Pig Data and Bayesian Inference on Multinomial Probabilities

John C. Kern
Duquesne University

*Journal of Statistics Education* Volume 14, Number 3 (2006), ww2.amstat.org/publications/jse/v14n3/datasets.kern.html

Copyright © 2006 by John C. Kern II, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words:

Abstract

Bayesian inference on multinomial probabilities is conducted based on data collected from the game *Pass the Pigs®*. Prior information on these probabilities is readily available from the instruction manual, and is easily incorporated in a Dirichlet prior. Posterior analysis of the scoring probabilities quantifies the discrepancy between empirical and prior estimates, and yields posterior predictive simulations used to compare competing extreme strategies.

1. Introduction

The twenty-five-year-old game *Pass the Pigs®*, created by David Moffat and currently marketed by Winning Moves© Games, requires a player to roll a pair of rubber, pig-shaped dice. The player earns—or as the case may be, loses—points based on the configuration of the rolled pigs. Due to the unusual shape of these dice, however, it is difficult to intuit the probability of any particular configuration. Indeed, exact probabilities for each configuration are unknown.

In this research, we use the data collected from 6000 rolls in combination with a multinomial - Dirichlet model to make Bayesian inference on the configuration probabilities. Our analysis is intended for students of Bayesian inference, usually advanced undergraduate mathematics/statistics majors or first-year statistics graduate students - as well as their instructors. Aside from providing an entertaining multinomial-Diachlet application, this analysis is of particular pedagogical interest because specification of prior parameters comes so naturally: The point values for each configuration as defined on the game’s packaging, scorepad, and instructions are used to directly determine the parameter values of the Dirichlet prior distribution. Before providing the details of this model and prior specification, it is necessary to clarify both the rules of the game and the data collection process.

1.1 The Rules

In a game of *Pass the Pigs®*, two or more people compete against one another to be the first to earn 100 points. The game progresses on a turn-by-turn basis through a fixed player ordering, whereby any points a player earns on a turn are added to their points earned on all previous turns. The advantageous first turn is randomly awarded. A player’s turn—which requires use of the pigs—is over when they “pass the pigs” to the next player.

Any turn begins with the rolling of both pig-shaped dice. The configuration of the pigs in this roll, or any other, must fall under exactly one of the following three categories:

- A positive-scoring roll.
- A zero-scoring roll.
- A roll in which the pigs are at rest and in physical contact with each other, regardless of configuration.

If the initial roll on a turn is positive-scoring, the player may choose to immediately roll the pair of pigs again. Such a choice remains available to the player provided the previous roll in that turn was positive scoring. In this way, the points a player earns on their turn is the sum of the point values of an unbroken string of positive-scoring rolls.

The end of a player’s turn is determined by the first occurrence of the following three events:

**End-Turn Event I:** The roll is zero-scoring. In this case the player loses all points accumulated on that turn and must pass the pigs to the next person.

**End-Turn Event II:** The roll is positive-scoring and the player chooses to pass the pigs to the next person. In this case the player retains all points accumulated on that turn.

**End-Turn Event III:** The roll finds the pigs in physical contact with each other. In this case the player loses all points accumulated on that turn as well as the points accumulated on all previous turns. The pigs are then passed to the next person.

It is worthwhile to note that this game does allow a player to incorporate strategy, but only through End-Turn Event II. For example, a player may choose to roll the pigs only once per turn. This is the most conservative strategy in the sense that points earned on a turn are never at risk of being lost to a zero-scoring roll. Conversely, the extreme risk strategy views each turn as an all-or-nothing opportunity. Those adopting this strategy will continue to roll until either a non-positive-scoring roll is obtained or at least 100 points are accumulated. Each turn taken under this strategy will end with the player earning either zero points or victory. Analysis of both extreme strategies is given in Section 2.3. Before presenting the data from these 6000 rolls (and the collection method), we now detail the configuration-to-point-value mapping.

1.2 Scoring

The pair of pigs that come in a new package (available on-line and in most toy stores for roughly 10 U.S. dollars) are virtually indistinguishable from each other. Each pig is molded in the same “trotting” position, reminiscent of the pose you might expect to see in a snapshot of a walking pig. When a single pig is rolled on a smooth, level, unobstructed surface, it will invariably come to rest in one of the six positions listed in Table 1. Table 2 provides pictorial representation of the positions described in Table 1. The names given to the positions come directly from the game, except for the Dot Up and Dot Down labels. These descriptors were chosen for the simple reason that the fleshy-pink colored pigs are marked by a noticeable black dot on the right side of their bodies. Thus, for example, when a pig is resting on its left side, the black dot is in the “up” position. Throughout this paper, we will use the position names and numbers interchangeably.

| Position | Name | Description |
|----------|------|-------------|
|          |      |             |
Table 2. Pictorial representations of the six single-pig positions described in Table 1.

