Bayesian methods for testing the randomness of lottery draws

David F. Percy

Centre for Operational Research and Applied Statistics,
Salford Business School,
University of Salford,
Manchester,
M5 4WT, UK.

email d.f.percy@salford.ac.uk

Salford Business School Working Paper Series

Paper no. 325/06

Bayesian Methods for Testing the Randomness of Lottery Draws

David F. Percy
Centre for Operational Research and Applied Statistics
University of Salford
Greater Manchester
M5 4WT

d.f.percy@salford.ac.uk

Summary

Lottery regulators intermittently commission statistical analyses of lottery draw histories to ascertain whether the drawings appear to be conducted randomly. The primary purpose is to reassure the public that any unusual occurrences actually concur with random drawings, though it is also possible to detect sources of non-randomness from these analyses. The methodology adopted usually consists of applying several frequentist significance tests for randomness. We argue that, although this approach has served its purpose well, it has limitations that a Bayesian analysis would avoid.

We describe appropriate Bayesian methodology for testing the randomness of lottery draws, with particular emphasis on "m from n" lotteries. In particular, we develop methods for testing simple and composite hypotheses of randomness, before comparing their advantages and disadvantages for a special case. We also consider how to specify regions of randomness for composite hypotheses and how to specify and elicit subjective prior distributions for non-randomness parameters. Finally, we demonstrate these methods on draw history data pertaining to the first 1,101 draws of the main U.K. National Lottery game.

Keywords: Bayesian; elicitation; hypothesis; lottery; randomness; subjective.

1. Introduction

Lottery regulators continually monitor the operating procedures to check that winning selections are drawn randomly. An important part of this process involves physical checks of the mechanical draw equipment; another involves supervising and verifying the conduct of the draws; a third involves statistical analyses to assess the numbers drawn. This paper focuses on the last of these issues. There are three main purposes for the monitoring and testing: to identify possible sources of bias, to warn of possible corruption and to reassure the public that the draws are random. Although statistical analyses might be able to detect very specific sources of bias or corruption, their purpose is mainly to reassure the public that the numbers drawn accord with the assumption of randomness.

The published literature on, and practical applications of, statistical analyses for assessing lottery randomness have almost exclusively adopted frequentist methods of inference, in the form of goodness-of-fit tests; see Joe (1993), Haigh (1997) and the University of Salford (2004-2005). While this approach addresses the problem satisfactorily and generally achieves the goal admirably, our belief is that it has serious flaws that Bayesian methods of inference would avoid. In particular, the frequentist approach suffers from problems of meaning and interpretation, arbitrary size and power and multiplicity of testing. Lindley (1957) and others identified these flaws and we elaborate on them below. There are also published papers on aspects of lotteries related to forecasting the market demand and modelling player selection bias, including those by Stern and Cover (1989), Beenstock and Haitovsky (2001) and Koning and Vermaat (2002).

1.1 Meaning and Interpretation

A significance test of the null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative hypothesis $H_1 : \theta \in \Theta_1$ is based on an observable test statistic T and a rejection region R, with associated probabilities of type I error

$$\alpha = P(T \in R | \theta \in \Theta_0) \tag{1}$$

and type II error

$$\beta = P(T \notin R | \theta \in \Theta_1). \tag{2}$$

However, non-rejection of H_0 does not imply acceptance, rejection of H_0 says nothing about how far θ is from Θ_0, and tail areas (p-values) are not probabilities about θ but statements about the data assuming H_0 to be true, so this is a poor procedure!

Moreover, classical significance tests of a simple null hypothesis

$$H_0 : \theta = \theta_0 \tag{3}$$

against a composite alternative

$$H_1 : \theta \neq \theta_0 \ , \quad H_1 : \theta < \theta_0 \quad \text{or} \quad H_1 : \theta > \theta_0 \tag{4}$$

are often nonsense, because it is often the case that $P(H_0) = 0$. In particular, lottery draws cannot be completely random. The mechanical nature of ball selection behaves as a dynamic system and must necessarily induce some non-randomness, even if it is negligible and of no practical importance. For example, the balls cannot be of exactly equal weights and there might be a tendency for the draw machines to select either the lighter or the heavier balls, though such effects might not be noticeable.

Therefore, it is nonsense to test the hypothesis that the draws are completely random. If the test result is significant, it merely confirms what we already know: that the null hypothesis is false. If the test result is not significant, it merely indicates that there are insufficient data for us to reject H_0! This reasoning is analogous to the situation of hierarchical model selection in regression analysis, whereby we test whether regression coefficients are zero in order to omit some predictor variables for the sake of robustness and parsimony, even though correlations are rarely zero; see Draper (1995).

1.2 Arbitrary Size and Power

The size of a significance test is α, the probability of rejecting a true null hypothesis. The statistical analyst is free to choose whatever size he or she deems appropriate and the default is generally taken as $0 \cdot 05$, corresponding to 5%. However, this is an arbitrary value and we could, albeit wrongly, choose a larger or smaller value depending on whether we do or do not wish to reject the null hypothesis of randomness. The power of a significance test is $1-\beta$, the probability of rejecting a false null hypothesis. The type of test used and the draw history or sample size available determine the power for a given scenario. With few data, the power is small and significant results indicating non-randomness are rare. Conversely, with many data, the power is large and significant results are common.

1.3 Multiplicity of Testing

Multiplicity of testing adversely inflates the number of false positives observed and arises in the context of lottery games in two ways. Firstly, the usual approach for statistical analysis of historical draw data involves conducting a battery of complementary significance tests. Secondly, the draw history develops continually over time and we could repeat any test as often as we like. In this case, the tests are dependent if new data merely supplement old data on each occasion, or independent with reduced power if we only analyse the new data on each occasion. If each of these tests has size $\alpha \in (0,1)$ then the overall size, or probability of rejecting at least one of k true null hypotheses, is $1-(1-\alpha)^k > \alpha$. For example, with $\alpha = 0 \cdot 05$ and $k = 10$, this probability is $0 \cdot 40$.

To avoid spurious significant results by over-testing, we can use multiple comparisons procedures such as the Bonferroni adjustment. This involves conducting each individual test with a specified size of α/k, which gives an overall size of testing

$$\alpha_k = 1-\left(1-\frac{\alpha}{k}\right)^k = \alpha - \frac{k-1}{2k}\alpha^2 + \dots + \left(-\frac{1}{k}\right)^k \alpha^k \approx \alpha \tag{5}$$

on applying the binomial theorem and noting that successive terms in the summation decrease in absolute value and have alternating signs, with $\alpha \approx 0 \cdot 05$. Moreover, suppose we define the smooth function

$$f(x)=1-\left(1-\frac{\alpha}{x}\right)^{x} \tag{6}$$

for $x>\alpha$, which clearly interpolates the discontinuous function α_k of Equation (5). We show that $f(x)$ is a strictly decreasing function of x, and hence α_k is a strictly decreasing function of k, by considering its derivative

$$f'(x)=-\left(1-\frac{\alpha}{x}\right)^{x}\left\{\frac{\alpha/x}{1-\alpha/x}+\ln\left(1-\frac{\alpha}{x}\right)\right\}. \tag{7}$$

Noting the Taylor series expansions

$$\frac{y}{1-y}=\sum_{i=1}^{\infty}y^{i} \tag{8}$$

for $0<y<1$ and

$$\ln(1-y)=-\sum_{i=1}^{\infty}\frac{y^{i}}{i} \tag{9}$$

for $-1\leq y<1$, we see that the sum in the derivative is positive and so $f'(x)<0$ for all $x>\alpha$. Consequently, $f(x)$ is indeed a strictly decreasing function of x and therefore α_k is a strictly decreasing function of k.

Hence, the largest value α_k can assume arises when $k=1$ and is equal to $\alpha_1=\alpha$, from Equation (5). The smallest value that α_k can assume arises asymptotically and is equal to

$$\alpha_{\infty}=\lim_{k\to\infty}\left\{1-\left(1-\frac{\alpha}{k}\right)^{k}\right\}=1-e^{-\alpha}\ , \tag{10}$$

so we have simple bounds of the form $1-e^{-\alpha}<\alpha_k\leq\alpha$. If we set $\alpha=0\cdot05$, it follows that $0\cdot04877<\alpha_k\leq0\cdot05$ for all $k=1,2,3,\ldots$ so this simple, Bonferroni multiple comparisons procedure is very good in ensuring that the overall size of the test is about $0\cdot05$. Unfortunately, this benefit comes at the expense of reduced power for individual tests; other multiple comparisons procedures fare better in this regard.

This power reduces further if the tests are dependent, which is a major problem with longitudinal or sequential testing. In the extreme case, all k tests might be identical and so rejecting one null hypothesis effectively rejects them all. The modified significance level of α/k is now the overall size of the tests, which considerably diminishes the power. This

effect is of less concern when a cross-sectional battery of complementary tests is applied at a single moment in time.

1.4 Summary

As a form of scientific inference, significance testing has inherent flaws despite its universal application throughout the twentieth century. Many simple null hypotheses cannot physically be true, so why bother to test them? In general, though, we might "not reject" a null hypothesis because (a) it is true or (b) it is false but the power of the test is too small. Conversely, we might "reject" a null hypothesis because (a) it is false or (b) it is true but the size of the test is too large. It is impossible to distinguish between these explanations for any individual test result.

Moreover, even if we reject a null hypothesis of randomness, the practical implications of this observation might be inconsequential. For example, the U.K. National Lottery selects six balls at random from forty-nine. It also selects a seventh, bonus ball, though this only differentiates between two classes of "match five" winners and we ignore it here. If the probability of selecting one particular number on a given draw were $6/50$ rather than $6/49$, the practical implications of this might be negligible.

2. Bayesian Tests for Randomness

Clearly, the results of significance testing are difficult to interpret. The usual practice of quoting p-values rather than binary conclusions to significance tests is of little use here, because the lottery regulator's interest is binary according to whether the numbers drawn are compatible with the hypothesis of randomness. Parallel arguments apply to the use of confidence intervals rather than significance tests.

The output from a Bayesian test takes the form of a conditional probability, p, of randomness given the available draw history, which we can evaluate after each draw in order to construct a control chart. Depending on the utility function associated with declaring draws to be random or not, we can declare thresholds for this conditional probability below which action should be taken to check or modify the lottery operating procedures. For example, we might decide to carry out a suitable detailed examination of the equipment and draw procedures if $p < 0 \cdot 1$. Similarly, we might decide to change the equipment or draw procedures with immediate effect if $p < 0 \cdot 01$.

