IN DEFENCE OF THE REVERSE GAMBLER’S BELIEF

B. O’NEILL * ** AND
B. D. PUZA,* *** Australian National University

Abstract

We investigate the problem of predicting the outcomes of a sequence of discrete random variables that are almost uniform, in the sense that they are generated from a random process that is designed to produce independent uniform outcomes, but may not do so exactly. Using assumptions based around this notion, we derive a useful stochastic ordering for prediction. This leads us to reject the gambler’s belief as unsound and conclude that the reverse gambler’s belief is the optimal prediction method.

Keywords: Gambler’s fallacy; reverse gambler’s fallacy; frequent outcome approach; exchangeability; stochastic ordering; Bayesian theory

2000 Mathematics Subject Classification: Primary 60G09
Secondary 62F15

1. The gambler’s fallacy and the reverse gambler’s fallacy

If an abundance of heads come up when a coin is tossed, then observers may be heard to assert that a tail is due; that it is more likely to come up than another head. This kind of belief (or assertion) is often called the gambler’s fallacy (or the Monte Carlo fallacy), though it should more properly be called the gambler’s belief when not accompanied by any justification. On the contrary, some observers may assert that a head is the more likely. This belief is often called the reverse gambler’s fallacy, though again it should more properly be called the reverse gambler’s belief.

In statistical literature both beliefs have been attributed to a failure to understand statistical independence. The former has also been attributed to a misunderstanding of informal principles akin to Kolmogorov’s strong law of large numbers (that asserts the almost sure convergence of expectation and sample mean for independent random variables). However, such criticisms often show a lack of understanding of statistical independence; many confuse causal independence with statistical independence, and incorrectly use the former to assert the latter with simplistic (and flawed) arguments along the lines of ‘dice have no memories’. Moreover, as Cowan (1969) rightly warned, the reasoning behind the beliefs are rarely explicated and are therefore imputed by the logician.

2. The Bayesian approach

In Bayesian statistics, the notion that successive outcomes are produced in an identical manner is, quite literally, described by the assumption of exchangeability, i.e. that the probabilities...
of outcomes in a sequence are invariant under permutations of their order. In random processes with a finite number of possible outcomes, this gives rise to a multinomial model with unknown long-run proportions of outcomes.

When modelling processes that are designed to produce independent random outcomes, it is common to assume that these processes generate exactly that. This is equivalent to assuming that the long-run proportions of outcomes are almost surely equal. In this case, if the sequence of outcomes is exchangeable then the outcomes are independent with equal marginal probability, so that prediction is arbitrary.

However, under standard Bayesian multinomial models using either the reference prior or symmetric conjugate prior (both of which are special cases of the Dirichlet prior), it can be shown that the reverse gambler’s belief arises as the correct posterior conclusion. The implications of these models are well known. Below we show that this conclusion actually arises under much wider and more intuitive assumptions; to sceptics this should be more palatable than existing Bayesian models.

Following the notation of Bernardo and Smith (1994), let $x = (x_1, x_2, \ldots)$ be a sequence of values, each with the same finite range $1, 2, \ldots, m$, and let $x_k = (x_1, x_2, \ldots, x_k)$ be the observed outcomes. We denote the observed counts by $n = n(x_k) = (n_1, n_2, \ldots, n_m)$ with $n_i = n_i(x_k) = \sum_{j=1}^{k} 1_{x_j = i}$ (where $1_{\cdot}$ is the indicator function), and the long-run proportions by $\theta = \theta(x) = (\theta_1, \theta_2, \ldots, \theta_m)$ with $\theta_i = \theta_i(x) = \lim_{k \to \infty} n_i(x_k)/k$.

We consider the same problem as is faced by the gambler, that of predicting $x_{k+1}$. Interest therefore lies in the probability mass function $p(x) = P(x_{k+1} = x | x_k)$.

Theorem 1(a) below shows that if both $x$ and $\theta$ are exchangeable and if $n_a \geq n_b$, then $\theta_a$ is stochastically greater than or equal to $\theta_b$ a posteriori (given $x_k$). It follows that $p(a) \geq p(b)$, which precludes the gambler’s belief. Note that if $\theta = (1/m, \ldots, 1/m)$ almost surely then $\theta$ is exchangeable, so we have not precluded the common approach.

We now diverge from the common approach by considering a nondegeneracy condition that corresponds to our contemplation of the possibility of bias in the random process. Either of the following definitions captures this notion.

**Definition 1.** (Prior nondegeneracy.) If a priori there is some possibility that the elements of $\theta$ are all positive and not all equal, then we say that $\theta$ is nondegenerate.

**Definition 2.** (Posterior nondegeneracy.) If a posteriori there is some possibility that the positive elements of $\theta$ are not all equal, then we say that $\theta|x_k$ is nondegenerate.