| Pig 1 Position | Pig 2 Position |
|----------------|----------------|
| 1 (Dot Up)     | 2 (Dot Down)   |
| 3 (Trotter)    | 4 (Razorback)  |
| 5 (Snouter)    | 6 (Leaning Jowler) |

Table 3. Scoring table for all possible positive-scoring and zero-scoring configurations.

| Pig 1 Position | Pig 2 Position |
|----------------|----------------|
| 1 (Dot Up)     | 2 (Dot Down)   |
| 3 (Trotter)    | 4 (Razorback)  |
| 5 (Snouter)    | 6 (Leaning Jowler) |

From Table 3 we see that only two of the thirty-six position combinations are zero-scoring: any roll that finds the pigs lying on opposite sides results in End-Turn Event I. Aside from the two configurations (both pigs lying on the same side) worth 1 point, all other configurations are positive-scoring and worth some multiple of 5 points. Note that a positive-scoring roll must yield a point value from the set \{1, 5, 10, 15, 20, 25, 40, 60\}. We will refer to Table 3 often, especially in Section 2 when determining a prior distribution for a Bayesian multinomial data model. Before discussing this model, we finish this introduction by describing the data and the method by which it was collected.

1.3 The Data

Data collected from 6000 rolls of a pair of pigs was generated by two people. Both people rolled their own, brand new pair of pigs 3000 times, and recorded the position of the pigs after each roll. To better enhance our roll-of-the-pigs understanding, one pig from both pairs was randomly selected and marked with a small black dot (made gently with a permanent marker) on its snout. The other was left unmarked. In this way, the roller would record the position number (from Table 1) of the marked - or pink - pig. Table 3 would then be used to determine the score of the roll.

Due to variability in rolling technique across people, we decided to standardize the rolling technique by using a trap-door style rolling apparatus. This apparatus was constructed in such a way as to impart on the pigs realistic rolling movement. It consisted of nothing more than a four-inch square sheet of sturdy cardboard, well-creased to divide its area into two equal size rectangles. This sheet was then placed on a level, eight-inch tall wooden platform, such that the crease was parallel to an edge of the platform. Rolling the pigs was accomplished by placing the pigs on one half of the crease-divided cardboard (in the trotting position, 0.25 inches apart, facing away from the crease and toward the parallel platform edge), and using the other half of the cardboard as a handle to push-slide the cardboard toward the parallel platform edge - making sure to always keep the crease and platform edge parallel. When the cardboard is moved far enough for the crease to overlap the edge of the platform, the pushing-sliding stops, and the weight of the pigs cause their half of the creased cardboard to drop in trap-door fashion. Even with no pigs on the cardboard, the crease was such that the cardboard weight itself would cause the drop. The other half of the cardboard is anchored securely under the fingers of the roller; hence only the pigs tumble to the table below. In this way, the pigs are not simply dropped to the table. Rather, they are dropped with the forward momentum gained from the pushing-sliding of the creased cardboard. Rolls that saw either pig touch a platform support were ignored.

Variation in the rolling technique is introduced from a variety of sources. A source of variation we intentionally impose is that of platform height: One person rolled the pair of pigs 3000 times from the aforementioned eight-inch tall platform, while the other pair were rolled 3000 times from a similar five-inch tall platform. Natural sources of variation not imposed by the author include:

- Any dissimilarity between the two pairs of pigs.
- Dissimilarity between the two rolling surfaces. (The rolls from eight inches were onto a Formica surface; the rolls from five inches were onto a hardwood surface.)
- Dissimilarity in the speeds at which the pig-carrying cardboard was pushed, both within roller and between rollers.
- Any pig rubber wear-and-tear associated with 3000 rolls.
This analysis treats these sources of variation - imposed or natural - as negligible.

Shown in Table 4 are the number of times each of the thirty-six possible position combinations were observed in the 6000 rolls. When comparing this data with the scores from Table 3, we see combinations that occur more often are generally associated with lower point values. Note that the frequencies in this table sum to 5977; exactly 23 of the 6000 rolls resulted in the two pigs touching each other (corresponding to End-Turn Event III). If on a given roll we treat the positions of the black and pink pigs as discrete random variables, then Table 4 can be viewed as their empirical joint distribution by dividing each cell by 6000.

| Black Pig Position | Pink Pig Position |
|--------------------|-------------------|
| 1 (Dot Up)         | 1 (Dot Up)        |
| 2 (Dot Down)       | 2 (Dot Down)      |
| 3 (Trotter)        | 3 (Trotter)       |
| 4 (Razorback)      | 4 (Razorback)     |
| 5 (Snouter)        | 5 (Snouter)       |
| 6 (Leaning Jowler) | 6 (Leaning Jowler)|

Table 4. Raw frequencies for the black-pink pig positions, based on 6000 rolls.

This data is used in the following section to estimate the probabilities of observing each of the eight possible positive scores, a zero-scoring roll, and a roll in which the pigs are touching each other.