We now investigate two alternative Bayesian approaches to testing for randomness, involving simple or composite hypotheses of randomness, with particular application to lotteries of the common "m from n" variety. The U.K. National Lottery is of this format, with $m = 6$ and $n = 49$.

2.1 Simple Hypothesis of Randomness

In this subsection, we suppose that true randomness is possible. Defining the events R: "draws are random" and D: "observed draw history", Bayes' theorem gives

$$P(R|D) = \frac{P(D|R)P(R)}{P(D)} \tag{11}$$

where

$$P(D) = P(D|R)P(R) + P(D|R')P(R') \tag{12}$$

from the law of total probability. If preferred, we can work with the posterior odds

$$\frac{P(R|D)}{P(R'|D)} = \frac{P(D|R)P(R)}{P(D|R')P(R')} \tag{13}$$

instead, as there is a one-to-one correspondence between this quantity (o) and the posterior probability (p) in Equation (11), whereby

$$o = \frac{p}{1-p} \quad \text{and} \quad p = \frac{o}{o+1}. \tag{14}$$

This procedure enables us to evaluate the probability that the draws are random given the draw history, which is in precisely the required form for constructing a control chart, informing regulators, making inference and drawing conclusions.

In order to calculate $P(R|D)$ using Equation (11), we need to evaluate the four probabilities in Equation (12) although, as randomness and non-randomness are complementary events, we note that $P(R') = 1 - P(R)$. The lottery regulator might subjectively choose a large value for the prior probability of randomness, such as $P(R) = 0 \cdot 9$ or $P(R) = 0 \cdot 99$, to reflect other supporting evidence based upon physical checks of the equipment and draw procedures. Such a large prior probability yields high specificity but low sensitivity of our statistical tests for randomness, avoiding excessive false alarms but making it difficult to identify problems. Agreement between the theory of reference priors, Jaynes' maximum entropy and the Bayes-Laplace postulate for finite discrete parameters (see Bernardo and Smith, 1993), suggests that we should set $P(R) = 0 \cdot 5$ in the absence of any specific prior beliefs. Although a suitable subjective prior is technically preferable, this objective approach would prove less contentious under public scrutiny. However, it would generate more false alarms, though this hitch could be overcome by adjusting the action thresholds so that they are more extreme.

Our difficulties lie in calculating the terms $P(D|R)$ and $P(D|R')$ in Equation (12) for a "m from n" lottery. We attempt to derive a formula for the joint probability mass function

of the observed data under randomness in Section 4, which would then allow us to evaluate the first of these two probabilities. The second probability is awkward to calculate because the lottery could be non-random in infinitely many ways and we need to consider them all. Our solution is to specify a parameter vector of non-randomness $\boldsymbol{\theta}$ with prior probability density function $g(\boldsymbol{\theta}|R')$ and evaluate the multiple integral

$$P(D|R') = \int_{-\infty}^{\infty} P(D|R',\boldsymbol{\theta})g(\boldsymbol{\theta}|R')d\boldsymbol{\theta} \tag{15}$$

where $P(D|R',\boldsymbol{\theta})$ can be a very complicated function of the observed draw frequencies and unknown parameters. We consider a detailed derivation of this probability in Section 4. Different sampling distributions and parameter vectors of non-randomness are appropriate for testing for different types of non-randomness.

2.2 Composite Hypothesis of Randomness

In this subsection, we no longer consider the state of true randomness to exist. Suppose we specify a parameter vector of non-randomness $\boldsymbol{\theta}$ with prior probability density function $g(\boldsymbol{\theta})$, which might resemble the conditional prior $g(\boldsymbol{\theta}|R')$ of Subsection 2.1. For this scenario, we also need to declare that the draws are effectively random if the values of $\boldsymbol{\theta}$ lie within a particular region of randomness R. In principle, we use these data to formulate a likelihood function $L(\boldsymbol{\theta};D)$ and then Bayes' theorem enables us to evaluate the posterior probability density function

$$g(\boldsymbol{\theta}|D) \propto L(\boldsymbol{\theta};D)g(\boldsymbol{\theta}). \tag{16}$$

This enables us to evaluate the probability of randomness given the observed data as

$$P(R|D) = \int_{\boldsymbol{\theta}\in R} g(\boldsymbol{\theta}|D)d\boldsymbol{\theta}. \tag{17}$$

As in Subsection 2.1, different forms of non-randomness require their own sampling distributions and parameter vectors.

In the case of a "m from n" lottery, the available data D consist of the observed draw history and θ_i might represent the probability that ball i is drawn if all n balls are available. We might try to simplify the problem by introducing a scalar measure of non-randomness θ, to avoid the multi-dimensional analysis that would be necessary if we were to deal with the parameter vector $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_n\}$ directly. For example, we might declare that the draws are effectively random if the values of $\boldsymbol{\theta}$ lie within a particular region $R : \theta < \varepsilon$, where

$$\theta = \max_i \left| \theta_i - \frac{1}{n} \right|; \quad i = 1,2,\ldots,n \tag{18}$$

and $\varepsilon = 0 \cdot 001$ for example. As additive differences are not ideal for probability measures $\theta_i \in [0,1]$, we might prefer to map them onto the set of real numbers first, using an inverse logit transformation for example. Our acceptable region for randomness would then be of the form $R : \theta < \varepsilon$, where

$$\theta = \max_i \left| \ln \frac{\theta_i}{1-\theta_i} - \ln \frac{\frac{1}{n}}{1-\frac{1}{n}} \right| = \max_i \left| \ln \frac{(n-1)\theta_i}{1-\theta_i} \right|; \quad i = 1,2,\ldots,n \tag{19}$$

and $\varepsilon = 0 \cdot 05$ for example. Figure 1 demonstrates that there is very little to choose between these two criteria for the U.K. National Lottery, for which $m = 6$ and $n = 49$. Consequently, we would prefer to use the criterion defined by Equation (18) for simplicity of computation and interpretation, if we were to use a scalar performance measure.

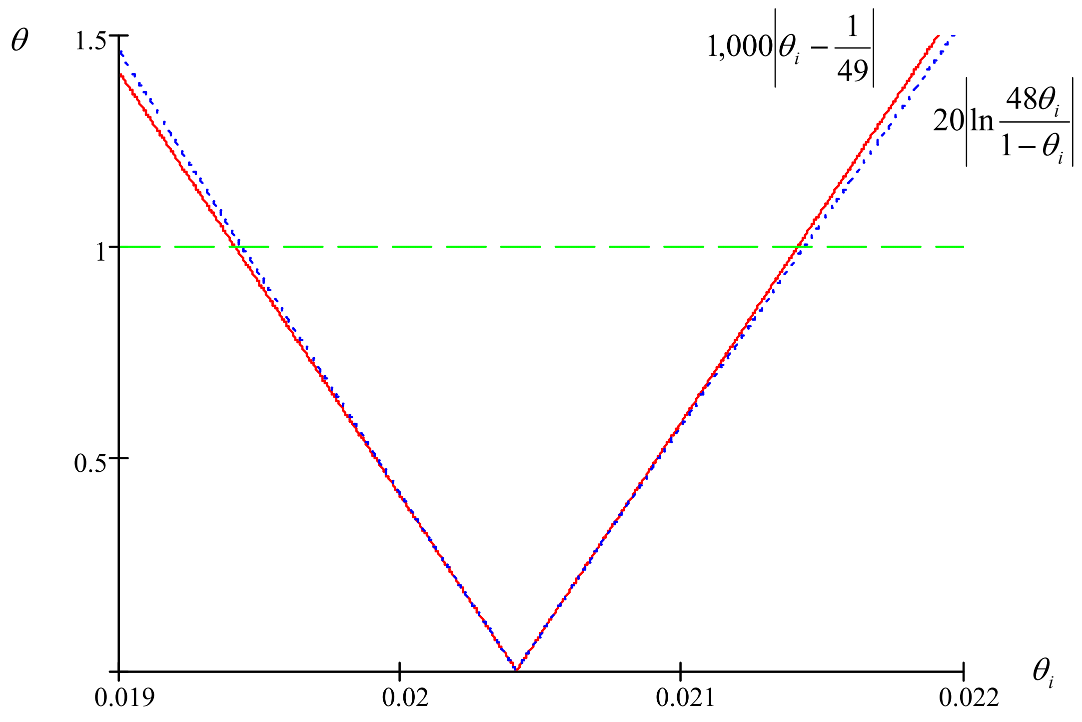

Figure 1: graphs of θ under Definitions (18) and (19), scaled to have thresholds of $\varepsilon = 1$.

Alternative measures to those in Definitions (18) and (19) include the mean absolute deviation

$$\theta = \frac{1}{n} \sum_{i=1}^{n} \left| \theta_i - \frac{1}{n} \right| \tag{20}$$

and the root mean square deviation

$$\theta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\theta_i - \frac{1}{n}\right)^2}\ , \tag{21}$$

though these are less easy to interpret than the maximum deviation of Equation (18). Although such scalar measures of non-randomness θ avoid the need to specify a multi-dimensional prior and evaluate the multiple integral associated with the parameter vector $\boldsymbol{\theta}$, the conditional distribution of $\mathbf{x}$ given θ is not well defined and multiple integration is still required. Consequently, such summary measures for $\boldsymbol{\theta}$ are relevant only for determining suitable regions of randomness if we assume composite hypotheses.

As with simple hypotheses of randomness, we can evaluate the conditional probability in Equation (17) after each draw in order to construct a control chart. We can again declare thresholds for this value below which action should be taken to check or modify the lottery operating procedures. Indeed, the methodology is very similar for simple and composite hypotheses of randomness, in that the former is a special limiting case of the latter where the region of randomness has negligible size and we use a mixture of two priors for $\boldsymbol{\theta}$. We consider the calculation and application of both strategies for a simple binary problem in Section 3, in order to extrapolate the observations and recommendations to the more general problem of lottery randomness testing in Section 4.

2.3 Further Observations

Both of these Bayesian analytical methods lead to test outcomes in the form of probabilities of randomness given the available data. These are easy to interpret and are precisely the results that we need to inform or reassure the public and to act upon when making decisions. They avoid the ambiguities of size and power that are features of significance testing and its variants. Moreover, they resolve the problem associated with sequential multiplicity of testing, as there is no objection to determining probabilities of randomness occasionally or after each draw. Indeed, we should consider all available data in order to make the most informed decisions. By evaluating probabilities that can be meaningfully combined, rather than p-values that cannot, these approaches can also resolve the multiplicity associated with applying a battery of randomness tests on each occasion. Subsections 2.1 and 2.2 consider joint prior and posterior distributions for parameter vectors $\boldsymbol{\theta}$ that measure departures from randomness in the restricted case where components relate to parameters from the same model for non-randomness.

