

Is a Binomial Process Bayesian?

Robert L. Andrews, Virginia Commonwealth University
Department of Management, Richmond, VA. 23284-4000
804-828-7101, rlandrew@vcu.edu

Jonathan A. Andrews, United States Navy
Dahlgren, VA 22448
jonathan.a.andrews@navy.mil

Steve Custer, Virginia Commonwealth University
Department of Management, Richmond, VA. 23284-4000
804-828-7170, swcuster@vcu.edu

ABSTRACT

This paper discusses whether a binomial process for a dichotomous variable with π as the probability of success can correctly be modeled as a Bayesian process. The question of interest is whether the value of π remains fixed for the phenomenon being observed or whether the value of π actually varies and has its own probability distribution. If the later is the case then the process can be modeled mathematically as a Bayesian process with a prior distribution for the probability of success a binomial conditional distribution. The paper considers two example situations where Bayesian modeling could be applied. One is shooting free throws in a basketball game and the other is shooting a missile at a military target. Graphical and ad hoc testing methods are proposed and tested using the basketball example. These methods were not able to support the modeling of free throw shooting with a Bayesian model.

INTRODUCTION AND OVERVIEW

The primary focus will be on a dichotomous variable for which there are two possible observed outcomes that can be modeled with a binomial distribution using π as the probability of success. In many such situations one can present a credible rationale to state that the probability of success can vary and as such has a probability distribution. One example we will examine will be shooting of free throws in a basketball game. The popular phrase “when you are hot you are hot and when you are not you are not” supports the concept of a varying probability of success. Another example would be shooting a missile at a military target. In this case one can also rationalize that there are forces which vary from situation to situation so that the probability of hitting the target would vary and qualify this situation for Bayesian modeling. However, for something to be worthwhile one must show that going through the extra effort to do calculations based on a Bayesian model actually adds value to a decision making process. Hence this paper addresses identifying circumstances for which knowing that a process is Bayesian would be of value. It also addresses how one could use actual data from a process to determine if there is evidence that the process is Bayesian. The methodology used in this paper can be used to address numerous processes but we will focus on the two shooting examples.

In a Bayesian process there is an observable variable denoted by X . The probability distribution for X denoted by $f(X|\theta)$ depends on one or more parameters with one of the parameters being

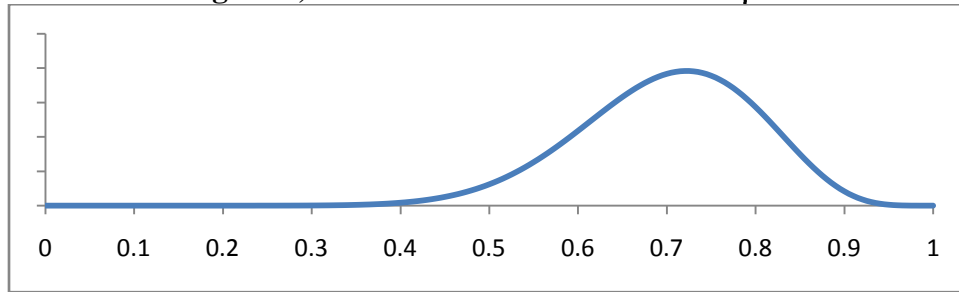
θ . Hence the value of X is conditional on the value of θ . For a Bayesian situation, θ has a probability distribution denoted by $g(\theta)$ and is referred to as the prior distribution, because this is the distribution of θ prior to obtaining any knowledge from an observable X . The joint probability distribution for X and θ is $f(X,\theta) = f(X|\theta) * g(\theta)$. In similar fashion, the joint probability distribution for X and θ can be expressed as $f(X,\theta) = f(\theta|X) * f(X)$. Using this expression then $f(\theta|X) = f(X,\theta) / f(X)$, and using the first expression to replace $f(X,\theta)$ one gets $f(\theta|X) = f(X|\theta) * g(\theta) / f(X)$. This distribution, $f(\theta|X)$, is referred to as the posterior distribution because it is the distribution of θ after or posterior to observing a value of X . If g is a continuous variable then $f(X) = \int f(X|\theta) * g(\theta) d\theta$.

