for every n , Θ has a Uniform $(0, 1)$ distribution.

Remarks. We avoid a formal proof of the theorem but outline the main ideas involved. The first part of the theorem about the existence of Θ is De Finetti's representation theorem. See, for example, Hewitt and Savage (1955) and Aldous (1985). Letting \mathcal{S} denote the symmetric sigma-field, Θ can be taken to be $E(X_1 \mid \mathcal{S})$ which is also the almost sure limit of \bar{X}_n . We elaborate on this a bit further. Letting $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ be a sequence of 0's and 1's, the crux of the

argument lies in the fact that:

$$P(X_1 = \epsilon_1, X_2 = \epsilon_2, \dots, X_n = \epsilon_n \mid \mathcal{S}) = \prod_{i=1}^n P(X_i = \epsilon_i \mid \mathcal{S}).$$

Now, $P(X_i = \epsilon_i \mid \mathcal{S}) = E(X_i \mid \mathcal{S})^{\epsilon_i} (1 - E(X_i \mid \mathcal{S}))^{1-\epsilon_i}$ and by exchangeability $E(X_1 \mid \mathcal{S}) = E(X_i \mid \mathcal{S})$ for each i . Using these facts,

$$P(X_1 = \epsilon_1, X_2 = \epsilon_2, \dots, X_n = \epsilon_n \mid \mathcal{S}) = E(X_1 \mid \mathcal{S})^{\sum \epsilon_i} (1 - E(X_1 \mid \mathcal{S}))^{n - \sum \epsilon_i},$$

showing that Θ can be taken to be $E(X_1 \mid \mathcal{S})$.

That the weak assumption (for every n) implies a uniform prior on Θ follows easily, since only the corresponding conditions were used in the proof of Theorem 2.1. Thus, the weak assumption implies Bayes' postulate for an infinite exchangeable sequence. This implication also holds true for a finite exchangeable sequence as shown next.

3. Bayes' postulate, the weak assumption and finitely many exchangeable Bernoulli variables

Our goal in this section is to explore the connection between the Bayes' postulate and the weak assumption when we have a finite sequence of exchangeable Bernoulli random variables X_1, X_2, \dots, X_n .

With a finite sequence, it may no longer be assumed that there is a 'mixing' random variable Θ assuming values in $[0, 1]$, conditional on which, the X_i 's are conditionally independent Bernoulli(Θ) random variables. In other words, exchangeability can no longer be viewed in the classical Bayesian set-up. *Indeed, for every n , it is possible to construct an exchangeable sequence of n Bernoulli random variables, such that no random Θ assuming values in $[0, 1]$ can act as a mixing variable.* For example, consider simple random sampling without replacement from a population of n individuals where np are of Type A and the remaining are of Type B, $0 < p < 1$. Let the size of the sample be n as well (we exhaust the population) and let X_1, X_2, \dots, X_n be the random sample obtained, with $X_i = 1$ if the i 'th draw yields an individual of Type A and 0 otherwise. It is not difficult to show that the Bernoulli random variables X_1, X_2, \dots, X_n are exchangeable, i.e. any permutation of the X_i 's has the same distribution as the original sequence. However, a mixing variable Θ cannot exist in this situation. For if it did, given Θ , the X_i 's would be conditionally independent and identically distributed Bernoulli(Θ) and it would follow that $\text{Cov}(X_i, X_j) = \text{Var}(\Theta) \geq 0$ for all $i \neq j$. It is easy to check that in the given situation $\text{Cov}(X_i, X_j) < 0$ for $i \neq j$. See, also, the example in the introduction in Diaconis and Freedman (1980).

However, as the following Theorem shows, the weak assumption does imply Bayes postulate in this setting as well. The equivalence of the weak assumption and Bayes' postulate is *purely a consequence of exchangeability* and not of a De Finetti type representation. For finite forms of De Finetti's theorem, see Diaconis (1977); for approximations to the distribution of a finite exchangeable

Bernoulli sequence by a mixture of i.i.d. random variables under appropriate conditions, see Diaconis and Freedman (1980); and for finite exchangeability in the context of binomial mixtures, see Wood (1992).

Theorem 3.1. *Let X_1, X_2, \dots, X_N be a sequence of exchangeable Bernoulli random variables. Assume that $P(S_n = n) = 1/(n+1)$ for $n = 1, 2, \dots, N$, where $S_n = \sum_{i=1}^n X_i$ is the number of successes in n trials. Then S_n has a discrete uniform distribution for each $n \leq N$.*

Proof. Let $p_k^{(n)}$ denote the probability of k successes in n trials for generic n and $k \leq n$. By exchangeability,

$$P(X_1 = 1, X_2 = 1, \dots, X_k = 1, X_{k+1} = 0, X_{k+2} = 0, \dots, X_n = 0) = \frac{1}{\binom{n}{k}} p_k^{(n)}.$$

Now the left side of the above display is simply:

$$P(X_1 = 1, X_2 = 1, \dots, X_k = 1, X_{k+1} = 0, X_{k+2} = 0, \dots, X_n = 0, X_{n+1} = 0) \\ + P(X_1 = 1, X_2 = 1, \dots, X_k = 1, X_{k+1} = 0, X_{k+2} = 0, \dots, X_n = 0, X_{n+1} = 1),$$

which, by exchangeability, equals

$$P(X_i = 1 \forall 1 \leq i \leq k, X_j = 0 \forall k < j \leq n+1) \\ + P(X_i = 1 \forall 1 \leq i \leq k+1, X_j = 0 \forall k+1 < j \leq n+1)$$

which, by exchangeability again, is

$$\frac{1}{\binom{n+1}{k}} p_k^{(n+1)} + \frac{1}{\binom{n+1}{k+1}} p_{k+1}^{(n+1)}.$$