If we could assume objective priors rather than subjective priors for the model parameters, we would avoid the need to elicit corresponding hyperparameters. Unfortunately, objective priors are not suitable for composite hypotheses, as the imposed prior probabilities of randomness are implausibly small and are actually zero if the prior is improper. This subsequently leads to posterior probabilities of randomness that are unrealistically small, especially in multi-parameter situations. With simple hypotheses, it seems reasonable to specify objective priors for non-randomness parameters. However, they are again highly

dispersed for multi-parameter situations and cause serious problems because the prior probability mass of randomness dominates the computation and results. Furthermore, objective priors can lead to improper prior predictive distributions that can cause problems here; see O'Hagan (2004). It seems that we cannot avoid the need for subjective priors in this context.

The notion of true randomness appears to be purely conceptual, leading us to use composite hypotheses of randomness for our analyses, despite the additional requirement that we specify acceptable regions of randomness in addition to eliciting hyperparameters for subjective prior distributions of the randomness parameters. However, we could justify adopting simple hypotheses by interpreting them as special cases of composite hypotheses, with negligible regions of randomness and priors that are mixtures of discrete and continuous components. These observations suggest that we can implement either approach. It is also worth noting that both approaches demand multiple integration over the parameter space, which might be difficult in high dimensions; refer to Equations (15) and (17). However, the integral is analytically tractable for simple hypotheses but not for composite hypotheses, due to the regions of integration involved. In order to investigate their properties further, we now apply both methods to a simple case of randomness testing corresponding to coin-tossing experiment, which is equivalent to a "1 from 2" lottery.

3. Bernoulli Outcomes

In order to demonstrate the methodology clearly and to identify the problems and rewards of these two Bayesian approaches, we consider the simplest scenario where there are only two outcomes. This would be the situation for a lottery where we only select one of two numbers at each draw, though an equivalent and more realistic situation would be testing whether a particular coin is biased when it is tossed to decide on an outcome of heads or tails.

3.1 Simple Hypothesis of Randomness

Suppose that a coin has probability θ of showing a head when it is tossed and that the coin is tossed d times independently. Then, the number of heads X observed has a binomial distribution with probability mass function

$$p(x|\theta) = \frac{d!}{x!(d-x)!}\theta^x(1-\theta)^{d-x} \;;\quad x = 0,1,\ldots,d\,. \tag{22}$$

The data D consist of d independent tosses of the coin in which we observe x heads and $d-x$ tails. We define the event R: "coin is unbiased" and suppose that there exists a prior probability $p \in (0,1)$ that this event is true, so $p = P(R)$. In this case, $\theta = 0\cdot 5$ and the probability of observing the data given randomness becomes

$$P(D|R) = \frac{d!}{x!(d-x)!2^d}\,. \tag{23}$$

If the coin is biased, which has prior probability $1-p=P(R')$ of being true, we model the extent of the bias using a suitable prior for θ. For a subjective analysis, the natural conjugate prior is a beta $Be(a,b)$ distribution with probability mass function

$$g(\theta)=\frac{1}{\mathrm{B}(a,b)}\theta^{a-1}(1-\theta)^{b-1}\ ; \quad 0<\theta<1\,. \tag{24}$$

For an objective analysis, the invariant prior is a $Be\left(\frac{1}{2},\frac{1}{2}\right)$ distribution with probability density function

$$g(\theta)=\frac{1}{\pi\sqrt{\theta(1-\theta)}}\ ; \quad 0<\theta<1\,. \tag{25}$$

We adopt the subjective prior for our analyses, as recommended in Subsection 2.3.

From Equations (15), (22) and (24), the prior predictive probability mass function for X under the assumption of a biased coin is the binomial-beta form

$$p(x)=\int_{-\infty}^{\infty}p(x|\theta)g(\theta)d\theta=\frac{d!\,\mathrm{B}(x+a,d-x+b)}{x!(d-x)!\,\mathrm{B}(a,b)}\ ; \quad x=0,1,\ldots,d \tag{26}$$

as described by Bernardo and Smith (1993). Consequently, the probability of observing the data under the assumption of bias is

$$P(D|R')=\frac{d!\,\mathrm{B}(x+a,d-x+b)}{x!(d-x)!\,\mathrm{B}(a,b)} \tag{27}$$

from Equation (26). We can now evaluate the posterior odds as

$$\frac{P(R|D)}{P(R'|D)}=\frac{p\mathrm{B}(a,b)}{(1-p)2^{d}\,\mathrm{B}(x+a,d-x+b)} \tag{28}$$

from Equations (13), (23) and (27).

For example, suppose we have equal prior beliefs that the coin is and is not biased, so $p=0\cdot5$. Suppose also that our prior beliefs concerning θ if there is bias are symmetric about $\theta=0\cdot5$, which implies that $a=b$. Finally, suppose that we toss the coin four times, observing two heads and two tails, so $d=4$ and $x=2$. Under these illustrative circumstances, we can determine the posterior odds as

$$\frac{P(R|D)}{P(R'|D)} = \frac{(2a+3)(2a+1)}{4a(a+1)} \tag{29}$$

from Equation (28), corresponding to a posterior probability of the coin's being unbiased of

$$P(R|D) = \frac{(2a+3)(2a+1)}{(2a+3)(2a+1)+4a(a+1)}, \tag{30}$$

from Equations (14). In this case, the data have increased our prior probability that the coin is fair from $\frac{1}{2}$ to a value in the interval $\left(\frac{1}{2},1\right)$ according to our specified value of the hyperparameter a, which represents the strength of our prior knowledge about the bias parameter θ.

For a contrast, suppose that our data showed four heads rather than two heads and two tails. In this case, we can determine the posterior odds as

$$\frac{P(R|D)}{P(R'|D)} = \frac{(2a+3)(2a+1)}{4(a+3)(a+2)} \tag{31}$$

from Equation (28), corresponding to a posterior probability of the coin's being unbiased of

$$P(R|D) = \frac{(2a+3)(2a+1)}{(2a+3)(2a+1)+4(a+3)(a+2)}, \tag{32}$$

from Equations (14). In this case, the data have decreased our prior probability that the coin is fair from $\frac{1}{2}$ to a value in the interval $\left(\frac{1}{9},\frac{1}{2}\right)$ according to our specified value of a. Figure 2 illustrates the relationships between $P(R|D)$ and a for both of these examples.

Our choice of $p = \frac{1}{2}$ gives equal prior weighting to the hypotheses that the coin is fair and unfair, and is appropriate in some circumstances. However, suppose we select the coin at random from all coins in general circulation. Then a better choice would be $p = 1-q$ where q is a small value equal to our guess (perhaps 10^{-6}) at the proportion of all such coins that is biased. We would then need more data to overturn the hypothesis that the coin is fair, if it is not. It remains to decide how to elicit suitable hyperparameter values a and b for the natural conjugate beta prior distribution for θ under the alternative hypothesis, as defined by Equation (24). As in the illustration, it is reasonable to suppose that $a = b$ due to symmetry. Figure 2 demonstrates the sensitivity of the output to the value of a and b for this example.

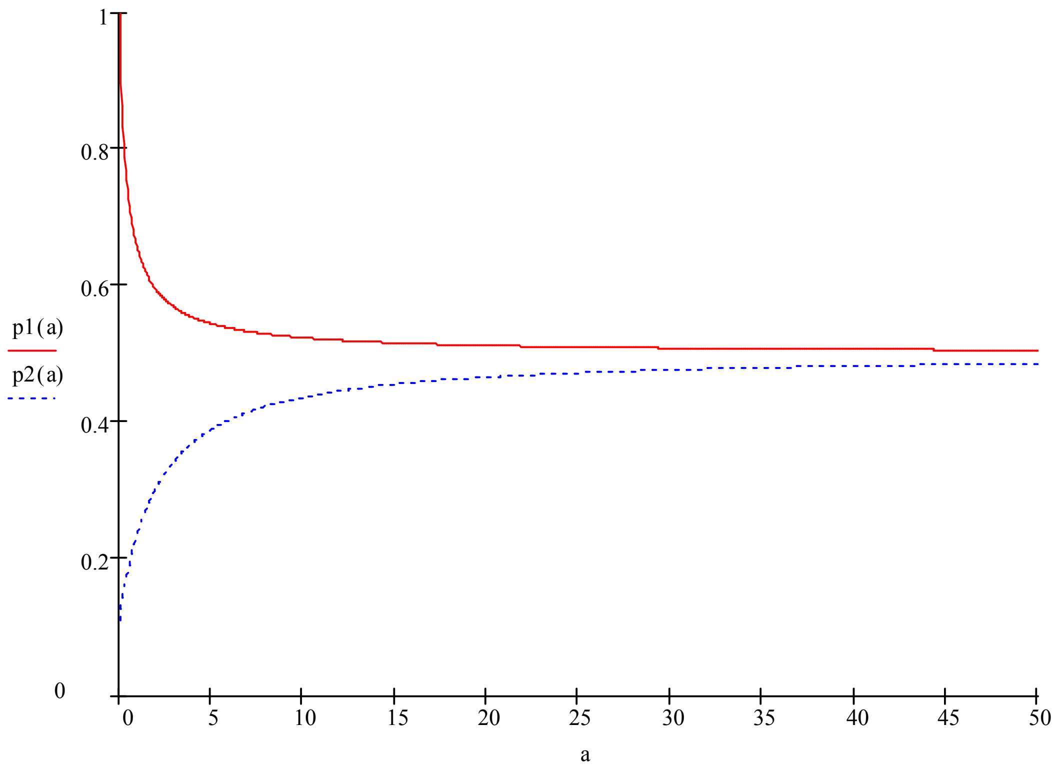

Figure 2: graphs of posterior probability of randomness against hyperparameter value a for coin examples with two heads, $p1(a)$, and four heads, $p2(a)$, from four tosses.

A proven strategy for specifying the hyperparameters of a binomial model with natural conjugate prior is by eliciting the tertiles, T_1 and T_2, corresponding to the $33\frac{1}{3}$ and $66\frac{2}{3}$ percentiles of the prior distribution, whose cumulative distribution function takes the form of an incomplete beta function

$$G(\theta)=\frac{1}{\mathrm{B}(a,b)}\int_0^\theta \theta^{a-1}(1-\theta)^{b-1}\,d\theta\,;\quad 0<\theta<1. \tag{33}$$

This leads to a pair of simultaneous nonlinear equations of the form

$$G(T_1)=\frac{1}{3}\quad\text{and}\quad G(T_2)=\frac{2}{3} \tag{34}$$

which we solve numerically for a and b. Percy (2003) illustrated this approach using Mathcad for an application arising in on-line quality control.

