

Sankhyā : The Indian Journal of Statistics
Special issue in memory of D. Basu
2002, Volume 64, Series A, Pt. 3, pp 820-841

A BAYESIAN PEEK INTO FELLER VOLUME I

By PERSI DIACONIS
Stanford University, USA

and

SUSAN HOLMES
Stanford University, USA and INRA, France

SUMMARY. We develop Bayesian versions of three classic probability problems: the birthday problem, the coupon collector's problem and the matching problem. In each case, the Bayesian component involves a prior on the underlying probability mechanism. Sometimes this appreciably changes the answer, sometimes not.

1. Introduction

We are going to re-analyse some basic problems of elementary probability from a Bayesian point of view. As an example, consider the familiar birthday problem. If k balls are randomly dropped into n boxes, what is the chance of a *match*, that is, that 2 or more balls are dropped into the same box? The classical answer assumes that the balls are dropped uniformly and independently into each box. Under these assumptions, if $n = 365$, the chance of a match is approximately $1/2$ when $k = 23$ balls are dropped. In contrast, consider a practical demonstration of the birthday problem on the first day of a class. Are the birthdays of the students reasonably considered as balls dropped uniformly into 365 categories? After all there are well-known weekend/weekday effects, and perhaps seasonal and lunar trends for birth rates. On reflection, the instructor may conclude that the chance p_i of a student being born on day i is not uniform and in fact is not known.

A Bayesian approach puts a prior distribution on the p_i and computes the chance of a match. We treat this problem with a variety of priors in Section 2. We also develop a general result for Dirichlet priors using Stein's method for Poisson approximation.

Paper received December 2001; revised July 2002.

AMS (2000) subject classification. 60A99, 60-02, 60E0.

Keywords and phrases. Birthday Problem, coupon collector's problem, matching problem, Dirichlet prior, Stein's Method, Bose-Einstein distribution.

Section 3 treats the coupon-collector's problem. If k balls are dropped randomly into n boxes what is the chance that each box contains at least one ball? (an event we call *covering*). Using classical computations, if $k = n \log n + \theta n$ for some θ the chance of covering is approximately $e^{-e^{-\theta}}$ when n is large. For example, when $n = 365$, $k = 2287$ there is about a 50% chance of covering. We give useful approximations to the probability of covering for a variety of prior distributions. For example, under a uniform prior, when $n = 365$, $k = 191,844$ balls are required to have probability 50% of covering. This dramatic change shows that naive use of a uniform prior in high dimensional problems can have unforeseen consequences whereas adding a prior only makes a small change in the birthday problem.

Section 4 treats the matching problem. A deck of cards labelled $1, 2, \dots, n$ is shuffled. What is the chance that a card labelled i is in position i for some i ? Classically, the chance of at least one match is $1 - 1/e$ to very good approximation, assuming that the cards are uniformly mixed. Here we first argue that the classical answer is also a Bayesian answer for some natural priors. We then develop some priors that incorporate notions of arrangements that are not uniform, and assess the chance of a match using Monte Carlo Markov chain methods.

The birthday problem, coupon collectors problem, and matching problem are three principal examples of Feller's great Volume I. Our examples show that a Bayesian view can lead to new perspectives on these classical problems. In the final section we review available Bayesian work on some other basic probability tools and point to some open problems. The appendix contains some standard facts about the Dirichlet distribution which are used in our treatment.

2. The Birthday Problem

We begin with a summary of the classical treatment of the birthday problem. Following this, a Bayesian analysis is given for a uniform prior, a symmetric Dirichlet prior, and then a more general Dirichlet prior.

2.1 Classical birthdays. Von Mises (1932) introduced the birthday problem: If k balls are dropped independently and uniformly into n boxes, the chance that at least one box contains two or more balls is

$$P(\text{at least one match}) = P(\text{match}) = 1 - P(\text{no matches}) = 1 - \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right).$$

The following asymptotic approximation is useful.

PROPOSITION 2.1 *If n and k are large in such a way that: $\binom{k}{2}/n \rightarrow \lambda$ then in the classical birthday problem:*

$$P(\text{match}) \cong 1 - e^{-\lambda}$$

PROOF. Write

$$\prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) = \exp\left(\sum_{i=1}^{k-1} \log\left(1 - \frac{i}{n}\right)\right)$$

Expand the log using $\log(1-x) = -x + O(x^2)$ to see that the exponent is

$$-\frac{\binom{k}{2}}{n} + O\left(\frac{k^3}{n^2}\right). \quad \square$$

REMARKS.