2. Multinomial - Dirichlet Inference

We now turn our attention to estimating, in Bayesian fashion, the probability that a single roll of the pigs will yield $k$ points, for $k = 0, 1, 5, 10, 15, 20, 25, 40,$ and $60$. These nine possible scoring outcomes exclude only End-Turn Event III, wherein the pigs are touching each other. To this outcome we assign an artificial point value of $k = -1$. In this way, we can now let $\theta_k$ represent the probability that a roll of the pigs yields $k$ points, where

$$\sum_{k \in \mathcal{S}} \theta_k = 1$$  \hspace{1cm} (1)

and $k$ is restricted to the ten integer set $\mathcal{S}$ defined by

$$\mathcal{S} = \{-1, 0, 1, 5, 10, 15, 20, 25, 40, 60\}.$$

Thus, for example, $\theta_{-1}$ represents the probability of observing a 15-point roll, $\theta_{0}$ the probability of a 0-point roll (End-Turn Event I), and $\theta_{60}$ the probability of End-Turn Event III.

Bayesian analyses distinguish themselves from their classical counterparts by incorporating into parameter inference information supplied by the researcher before data have been observed. This information is presented in the form of a probability distribution - called the prior distribution - on the parameter(s) of interest. In our application, the parameters of interest are the scoring probabilities $\theta = (\theta_1, \theta_2, \ldots, \theta_{60})$ and the data are the scoring outcomes of the 6000 rolls. What makes this analysis so appealing from a Bayesian pedagogical perspective is that through the very scores it assigns to each configuration, the game provides prior information about $\theta$ that can be easily expressed through a probability distribution on $\theta$. Too often are students of Bayesian inference confronted with examples where prior distributions are presented with no motivation/explanation except a distancing phrase such as “suppose our prior distribution for $\theta$ is ...,” or, “expert opinion suggests the prior distribution for $\mu$ to be ....” In this example, simply reading the instructions empowers one to give “expert opinion” on the unknown scoring probabilities $\theta$.

The method by which the provided scores are used to make a priori statements about $\theta$ may differ according to individual philosophy; we believe readers will find the approach presented here among the most reasonable and intuitive. Based on three assumptions, we translate the provided scores into point estimates for the scoring probabilities. The first of these assumptions is the least imposing (and most intuitive):

**Assumption 1:** Positive roll scores and their corresponding roll probabilities are inversely related.

Thus, a 60-point roll is less likely than any other positive-scoring roll (as it has the highest point value); a 1-point roll is more likely than any other positive-scoring roll (as it has the lowest point value); the 5-point roll is more likely than the 10-point roll, but less likely than the 1-point roll, and so on. Our second assumption about $\theta$ tells exactly how much more likely a $k$-point roll is, relative to a 1-point roll for $k > 0$:

**Assumption 2:** The probability of observing a 1-point roll is $c$ times greater than the probability of observing a $k$-point roll, for $k > 0$.

The yet unmentioned zero-scoring and End-Turn Event III rolls are accounted for in our final assumption:

**Assumption 3:** 1-point rolls and zero-scoring rolls have the same probability, while a 1-point roll is $c$ times more likely ($c > 0$) than a $k = -1$ point roll.

That $\theta_1 = \theta_60$ follows naturally from our willingness to assume the dot-up position has the same, or nearly the same probability as the dot-down position. If one assumes the dot-up and dot-down probabilities differ greatly, then take $\theta_1 = \alpha \theta_60$ for some $\alpha > 0$. As for the relationship between $\theta_1$ and $\theta_{-1}$, we are inclined to set $\alpha$ equal to a number near 60 in an effort to equate the chances that the most detrimental event (score of $k = -1$) and the most beneficial event (score of $k = 60$) occur.

From these assumptions, we formally recognize the relationship between $\theta_{-1}$ and each of the nine remaining probabilities as follows:

$$\theta_k = \theta_{-1} \cdot \theta_{\epsilon} \quad \text{for} \quad k \in \mathcal{S} \setminus \mathcal{L}$$ \hspace{1cm} (2)

$$\theta_{-1} = \theta_0$$ \hspace{1cm} (3)
\[ \theta_i = c \cdot \theta_i \quad \text{for} \quad c > 0. \] (4)

Using these relationships, we express the sum in (1) in terms of only \( \Theta_1 \) to yield

\[ \Theta_1 \left( \frac{1}{c} + 1 + \frac{1}{5} + \frac{1}{10} + \frac{1}{15} + \frac{1}{20} + \frac{1}{25} + \frac{1}{40} + \frac{1}{60} \right) = 1. \] (5)

Solving (5) for \( \Theta_1 \) gives our a priori point estimate of \( \Theta_1 \) (and hence of \( \Theta_2 \)):

\[ \Theta_1 = \Theta_2 = \frac{600}{1.099 + 600/c}. \] (6)