It follows that for each k with $0 \leq k \leq n$,

$$\frac{1}{\binom{n}{k}} p_k^{(n)} = \frac{1}{\binom{n+1}{k}} p_k^{(n+1)} + \frac{1}{\binom{n+1}{k+1}} p_{k+1}^{(n+1)}$$

which implies

$$p_k^{(n)} = \frac{n-k+1}{n+1} p_k^{(n+1)} + \frac{k+1}{n+1} p_{k+1}^{(n+1)}. \quad (3.1)$$

The proof now follows by induction. For $n = 1$, it is trivially seen that $S_n \equiv X_1$ assumes the values 0 and 1 with equal probability $1/2$. So we assume that the result holds for n and establish it for $n+1$. From (3.1), it is easily verified that if $p_{k+1}^{(n+1)} = 1/(n+2)$ and the induction hypothesis holds, so that $p_k^{(n)} = 1/(n+1)$, then $p_k^{(n+1)} = 1/(n+2)$ as well. But, by the weak Bayes' postulate, $p_{n+1}^{(n+1)} = 1/(n+2)$, showing that $p_k^{(n+1)} = 1/(n+2)$ for all $0 \leq k \leq n+1$. \square

Remark. The weak assumption could also have been formulated as: $P(S_n = 0) = 1/(n+1)$ for $1 \leq n \leq N$. This would imply that $P(\tilde{S}_n = n) = 1/(n+1)$ for all n , where $\tilde{S}_n = Y_1 + Y_2 + \dots + Y_n$ with $Y_i = 1 - X_i$. Since the X_i 's are exchangeable Bernoullis, so are the Y_i 's. Theorem 3.1 then leads to the conclusion that each \tilde{S}_n is uniformly distributed on $\{0, 1, \dots, n\}$. Hence, each $S_n = n - \tilde{S}_n$ also has a discrete uniform distribution.

4. A characterization of exchangeable distributions on $\{0, 1\}^N$

The considerations of the previous section raise a natural question: for which sequences $1 \geq p_1 \geq p_2 \geq \dots \geq p_N \geq 0$ do there exist bona-fide probability distributions for a sequence X_1, X_2, \dots, X_N of exchangeable Bernoulli random variables such that $P(S_n = n) = p_n$ for all $1, 2, \dots, N$? We have already seen that there is at least one sequence that works: namely, $p_n = 1/(n+1)$ for $1 \leq n \leq N$.

We show below that such sequences are well-characterized and that, in fact, there is a one-to-one correspondence between such sequences and the class of all exchangeable probability distributions on $\{0, 1\}^N$. So, consider the set $\{(q_0, q_1, \dots, q_N), q_i \geq 0, \sum q_i = 1\}$. Each vector in this set corresponds to a unique exchangeable distribution on $\{0, 1\}^N$ which is completely determined by the requirement that $P(S_N = n) = q_n$, for $0 \leq n \leq N$, the probability mass being assigned to a generic sequence $(\epsilon_1, \epsilon_2, \dots, \epsilon_N) \in \{0, 1\}^N$ by this distribution being $q_n / \binom{N}{n}$ where $n = \sum \epsilon_i$.

We now compute $P(S_n = n)$ for this distribution. This is simply the sum, over $0 \leq k \leq N - n$, of the probability that $X_1 = X_2 = \dots = X_n = 1$, k of the remaining X_i 's are 1 and the remaining $N - (n+k)$ are all 0. For a fixed k , there are $\binom{N-n}{k}$ such points in $\{0, 1\}^N$ and the probability of each such sequence is simply $q_{n+k} / \binom{N}{n+k}$. It follows that

$$P(S_n = n) = \sum_{k=0}^{N-n} \frac{\binom{N-n}{k}}{\binom{N}{n+k}} q_{n+k} = \sum_{k=0}^{N-n} C_{n,k} q_{n+k},$$

where

$$C_{n,k} = \frac{(k+1)(k+2)\dots(k+n)}{(N-n+1)(N-n+2)\dots N}.$$

Letting $\mathbf{q} = (q_1, q_2, \dots, q_N)$ and $\mathbf{p} = (p_1, p_2, \dots, p_N)$ (where p_n denotes $P(S_n = n)$), we see that the two vectors are connected by the equation

$$\mathbf{p} = \mathbf{A}_N \mathbf{q}$$

where \mathbf{A}_N is an $N \times N$ upper triangular matrix with the n 'th row given by $(\mathbf{0}_{1 \times (n-1)}, C_{n,0}, C_{n,1}, C_{n,2}, \dots, C_{n,N-n})$. Since A_N is (clearly) non-singular, it follows that the family of exchangeable probability distributions on $\{0, 1\}^N$ can equally well be characterized by (p_1, p_2, \dots, p_N) . The equation in the above display therefore provides a complete characterization of all sequences $1 \geq p_1 \geq \dots \geq p_N \geq 0$ that correspond to (and determine) valid exchangeable distributions on $\{0, 1\}^N$.