1. Setting $p = 1 - e^{-\lambda}$ and solving the resulting equation for k gives a good approximation to the k needed to obtain probability p of a match. For example, when $p = \frac{1}{2}$. We get $k = \sqrt{2n \log 2} \doteq 1.2\sqrt{n}$. When $n = 365$, $k = 1.2\sqrt{365} \doteq 22.9$.
2. A second way to see Proposition 1 (and somewhat more) is to prove that the number of matches has an approximate Poisson(λ) distribution. From this, $P(\text{No Match}) \doteq e^{-\lambda}$. For careful treatments using Stein's method see Arratia, Goldstein and Gordon (1989), Barbour, Holst and Janson (1992, pp.104-107) or Stein (1990) who treats the non-uniform case. A review of many variations of the birthday problem: varying cell probabilities, multiple matches, close matches, etc is in Aldous (1989, pp.108-110) and Diaconis and Mosteller (1989).
3. It is certainly possible for a Bayesian to accept classical computations based on a multinomial with all $p_i = 1/n$, by using a point prior at this p_i . Cases in which the underlying probability mechanism is unknown are treated next.

2.2 *Birthdays under a uniform prior.* In coin tossing a standard prior is the uniform on $[0, 1]$. A standard prior on the n -dimensional simplex Δ_n is the uniform distribution U where all vectors \tilde{p} have same density.

The probability of a match, averaged over (p_1, p_2, \dots, p_n) , represents the chance of success to a Bayesian statistician who has chosen the uniform prior. As is well known, (Bayes, 1764, Good, 1979, Diaconis and Efron, 1987) such a

uniform mixture of multinomials results in Bose-Einstein allocation of balls in boxes, each configuration, or composition (k_1, k_2, \dots, k_n) being equally likely with chance $1/\binom{k+n-1}{k}$. For this simple prior, it is again possible to do an exact calculation:

PROPOSITION 2.2 *Under a uniform prior on Δ_n*

$$P_u(\text{match}) = 1 - \prod_{i=1}^{k-1} \left(\frac{1 - i/n}{1 + i/n} \right). \quad (2.1)$$

If n and k are large in such a way that $k^2/n \rightarrow \lambda$, then

$$P(\text{match}) \cong 1 - e^{-\lambda}. \quad (2.2)$$

PROOF. Represent the uniform mixture of multinomials using Polya's urn (see Johnson and Kotz, 1975).

Thus consider an urn containing n balls labelled $1, 2, 3, \dots, n$. Each time a ball is chosen at random and replaced along with an extra ball having the same label.

The chance that the first k balls have different labels is

$$\frac{n-1}{n+1} \times \frac{n-2}{n+2} \cdots \times \frac{n-(k-1)}{n+(k-1)}.$$

This gives (2.1) and (2.2) follows by writing the product as the exponential of $\sum_{i=1}^{k-1} \log(1 - i/n)/(1 + i/n)$, expanding the logs and using $\log(1 - x) = -x + O(x^2)$ as in the proof of Proposition 1. \square

Thus in order to obtain a 50-50 chance of a match under a uniform prior k must be $.83\sqrt{n}$. When $n = 365$, this becomes $k = 16$, and for $k = 23$, $P_u(\text{match}) \doteq .75$.

The uniform prior allows some mass far from $(1/n, 1/n, \dots, 1/n)$ and such "lumpy" configurations make a match quite likely.

The uniform prior studied above is a special case of a symmetric Dirichlet prior D_c on Δ_n , with $c = 1$. We next extend the calculations above to a general c . For c increasing to infinity, the prior converges to point mass $\delta_{(1/n, 1/n, \dots, 1/n)}$ and thus gives the classical answer. When c converges to 0, D_c becomes an improper prior giving infinite mass to the corners of the simplex, thus for small c , the following proposition shows that matches are judged likely when $k = 2$.

PROPOSITION 2.3 *Under a symmetric Dirichlet prior D_c on Δ_n ,*

$$P_c(\text{match}) = 1 - \prod_{i=1}^{k-1} \frac{(n-i)c}{nc+i}.$$

This is proved like Proposition 2.2, using the Polya urn description of the Dirichlet, starting with nc balls in the urn.

In order for the probability of a match to be about $1/2$; $k_c \doteq \lceil 1/2 + \sqrt{2 \log(2)nc/(c+1)} \rceil$. The following table shows how k_c depends on c when $n = 365$:

c	.5	1	2	5	20	∞
k_c	14	17	19	21	23	23

The next section treats a similar computation with a general Dirichlet prior for which no neat formula is available thus making Poisson approximation necessary.

2.3 Birthdays with general Dirichlet priors. Consider the birthday problem in the classroom setting described above in a class with students born the same year. Should one suppose that they are uniformly distributed in that year? A first reply can be motivated by the additional fact that in recent years there are an increasing number of induced births most often performed on weekdays, when a full staff is available. This decreases to about 70% the rate of births on weekend days as compared to weekdays. Less precisely there may also be lunar and seasonal trends.