Using this value of \( \Theta_1 \) with (2) and (4) gives the following prior point estimates for the remaining eight probabilities:

\[ \begin{align*}
\Theta_4 &= \frac{600}{1.099 + 600/c} \quad \Theta_5 = \frac{120}{1.099 + 600/c} \\
\Theta_9 &= \frac{60}{1.099 + 600/c} \quad \Theta_{10} = \frac{40}{1.099 + 600/c} \\
\Theta_{19} &= \frac{30}{1.099 + 600/c} \quad \Theta_{20} = \frac{24}{1.099 + 600/c} \quad \Theta_{25} = \frac{15}{1.099 + 600/c} \\
\Theta_{40} &= \frac{10}{1.099 + 600/c} 
\end{align*} \] (7)

We remind the reader that these point estimates come directly from the scores supplied by the game as well as our assumptions (2), (3), and (4) about the relationships among the scoring probabilities \( \Theta \). When specifying the prior distribution on \( \Theta \), we will choose parameter values that set the marginal expectation of each \( \Theta_k \) equal to the corresponding point estimate in (6) or (7). We choose the functional form of the prior distribution (i.e. the family of prior distributions) to be Dirichlet; the motivation and density for this distribution will be presented upon explicit presentation of the multinomial data model.

2.1 Multinomial Likelihood

Let \( X_k \) represent the number of \( k \)-point rolls observed among the \( n = 6000 \), where \( k \) is again restricted to the set \( S \). Given \( \Theta \), the joint distribution of the ten random variables \( \{X_1, X_2, \ldots, X_{60}\} \) is then multinomial, provided we make the mild assumptions that rolls are independent and the scoring probabilities \( \Theta \) remain constant from one roll to another. The multinomial mass function, denoted by

\[ p(X_1 = x_1, X_2 = x_2, \ldots, X_{60} = x_{60} | \Theta) \]

for a vector \((x_1, x_2, \ldots, x_{60})\) of nonnegative integers satisfying \( \sum_{k=1}^{60} x_k = n \), is

\[ p(X_1 = n_1, X_2 = n_2, \ldots, X_{60} = n_{60} | \Theta) = \frac{n!}{n_1! n_2! \cdots n_{60}!} \cdot \theta_1^{n_1} \cdot \theta_2^{n_2} \cdots \theta_{60}^{n_{60}}. \] (8)

When we view the right hand side of (8) as a function of \( \Theta \) and treat \((n_1, n_2, \ldots, n_{60})\) as known, we obtain the multinomial likelihood function for \( \Theta \) denoted by \( L(\Theta) \):

\[ L(\Theta) \propto \theta_1^{n_1} \cdot \theta_2^{n_2} \cdots \theta_{60}^{n_{60}}. \] (9)

The proportionality symbol is used to remind the reader that any function of \( \Theta \) proportional to the right hand side of (9) can be labeled “the” likelihood function, and will yield the same inference on \( \Theta \).

Standard classical inference finds the values \( \hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_{60} \) that maximize the likelihood function. These values are given by

\[ \hat{\Theta}_i = \frac{n_i}{n} \]

as shown in various texts (see Berry and Lindgren 1996, or Lange 1999). Thus, the maximum likelihood estimates for the \( \Theta_k \)'s are simply the respective proportions of \( k \)-point rolls observed among the 6000. These proportions, as obtained from the data in Table 4, are:

\[ \begin{align*}
\hat{\Theta}_4 &= \frac{23}{6000} \quad \hat{\Theta}_5 = \frac{1279}{6000} \\
\hat{\Theta}_9 &= \frac{1304}{6000} \quad \hat{\Theta}_{10} = \frac{2337}{6000} \\
\hat{\Theta}_{19} &= \frac{508}{6000} 
\end{align*} \] (10)

Note that these estimates are based completely on the likelihood function for \( \Theta \). Bayesian inference on \( \Theta \), however, is based on the conditional distribution of \( \Theta \) given the data. This distribution - called the posterior distribution and denoted here by \( \pi(\Theta | \text{data}) \) - combines information in the data (through \( L(\Theta) \)) with the prior distribution for \( \Theta \). We denote the prior distribution by \( \pi_0(\Theta) \). Once the family and parameter values of the prior distribution have been chosen, we use Bayes’s Theorem to give the posterior distribution \( \pi_1(\Theta) \) as a function of the likelihood \( L(\Theta) \) and prior \( \pi_0(\Theta) \):

\[ \pi_1(\Theta) \propto L(\Theta) \cdot \pi_0(\Theta). \] (11)
Thus the conditional distribution of $\boldsymbol{\theta}$ given the data is proportional to the product of the likelihood function and prior distribution for $\boldsymbol{\theta}$. Estimates analogous to (10) can be obtained from the posterior distribution by finding the marginal expected value of each $\theta_k$, and will reflect a balance between the maximum likelihood estimates and the prior point estimates for each $\theta_k$. The following section details our choice of $\pi_0(\boldsymbol{\theta})$.