Remark 4.1. It should be noted that not all decreasing sequences $1 \geq \tilde{p}_1 \geq \dots \geq \tilde{p}_N \geq 0$ correspond to *bona fide* exchangeable distributions on $\{0, 1\}^N$. To see this, take $N = 2$, and use the general relations established above (or direct calculation) to see that $q_1 = 2p_1 - 2p_2$. Since $q_2 = p_2$ and $q_1 + q_2 \leq 1$, this forces $2p_1 - p_2 \leq 1$ in addition to $p_1 \geq p_2$.

Remark 4.2. Given a sequence, $1 \geq \tilde{p}_1 \geq \dots \geq \tilde{p}_N \geq 0$, it can be easily ascertained whether this corresponds to a *bona fide* exchangeable distribution on $\{0, 1\}^N$ by computing $\tilde{\mathbf{q}} = \mathbf{A}_N^{-1} \tilde{\mathbf{p}}$ (where $\tilde{\mathbf{p}} = (\tilde{p}_1, \dots, \tilde{p}_N)^T$) and checking that the entries of $\tilde{\mathbf{q}}$ are non-negative and sum up to at most 1.

The next result provides an explicit characterization of the distribution of S_n for each n in terms of the vector \mathbf{p} .

Proposition 4.1. Let $\{p_0^{(n)}, p_1^{(n)}, \dots, p_n^{(n)}\}$ be the pmf of S_n , for $n = 1, 2, \dots, N$, so that $\mathbf{q} = \{p_0^{(N)}, \dots, p_n^{(N)}\}$. In terms of the vector \mathbf{p} , $1 \equiv p_0 \geq p_1 \geq \dots \geq p_N \geq 0$, where, for each $m > 0$, $p_m = p_m^{(m)}$, we may obtain the pmf for S_n as follows:

$$p_k^{(n)} = \sum_{t=0}^{n-k} (-1)^t \binom{n}{(k, t)} p_{k+t}. \quad (4.1)$$

See also de Finetti (1964), p. 122, Eq. (5) for an alternative formulation in terms of finite difference operators.

Proof.

$$\begin{aligned} p_k^{(n)} &= P\left(\sum_{i=1}^n X_i = k\right) \\ &= \binom{n}{k} P(X_1 = \dots = X_k = 1, X_{k+1} = \dots = X_n = 0) \\ &= \binom{n}{k} \sum_{C \subseteq \{k+1, \dots, n\}} (-1)^{|C|} P(X_i = 1, i \in \{1, \dots, k\} \cup C) \\ &= \binom{n}{k} \sum_{C \subseteq \{k+1, \dots, n\}} (-1)^{|C|} p_{|C|+k} \\ &= \binom{n}{k} \sum_{t=0}^{n-k} (-1)^t \binom{n-k}{t} p_{k+t} \\ &= \sum_{t=0}^{n-k} (-1)^t \binom{n}{(k, t)} p_{k+t}. \end{aligned}$$

Here the second equation follows by exchangeability; the third via Möbius inversion (see, e.g. Proposition 1 in Drton and Richardson (2008)); the fourth via exchangeability; the fifth by counting subsets of a given size; the conclusion follows by definition of a trinomial coefficient: $\binom{n}{(k, t)} := n! / (k! t! (n - k - t)!)$. Note that in the third line, when $k = 0$, the term in the sum corresponding to $C = \emptyset$ is $(-1)^{|C|} P(X_i = 1, i \in C) = 1$, since the event is vacuously true. \square

Remark 4.3. Notice that taking $n = N$, in (4.1) gives:

$$p_k^{(N)} = \binom{N}{k} p_k + \sum_{s=k+1}^N (-1)^{s-k} \binom{N}{k, s-k} p_s. \quad (4.2)$$

Thus, $p_k^{(N)}$ is a linear function of p_k, \dots, p_N . Further note that for $k \geq 0$ we have

$$\begin{aligned} 1 \geq \sum_{r=k}^N p_r^{(N)} &= \sum_{s=k}^N p_s \sum_{r=k}^s (-1)^{s-r} \binom{N}{r \quad s-r} \\ &= \binom{N}{k} p_k + \sum_{s=k+1}^N p_s \sum_{r=k}^s (-1)^{s-r} \binom{N}{r \quad s-r}, \end{aligned}$$

with equality when $k = 0$. Hence given valid values for p_N, \dots, p_{k+1} we have the following upper bound:

$$p_k \leq \binom{N}{k}^{-1} \left(1 - \sum_{s=k+1}^N p_s \sum_{r=k}^s (-1)^{s-r} \binom{N}{r \quad s-r} \right). \tag{4.3}$$

Note that this bound is achievable, since given valid values of p_{k+1}, \dots, p_N , we can first determine $\{p_j^{(N)}\}_{j=k+1}^N$ uniquely, using Proposition 4.1, and then set $p_k^{(N)} = 1 - \sum_{r=k+1}^N p_r^{(N)}$, and $p_j^{(N)} = 0$ for $j < k$. Conversely, since $p_k^{(N)} \geq 0$ we have the following lower bound on p_k :

$$p_k \geq \frac{1}{\binom{N}{k}} \sum_{s=k+1}^N (-1)^{s-k+1} \binom{N}{k \quad s-k} p_s = \sum_{s=k+1}^N (-1)^{s-k+1} \binom{N-k}{s-k} p_s. \tag{4.4}$$