It can be shown that prior nondegeneracy implies posterior nondegeneracy and, in fact, corresponds to a belief that posterior nondegeneracy would hold for all possible outcomes that could be observed. Lemma 1 and Theorem 1 below show that the addition of either nondegeneracy condition implies that if $n_a > n_b$ then $\theta_a$ is stochastically greater than $\theta_b$ a posteriori. It then follows that $p(a) > p(b)$, which is the reverse gambler’s belief.

This immediately gives us an optimal prediction method, namely that we should predict one of the outcomes that has occurred the most in our observations. We call this method the frequent outcome approach.

It is easy to show that under this approach, unless we have observed all outcomes the same number of times (including none, as is the case a priori), the probability of successful prediction is strictly greater than $1/m$. This result calls into question the fairness of some so-called fair bets that are predicated on an assumption of independence between outcomes.
3. Conclusion

In scrutinising probabilistic beliefs, as with any application of logic, it is rather unfair to impute to the believer an unsound argument (which they have not explicated) as a basis for rejecting their beliefs. It may well be that a gambler relies on the gambler’s belief or the reverse under erroneous logical arguments. However, sometimes the reverse gambler’s belief is the optimal rational behaviour. In particular, if we contemplate the possibility of bias in a random process designed to produce independent uniform random outcomes then we should reject the common belief that prediction is arbitrary in favour of the frequent outcome approach.

**Lemma 1.** If \( x \) and \( \theta \) are exchangeable and either \( \theta \) or \( \theta | x_k \) is nondegenerate, then \( P(\theta_a > \theta_b | n) > 0 \) for all \( a \neq b \) such that \( n_a > 0 \).

A proof of this lemma is available from the authors on request.

**Theorem 1.** Suppose that \( x \) and \( \theta \) are exchangeable. Then we have

(a) if \( n_a \geq n_b \) then \( \theta_a \) is stochastically greater than or equal to \( \theta_b \) a posteriori (given \( x_k \)),

(b) if \( n_a > n_b \) and \( P(\theta_a > \theta_b | n) > 0 \) then \( \theta_a \) is stochastically greater than \( \theta_b \) a posteriori (given \( x_k \)).

**Proof.** Since \( x \) is exchangeable, it follows from the representation theorem of de Finetti (1980) that \( p(\theta | x_k) = p(\theta | n) \) and \( n | \theta \sim \text{multinomial}(k, \theta) \), so that

\[
p(n | \theta) = \binom{k}{n} \prod_{i=1}^{m} \theta_i^{n_i}.
\]

Let \( \hat{\theta} \) be the permutation of \( \theta \) with \( \theta_a \) swapped with \( \theta_b \). Since \( \theta \) is exchangeable then, for all \( \theta_a > 0 \), we have

\[
p(\hat{\theta} | n) = \frac{p(n | \hat{\theta}) p(\hat{\theta})}{p(n)} = \left( \frac{\theta_b}{\theta_a} \right)^{n_a-n_b} \frac{p(n | \theta) p(\theta)}{p(n)} = \left( \frac{\theta_b}{\theta_a} \right)^{n_a-n_b} p(\theta | n).
\]

It then follows that

\[
P(\theta_a > t | x_k) = P(\theta_b > t | x_k) = P(\theta_a > t | n) - P(\theta_b > t | n)
\]

\[
= P(\theta_a > t | n) - P(\theta_b > t | n)
\]

\[
= E(1_{\{\theta_a > t \geq \theta_b\}} | n) - E(1_{\{\theta_b > t \geq \theta_a\}} | n)
\]

\[
= E(1_{\{\theta_a > t \geq \theta_b\}} | n) - E\left( 1_{\{\theta_a > t \geq \theta_b\}} \left( \frac{\theta_b}{\theta_a} \right)^{n_a-n_b} \right) | n
\]

\[
= E\left( 1_{\{\theta_a > t \geq \theta_b\}} \left( 1 - \left( \frac{\theta_b}{\theta_a} \right)^{n_a-n_b} \right) | n \right).
\]

(1)

For part (a), since \( n_a \geq n_b \), the integrand implicit in (1) is nonnegative over the range of the integral (that is, the range of the indicator), so that \( P(\theta_a > t | x_k) \geq P(\theta_b > t | x_k) \) for all \( t \). The result follows.

For part (b), since \( n_a > n_b \), the integrand implicit in (1) is strictly positive over the range of the integral. Since \( P(\theta_a > \theta_b | n) > 0 \) it can be shown that \( P(\theta_a > t \geq \theta_b | n) > 0 \) for
all $t$ in some neighbourhood. Therefore $P(\theta_a > t | x_k) > P(\theta_b > t | x_k)$ for all $t$ in that
neighbourhood, and the result follows.

This paper is an abridged version of O’Neill and Puza (2004). This unpublished report,
containing all relevant proofs, can be obtained from the authors.

References

O’Neill, B. and Puza, B. D. (2004). Dice have no memories but I do: a defence of the reverse gambler’s