With the further constraint that $a=b$, this simplifies to a single nonlinear equation that conveniently only requires us to elicit one fractile of the prior, perhaps the lower and upper quartiles Q_1 and Q_3 subject to $Q_1+Q_3=1$, due to symmetry as the median Q_2 is fixed at

$\theta = \frac{1}{2}$ by this constraint. We thus determine the hyperparameter values by solving the equation

$$\frac{1}{\mathrm{B}(a,a)} \int_{Q_1}^{Q_3} \{\theta(1-\theta)\}^{a-1} d\theta = \frac{1}{2} \qquad (35)$$

for a numerically. For the coin example, we might be willing to express our prior quartiles as $0 \cdot 45$ and $0 \cdot 55$, which correspond to an arbitrary 10% below and above the true value under randomness. We interpret this choice as follows: given that the coin is biased, we regard the probability of a head as equally likely to be in the interval $(0 \cdot 45, 0 \cdot 55)$ as it is to be outside this interval. Mathcad then gives the solution to Equation (35) as $a \approx 23$, which we demonstrate in Figure 3.

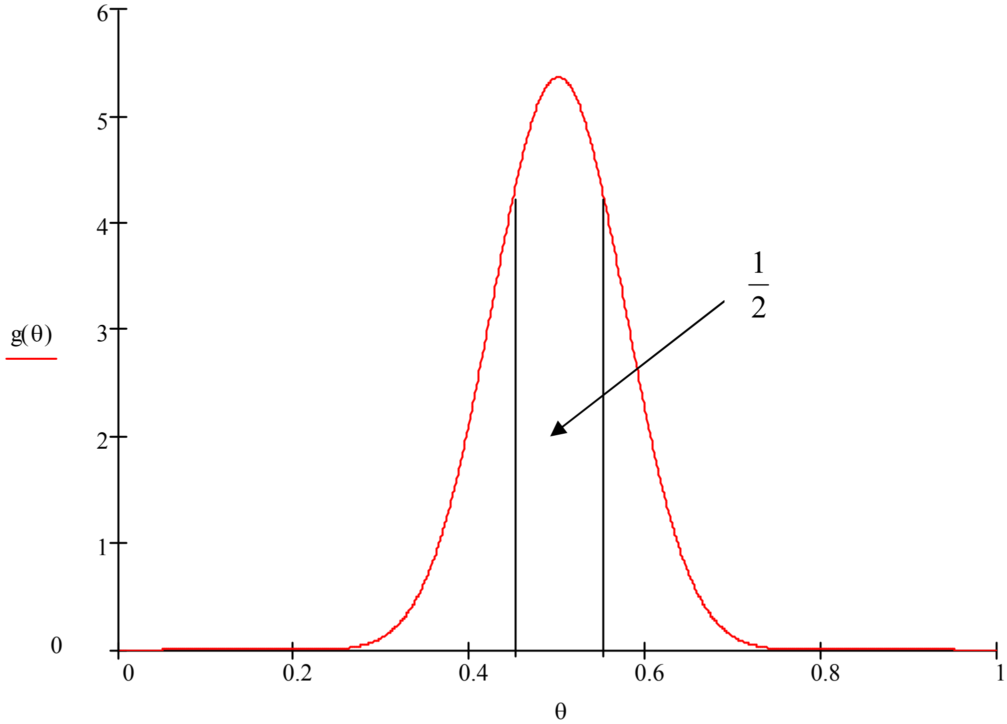

Figure 3: $Be(23,23)$ prior probability density function for the bias parameter.

We can also assess the sensitivity of the posterior probabilities of randomness $P(R|D)$ to the choice of prior quartiles Q_1 and Q_3 for the bias parameter θ. For the scenario where we observe four heads and no tails in four tosses of the coin, Equation (32) gives the results displayed in Table 1. The choice of prior quartiles is clearly influential, yet is subjective and so is not easy to specify in a way that would be broadly acceptable to lottery regulators, operators, media and players. We reconsider this problem when recommending a suitable approach for analysing general lottery situations in Section 4.

Q_1	Q_3	a	$P(R\|D)$
$0 \cdot 50$	$0 \cdot 50$	∞	$0 \cdot 5000$
$0 \cdot 49$	$0 \cdot 51$	$568 \cdot 8$	$0 \cdot 4987$
$0 \cdot 45$	$0 \cdot 55$	$22 \cdot 88$	$0 \cdot 4696$
$0 \cdot 40$	$0 \cdot 60$	$5 \cdot 816$	$0 \cdot 4014$
$0 \cdot 25$	$0 \cdot 75$	$1 \cdot 000$	$0 \cdot 2381$
$0 \cdot 00$	$1 \cdot 00$	0	$0 \cdot 1111$

Table 1: sensitivity of posterior probability of randomness to prior upper quartile.

3.2 Composite Hypothesis of Randomness

We retain the binomial model specified in Equation (22) but now suppose that the state of randomness does not exist, except as a limiting case whereby an acceptable region of randomness based on the parameter θ becomes small. An improper prior for θ would give $P(R) = 0$, which does not fairly reflect anybody's prior beliefs about randomness. If it did, there would be no need to test whether coin tosses or lottery draws are random. Consequently, we assume a natural conjugate prior for θ, which represents the probability that the coin shows a tail. This takes the same form as the beta distribution in Subsection 3.1, with probability mass function as specified in Equation (24).

The acceptable region of randomness referred to in Subsection 2.2 might take the simple form

$$R : \left|\theta - \frac{1}{2}\right| < \varepsilon \tag{36}$$

where $0 < \varepsilon < \frac{1}{2}$. As we only have one binomial observation, the number of heads x in d draws, the likelihood function is

$$L(\theta; D) \propto p(x|\theta) \propto \theta^x (1-\theta)^{d-x} \tag{37}$$

from Equation (22) The posterior probability density function for θ can then be calculated using Relation (16) as

$$g(\theta|D) \propto L(\theta; D) g(\theta) \propto \theta^{x+a-1}(1-\theta)^{d-x+b-1} \ ; \quad 0 < \theta < 1 \tag{38}$$

so $\theta|D \sim Be(x+a, d-x+b)$ with probability density function

$$g(\theta|D) = \frac{1}{\mathrm{B}(x+a, d-x+b)} \theta^{x+a-1}(1-\theta)^{d-x+b-1} \ ; \quad 0 < \theta < 1 . \tag{39}$$

Inequality (36) and Equation (39) now lead us to calculate the probability that the coin is effectively fair as

$$P(R|D) = \int_{\frac{1}{2}-\varepsilon}^{\frac{1}{2}+\varepsilon} g(\theta|D)d\theta . \tag{40}$$

As the integrand is a beta probability density function, we evaluate the integral numerically in general.

Suppose that, as in the previous section, we toss the coin four times, observing two heads and two tails, so $d = 4$ and $x = 2$. Also suppose that our prior distribution for the probability of a tail θ is a natural conjugate $Be(a,b)$ form and that we regard any value of θ in the interval $(0 \cdot 45, 0 \cdot 55)$ as indicating that the coin is reasonably fair. This interval arbitrarily allows for an error of up to 10% either side of the value that would represent true randomness. In order to specify suitable values for the hyperparameters a and b, we might suppose that the prior is symmetric about $\theta = 0 \cdot 50$, in which case $a = b$. We might also suppose that our prior belief that the coin is reasonably fair is equal to one half, which leads us to solve

$$P(R) = \int_{0\cdot 45}^{0\cdot 55} \frac{\theta^{a-1}(1-\theta)^{a-1}}{\mathrm{B}(a,a)} d\theta = \frac{1}{2} \tag{41}$$

for a. This is equivalent to Equation (35) and we can again evaluate the result of this prior elicitation numerically using Mathcad, for example, to give $a = b \approx 23$ and so $\theta \sim Be(23,23)$ as previously demonstrated in Figure 3.

However, the posterior distribution for θ becomes $\theta|D \sim Be(25,25)$, from which we obtain

$$P(R|D) = \int_{0\cdot 45}^{0\cdot 55} \frac{\theta^{24}(1-\theta)^{24}}{\mathrm{B}(25,25)} d\theta \approx 0 \cdot 52 \tag{42}$$

so the data have slightly increased our belief in the coin's tendency to show heads and tails equally often. Equivalently, the odds in favour of the coin's being fair have increased from $0 \cdot 50/(1 - 0 \cdot 50) = 1 \cdot 00$ to about $0 \cdot 52/(1 - 0 \cdot 52) \approx 1 \cdot 08$. For a contrast, as in the previous subsection, suppose that our data showed four heads rather than two heads and two tails. In this case, the posterior distribution for θ becomes $\theta|D \sim Be(23,27)$, from which we obtain

$$P(R|D) = \int_{0\cdot 45}^{0\cdot 55} \frac{\theta^{22}(1-\theta)^{26}}{\mathrm{B}(23,27)} d\theta \approx 0 \cdot 45 \tag{43}$$

so the data have decreased our belief in the coin's tendency to show heads and tails equally often. Equivalently, the odds in favour of the coin's being fair have decreased from $1 \cdot 00$ to about $0 \cdot 45/(1-0 \cdot 45) \approx 0 \cdot 82$, confirming that we now marginally prefer to believe that the coin is not fair.

Table 2 summarizes the posterior probabilities of randomness under a simple hypothesis and under a composite hypothesis for this simple example, using the same prior distributions, in the case where we observe two heads and two tails and in the case where we observe four heads. The prior probability of randomness was one half in both of these situations. The results are very similar, though it is interesting to note that if were to take the region of randomness for the composite approach as $(0 \cdot 46, 0 \cdot 54)$ rather than $(0 \cdot 45, 0 \cdot 55)$, the posterior probabilities of randomness would be the same as for the simple approach to two decimal places.

Randomness Hypothesis	Observed Data: 2 heads, 2 tails	Observed Data: 4 heads
simple	$0 \cdot 51$	$0 \cdot 47$
composite	$0 \cdot 52$	$0 \cdot 45$

Table 2: posterior probabilities of randomness for simple and composite hypotheses.

4. Lottery Randomness Testing

In Section 3, we performed Bayesian analyses for testing whether the probability of success in a sequence of Bernoulli trials is equal to a specified value, there taken as $0 \cdot 50$ to assess whether a coin is fair when tossed to show a head or a tail. Section 4 draws on observations and recommendations from that work and extends the ideas to randomness testing for "m from n" lotteries. It is clear from our investigations so far that a Bayesian approach appears to be suitable for randomness testing and provides more direct and informative inference than the frequentist approach. Moreover, for the simple binomial model, appropriate analytical methods are easily established, readily implemented and clearly interpreted.