Again, for $k > 0$ the bound is achievable, as it corresponds to setting $p_k^{(N)} = 0$. (Note that $p_0^{(N)}$ is determined by $p_1^{(N)}, \dots, p_N^{(N)}$.) Lastly, simple algebra shows that the difference between the upper and lower bounds is:

$$\binom{N}{k}^{-1} \left(1 - \sum_{r=k+1}^N p_r^{(N)} \right) \geq 0.$$

These bounds provide a means of specifying a joint distribution over the set of permissible vectors \mathbf{p} : first, one specifies the marginal distribution of p_N , and conditional on a value of p_N , one specifies the conditional distribution of p_{N-1} so that it is concentrated on the interval given by the lower and upper bounds (which obviously depend on p_N) and so on.

Remark 4.4. Theorem 3.1 can be derived from Proposition 4.1, but is of independent interest, as it quantifies the extent to which the conditions of Bayes' postulate can be relaxed under finite exchangeability.

Remark 4.5. The correspondence that we have established between \mathbf{p} and \mathbf{q} also bears similarity to the mapping between distributions on the number of successes out of n and distributions over exchangeable sequences considered by Wood (1992).

5. Bayes' postulate and multinomial trials

Good (1979) observed that Bayes' billiard ball argument could be easily extended to a multinomial context. More recently, Diniz and Polpo (2012) considered Bayes' postulate in the context of trinomial trials. They show, using the strategy in de Finetti (1930), that the assumption of equiprobability *a priori* for the possible outcomes of trinomial trials, in the spirit of Bayes' postulate, implies that the parameter vector must have a Dirichlet(1, 1, 1) prior, which can be viewed as a uniform distribution on the 2 dimensional simplex. Their proof relies on using a recursive equation for multinomial probabilities under exchangeability to obtain a differential equation for the probability generating function, the solution to which is then deployed to obtain a limiting characteristic function. In this section, we provide a proof of a general version of their result in the multinomial setting, using a somewhat different argument that relies on moments.

To set up the problem, let $\tilde{X}_1, \tilde{X}_2, \dots$ be an infinite sequence of exchangeable multinomial vectors with $\tilde{X}_1 \sim \text{Mult}(1, p_1, p_2, \dots, p_k, p_{k+1})$ with $p_{k+1} = 1 - \sum_{i=1}^k p_i$. Each \tilde{X}_i is then a vector containing 1's and 0's, of length $k + 1$, with a solitary 1 at the l 'th position if outcome l was realized among the $k + 1$ exhaustive and mutually exclusive outcomes 1 through $k + 1$. Let $\mathbf{S}_n \equiv (S_{n,1}, S_{n,2}, \dots, S_{n,k}, n - \sum_{j=1}^k S_{n,j})$ be defined as $\sum_{j=1}^n \tilde{X}_j$. Then, \mathbf{S}_n is simply the status of the multinomial trials at Stage n . The natural extension of Bayes' postulate to this situation proceeds thus: under the assumption that *we absolutely know nothing* of the underlying mechanism, we should postulate that, for each n , all outcomes of \mathbf{S}_n are equally likely, i.e. for each n , \mathbf{S}_n has a discrete uniform distribution on the set $\{(x_1, x_2, \dots, x_k, n - \sum_{j=1}^k x_j) : x_i \in \mathbb{Z}^+, \sum_{j=1}^k x_j \leq n\}$. Since the cardinality of this set is $\binom{n+k}{k}$, $P(\mathbf{S}_n = (x_1, x_2, \dots, x_k, n - \sum_{j=1}^k x_j)) = 1/\binom{n+k}{k}$ for a generic vector in the above set.

Invoking the general De Finetti representation (Hewitt and Savage, 1955)², there exists a random vector $\Theta = (\Theta_1, \Theta_1, \dots, \Theta_k, 1 - \sum_{j=1}^k \Theta_j)$ such that $\{\tilde{X}_i\}$ are i.i.d. conditional on Θ ; further, $\tilde{X}_1 | \Theta \sim \text{Mult}(1, \Theta)$ and $\mathbf{S}_n/n \rightarrow_{a.s.} \Theta$. Letting G denote the distribution of $(\Theta_1, \Theta_2, \dots, \Theta_k)$ on the k dimensional simplex $\mathcal{S}_k \equiv \{(\theta_1, \theta_2, \dots, \theta_k) : \sum_{i=1}^k \theta_i \leq 1, \theta_i \geq 0\}$, we can write:

$$\begin{aligned} \frac{1}{\binom{n+k}{k}} &= P(\mathbf{S}_n = (x_1, x_2, \dots, x_k, n - \sum_{j=1}^k x_j)) \\ &= \int_{\mathcal{S}_k} P(\mathbf{S}_n = (x_1, x_2, \dots, x_k, n - \sum_{j=1}^k x_j) | \{\Theta_i\} = \{\theta_i\}) dG(\theta_1, \dots, \theta_k) \\ &= \int_{\mathcal{S}_k} \binom{n}{(x_1, x_2, \dots, x_k)} \prod_{i=1}^k \theta_i^{x_i} (1 - \sum_{i=1}^k \theta_i)^{n - \sum_{i=1}^k x_i} dG(\theta_1, \dots, \theta_k). \end{aligned}$$