If the conditional distribution $f(X|\theta)$ is some known probability distribution then one would like to find a prior probability distribution so that the posterior distribution is also some known distribution. Such a prior distribution is referred to as a conjugate prior distribution. For example, if the variable X is a continuous variable and follows a normal distribution with mean denoted by μ_c and standard deviation denoted by σ_c with the value of μ_c being the parameter that has a prior probability distribution then the form of the conjugate prior is also a normal distribution. If the prior mean is μ_p and standard deviation is σ_p , then the posterior distribution mean is $(\sigma_p^2 * X + \sigma_c^2 * \mu_p) / (\sigma_p^2 + \sigma_c^2)$ and the variance is $(\sigma_p^2 * \sigma_c^2) / (\sigma_p^2 + \sigma_c^2)$. The standard deviation is the square root of the variance for this normal posterior distribution.

If the conditional distribution $f(X|\theta)$ is the binomial distribution then the parameters are the number of trials denoted by n and the probability of success on a single trial denoted by π . Since X is a discrete variable taking on integer values then one can directly calculate the probability of a specific integer value of X and the distribution will be denoted using a P rather than an f . For the binomial, $P(X|\pi, n) = n! / [X! * (n - X)!] * \pi^X * (1 - \pi)^{(n-X)}$. π will be assumed to be constant for a set of n trials but will be subject to varying from one set of trials to another set of trials.

If the beta distribution is used as a prior for the binomial distribution then mathematically it can be shown that the posterior distribution is also a beta distribution. Hence the beta is a conjugate prior for a binomial distribution. Values for the beta distribution vary over the range from 0 to 1 and the parameters are denoted by α and β , which must both be positive. The beta can take on a variety of shapes over the range of 0 to 1. For $\alpha=\beta$, the beta distribution is symmetric and if $\alpha=\beta=1$ then the distribution is a continuous uniform from 0 to 1. For α and β both less than 1, the distribution is U-shaped. For either α or β less than 1 and the other greater than 1, the distribution is strictly decreasing ($\alpha < 1$) or strictly increasing ($\beta < 1$). For α & β both greater than 1, the distribution is unimodal with a peak between 0 and 1. In Excel, one can easily find beta probabilities with the BETADIST function or beta quantile values with the BETAINV function. These characteristics make the beta a reasonable probability distribution to use for π . Figure 1 below shows a beta distribution with $\alpha=14$ and $\beta=6$. For this beta distribution, the mean is .70, mode is .72, standard deviation is .10 and skewness is -.36.

Figure 1, Beta Distribution with $\alpha=14$ & $\beta=6$



The mean of the beta distribution is $\alpha/(\alpha+\beta)$ and the variance is $\alpha*\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$.

The coefficient of skewness for the beta distribution is $\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha*\beta}}$. From this expression for skewness one can see that the skewness measure for the beta distribution is zero when $\alpha=\beta$, which indicates that the distribution is symmetrical with mean = .5. If the mean of the beta is greater than .5, $\alpha > \beta$, then the distribution is skewed left and correspondingly if the mean is less than .5, $\alpha < \beta$, then the distribution is skewed right. If the prior distribution is beta with parameters α and β and if x successes have been observed in n trials for a binomial variable then the posterior distribution will be a beta distribution with parameters $\alpha+x$ and $\beta+(n-x)$. Hence the mean of the posterior distribution is $(\alpha+x)/(\alpha+\beta+n)$ and the variance of the posterior distribution is $\frac{(\alpha+x)(\beta+n-x)}{[(\alpha+\beta+n)^2(\alpha+\beta+n+1)]}$.

The value of the posterior mean ends up being a weighted average of the mean of the prior distribution and the observed value from the conditional distribution used to estimate the parameter. For the normal conditional distribution, if the sample information is used exclusively then x would be the estimate of the mean. If the prior is the only information used for estimating the mean then the estimate would be μ_p , the mean of the prior. For the binomial conditional distribution, $p=x/n$ would be the estimate of the proportion π exclusively using the observed sample information. The mean of the beta prior distribution is $\alpha/(\alpha+\beta)$, which would be the estimate if only the prior is used. The posterior means for the two different situations are shown below in a format that illustrates that the posterior mean is a weighted average of the estimate using only the prior and the estimate based on the sample from the conditional distribution.

Expressing the posterior mean as $\frac{\sigma_p^2}{(\sigma_p^2+\sigma_c^2)} * x + \frac{\sigma_c^2}{(\sigma_p^2+\sigma_c^2)} * \mu_p$ for the normal makes this clear.

The sum of the two weights is one. Correspondingly the posterior mean for the binomial situation can be expressed as $\left(\frac{\alpha+\beta}{\alpha+\beta+n}\right) * \left(\frac{\alpha}{\alpha+\beta}\right) + \left(\frac{n}{\alpha+\beta+n}\right) * \left(\frac{x}{n}\right)$. As with the previous situation the sum of the two weights is one.