Even though the notion of true randomness appears to be purely conceptual, we would need to specify acceptable regions of randomness and perform multiple integrals over these regions if we were to adopt composite hypotheses of randomness. Having done so though, Subsection 3.2 described an objective procedure for specifying the hyperparameters of the subjective prior distribution for the randomness parameter. In contrast, the simple hypotheses of randomness considered in Subsection 3.1 do not require us to specify acceptable regions of randomness or evaluate awkward multiple integrals but do require us to elicit subjective beliefs in order to determine suitable hyperparameter values. We noted the difficulty in doing this in a manner that pleases all interested parties and we observed that the resulting posterior probabilities are sensitive to this subjective input.

Although simple hypotheses of randomness appear to be inappropriate here as a matter of principle, we note the following analogous point. Even if we assume a composite hypothesis of randomness, any analysis requires us to assume that a specified model is correct despite this event's having zero probability. Draper (1995) investigated such issues relating to model selection. Consider the coin tossing experiments of Section 3. The probability of a tail varies from trial to trial, however slightly, due to temporal variations of the tossing action and coin's centre of mass, due to gradual wear and the accumulation of dust and grease. Similar serial dependence arises for lottery drawings, so the model is strictly invalid. However, the above frequentist and Bayesian methods assume that the model is correct with a probability of one, so it is implicitly assumed as a very good approximation to the true model.

The situation concerning a simple or composite null hypothesis of randomness is similar to this. Following a similar argument for justification, we conclude that it is reasonable to assume that true randomness has a non-zero probability of occurrence, as a very good approximation to the true situation. We must also consider the impact of our subjective assumptions, which are the prior distribution for simple hypotheses and the region of randomness for composite hypotheses. Such subjectivity might attract misdirected criticism from lottery players and the media, which would do little to reassure the public that the operators conduct the lotteries with impartiality. However, it is important to note that these two approaches appear to give very similar results based on the coin-tossing experiments.

On balance, we prefer to work with simple null hypotheses of randomness, as they avoid the need to specify acceptable regions of randomness and evaluate awkward integrals without compromising our results. We now consider suitable measures of non-randomness for "m from n" lotteries and investigate reasonably objective Bayesian analyses for these measures, based on the assumption of simple hypotheses of randomness as described in Section 2 and recommended above.

4.1 Equality of Frequencies of the Numbers Drawn

Extending the approach of Subsection 3.1, simple measures of non-randomness in the case of a "m from n" lottery might be the relative frequencies that each of the numbers 1 to n are selected, in which case

$$\theta_i = P(\text{ball } i \text{ is selected when one of the } n \text{ balls is drawn}) \tag{44}$$

for $i = 1,2,\ldots,n$ subject to

$$\sum_{i=1}^{n} \theta_i = 1. \tag{45}$$

In this case, the event R: "draws are random" would refer to randomness only in the sense that each ball is equally likely to be drawn, so $\theta_i = \frac{1}{n}$ for $i = 1,2,\ldots,n$. Although this is the most important measure, non-randomness might also arise by clustering within and among draws. Furthermore, bias might arise according to the order that balls enter the draw machine

or some other natural ordering, rather than simply the numerical values of the balls. We consider these aspects later in Section 4.

In order to proceed with the analysis, we must determine the joint probability distribution of $\mathbf{x} = \{X_1, X_2, \ldots, X_n\}$ under the assumptions of randomness (equality of frequencies) and non-randomness (inequality of frequencies), where $X_i = 0,1,\ldots,d$ is the number of times that ball i is drawn in the d draws subject to

$$\sum_{i=1}^{n} X_i = md \,. \tag{46}$$

When we have specified the joint probability distribution of $\mathbf{x}$ under the assumptions of randomness and non-randomness, we can formulate the likelihoods of the observed draw history D under these assumptions, $P(D|R)$ and $P(D|R',\boldsymbol{\theta})$ respectively. Equation (15) then takes the expectation of the latter over the prior distribution for $\boldsymbol{\theta}$ to give $P(D|R')$. Bayes' theorem in Equation (11) and the law of total probability in Equation (12) then combine this information with our prior probability of randomness $P(R)$ to calculate the quantity of interest, which is the probability of randomness given the draw history, $P(R|D)$.

However, the distribution of $\mathbf{x}$ given $\boldsymbol{\theta}$ is difficult to specify. We might consider determining its form by writing

$$p(\mathbf{x}|\boldsymbol{\theta}) = p(x_1|\boldsymbol{\theta})p(x_2|x_1,\boldsymbol{\theta})\ldots p(x_n|x_1,x_2,\ldots,x_{n-1},\boldsymbol{\theta}) \tag{47}$$

but the individual terms are complicated functions of both $\mathbf{x}$ and $\boldsymbol{\theta}$, which are not amenable to further algebraic development. The simplest of these terms corresponds to the marginal distribution of X_1 given $\boldsymbol{\theta}$, which is binomial with probability mass function

$$p(x_1|\boldsymbol{\theta}) = \frac{d!}{x_1!(d-x_1)!}\phi_1^{x_1}(1-\phi_1)^{d-x_1} \; ; \quad x_1 = 0,1,\ldots,d \tag{48}$$

where ϕ_1 is the probability that ball 1 is selected in a particular draw. Similar expressions can be specified for the marginal distributions of $X_2, X_3, \ldots, X_n$ given $\boldsymbol{\theta}$, in terms of corresponding parameters $\phi_2, \phi_3, \ldots, \phi_n$. Under the assumption of randomness, symmetry demands that $\phi_i = \frac{m}{n}$ for $i = 1,2,\ldots,n$. Assuming non-randomness, ϕ_i is a complicated function of $\boldsymbol{\theta}$, as it involves nested summations. However, even in terms of the parameterisation $\phi_1, \phi_2, \ldots, \phi_n$, the conditional distributions in Equation (47) are difficult to specify and so the expression for $p(\mathbf{x}|\boldsymbol{\theta})$ is algebraically cumbersome.

We next consider whether scalar simplifications are appropriate for the draw history $\mathbf{x}$, as they might also avoid the need to specify such a multi-dimensional distribution. Indeed,

the current frequentist approach involves test statistics X, which are summaries of the draw history and whose distributions we know, either exactly or asymptotically, under the hypothesis of randomness. For testing equality of frequencies of the numbers drawn, the standard summary is the modified chi-square goodness-of-fit test statistic

$$X = \frac{n-1}{n-m}\sum_{i=1}^{n}\frac{\left(X_i - \frac{md}{n}\right)^2}{\frac{md}{n}} = \frac{n-1}{n-m}\left(\frac{n}{md}\sum_{i=1}^{n}X_i^{\,2} - md\right) \dot{\sim} \chi^2(n-1), \tag{49}$$

which was proposed by Joe (1993) and subsequently applied by Haigh (1997) and the University of Salford (2004-2005) to the U.K. National Lottery. Consequently,

$$f(x) \approx \frac{\left(\frac{1}{2}\right)^{(n-1)/2}}{\Gamma\left(\frac{n-1}{2}\right)} x^{(n-1)/2-1} e^{-x/2} \; ; \quad x > 0 \tag{50}$$

under the assumption of randomness, which is the continuous equivalent of $P(D|R)$ in Equations (11), (12) and (13).

Under the assumption that the draws are not random,

$$\frac{n-1}{n-m}\sum_{i=1}^{n}\frac{\left(X_i - md\theta_i\right)^2}{md\theta_i} = \frac{n-1}{n-m}\left(\frac{1}{md}\sum_{i=1}^{n}\frac{X_i^{\,2}}{\theta_i} - md\right) \dot{\sim} \chi^2(n-1) \tag{51}$$

but the presence of the non-randomness parameters θ_i in this expression prevents our expressing this statistic as a simple function of X, as defined in Equation (49). Consequently, we cannot readily determine $f(x|\boldsymbol{\theta})$ and so are equally unable to find an expression for the continuous equivalent of $P(D|R')$ as required for Equations (12) and (13). Unfortunately, we still need to determine the multivariate distribution of the random vector $\mathbf{x}$ in order to determine the distribution of the test statistic in Equation (49) under the assumption of non-randomness. Consequently, there appears to be no benefit in dealing with scalar summaries of $\mathbf{x}$.

In order to make any progress, we note that for the special case when $m = 1$, the joint distribution of $\mathbf{x}$ given $\boldsymbol{\theta}$ under the assumption of non-randomness is multinomial, with probability mass function

$$p(\mathbf{x}|\boldsymbol{\theta}) = d!\prod_{i=1}^{n}\frac{\theta_i^{\,x_i}}{x_i!} \; ; \quad \mathbf{x} = \{0,1,\ldots,d\}^n : \sum_{i=1}^{n} x_i = d \, . \tag{52}$$

If we assume randomness, then $\theta_i = \frac{1}{n}$ for $i = 1,2,\ldots,n$ and this multinomial probability mass function simplifies to give

$$p(\mathbf{x}) = \frac{d!}{n^d \prod_{i=1}^{n} x_i!} \; ; \quad \mathbf{x} = \{0,1,\ldots,d\}^n : \sum_{i=1}^{n} x_i = d \,. \tag{53}$$

As these distributions are easy to manipulate, we reconsider our criterion for randomness when testing for equality of frequencies of the numbers drawn. Rather than simultaneously testing all of the numbers selected across all draws, we introduce m separate test criteria based on the orders of the balls drawn. Although these tests are dependent, multiplicity of testing is no longer a concern because it is meaningful to evaluate an overall posterior probability of randomness. Such calculations are not appropriate for p-values.

Define $\mathbf{y}_k = \{Y_{k1}, Y_{k2}, \ldots, Y_{kn}\}$ where $Y_{ki} = 0,1,\ldots,d$ is the number of times that ball i is drawn in position k over the d draws for $i = 1,2,\ldots,n$ and $k = 1,2,\ldots,m$ subject to

$$\sum_{i=1}^{n} Y_{ki} = d \,. \tag{54}$$

For this "m from n" lottery, the relevant draw history can be partitioned into subsets that consist of the actual ball frequencies observed in each of the m draw positions over the d draws. Suppose the actual ball frequencies for draw position k are contained in the data set $E_k = \{y_{k1}, y_{k2}, \ldots, y_{kn}\}$. From Equations (52) and (53), it follows that

$$P(E_k | R) = \frac{d!}{n^d \prod_{i=1}^{n} y_{ki}!} \tag{55}$$

under the assumption of randomness and

$$p(\mathbf{y}_k | \boldsymbol{\theta}) = d! \prod_{i=1}^{n} \frac{\theta_i^{\,y_{ki}}}{y_{ki}!} \; ; \quad \mathbf{y}_k = \{0,1,\ldots,d\}^n : \sum_{i=1}^{n} y_{ki} = d \tag{56}$$

under the assumption of non-randomness.