²but see also Diniz and Polpo (2013)

In the above display $\binom{n}{(x_1, x_2, \dots, x_k)} := n! / (x_1! x_2! \dots x_k! (n - \sum_{i=1}^k x_i)!)$ is the usual multinomial coefficient. The above display implies that for all $\{(y_1, y_2, \dots, y_k) : y_i \geq 0, \sum y_i = n\}$,

$$\frac{1}{\binom{n+k}{k} \binom{n}{(y_1, y_2, \dots, y_k)}} = \int_{\mathcal{S}_k} \theta_1^{y_1} \theta_2^{y_2} \dots \theta_k^{y_k} dG(\theta_1, \theta_2, \dots, \theta_k) = E(\Theta_1^{y_1} \Theta_2^{y_2} \dots \Theta_k^{y_k}).$$

It follows that all possible moments of $(\Theta_1, \dots, \Theta_k)$ are determined by the postulate and provided that these correspond to the moments of a valid distribution G_0 on \mathcal{S}_k , the G in question must be G_0 , since distributions supported on bounded domains of Euclidean space are completely determined by their moments. Note that we defined the k dimensional simplex, above, as a subset of \mathbb{R}^k , rather than \mathbb{R}^{k+1} as is usually done. Of course, there is a 1-1 correspondence between the two sets.

We show that choosing G_0 to be the Dirichlet(1, 1, ..., 1) distribution on \mathcal{S}_k works. Recall that the Dirichlet($\beta_1, \beta_2, \dots, \beta_{k+1}$) distribution on \mathcal{S}_k is given by the following density function:

$$f(\theta_1, \theta_2, \dots, \theta_k; \{\beta_i\}) = \frac{\Gamma(\beta_1 + \beta_2 + \dots + \beta_{k+1})}{\prod_{i=1}^{k+1} \Gamma(\beta_i)} \theta_1^{\beta_1-1} \dots \theta_k^{\beta_k-1} \left(1 - \sum_{i=1}^k \theta_i\right)^{\beta_{k+1}-1}.$$

Now we take G_0 to be the Dirichlet with all β_i 's equal to 1 and compute a generic moment of $(\Theta_1, \dots, \Theta_k)$ under G_0 :

$$\begin{aligned} E_{G_0}(\Theta_1^{y_1} \dots \Theta_k^{y_k}) &= \int_{\mathcal{S}_k} \theta_1^{y_1} \dots \theta_k^{y_k} k! d\theta_1 \dots d\theta_k \\ &= k! \int_{\mathcal{S}_k} \prod_{i=1}^k \theta_i^{y_i+1-1} (1 - \sum_{i=1}^k \theta_i)^{1-1} \theta_1 \dots d\theta_k \\ &= k! \frac{\prod_{i=1}^k \Gamma(y_i + 1)}{\Gamma(n + k + 1)} \\ &= \frac{k! \prod_{i=1}^k y_i!}{(n + k)!} = \frac{1}{\binom{n+k}{k} \binom{n}{(y_1, y_2, \dots, y_k)}}. \end{aligned}$$

We conclude that the prior distribution implied by Bayes' postulate in the multinomial setting is precisely the uniform distribution over \mathcal{S}_k (note that \mathcal{S}_k has volume $1/k!$).

Multinomial random variables and insights into the ‘weak assumption’

Recall that for a finite exchangeable sequence of Bernoullis, say X_1, X_2, \dots, X_N , the weak assumption specified the distribution of S_n at the point n for each $1 \leq n \leq N$. Our assumption was that $P(S_n = n) = 1/(n + 1)$ and we showed that this forced the distribution of each S_n to be uniform. The argument relied on the recursion formula (3.1) which can be viewed as a system of $(n + 1)$ linear equations in $\{p_{k+1}^{(n+1)}\}_{k=0}^{n+1}$. However, as $\sum_{k=0}^n p_k^{(n)} = 1$, there are only n

linearly independent equations involving $(n + 1)$ variation dependent unknowns (since $\sum_{k=0}^{n+1} p_k^{(n+1)} = 1$). It follows that the distribution of S_{n+1} cannot be uniquely recovered from that of S_n (via these equations) unless one of the $p_k^{(n+1)}$ is pre-specified. And this, by specifying $\{p_n^{(n)}\}_{n=1}^N$, is precisely what the weak assumption does. Specifying $p_1^{(1)}$ pins down the distribution of $S_1 \equiv X_1$ as Bernoulli(1/2), and now, using the pre-specified values of $p_2^{(2)}, p_3^{(3)}, \dots$, the distribution of each S_n can be recovered uniquely. Note that alternative specifications would also have worked, as the remark prior to Section 4 shows. In fact, all that is needed for unique recovery is a specification of the vector $\{p_{k_n}^{(n)}\}_{n=1}^N$ where (k_1, k_2, \dots, k_N) is an arbitrary set of integers with $0 \leq k_n \leq n$; and, provided $p_{k_n}^{(n)}$ is set to $1/(n + 1)$ for each n , the distributions of the S_n 's will be seen to be discrete uniforms as before. The specifications we discussed in Section 3, where we set all k_n 's to be n (or 0), are aesthetically somewhat more pleasing as they correspond to probability statements about all n trials resulting in successes (failures).