### 2.2 Dirichlet Prior

We choose $\pi_0(\boldsymbol{\theta})$ to be Dirichlet with parameter vector $\alpha = a_{-1}, a_0, ..., a_{60}$. The corresponding density is

$$
\pi_0(\boldsymbol{\theta}) = \Gamma(\mathcal{A}) \prod_{k=1}^{m} \left( \frac{\theta_k^{a_k-1}}{\Gamma(a_k)} \right)
$$

(12)

where $\mathcal{A} = \sum_{k=0}^{m} a_k$. Recognizing the marginal expected value and variance of any $\theta_k$ as

$$
E[\theta_k] = \frac{a_k}{\mathcal{A}} \quad \text{and} \quad \text{Var}[\theta_k] = \frac{a_k (\mathcal{A} - a_k)}{\mathcal{A}^2 (\mathcal{A} + 1)}
$$

(13)

makes the task of specifying $\alpha$ straightforward. Specifically, we set these marginal expected values equal to their corresponding prior point estimates in (6) or (7) by choosing $c = 50$, thereby assuming the a priori probability of $k = -1$ to fall between that of $k = 40$ and $k = 60$ and letting

$$
\begin{align*}
\alpha_0 &= 12 \cdot m \\
\alpha_0 &= 600 \cdot m \\
\alpha_2 &= 600 \cdot m \\
\alpha_2 &= 120 \cdot m \\
\alpha_2 &= 60 \cdot m \\
\alpha_2 &= 40 \cdot m \\
\alpha_2 &= 30 \cdot m \\
\alpha_2 &= 30 \cdot m \\
\alpha_2 &= 15 \cdot m \\
\alpha_2 &= 10 \cdot m
\end{align*}
$$

(14)

for some $m > 0$. Note that the choice of $m > 0$ does not influence the marginal variance of any $\theta_k$. The choice of $m$ does, however, influence the marginal variance of every $\theta_k$. In particular, because the marginal variance of $\theta_k$ is roughly proportional to $1/m$, we can express greater prior uncertainty in a $\theta_k$ with smaller values of $m$, and greater prior certainty in a $\theta_k$ with larger values of $m$.

It is now clear that one of the factors motivating our Dirichlet prior choice is our ability to use the available prior point estimates for $\boldsymbol{\theta}$ to easily determine the values of $\alpha$, up to a strength-of-prior-belief parameter $m$. Inference on $\boldsymbol{\theta}$ in the following sections will be a function of this (now) lone prior parameter $m$. We will also refer to $m$ as a hyperparameter, as is common for parameters of a prior distribution.

That the Dirichlet prior for $\boldsymbol{\theta}$ is conjugate for the multinomial likelihood is a second reason for its use. A prior density $\pi_0(\mathcal{H})$ is conjugate for a likelihood if the resulting hyperparameter distribution for $\mathcal{H}$ in (11) our likelihood from (9) and our prior density from (12) gives the following posterior distribution on $\mathcal{H}$:

$$
\pi_0(\mathcal{H}) \propto \prod_{k=0}^{m} \left( \frac{\theta_k^{a_k-1}}{\Gamma(a_k)} \right)
$$

(15)

Comparing this posterior density with the prior density in (12), we see $\pi_0(\mathcal{H})$ is indeed Dirichlet, with parameter vector $\beta = (b_0, b_1, ..., b_{60})$, where $b_k = (n_k + a_k)$. While a Dirichlet prior on $\mathcal{H}$ does afford an easily obtainable (and manageable) posterior distribution for $\mathcal{H}$, it necessarily reflects a somewhat limited range of prior beliefs (O’Hagan, 1994). For example, the correlation between a particular $(\theta_k, \mathcal{H})$ pair, as given by our Dirichlet prior, is fixed for every $m > 0$. Other prior distributions on $\mathcal{H}$, such as the multivariate normal family, can eliminate such a limitation at the cost of introducing additional hyperparameters. Ultimately, limitations associated with use of the Dirichlet prior are deemed here as insignificant in comparison with the ease-of-implementation advantage its use carries. Having recognized the Dirichlet posterior distribution for $\mathcal{H}$ in (15), we are now ready to make Bayesian inference on these ten scoring probabilities.

### 2.3 Posterior Analysis

The marginal posterior expectation $E[\theta_k]$ of any $\theta_k$ is found easily from the Dirichlet posterior in (15):

$$
E[\theta_k] = \frac{n_k + a_k}{n + \mathcal{A}}.
$$

(16)

Note from (14) that each $\alpha_k$ value (and hence $A$) is proportional to the hyperparameter $m$. Thus, as this strength-of-prior-belief parameter $m$ goes to zero, the posterior point estimate of $\theta_k$ in (16) converges to $\theta_k$. As $m \to 0$, the posterior is dominated by the prior and $E[\theta_k]$ converges to $a_k/\mathcal{A}$.