As in Subsection 3.1, we might set the prior probability of randomness as $P(R) = \frac{1}{2}$ and use a subjective distribution, such as the natural conjugate prior, or an objective distribution, such as Jeffreys' invariant prior, on the parameter vector $\boldsymbol{\theta}$ if the draws are not random. Under this assumption of non-randomness, the natural conjugate prior is of the Dirichlet form

$$g(\boldsymbol{\theta}) = \frac{\Gamma\left(\sum_{i=1}^{n} a_i\right)}{\prod_{i=1}^{n} \Gamma(a_i)} \prod_{i=1}^{n} \theta_i^{\,a_i - 1} \; ; \quad \boldsymbol{\theta} \in (0,1)^n : \sum_{i=1}^{n} \theta_i = 1 \tag{57}$$

and the invariant prior takes the form

$$g(\boldsymbol{\theta}) = \frac{\Gamma\left(\frac{n}{2}\right)}{\pi^{n/2}} \prod_{i=1}^{n} \theta_i^{-1/2} \; ; \quad \boldsymbol{\theta} \in (0,1)^n : \sum_{i=1}^{n} \theta_i = 1 . \tag{58}$$

As recommended in Subsection 2.3, we prefer to use the subjective prior.

Based on the result in Equation (15), the natural conjugate prior in Equation (57) and the multinomial sampling distribution in Equation (56), the prior predictive probability mass function for $\mathbf{y}_k$ under the assumption of non-random drawings is

$$\begin{aligned} p(\mathbf{y}_k) &= \int_{-\infty}^{\infty} p(\mathbf{y}_k|\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta} = \frac{d!\Gamma\left(\sum_{i=1}^{n} a_i\right)}{\prod_{i=1}^{n} \{y_{ki}!\Gamma(a_i)\}} \int_{\mathbf{0}}^{\mathbf{1}} \prod_{i=1}^{n} \theta_i^{y_{ki}+a_i-1} d\boldsymbol{\theta} \\ &= \frac{d!\Gamma\left(\sum_{i=1}^{n} a_i\right)}{\Gamma\left(d + \sum_{i=1}^{n} a_i\right)} \prod_{i=1}^{n} \frac{\Gamma(y_{ki} + a_i)}{y_{ki}!\Gamma(a_i)} ; \quad \mathbf{y}_k = \{0,1,\ldots,d\}^n : \sum_{i=1}^{n} y_{ki} = d \end{aligned} \tag{59}$$

by comparing the integrand with the probability density function of a Dirichlet distribution and after some simplification. This corresponds to a multinomial-Dirichlet distribution; see Bernardo and Smith (1993). It enables us to evaluate the probability of the observed draw history for position k under the assumption of non-random drawings as

$$P(E_k|R') = \frac{d!\Gamma\left(\sum_{i=1}^{n} a_i\right)}{\Gamma\left(d + \sum_{i=1}^{n} a_i\right)} \prod_{i=1}^{n} \frac{\Gamma(y_{ki} + a_i)}{y_{ki}!\Gamma(a_i)} \tag{60}$$

for draw positions $k = 1,2,\ldots,m$. These probabilities and their counterparts in Equation (55) are analogous to, and extensions of, Equations (23) and (26) for the binomial coin tossing experiment where $n = 2$.

We are now able to apply Bayes' theorem to calculate the posterior probabilities that the draws are random from Equations (55) and (60), at least in terms of the relative frequencies of the balls drawn in each of the m ordered selections. With prior probability of randomness equal to one half, we obtain the posterior odds in favour of randomness

$$\frac{P(R|E_k)}{P(R'|E_k)} = \frac{\Gamma\left(d+\sum_{i=1}^{n} a_i\right)}{n^d \Gamma\left(\sum_{i=1}^{n} a_i\right)} \prod_{i=1}^{n} \frac{\Gamma(a_i)}{\Gamma(y_{ki}+a_i)} \tag{61}$$

from Equation (13) and then the posterior probability of randomness for each draw position can be determined from Equations (14).

Although these tests are dependent, multiplicity is not a problem with this Bayesian analysis because we can calculate the overall posterior odds of randomness by Bayesian updating of the individual posterior odds, which is not possible for frequentist p-values. This allows us to calculate a reasonable measure of overall posterior probability of randomness, based on the equality of frequencies of balls drawn. To do this, we make the simplifying assumption that $E_1, E_2, \ldots, E_m$ are mutually independent given either R or R'. Although it is only an approximation, it is a reasonable assumption to make for this analysis. From Equation (13), we then have

$$\frac{P(R|E_1,E_2,\ldots,E_m)}{P(R'|E_1,E_2,\ldots,E_m)} = \frac{P(E_1,E_2,\ldots,E_m|R)P(R)}{P(E_1,E_2,\ldots,E_m|R')P(R')} = \frac{P(R)}{P(R')}\prod_{k=1}^{m}\frac{P(E_k|R)}{P(E_k|R')} \tag{62}$$

so, with prior odds of one as above, an approximate value for the overall posterior odds of randomness is calculated as the product of the marginal odds determined by Equation (61). Finally, we can convert this into an approximate overall posterior probability of randomness based on equality of frequencies using Equations (14).

When evaluating the expressions in Equations (61), (62) and similar formulae, the numerical computation is much easier if we calculate the natural logarithm of the odds and exponentiate the answer than it is if we attempt to calculate the odds directly. Another useful computational tip for large, non-integer x is that

$$\Gamma(x) = \Gamma(x - \lfloor x \rfloor)\prod_{i=1}^{\lfloor x \rfloor}(x-i) \tag{63}$$

where $\lfloor x \rfloor$ is the integer part of x. Combined with the use of logarithms and exponentials, this enables us to evaluate expressions involving gamma functions of large numbers.

It remains to propose suitable values for the hyperparameters $a_1, a_2, \ldots, a_n$. As in Subsection 3.1, symmetry of our prior beliefs about the parameters $\theta_1, \theta_2, \ldots, \theta_n$ suggests that we should set the hyperparameters to be equal. If $a_i = a$ for $i = 1,2,\ldots,n$ then the expression for the posterior odds in Equation (61) simplifies to give

$$\frac{P(R|E_k)}{P(R'|E_k)} = \frac{\Gamma(d+na)\{\Gamma(a)\}^n}{n^d \Gamma(na)\prod_{i=1}^{n}\Gamma(y_{ki}+a)} = \frac{\{\Gamma(a)\}^n \prod_{i=1}^{d}\left(\frac{d-i}{n}+a\right)}{\prod_{i=1}^{n}\Gamma(y_{ki}+a)} \tag{64}$$

and we only need to determine a suitable value for a. In Subsection 3.1, there were only one parameter and one hyperparameter, so we only required a single prior elicitation. Here, we have $n-1$ parameters and one hyperparameter, so we can simplify the multivariate prior distribution by considering only the marginal distribution of any one of its parameters, such as θ_1. As Bernardo and Smith (1993) point out, the marginal distribution of θ_1 based on the joint distribution in Equation (57) is a beta distribution

$$\theta_1 \sim Be\left(a_1, \sum_{i=2}^{n} a_i\right), \tag{65}$$

which simplifies to a $Be\{a,(n-1)a\}$ distribution under the symmetry assumption above Equation (64).

As defined by Equation (44), the parameter θ_1 represents the relative frequency that ball 1 is selected when one ball is drawn from the full set of n balls and is thus equal to $\frac{1}{n}$ under the hypothesis of randomness. In a similar manner as for the coin example in Subsection 3.1, we might be willing to express a suitable fractile for θ_1 to express our prior beliefs. Previously, we used the constrained lower and upper quartiles, Q_1 and $Q_3 = 1 - Q_1$ respectively, because the median, Q_2, was necessarily $\frac{1}{2}$ due to symmetry. However, the symmetry here does not fix the median of θ_1 even though it constrains the mean to be $\frac{1}{n}$ and the median to be less than $\frac{1}{n}$ because the prior has positive skew.

A slightly modified procedure is useful here. Suppose we elicit unspecified lower and upper fractiles, L and U, such that our prior beliefs satisfy

$$P\{\theta_1 \in (L,U)\} = P\{\theta_1 \notin (L,U)\} = \tfrac{1}{2}. \tag{66}$$

Note that L and U do not represent the lower and upper quartiles in general. We can then solve the nonlinear equation

$$\frac{1}{\mathrm{B}\{a,(n-1)a\}} \int_L^U \theta^{a-1}(1-\theta)^{(n-1)a-1}\, d\theta = \frac{1}{2} \tag{67}$$

numerically to determine an appropriate value for the hyperparameter a. Reasonable values for these fractiles might take the form $L = (1-q)\frac{1}{n}$ and $U = (1+q)\frac{1}{n}$, corresponding to $100q\%$ either side of the true value of $\frac{1}{n}$ under the hypothesis of randomness. For example,

$q = 0 \cdot 1$ suggests that we regard it equally likely that the probability of ball 1's being selected lies inside and outside 10% of this ideal value.

4.2 Independence of Consecutive Draws

Haigh (1997) proposed a method to test for dependence between successive drawings in a "m from n" lottery, based on analysing the gaps between repeated appearances of particular numbers. For any of the numbers $k = 1,2,\ldots,n$ redefine $\mathbf{y}_k = \{Y_{k1}, Y_{k2}, Y_{k3}, \ldots\}$ where Y_{ki} is the number of draws between appearances $i-1$ and i of the number k for $i = 1,2,3,\ldots$ and appearance 0 represents the start of data collection. Under the assumption of randomness (serial independence) for any fixed k, these gaps are independent geometric random variables, each with a probability mass function of the form

$$p(y_k) = \left(\frac{n-m}{n}\right)^{y_k - 1} \frac{m}{n}; \quad y_k = 1,2,3,\ldots \,. \tag{68}$$

This leads to k dependent chi-square goodness-of-fit tests, based on the observed gap frequencies for each of the ball numbers over the course of d draws. As in Subsection 4.1, and for equivalent reasons, we deal with the joint distribution of the gaps for each fixed k rather than the summary chi-square statistic. From Equation (68) and allowing for the final right-censored gap observation, the probability of the observed draw history of gaps for ball k under the assumption of random drawings becomes

$$P(E_k|R) = \prod_{i=1}^{g_k} \left\{ \left(\frac{n-m}{n}\right)^{y_{ki}-1} \frac{m}{n} \right\} \left(\frac{n-m}{n}\right)^{d - \sum_{i=1}^{g_k} y_{ki}} = \left(\frac{n-m}{n}\right)^{d-g_k} \left(\frac{m}{n}\right)^{g_k} \tag{69}$$

where a total of g_k gaps are observed for ball k in the d draws, which is equivalent to defining g_k as the number of times that ball k is selected in the d draws.