It is natural to ask whether there are analogous weak assumptions in the general multinomial setting. We discuss this briefly in the trinomial setting as considered in Diniz and Polpo (2012). The analogue of (3.1) in the trinomial context is equation (14) of their paper:

$$\omega_{x_1, x_2}^{(n-1)} = \omega_{x_1+1, x_2}^{(n)} \left(\frac{x_1 + 1}{n} \right) + \omega_{x_1, x_2+1}^{(n)} \left(\frac{x_2 + 1}{n} \right) + \omega_{x_1, x_2}^{(n)} \left(\frac{n - x_1 - x_2}{n} \right), \quad (5.1)$$

where $\omega_{x_1, x_2}^{(n)} = P(\mathbf{S}_n = (x_1, x_2, n - x_1 - x_2))$, using our notation from the general multinomial setting with $k = 2$. The distribution of \mathbf{S}_{n-1} is determined by the vector $\{\omega_{x_1, x_2}^{(n-1)}\}$, of dimension $\binom{n+1}{2}$, while that of \mathbf{S}_n is determined by $\{\omega_{x_1, x_2}^{(n)}\}$, of dimension $\binom{n+2}{2}$. If the distribution of \mathbf{S}_{n-1} is known, and the $\omega_{x_1, x_2}^{(n)}$ are treated as unknowns, this gives us a system of $\binom{n+1}{2}$ equations in $\binom{n+2}{2}$ variables, and for unique recovery of the distribution of \mathbf{S}_n , $\binom{n+2}{2} - \binom{n+1}{2} = n + 1$ of the $\omega_{x_1, x_2}^{(n)}$ would, at least, need to be specified.³ Thus, a 'weak assumption' in this setting would need to specify the distribution of \mathbf{S}_1 at 2 points, the distribution of \mathbf{S}_2 at 3, and so on till \mathbf{S}_N . A 'weak assumption' that, for each $n \leq N$, sets the value of each of some $n + 1$ pre-specified $\omega_{x_1, x_2}^{(n)}$'s to be $1/\binom{n+2}{2} = 2/(n+1)(n+2)$, would force the distribution of each $\mathbf{S}_n, 1 \leq n \leq N$ to be the discrete uniform, as can be readily verified via an inspection of the equation (5.1): it is satisfied if we set $\omega_{x_1, x_2}^{(n-1)} = 1/\binom{n+1}{2} = 2/n(n+1)$ and $\omega_{x_1+1, x_2}^{(n)} = \omega_{x_1, x_2+1}^{(n)} = \omega_{x_1, x_2}^{(n)} = 2/(n+1)(n+2)$. We note, as before, that with finitely many exchangeable trinomials, it is not possible to talk about a prior distribution on the trinomial parameter and any analysis of the distributions of the \mathbf{S}_n 's has to start from (5.1), which follows *purely from the exchangeability hypothesis*.

³Note that as the variables have to add up to 1, there are only $\binom{n+2}{2} - 1$ 'free' variables; on the other hand, there are only $\binom{n+1}{2} - 1$ linearly independent equations, since the ω_{x_1, x_2}^{n-1} have to add up to 1. So, the difference between the number of 'free' variable and the number of independent equations is $n + 1$.

The linear recurrence relations (3.1) and (5.1) can be generalized, without much difficulty, to the generic multinomial setting, based on exchangeability considerations as in Section 3 of this paper or Section 4 of Diniz and Polpo (2012). For the general $k + 1$ compartment multinomial of this section, letting $p(x_1, x_2, \dots, x_k, n)$ denote the probability of the event that outcome 1 is obtained in the first x_1 , outcome 2 in the next x_2, \dots , outcome k in the next x_k and outcome $k + 1$ in the final $n - \sum_{i=1}^k x_i$, the following relation is an easy outcome of exchangeability:

$$\begin{aligned}
 p(x_1, x_2, \dots, x_k, n - 1) &= p(x_1, x_2, \dots, x_k, n) + p(x_1 + 1, x_2, \dots, x_k, n) + \dots \\
 &\quad + p(x_1, x_2, \dots, x_{i-1}, x_i + 1, x_{i+1}, \dots, x_k, n) + \dots \\
 &\quad + p(x_1, \dots, x_{k-1}, x_k + 1, n). \tag{5.2}
 \end{aligned}$$

By exchangeability,

$$p(y_1, y_2, \dots, y_k, n) = \frac{\omega(y_1, y_2, \dots, y_k, n)}{\binom{n}{(y_1, y_2, \dots, y_k)}},$$

where $\omega(y_1, y_2, \dots, y_k, n) = P(\mathbf{S}_n = (y_1, y_2, \dots, y_k, n - \sum_{j=1}^k y_j))$. We can use this equation to rewrite (5.2) in terms of the pmf's of \mathbf{S}_n and \mathbf{S}_{n-1} as follows:

$$\begin{aligned}
 \omega(x_1, x_2, \dots, x_k, n - 1) &= \frac{n - \sum_{i=1}^k x_i}{n} \omega(x_1, x_2, \dots, x_k, n) \tag{5.3} \\
 &\quad + \sum_{j=1}^k \frac{x_j + 1}{n} \omega(\dots, x_{j-1}, x_j + 1, x_{j+1}, \dots, n).
 \end{aligned}$$

Following the discussion for the trinomial situation, it is not difficult to see, that for a $k + 1$ compartment multinomial, a unique recovery of the distribution of $\{\mathbf{S}_n\}_{n=1}^N$ using (5.3) would require specifying the distribution of each \mathbf{S}_n at $\binom{n+k}{k} - \binom{n-1+k}{k} = \binom{n-1+k}{k-1}$ points.