The Bayesian methodology provides a way to combine the previously obtained information that allowed for the specification of the prior distribution with current information obtained from the conditional distribution and is a valid methodology if the parameter for the conditional distribution does truly vary as described by the prior distribution. This means that some assessment must be made from data to try to determine if the data support that the underlying parameter for the distribution is not a fixed value for the observed situations. If it is then one should assess whether using the Bayesian model provides any real practical value.

TWO POTENTIAL AREAS OF APPLICATION

This paper will focus on two potential areas of application for processes that are binomial. One of these is in the sport of basketball. When a player shoots the basketball then the shot is either made or missed. For a series of shots under similar conditions, such as shooting a free throw, one can reasonably say that the process can be modeled by a binomial distribution. Another area would be in a military setting when a weapon is propelled toward or shot at a target. The result would either be that the weapon hit the target or missed the target. For the situation of shooting a basketball, there is a circular goal of fixed diameter and the ball either passes through the goal or does not. For the military situation there is a fixed target. If the launched weapon has an explosive device then the weapon does not have to exactly hit the point that is the center of the target but can effectively be considered a hit if falls in a circle around this center. The diameter of the circle around the target is determined by the power of the explosive in the weapon. Hence this situation with similar conditions for each weapon launch can effectively be modeled by a binomial distribution. The question at hand for both situations is whether they can correctly be modeled with a Bayesian model.

IS THE CONDITIONAL PARAMETER FIXED OR DOES IT VARY?

The real challenge in the situations mentioned is to determine if the process is truly Bayesian with a binomial proportion that varies from one series or set of trials to another series or set of trials. We will consider three different realities that could be the case for either of these applications. One would be that the process is truly Bayesian and the variation in the binomial proportion can be modeled using a probability distribution as has been discussed. Another reality would be that the binomial proportion is essentially the same for all trials and does not change from one set of trials to another. The third reality would be one for which the binomial proportion is not always the same from one set of trials to another but the change in proportion can be explained by one or more other factors. For example, the free throw percentage for a player may drop when she injures her hand. This change in value is due to a special cause and not due to random variation as described by a probability distribution. One can imagine numerous situations such as this for which the lack of stability and variability of the binomial proportion would not be appropriately described by a probability distribution.

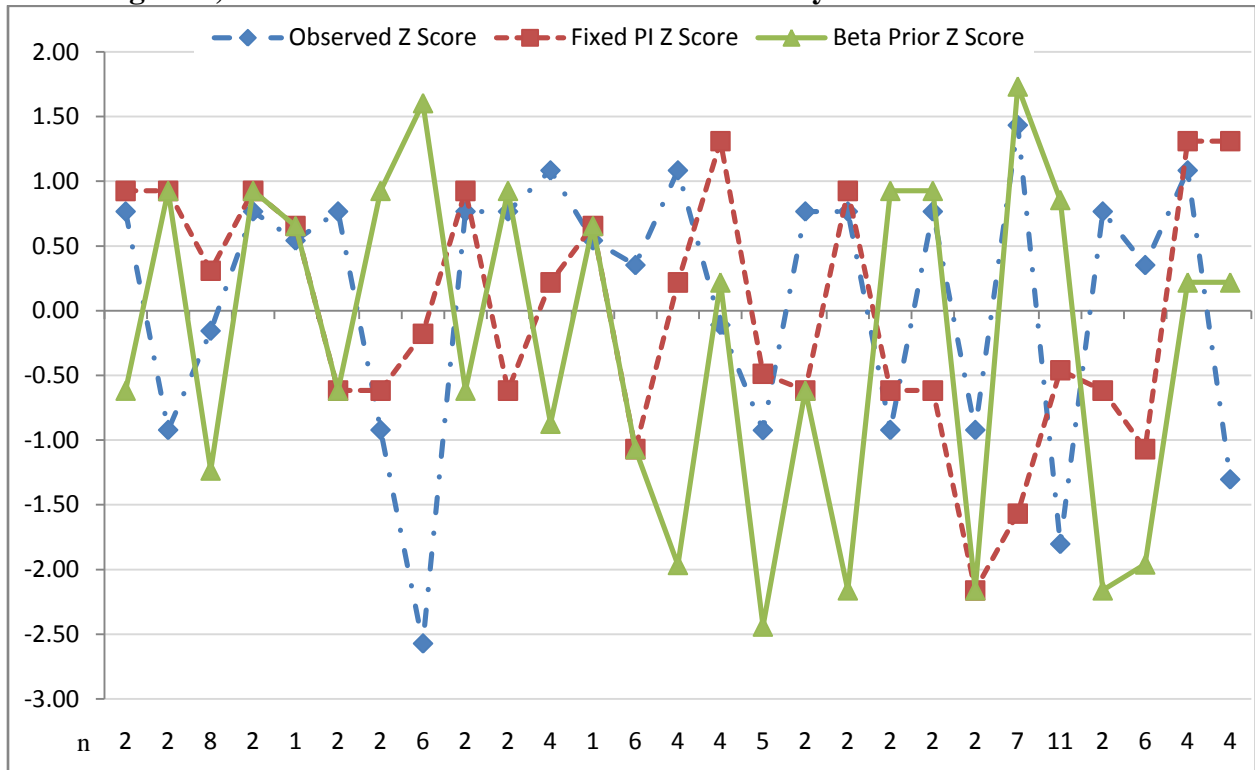
We will begin with an assumption that the value of a binomial proportion has the same fixed value for all sets of trials and will advocate using this model until there is adequate evidence to indicate that the binomial proportion is changing from one set of trials to another. To make a decision about the adequacy of the evidence one can observe the outcomes from several sets of trials to see if the variability is what one would expect if the proportion has the same value for all sets of trials. To do this one must define what constitutes a set of trials. For shooting free throws, we believe that a day should be considered as a set of trials. One could conduct an experiment and have a player to shoot a fixed number of free throws each day and track the number or proportion of observed successes each day. However, the desire would be to create a model that could be used in a game situation and most would agree that player's percentage in a game may be different from the percentage in practice. The number of free throws attempted in each game will vary from game to game. By tracking the proportion or percentage made each game rather than the number made, one has a statistic that is comparable from game to game. However, observing 100% or 0% made out of two attempts does not provide the same evidence