Under the assumption of non-randomness (serial dependence) and for any fixed k, the gaps are conditionally independent geometric random variables with probability mass function

$$p(y_k|\phi_k) = (1-\phi_k)^{y_k - 1} \phi_k \,; \quad y_k = 1,2,3,\ldots \tag{70}$$

where ϕ_k is the probability that ball k is selected in a particular draw, as defined in and below Equation (48). For this latter case, the natural conjugate prior distribution has a beta form, with probability density function

$$g(\phi_k) = \frac{1}{\mathrm{B}(a,b)} \phi_k^{\,a-1} (1-\phi_k)^{b-1} \,; \quad 0 < \phi_k < 1 \tag{71}$$

in terms of hyperparameters a and b to be specified, and Jeffreys' invariant prior distribution for ϕ_k has probability density function

$$g(\phi_k) \propto \frac{1}{\phi_k \sqrt{1-\phi_k}}\,; \quad 0 < \phi_k < 1\,. \tag{72}$$

As in Subsection 4.1, we adopt the subjective prior in Equation (71) in order to reflect the true state of our knowledge about ϕ_k and to avoid problems that can arise when using the objective prior.

We now consider how to specify the hyperparameters a and b for this analysis. As ϕ_k is the probability that ball k is selected in a particular draw, it is equal to $\frac{m}{n}$ under the hypothesis of randomness and so we might reasonably choose to set the prior mean $E(\phi_k)$ equal to this value, in which case $b = \frac{n-m}{m} a$. Following the procedure adopted in Subsection 4.1, we next elicit lower and upper fractiles of the form $L = (1-q)\frac{m}{n}$ and $U = (1+q)\frac{m}{n}$ respectively, such that

$$P\{\phi_k \in (L,U)\} = P\{\phi_k \notin (L,U)\} = \tfrac{1}{2}\,, \tag{73}$$

corresponding to $100q\%$ either side of the true value of $\frac{m}{n}$ under the hypothesis of randomness. We then solve the nonlinear equation

$$\frac{1}{\mathrm{B}(a,b)} \int_L^U \theta^{a-1} (1-\theta)^{b-1}\, d\theta = \frac{1}{2} \tag{74}$$

numerically to determine an appropriate value for the hyperparameter a, subject to $b = \frac{n-m}{m} a$.

Although the Y_k are conditionally independent given ϕ_k in Equation (70), they are not necessarily marginally independent. Consequently, we need to determine the joint prior predictive distribution of $\mathbf{y}_k$ rather than its constituent parts. Using the result in Equation (15), the natural conjugate prior in Equation (71) and the geometric sampling distribution in Equation (70), the probability of the observed draw history of gaps for ball k under the assumption of non-random drawings over the course of d draws becomes

$$\begin{aligned} P(E_k|R') &= \int_{-\infty}^{\infty} p(\mathbf{y}_k|\phi_k) g(\phi_k) d\phi_k \\ &= \int_0^1 \frac{\phi_k^{\,g_k+a-1} (1-\phi_k)^{d-g_k+b-1}}{\mathrm{B}(a,b)} d\phi_k = \frac{\mathrm{B}(g_k + a, d - g_k + b)}{\mathrm{B}(a,b)} \end{aligned} \tag{75}$$

by comparing the integrand with the probability density function of a beta distribution, where g_k is the number of times that ball k is selected in the d draws.

We are now able to apply Bayes' theorem to calculate the posterior probabilities that the draws are random from Equations (69) and (75), at least in terms of serial independence based on gaps between successive occurrences of the numbers drawn, for each of the n possible numbers. With prior probability of randomness equal to one half, we obtain the posterior odds in favour of randomness

$$\frac{P(R|E_k)}{P(R'|E_k)} = \frac{\left(\frac{n-m}{n}\right)^{d-g_k}\left(\frac{m}{n}\right)^{g_k} \mathrm{B}(a,b)}{\mathrm{B}(g_k + a, d - g_k + b)} \tag{76}$$

from Equation (13) and then the posterior probability of randomness can be calculated from Equations (14). It is interesting to note that g_k is a sufficient statistic for this analysis: we do not need to know the actual gap sizes for each ball number, just how many times each was selected over the available draw history. This suggests that the tests for equality of frequencies in the previous subsection are very similar to those for serial independence in this subsection.

As in Subsection 4.1, these tests are dependent but multiplicity is not a problem with this Bayesian analysis because we can calculate the overall posterior odds of randomness by Bayesian updating of the individual posterior odds, which is not possible for frequentist p-values. This allows us to calculate a reasonable measure of overall posterior probability of randomness, based on the independence of consecutive draws. To do this, we make the simplifying assumption that $E_1, E_2, \ldots, E_n$ are mutually independent given either R or R'. Although it is only an approximation, it is a reasonable assumption to make for this analysis. By similar construction to that in Subsection 4.1, and with prior odds of one again, an approximate value for the overall posterior odds of randomness is given by

$$\frac{P(R|E_1, E_2, \ldots, E_n)}{P(R'|E_1, E_2, \ldots, E_n)} = \frac{P(R)}{P(R')} \prod_{k=1}^{n} \frac{P(E_k|R)}{P(E_k|R')} \tag{77}$$

Finally, we can convert this into an approximate overall posterior probability of randomness based on serial independence using Equations (14).

4.3 Other Tests for Lottery Randomness

Haigh (1997) and the University of Salford (2004-2005) applied several other significance tests for randomness to data from the U.K. National Lottery. Two of these model the mean and standard deviation of the sums of numbers based on all observed draws and a third models the parity of the numbers drawn. These all generate summary statistics of the observed draw history, which we considered to be difficult and unnecessary in Subsection 4.1. Other tests commonly used for "m from n" lotteries are concerned with whether there is any

evidence of non-randomness in any combinations of ball sets and draw machines. These are specific to the particular lottery operating procedures and are merely sub-analyses of the preceding methods.

Yet another test investigates whether the frequency of jackpot winners is as expected. This can be tested by noting that if there are k entries in any draw and the probability of a jackpot for any entry is π, then the number of jackpot winners in any given draw, Y, has a binomial distribution $Y|\pi \sim Bi(k,\pi)$ where

$$\pi = \frac{m!(n-m)!}{n!} \qquad (78)$$

under the assumption of randomness. Using a natural conjugate beta $Be(a,b)$ prior for π under the assumption of non-randomness, we can develop a similar Bayesian analysis to those in Subsections 4.1 and 4.2. As noted by the above authors, conscious selection among lottery players should be accounted for. This involves supplementing our knowledge of the numbers of entries for all draws by observing the coverage rates, which are the percentages of all combinations selected for all draws, in order to modify the value of π accordingly; such information is not generally available for the U.K. National Lottery.

5. Bayesian Analysis of U.K. National Lottery

We now demonstrate an application of the Bayesian analyses of Subsections 4.1 and 4.2 to the complete draw history of the "6 from 49" U.K. National Lottery as at 24th June 2006 for comparison with the current frequentist analyses in use. At the time of writing, the full Lotto draw history is readily available for downloading from the National Lottery website and includes information on the sequential numbers drawn for the first 1,101 draws. We used Minitab and Mathcad for numerical computation.

5.1 Equality of Frequencies of the Numbers Drawn

In order to apply the m tests, one for each draw position, relating to equality of frequencies of the 49 possible numbers, we presented the odds in favour of randomness in Equation (61). Setting $a_i = a$ and $q = 0\cdot 1$ as proposed in Subsection 4.1 leads us to adopt $Be(44\cdot 49, 2{,}135)$ prior distributions for the parameters θ_i, for $i = 1,2,\ldots,n$. We then determine the probabilities of randomness given the available draw history from Equations (14). Table 3 shows the posterior odds and probability of randomness (equality of frequencies) based on the six draw positions.

All of these tests strongly support the hypothesis of random drawings, with posterior probabilities of randomness well in excess of one half. Perhaps more importantly, the approximate overall posterior odds of randomness according to Equation (62) is the product of these marginal odds, which is $1\cdot 4\times 10^{6}$. This corresponds to an approximate overall posterior probability of randomness with respect to equalities of frequencies of $1\cdot 0000$ to

four decimal places from Equations (14). Clearly, this draw history strongly supports the hypothesis of random drawings.

Draw Position	Odds	Probability
1	$16 \cdot 15$	$0 \cdot 9417$
2	$13 \cdot 93$	$0 \cdot 9330$
3	$9 \cdot 828$	$0 \cdot 9076$
4	$14 \cdot 24$	$0 \cdot 9344$
5	$3 \cdot 131$	$0 \cdot 7579$
6	$14 \cdot 47$	$0 \cdot 9354$

Table 3: posterior odds and probabilities of randomness based on equality of frequencies.

For comparison, the frequentist equivalent is the modified chi-square goodness-of-fit test in Equation (49), which gives a test statistic of $X = 60 \cdot 46$ on 48 degrees of freedom and a p-value of $0 \cdot 1071$. This implies that if the draws were random, there would be an 11% chance of observing data at least as extreme as those actually observed. It does not comment on the chance of observing these data if the draws were not random; nor does it take account of our prior belief of randomness or generate a probability measure of randomness as output. For these reasons, we believe that the Bayesian output in Table 3 is more informative for routine monitoring and policy planning.

Having earlier attempted to justify the use of simple hypotheses of randomness and otherwise adopted standard Bayesian methodology, our only possible concern with this innovative approach is in determining the hyperparameter value a by eliciting prior fractiles from a marginal prior as described around Equations (66) and (67). Although the author is content with this approach, critics might require further evidence of its suitability. We provide this by means of a sensitivity analysis displayed in Figure 4, which presents a graph of the posterior probability of randomness for the first draw position, $P(R|E_1)$ defined by Equations (64) and (14), as a function of the prior fractiles elicitation factor, q defined below Equation (67).