Remark. The proof in Section 4 of Diniz and Polpo (2012), of the fact that Bayes' postulate implies the uniform (Dirichlet) prior in the trinomial situation, was based upon their equation (14) which is what (5.3) reduces to, for $k = 2$. It would seem that their method, involving difference-differential equations, should be extendible to the multinomial case, where the relevant differential equations would be derived from (5.3).

6. Concluding remarks

As we have seen, the equal assignment of probabilities to all possible outcomes in an exchangeable multinomial experiment (with infinitely many trials) – i.e. postulating that each \mathbf{S}_n is uniformly distributed – forces a uniform distribution on the prior parameter. However, it is important to note the fundamental distinction between imposing uniformity on the outcome variable of an experiment and that of imposing uniformity on a prior, a point that has often been

glossed over. We have noted that with *finitely* many exchangeable Bernoullis (the simplest possible multinomial variables), many different prior distributions can produce uniformity of the outcome distribution. It would be interesting to investigate the connections between outcome distributions in a general infinitely exchangeable experiment and the prior distribution that would uniquely exist in this case. More specifically, can we characterize all infinitely exchangeable experiments for which the assumption of equiprobability of outcomes forces a uniform distribution on the prior? As Diniz and Polpo (2012) note in their conclusion, for certain experiments, uniformity of the outcome distribution might actually translate to a *non-uniform* prior, in which case, the uniform prior should not be regarded as a non-informative prior. Such considerations would go a long way towards extending our understanding of Bayes' postulate and its implications in general.

7. Appendix

We establish a slightly stronger version of the Proposition stated in Section 2.

Proposition. Let X_1, X_2, \dots, X_N be (exchangeable Bernoulli) random variables that are conditionally i.i.d. Bernoulli (Θ), given some random variable Θ taking values in $[0, 1]$. Let $S_n = X_1 + X_2 + \dots + X_n$. Then, there exist infinitely many (prior) distributions for Θ , for which S_n is uniformly distributed for each $1 \leq n \leq N$.

Proof. For a positive integer P , let g denote a density on $[0, 1]$ that assumes values L_1, L_2, \dots, L_P on $[0, 1/P], (1/P, 2/P], \dots$ respectively and such that S_n is uniformly distributed for each $1 \leq n \leq N$ under g . The constraints that

$$P(S_n = m) = \int_0^1 \binom{n}{m} \theta^m (1-\theta)^{n-m} g(\theta) d\theta = \frac{1}{n+1}, \quad \text{for } 0 \leq m \leq n, \quad 1 \leq n \leq N$$

and the fact that g integrates to 1 give rise to a system of linear equations in L_1, L_2, \dots, L_P , namely

$$\sum_{i=1}^P I_{m,n,i} L_i = 1/(n+1), \quad \text{for } 0 \leq m \leq n, \quad 1 \leq n \leq N$$

and $\sum_{i=1}^P L_i/P = 1$, where $I_{m,n,i} = \int_{(i-1)/P}^{i/P} \binom{n}{m} \theta^m (1-\theta)^{n-m}$. This can be written in matrix notation as $M\mathbf{1} = \mathbf{v}$ for a $(c(N)+1) \times P$ matrix M whose last row is the vector $(1/P, 1/P, \dots, 1/P)$ and the rows above are given by $(I_{m,n,1}, I_{m,n,2}, \dots, I_{m,n,P})$, stacked in order of increasing n , and increasing m within n ; \mathbf{v} is a $c(N)+1$ dimensional vector of the form $(1/2, 1/2, 1/3, 1/3, 1/3, 1/4, \dots, 1/(N+1), 1/(N+1), 1)^T$; and, $\mathbf{1} = (L_1, L_2, \dots, L_P)^T$. Here $c(N) = \sum_{n=1}^N (n+1)$.

Now, P can be chosen to be larger than $c(N)+1$ in which case the columns of M are linearly dependent and there must exist a non-zero vector \mathbf{x} such that $M\mathbf{x} = \mathbf{0}_{(c(N)+1) \times 1}$. Since \mathbf{x} is orthogonal to the last row of M , whose entries are all equal, the components of \mathbf{x} cannot all be equal. We know that there

is at least one piecewise constant density on the grid $[0, 1/P, 2/P, \dots]$, namely, the uniform density, whose corresponding $\mathbf{1} = (1, 1, \dots, 1)^T$ satisfies the matrix equation of the previous paragraph. But then the vector $(1, 1, \dots, 1)^T + \lambda \mathbf{x}$, for all λ sufficiently small, has strictly positive co-ordinates, not all equal, and also satisfies the matrix equation and corresponds to a (piecewise constant) *non-uniform density* on the grid $[0, 1/P, 2/P, \dots]$ that makes the distribution of each S_n uniform. \square