**Influence of $m$ on Posterior Point Estimates**

Recognizing that for fixed $m$

$$
\sum_{k=0}^{60} E[\theta_k] = 1
$$

allows us to partition the unit interval into ten segments whose lengths represent the posterior expectation of $\theta$. The curves shown in Figure 1 partition the unit interval - presented along the vertical axis - into ten segments so that the vertical space between two curves represents the posterior expected value of a $\theta_k$ at fixed $m$. As $m \to 0$, the vertical separations between the curves correspond directly to the maximum likelihood estimates in (10), and represent posterior estimates that ignore prior information. As $m \to 0$, the vertical separations between curves correspond directly to the prior point estimates in (6) and (7), and represent posterior estimates that ignore the data collected from the 6000 rolls. For values of $m$ greater than 15, the curves have practically converged to horizontal lines whose separations are equal to the prior point estimates.
By comparing the curve separations where \( m \) is near zero with those as \( m \to \infty \), reveals great discrepancies between the maximum likelihood (or “empirical”) estimates and the prior point estimates:

- Values of \( m \) less than 2.2 estimate \( \theta_2 \) as the largest of the scoring probabilities; this distinction is shared by \( \hat{\theta}_0 \) and \( \hat{\theta}_1 \) for values of \( m \) greater than 2.2.

- Prior estimates of \( \hat{\theta}_0 \) and \( \hat{\theta}_1 \) are roughly 1.8 times larger than the corresponding empirical estimates.

- The empirical estimate of \( \theta_2 \) is roughly 4.8 times larger than its prior estimate.

- The empirical estimate of \( \theta_0 \) is roughly 2.1 times larger than its prior estimate.

**Figure 2** shows a rescaled (i.e. magnified) version of **Figure 1**; only the curves presented in the \([0.9, 1.0]\) subset of the **Figure 1** unit interval are shown in **Figure 2**. From this magnified Figure we notice the following:

- Prior estimates of \( \theta_{25} \), \( \theta_{40} \), and \( \theta_{60} \) are at least 25 times larger than their corresponding empirical estimates.

- The empirical estimate of \( \theta_{25} \) is roughly 3 times larger than its prior point estimate.

- The empirical estimate of \( \theta_5 \) and its corresponding prior point estimate are in good agreement.

---

**Figure 2**: Partition of \((0.9,1.0)\) vertical axis as magnified from **Figure 1**. The partition is a function of the hyperparameter \( m \).
We conclude from these two Figures that the scores assigned by the game - in combination with our assumptions (2), (3), and (4) - do not reflect well the relative frequencies we observed in 6000 rolls.

**Posterior Predictive Distribution and Extreme Strategies**

Let $\hat{Y}$ be the yet to be recorded score of the next roll of the pigs. We use the posterior distribution to sample from the distribution of $\hat{Y}$ - called the posterior predictive distribution - as follows:

1. Sample $\theta$ from the Dirichlet posterior in (15). One way to easily accomplish such a simulation is to generate ten independent realizations from gamma densities (Gelman, Carlin, Stern, and Rubin 1995). For each $k \in S$, independently draw $w_k \sim \text{gamma}(b_k)$. A draw from the Dirichlet posterior is then obtained by setting $\theta_k = w_k / \sum w_k$.

2. Use this realization to simulate a single score $\hat{Y}$ from the multinomial distribution in (8). This is accomplished by partitioning the unit interval according to the realization, and observing in which partition a random uniform $(0, 1)$ draw falls. The score corresponding to this randomly selected partition is saved.

Repeating steps 1 and 2 will yield a distribution of predicted scores that incorporates uncertainty in $\theta$ as given by the posterior distribution - each prediction $\hat{Y}$ we simulate comes from a Dirichlet distribution with different $\theta$, as sampled from $\pi_\theta(\theta)$. We must also point out that such a simulation requires specification of the hyperparameter $m$.

In the following analysis we use $m = 1$, as it reflects our belief that data from the 6000 rolls are more influential in estimating $\theta$ than the accompanying scores.

We now use $10^5$ simulated values of $\hat{Y}$ to investigate the effectiveness of the two extreme strategies mentioned at the end of Section 1.1:

**Extreme-Conservative:** Player rolls the pigs only once per turn.

**Extreme-Risk:** Player continues rolling until either 100 points are accumulated, or a non-positive score is obtained.

Shown in Figure 3 are histograms representing the distribution of the number of turns necessary to obtain 100 points under the extreme-conservative strategy.

![Figure 3: Empirical distribution (left) and posterior predictive distribution (right) of the number of turns necessary to reach 100 points, assuming the conservative, one-roll-per-turn strategy.](image)

The histogram on the left is based only on the 6000 rolls; the histogram on the right is based on $10^5$ posterior predictive realizations of $\hat{Y}$. Five number summaries for these distributions are given in Table 5.

|      | Min | Q1  | Q2  | Q3  | Max  |
|------|-----|-----|-----|-----|------|
| Empirical | 10  | 18.75 | 22  | 26.25 | 61   |
| Predictive Posterior | 5   | 20  | 24  | 29  | 103  |

Table 5. Five number summary for the minimum number of extreme-conservative strategy turns necessary to reach 100 points, as based on the data and $10^5$ posterior predictive realizations.