as observing either out of ten attempts. The standard deviation or standard error for a sample proportion for n observations from a phenomenon with π as its proportion of success is $\sqrt{\pi * (1 - \pi)/n}$. Using the mean π and the standard error computed with π and n one can transform each sample proportion p into a z-score that will include the sample size as well as the observed proportion. These z-scores can be plotted to see if any pattern is visually apparent. In particular, are there more extreme scores than one would anticipate? In the Bayesian model there are two primary sources of variability for the observed proportion. One source is the random variation of the observed proportion around the true value of π and this is measured by the standard error of the sample proportion. The other is the variation of π as determined by the prior probability distribution for π . Hence one would expect more variability if the process is truly Bayesian than if the proportion has the same fixed value. We know of no formal test to be able to perform in this situation and use graphs and the distribution of the z-scores. Since the sample sizes will be relatively small the distribution of the z-scores will not be exactly standard normal but they should be somewhat close to a standard normal. Hence we will look at the graphs for any clear patterns in the z-scores and compare the proportions of extreme values with what one would expect for a standard normal distribution. If there are no obvious shifts in the graph and there are clearly more extreme values than anticipated then we will consider this evidence as supporting the use of a Bayesian model for the overall process.

We also propose an ad hoc testing procedure using the χ^2 distribution. If the value of z follows a standard normal distribution, then the sum of k values of z^2 follows a χ^2 distribution with k degrees of freedom. To apply this testing procedure for free throw shooting data from k games we will square each of the z-scores. The test statistic will be the sum of the k squared z-scores. As was stated above we would expect the distribution of the observed z-scores to be reasonably close to a standard normal distribution if the free throw percentage does not vary from game to game. If there is game to game variability then we would expect more extreme values for the z-scores which would result in a higher total for the sum of squared z-scores. Hence this ad hoc testing procedure will be a one-tail upper-tail test using the χ^2 distribution with k degrees of freedom.