These data conform so readily to the hypothesis of randomness that the posterior probability of randomness exceeds one half for all possible values of the prior fractiles elicitation factor. A sceptic might suggest that this is a consequence of assigning a prior probability of one half to the simple hypothesis of randomness. However, this is not so. If we were artificially to modify the data slightly so that balls 1 and 2 did not occur 29 and 21 times but rather 0 and 50 times, respectively, the posterior probability of randomness with $q = 0 \cdot 1$ as before would become $P(R|E_1) = 0 \cdot 0019$, rather than $P(R|E_1) = 0 \cdot 9417$. The data actually observed really do support the hypothesis of random drawings!

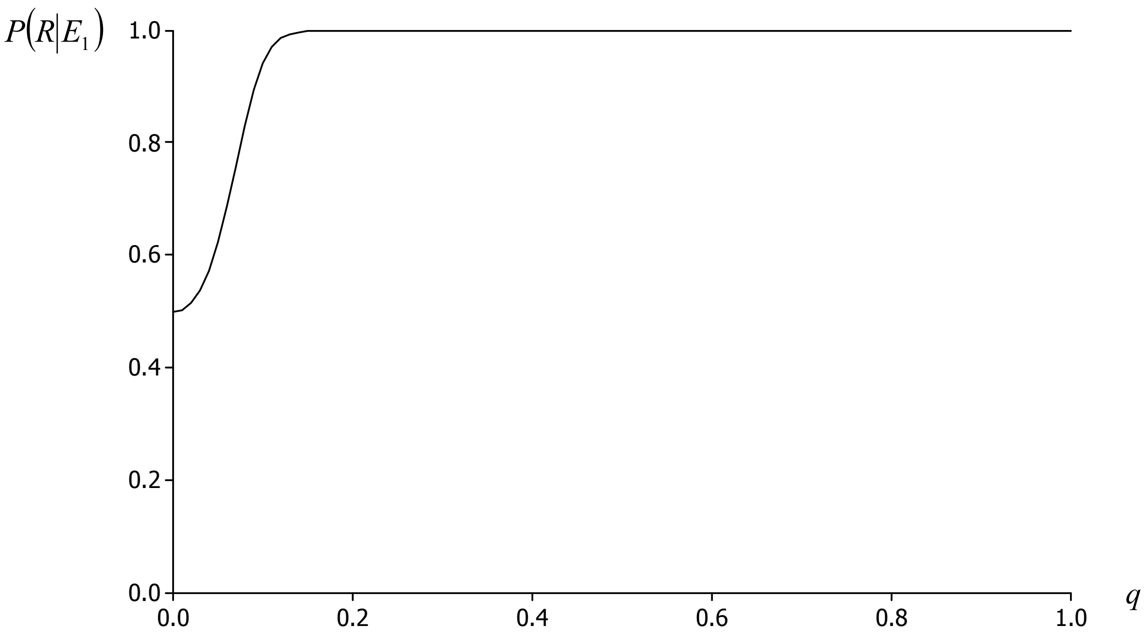

Figure 4: graph of probability of randomness $P(R|E_1)$ against elicitation factor q.

5.2 Independence of Consecutive Draws

In order to apply the n tests, one for each numbered ball, relating to serial independence across all draws, we presented an expression for the odds in favour of randomness in Equation (76). For calculation purposes, it is again helpful to evaluate the log odds and then exponentiate this number to determine the odds. This enables us to calculate the probability of randomness given the available draw history by Equations (14). We determine the hyperparameters for the prior distribution of the probability that ball k is selected in a particular draw using the method described around Equations (73) and (74). Applying numerical quadrature and a nonlinear equation solver in Mathcad, with $q = 0 \cdot 1$ as in Subsection 5.1, we obtain $a = 39 \cdot 90$ and $b = 285 \cdot 9$.

Table 4 shows the resulting posterior odds and probabilities of randomness (serial independence) based on the forty-nine ball numbers. None of these tests casts serious doubt about the hypothesis of random drawings, except the test relating to ball 20. This ball arose only 102 times in the 1,101 draws, compared with the expected frequency of $134 \cdot 8$, and was the largest discrepancy. Perhaps more importantly, the approximate overall posterior odds of randomness according to Equation (77) are the product of these marginal odds, which is $2 \cdot 0 \times 10^5$. This corresponds to an approximate overall posterior probability of randomness with respect to serial independence of $1 \cdot 0000$ to four decimal places from Equations (14). Clearly, this draw history strongly supports the hypothesis of random drawings.

Ball Number	Posterior Odds	Posterior Prob.	P-value	Ball Number	Posterior Odds	Posterior Prob.	P-value
1	2.09	0.6761	0.1024	26	2.06	0.6737	0.9027
2	2.10	0.6771	0.5403	27	2.00	0.6668	0.9257
3	1.85	0.6489	0.5527	28	2.04	0.6715	0.5401
4	1.58	0.6131	0.9003	29	2.03	0.6697	0.6184
5	1.77	0.6387	0.7421	30	1.63	0.6203	0.4334
6	1.95	0.6605	0.0609	31	1.63	0.6203	0.4032
7	1.63	0.6203	0.3575	32	1.72	0.6328	0.4572
8	1.77	0.6387	0.8194	33	1.72	0.6328	0.0993
9	1.72	0.6328	0.1024	34	1.58	0.6131	0.7484
10	1.95	0.6605	0.2727	35	2.00	0.6668	0.9308
11	1.24	0.5538	0.5102	36	1.68	0.6268	0.7170
12	2.10	0.6771	0.6081	37	1.17	0.5398	0.9569
13	0.39	0.2830	0.1032	38	0.26	0.2063	0.0865
14	1.92	0.6574	0.8799	39	1.38	0.5803	0.5763
15	1.28	0.5610	0.8939	40	1.54	0.6063	0.2203
16	0.61	0.3771	0.6641	41	0.20	0.1676	0.2202
17	1.77	0.6387	0.6240	42	2.03	0.6697	0.4698
18	1.98	0.6643	0.9821	43	0.77	0.4340	0.8144
19	1.95	0.6605	0.8769	44	0.53	0.3483	0.2486
20	0.05	0.0442	0.0344	45	2.04	0.6715	0.8556
21	0.87	0.4653	0.4051	46	1.98	0.6643	0.7892
22	2.09	0.6761	0.3842	47	0.68	0.4062	0.2990
23	1.04	0.5104	0.3613	48	1.14	0.5330	0.2450
24	1.38	0.5803	0.6810	49	1.98	0.6643	0.5992
25	0.77	0.4340	0.4570				

Table 4: posterior odds and probabilities of randomness based on serial independence, with frequentist p-values for comparison.

For comparison, the frequentist equivalents are the chi-square goodness-of-fit tests described below Equation (68). Based on the available data, reasonable categories for frequencies of gaps are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-15, 16-20 and >20. As the geometric parameter is known and there are thirteen categories, each of these tests has twelve degrees of freedom. Table 3 also presents the p-values obtained by applying these tests to the same data that were analysed to obtain the posterior odds and probabilities of randomness. These p-values represent the chances of observing data at least as extreme as those actually observed if the draws were random. As before, these significance tests do not comment on the chance of observing these data if the draws were not random, nor do they take account of our prior beliefs of randomness. Although there appears to be little similarity between the probabilities and p-values in Table 4, Spearman's rank correlation coefficient between these variables is $0 \cdot 335$. This is positive, as expected, and significantly non-zero at the 5% level.

6. Conclusions

Lottery randomness testing is an important aspect of monitoring and checking that draws are conducted fairly. Current practice generally involves applying a battery of significance tests at infrequent time intervals and we argue that a Bayesian approach would provide a natural analysis for this scenario. Although the standard procedures are effective, they raise many questions. By considering the events randomness and non-randomness, we develop reasonably objective Bayesian methods for calculating posterior probabilities of randomness, based on equality of frequencies for all ball numbers in each draw position and serial independence for all ball numbers across all draws.

We apply these methods to the draw history of the U.K. National Lottery as at 24th June 2006 and compare the results with those obtained using existing frequentist methods. Rather than concluding that there is insufficient evidence to reject the hypothesis of randomness, we are now able to calculate the probabilities that the draws were indeed random. For each of the two criteria considered, the posterior probabilities of randomness clearly indicate that the evidence supports the hypothesis of random drawings.

Our theoretical and practical analyses enable us to conclude that Bayesian updating is easier to interpret and more informative than significance testing. It removes the arbitrariness of size and power and avoids the multiplicity associated with sequential and cross-sectional testing. By choosing reasonably objective priors as described in the text, this analysis could become a standard reference approach when testing lotteries for randomness and supported by descriptive statistics and graphical displays.

Future research with this theme could investigate the influence of dependence between Bayesian tests for randomness, particularly the m tests in Subsection 4.1 and the n tests in Subsection 4.2. One might also consider the development and implementation of new test criteria, such as those suggested in Subsection 4.3, possibly including multivariate response distributions to model the above dependence. Another profitable exercise might involve the analysis of simulated lottery draws to assess the sensitivity of results to hyperparameter specifications.

Acknowledgements

The author is grateful to Dr. D. K. Forrest (University of Salford) and Prof. P. Grindrod (Quantisci Ltd.) for helpful suggestions that inspired this research.

References

Beenstock, M. and Haitovsky, Y. (2001) Lottomania and other anomalies in the market for lotto. *Journal of Economic Psychology*, 22, 721-744.

Bernardo, J.M. and Smith, A.F.M. (1993) Bayesian Theory. New York: Wiley.

Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *J. R. Statist. Soc.* B, 57, 45-97.
Haigh, J. (1997) The statistics of the National Lottery. *J. R. Statist. Soc.* A, 160, 187-206.
Jeffreys (1998) Theory of Probability. Oxford: University Press.
Joe, H. (1993) Tests of uniformity for sets of lotto numbers. *Statistics & Probability Letters*, 16, 181-188.
Koning, R.H. and Vermaat, M.B. (2002) A probabilistic analysis of the Dutch lotto. *University of Groningen research report*.
Lindley, D.V. (1957) A statistical paradox. *Biometrika*, 44, 187-192.
O'Hagan A. (2004) Bayesian Inference. London: Edward Arnold.
Percy, D.F. (2003) Subjective reliability analysis using predictive elicitation. In *Mathematical and Statistical Methods in Reliability* (eds. Lindqvist, B.H. and Doksum, K.A.), Singapore: World Scientific, 57-72.
Stern, H. and Cover, T.M. (1989) Maximum entropy and the lottery. *J. Am. Statist. Ass.*, 84, 980-985.
University of Salford (2004-2005) Reports on the randomness of U.K. lottery games. *National Lottery Commission website*.