Notice in this table the impact of our prior distribution. By choosing $m = 1$, we allow the prior scores to have a slight influence on our posterior for $\theta$. As a result, the posterior expectations of $\theta_1$ and $\theta_3$ will be higher, on average, than their corresponding maximum likelihood estimates. This greater probability of a non-positive scoring -
or penalty - roll will require more extreme-conservative turns to attain 100 points; hence the quartiles and maximum in Table 5 are greater for the distribution based on posterior predictive realizations.

The argument that our posterior with \( m = 1 \) estimates the probability of a penalty roll to be greater than our empirical estimate \( \left( \hat{\beta}_1 + \hat{\beta}_0 \right) \) also implies that extreme-risk turns should yield zero points more often if based on posterior predictive realizations. Figure 4 shows the distribution of the number of points obtained before the first penalty roll as observed under the extreme-risk strategy.

![Figure 4: Empirical distribution (left) and posterior predictive distribution (right) of the number of points accumulated on a single turn before the first penalty roll is observed.](image)

The histogram on the left is based only on the 6000 rolls; the histogram on the right is based on \( 10^5 \) posterior predictive realizations of \( \tilde{F} \). Notice from Table 6 that indeed, fewer points are accumulated before the first penalty roll when using posterior predictive realizations. Furthermore, the empirical proportion of turns resulting in 100 or more points before the first penalty roll is 0.022, while our posterior predictive realizations estimate this proportion to be 0.011.

Table 6. Five number summary for the number of points accumulated before the first penalty roll using the extreme-risk strategy, as based on the data and \( 10^5 \) posterior predictive realizations.

|          | Min | Q₁ | Q₂ | Q₃ | Max |
|----------|-----|----|----|----|-----|
| Empirical| 0   | 1.25 | 12 | 31 | 215 |
| Predictive Posterior | 0 | 0 | 10 | 25 | 199 |

3. Discussion

The game Pass the Pigs® provides an opportunity for students to conduct Bayesian inference on multinomial probabilities using data collected from an entertaining random phenomena. Furthermore, this game provides students an example where prior information on the parameters of interest clearly exists, and can be easily expressed through a (Dirichlet) prior density. This is in contrast to standard introductory examples of Bayesian inference where prior distributions are specified either without explanation or intuition, or with motivation to which students have a difficult time relating (such as “expert opinion suggests...”). Here, students can use the supplied scores to make a priori statements about the multinomial probabilities. What follows are our conclusions about the ten scoring probabilities as based on a multinomial likelihood using data from a rolling apparatus, a Dirichlet prior with \( c = 50 \), and resulting Dirichlet posterior with \( m = 1 \).

The Dirichlet posterior for the ten scoring probabilities found in (15) combines information from 6000 rolls with the provided scores, and reveals that these scores do not reflect well what happens in practice. For example, there are instances where a higher scoring roll is estimated to have a substantially higher probability than a roll that scores lower: a 5-point roll is more likely than a 1-point roll, and a 20 point roll is more likely than a 15-point roll. This does not agree with the basic intuition that a roll’s score and corresponding probability should be inversely related. In this same spirit, the 60-point roll is so rare that, relative to the other roll scores, it is quite undervalued. This same argument can be made for the 40- and 25-point rolls.

Our Dirichlet posterior estimates the chance of a positive scoring roll to be 74.5%. Thus the probability that your next roll is a penalty roll is approximately 1/4. Using our Dirichlet posterior to simulate posterior predictive realizations of the next, unobserved roll score allows us to investigate the effectiveness of various strategies. The extreme-conservative strategy will yield 100 points every 24 turns, on average, while those adopting the extreme-risk strategy are expected to attain 100 points only once per 100 turns.

3.1 A Test for Symmetry

Certainly we are not limited to Bayesian methods when analyzing the data from our 6000 rolls. There exist several pertinent questions that can be directly addressed with a classical hypothesis test. For example, we might ask: Is there a significant difference in the symmetry of the rolls, based on our arbitrary black/pink assignment? If not, then the observed frequencies displayed in Table 4 should be roughly symmetric about the main diagonal. A test for contingency table symmetry (Bowker, 1948) gives a non-significant \( p \)-value of 0.47; hence we do not have evidence to reject the hypothesis of symmetry (i.e., that the pink/black assignment matters) based on the data from this
3.2 Using Expected Gain to Define Strategy

The game Pig is a non-commercial analogue to Pass the Pigs® requiring a player to roll a single, fair, six-sided die. Rolling a 1 corresponds to our End-Turn Event I; a roll of any other value earns the player that many points. Aside from these differences play proceeds as in Pass the Pigs®. There is no End-Turn Event III in Pig. Literature analyzing effective Pig strategies (see Neller and Presser 2004, and Shi 2000) need not confront unknown score probabilities, as they are all 1/6. It is possible, however, to use the maximum likelihood estimates in (10) as the score probabilities for Pass the Pigs® and conduct strategy analyses (non-Bayesian) similar to those in Neller and Presser 2004, and Shi 2000.