Figure 2 shows a plot of z-scores for three series of data. The Observed Z Score values are the z-scores calculated from the 2007-08 season results for Lawrence McKenzie, a senior guard and leading scorer for the University of Minnesota's men's basketball team. He averaged playing about 28 minutes and shooting three free throws per game with a 77.3% free throw percentage for the season. The Observed Z Score values used .773 for the probability of making each free throw and the individual number of free throws he attempted each game. The line for Fixed PI Z Score used a fixed value for π of .7 which is the mean of the beta distribution shown in Figure 1 and the n for each was the same number of free throws shot by McKenzie. The actual number of free throws made was simulated using the binomial with $\pi=.7$. The z scores were calculated using the simulated number of made free throws for each game, $\pi=.7$ and the values of n for Lawrence McKenzie for that corresponding game. Note that the values of n appear at the bottom of the graph in Figure 2. The line for Beta Prior Z Score used a value for π for each game that was obtained through simulation using the beta prior in Figure 1 with $\alpha=14$ and $\beta=6$. The z scores were calculated using the mean of the beta distribution of .7 as the value for π .

Figure 2, Plot of Z-Scores for 2007-08 Data for a Player and Two Simulations



Our hope was that extra variability introduced by the prior distribution for π would manifest itself in the distribution of the z-scores. However the graph for the solid line showing the distribution with the beta prior does not provide clear visible evidence that its variability is greater than that for the dashed line representing data for a fixed value of π . For the data used to create Figure 2, the standard deviation for the Beta Prior Z Score values is .99 and the standard deviation for the Fixed PI Z Score values is .93. The sum of the squared z-scores for the Beta Prior Z Score values is 26.6, with a 1-tail p-value of .48 and the sum of the squared z-scores for the Fixed PI Z Score values is 23.5, with a 1-tail p-value of .66. To get an idea of the power of this ad hoc procedure to detect when a process was truly Bayesian, 100 simulations were performed for each of these. 6 out the 100 simulations for the Beta Prior Z Score values had a p-value less than $\alpha=.05$. 4 out the 100 simulations for the Fixed PI Z Score values had a p-value less than $\alpha=.05$. These simulation results do not indicate that this ad hoc test is not a valid method for determining if a binomial process is truly Bayesian with a value of π that varies from one set of trials to another.

The spread for the plot of the Observed Z Score values is not visually greater than that for the other two lines. This plot does not encourage the use of a Bayesian model for Lawrence McKenzie's free throw shooting. The 27 observed z-scores ranged from a minimum of -2.57 to a maximum of 1.43 with a mean of .10 and standard deviation of 1.02. These values are certainly reasonable for a sample of 27 observations from a standard normal distribution. The sum of the squared z-scores for the Observed Z Score values is 27.6, with a 1-tail p-value of .43. Neither the plot nor the statistics support the use of a Bayesian model for Lawrence McKenzie's free throw shooting. We also looked at data for a few additional players including some from the

NBA with varying free throw percentages ranging from Steve Nash with a high of 90% to Shaquille O'Neal with a low of 50%. None the information for them supported a Bayesian model for free throw shooting.

SUMMARY

Being able to correctly use a Bayesian model depends on the ability of the user to determine if the binomial parameter π does truly vary from one set of trials to another. Each of us has played basketball and have had the perception that we were “on” in our shooting some days and “off” some other days. If this was really true then the value of π was not constant and varied from game to game. However, we also know that perception and reality are not always the same. We cannot justify using a Bayesian model because of our perception. We were not able to find empirical evidence from the limited data we observed that free throw shooting was Bayesian. However, the problem may be with the methods we attempted to use. These methods were not effective for the simulations when one process was Bayesian with a value of π that varied from game to game according to a prescribed beta prior because out of 100 simulations only 6 had a p-value less than .05. This number is only slightly higher than 4 out 100 simulations with a p-value less than .05 for a process with fixed value for π , meaning that it was not Bayesian. With the data that were readily available for basketball we were not able to affirm that free throw shooting can be effectively modeled with a Bayesian model. Data for the military application are not so readily available. Before working on a Bayesian model for the military application we believe that we need to be able to demonstrate that a Bayesian model can be effective for a somewhat similar situation such as shooting a basketball.

REFERENCES

- [1] Canavos, George C., Applied Probability and Statistical Methods; Little, Brown and Company; Boston, MA, 1984.
- [2] Lee, Jack C. and Sabavala, Darius J., “Bayesian Estimation and Prediction for the Beta-Binomial Model,” Journal of Business & Economic Statistics, Vol. 5, No. 3 (Jul., 1987), pp. 357-367.
- [3] Holloway, Charles A., A Decision Making Under Uncertainty: Models and Choices, Prentice Hall, INC., Englewood Cliffs, NJ, 1979.
- [4] Wikipedia, “Posterior distribution of the binomial parameter” (5/18/2008)
http://en.wikipedia.org/wiki/Bayesian_inference#Posterior_distribution_of_the_binomial_parameter