Alternative proof of Theorem 3.1 using Proposition 4.1. Since it is assumed in Theorem 3.1, that $p_n \equiv p_n^{(n)} = 1/(n + 1)$ for each n , it suffices to prove that $p_k^{(n)} = p_{k+1}^{(n)}$ for all $k \leq (n - 1)$. From Proposition 4.1, it is not difficult to see that

$$p_{k+1}^{(n)} = - \sum_{t=1}^{n-k} \frac{(-1)^t n!}{(k+1)!(t-1)!(n-k-t)!} p_{k+t}.$$

Now,

$$\begin{aligned} p_k^{(n)} - p_{k+1}^{(n)} &= \sum_{t \in \{0\}} \frac{(-1)^t n!}{k! t! (n-k-t)!} p_{k+t} \\ &\quad + \sum_{t=1}^{n-k} \left(\frac{(-1)^t n!}{k! t! (n-k-t)!} + \frac{(-1)^t n!}{(k+1)!(t-1)!(n-k-t)!} \right) p_{k+t} \\ &= \frac{n!}{k!(n-k)!} p_k + \sum_{t=1}^{n-k} (-1)^t \frac{n!}{(k+1)! t! (n-k-t)!} (k+t+1) p_{k+t} \\ &= \frac{n!}{(n-k)!(k+1)!} \left[1 + \sum_{t=1}^{n-k} (-1)^t \binom{n-k}{t} \right] \\ &= 0, \end{aligned}$$

where, in the penultimate step, we used $p_{k+t} = 1/(k+t+1)$ for $t \geq 0$, which is assumed in Theorem 3.1. \square

References

- ALDOUS, D. (1985). Exchangeability and related topics. In Hennequin, P., editor, *École d'Été de Probabilités de Saint-Flour XIII 1983*, pages 1–198. Springer-Verlag, Heidelberg. Lecture Notes in Mathematics 1117. [MR0883646](#)
- BAYES, T. (1763). An Essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370–418.
- COOLEN, F.P.A. (1998A). Bayes' postulate repostulated. Contributed to *6th Valencia International Meeting on Bayesian Statistics*, Alcossebre, Spain (May/June 1998). Available at <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.1478>.
- COOLEN, F.P.A. (1998B). Low structure imprecise predictive inference for Bayes' problem. *Statistics and Probability Letters*, **36**, 349–357. [MR1621919](#)

- DE FINETTI, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Atti Reale Accademia Nazionale de Lincei, Mem.*, **4**, 86–133.
- DE FINETTI, B. (1964). Foresight: Its Logical Laws, its subjective sources, (English translation of *La prévision: ses lois logiques, ses sources subjectives*. *Annales de l'Institut Henri Poincaré*. 1937; VII: 1–68. trans. H.E. Kyburg) in *Studies in Subjective Probability*, H.E. Kyburg & H.E. Smokler (eds), Wiley, New York. [MR0179814](#)
- DIACONIS, P. (1977). Finite forms of De Finetti's theorem on exchangeability. *Synthese* **36**, 271–181. [MR0517222](#)
- DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *Annals of Probability* **8**, 745–764. [MR0577313](#)
- DINIZ, M.A. and POLPO, A. (2012). Bayes' postulate for trinomial trials. *AIP Conference Proceedings*, **1490**, 108–115.
- DINIZ, M.A. and POLPO, A. (2013). A simple proof for the multinomial version of the Representation Theorem. Available at <http://www.mi.imati.cnr.it/conferences/BAYSM2013/papers/baysm10.pdf>.
- DRTON, M. and RICHARDSON, T.S. (2008). Binary models for marginal independence. *Journal of the Royal Statistical Society Series B*, **70**, 287–309. [MR2424754](#)
- EDWARDS, A.W.F. (1978). Commentary on the arguments of Thomas Bayes. *Scandinavian Journal of Statistics*, **5**, 116–118. [MR0518602](#)
- GOOD, I.J. (1979). Bayes' Billiard-Table Argument Extended to Multinomials. *Journal of Statistical Computation and Simulation*, **9**(2), 161–163.
- HEWITT, E. and SAVAGE, L.J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, **80**, 470–501. [MR0076206](#)
- HILL, B.M. (1968). Posterior distribution of percentile Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, **63**, 677–691. [MR0238430](#)
- HILL, B.M. (1993). Parametric models for $A_{(n)}$: Splitting processes and mixtures. *Journal of the Royal Statistical Society Series B*, **55**(2), 423–433. [MR1224406](#)
- MOLINA, E.C. (1931). Bayes' Theorem: An expository presentation. *Annals of Mathematical Statistics*, **2**, 23–37.
- MURRAY, F.H. (1930). Note on a scholium of Bayes. *Bulletin of the American Mathematical Society*, **36**, 129–132. [MR1561904](#)
- SCHRIEBER, F. (1987). The Extended Bayes-Postulate, Its Potential Effect on Statistical Methods and Some Historical Aspects. *Probability and Bayesian Statistics*, pages 423–430. Springer, US.
- STIGLER, S.M. (1986). *The History of Statistics*. The Belknap Press of Harvard University Press. Cambridge, MA & London, England. [MR0852410](#)
- STIGLER, S.M. (1982). Thomas Bayes's Bayesian inference. *Journal of the Royal Statistical Society Series A*, **145**(2), 250–258. [MR0669120](#)
- WOOD, G.R. (1992). Binomial mixtures and finite exchangeability. *Annals of Probability*, **20**, 1167–1173 [MR1175255](#)