In the spirit of Shi 2000, we present here a brief, strategy-defining expected value calculation for Pass the Pigs®. Let $X$ be the random variable representing the gain in score on a particular player’s next roll. Note that this gain can be negative. Possible values of $X$ and their corresponding probabilities are given in Table 7, where $S$ represents the player’s score before starting the current turn, and $T$ the score gained, so far, by the player on the current turn.

Table 7. Distribution of the gain in score $X$ on the next roll, as based on maximum likelihood probability estimates.

| $X$ | $(S + T)$ | -$T$ | 1 | 5 | 10 | 15 | 20 | 25 | 40 | 60 |
|-----|-----------|------|---|---|----|----|----|----|----|----|
| $P(X=x)$ | 23/6000 | 1279/6000 | 1304/6000 | 2337/6000 | 508/6000 | 171/6000 | 372/6000 | 2/6000 | 2/6000 | 1/6000 |

We can now represent the expected gain in score - with just one additional toss of the pigs on the current turn - as the expected value of the random variable $X$. Using $EX$ to denote this expectation, straightforward calculation gives

$$EX = -\frac{S}{6000} - \frac{1202}{6000} + \frac{28284}{6000}.$$  \hspace{1cm} (17)

This expected value suggests a simple strategy: choose to roll again only when the expected score of the next roll is positive. Since (17) is positive if and only if

$$T < \frac{28284 - 23S}{1302},$$

this strategy dictates rolling again until $T \geq 22$ when starting with zero points ($S = 0$). Notice that the cutoff for $T$ does not change much with $S$, as End-Turn Event III is so rare. For example, starting a turn with 79 points ($S = 79$) dictates rolling again until $T \geq 21$. Players in situations where $S \geq 80$ would, of course, abandon this strategy, as (100 - $S$) is less than $\frac{28284 - 23S}{1302}$; that is, it is possible for both $(S + T) \geq 100$ and $EX > 0$ to occur simultaneously.

3.3 Further Analyses of Interest

The dataset accompanying this investigation provides an opportunity to conduct many more analyses than can be presented here. What is the influence of roll-height on the distribution of roll scores? Indeed, all analyses conducted thus far can be applied separately to the 3000 rolls from eight inches, and then to those from five inches. Tests for symmetry would be especially relevant, as a true difference between black and pink pigs may be masked in the combined data presented in Table 4 by the arbitrary black pig assignment for the five-inch rolls and the arbitrary black pig assignment for the eight-inch rolls.

Does the starting position of the pigs have an effect on score distribution? The first 1500 rolls from both platforms saw both pigs in a “face-forward” initial position. The second 1500 rolls from the eight-inch platform had one pig in the “face-forward” initial position, and the other in the “face-backward” initial position. The second 1500 rolls from the five-inch platform saw both pigs in the “face-backward” initial position.

As our final suggestion for further analysis, we point out that sampling from the posterior predictive distribution provides an endless stream of “next roll” scores that can be used to ascertain the effectiveness of any strategy—even those conditional upon the scores of other players. This is perhaps the most engaging analysis for students, as the personal strategy of each student can be combined in a program that simulates an entire game. Several game simulations reveal the student with the most successful strategy.

4. Getting the Data

The file pig.dat.txt is a text file containing 6000 rows. Each row corresponds to a roll of the two pigs. The file pig.txt is a documentation file describing the variables.

Acknowledgments

The author would like to acknowledge the Phillip H. and Betty L. Wimmer Family Foundation for contributing their generous support to this project. The author is grateful to Rachel Riberich for her contribution of 3000 pig rolls, and to the editor and referees for their comments and suggestions.

References

Berry, D. A., Lindgren, B. W., (1996), Statistics: Theory and Methods, 2nd ed., Belmont, CA: Duxbury Press.

Bowker, A.H., (1948), “A test for symmetry in contingency tables,” Journal of the American Statistical Association, 43(244), 572 - 574.

Gelman, A. B., Carlin, J. S., Stern, H. S., and Rubin, D. B. (1995), Bayesian Data Analysis, London; New York: Chapman and Hall.

Lange, K. (1999), Numerical Analysis for Statisticians, New York: Springer-Verlag.

Neller, T., and Presser, C. (2004), “Optimal play of the dice game Pig,” The UMAP Journal, 25(1), 25 - 47.
O’Hagan, A. (1994), Kendall’s Advanced Theory of Statistics: Bayesian Inference, New York; Toronto: Halsted Press.

Shi, Y. (2000), “The game PIG: Making decisions based on mathematical thinking,” Teaching Mathematics and its Applications, 19(1), 30 - 34.

John C. Kern II
Department of Mathematics and Computer Science
Duquesne University
Pittsburg, PA 15282
U.S.A.
kern@mathcs.duq.edu