If the students are judged exchangeable then de Finetti's theorem is in force: The birthdays may be regarded as chosen with probabilities $\tilde{p} = (p_1, p_2, \dots, p_n)$ where p is considered to have some distribution.

We carry out the calculation for a Dirichlet prior with parameter $\tilde{a} = (a_1, a_2, \dots, a_n)$ with $a_i > 0$. The next result shows that the number of matches has a Poisson distribution. Following this we specialize the choice of \tilde{a} to values that fit the considerations above.

PROPOSITION 2.4 *Consider a Dirichlet prior $D_{\tilde{a}}$ on Δ_n with $\tilde{a} = (a_1, a_2, \dots, a_n)$ and $A = a_1 + a_2 + \dots + a_n$. If k and $n \rightarrow \infty$ in such a way that:*

$$\frac{\binom{k}{2}}{A(A+1)} \sum_{i=1}^n a_i(a_i+1) \rightarrow \lambda,$$

then the number of boxes with two or more balls has a limiting Poisson(λ) distribution.

PROOF. Let $\tilde{K} = (K_1, K_2, \dots, K_n)$ be the box counts after dropping k balls into n boxes using the Multinomial/Dirichlet law with parameter $\tilde{a} = (a_1, a_2, \dots, a_n)$. Then

$$P_{\tilde{a}}(K_1=k_1, K_2=k_2, \dots, K_n=k_n) = P_{\tilde{a}}(\tilde{k}) = P(Y_1=k_1, \dots, Y_n=k_n \mid \sum Y_i=k) \tag{2.3}$$

where the Y_i 's are independent negative binomials (a_i, p_i) with distribution

$$P(Y_i = j) = \binom{j-1}{a_i-1} p_i^{a_i} (1-p_i)^{j-a_i}.$$

For $a_i \geq 1$, the laws of Y_i 's are log-concave. Thus from Joag-Dev and Proschan (1983), the law of (K_1, K_2, \dots, K_n) is negatively associated, that is: if $f, g : \mathbb{N}^n \rightarrow \mathbb{R}$ are bounded and increasing in each coordinate,

$$E[f(\tilde{K})g(\tilde{K})] \leq E(f(\tilde{K}))E(g(\tilde{K})).$$

Let $Z_i = \mathbb{I}_{\{K_i \leq 2\}}$. This is an increasing function and so (Z_1, Z_2, \dots, Z_n) are negatively associated. Now Barbour, Holst and Janson (1992) provide a Poisson approximation for $W = Z_1 + Z_2 + \dots + Z_n$ through their Corollary 2.c.2, with bounds on the error. Let $\lambda = E(W)$, $\sigma^2 = var(W)$, they show:

$$\|\mathcal{L}(W) - Po_\lambda\|_{T.V.} \leq (1 - e^{-\lambda})(1 - \sigma^2/\lambda)$$

To evaluate means and variances we denote as before $x_{(j)} = x(x+1)\dots(x+j-1)$ and use:

$$P_{\tilde{a}}(k_1, \dots, k_n) = \frac{\prod_{i=1}^n (a_i)_{(k_i)}}{A_{(k)}} \binom{k}{k_1 \dots k_n}.$$

$$P(K_i = j) = \frac{(a_i)_{(j)}(A - a_i)_{(k-j)} \binom{k}{j}}{A_{(k)}}.$$

$$\begin{aligned} \pi_i &= P(\text{Box } i \geq 2) = 1 - P(\text{Box } i \leq 1) \\ &= 1 - \frac{(A - a_i)_{(k)}}{A_{(k)}} - \frac{(a_i)_i (A - a_i)_{(k-1)}}{A_{(k)}}. \\ \pi_{ij} &= P(\text{Box } i \geq 2 \text{ and Box } j \geq 2) \\ &= 1 - \{1 - \pi_i + 1 - \pi_j - P((0, 0), (0, 1), (1, 0), (1, 1))\} \\ &= \pi_i + \pi_j + \frac{(A - a_i - a_k)_{(k)}}{A_{(k)}} + \frac{a_i k (A - a_i - a_k)_{(k-1)}}{A_{(k)}} \\ &\quad + \frac{a_j k (A - a_i - a_k)_{(k-1)}}{A_{(k)}} + \frac{a_i a_j k (k-1) (A - a_i - a_k)_{(k-2)}}{A_{(k)}} - 1, \end{aligned}$$

with $\lambda = \pi_1 + \pi_2 + \cdots + \pi_n$ and $\sigma^2 = \sum_{i=1}^n \pi_i(1 - \pi_i) + \sum_{i \neq j} (\pi_{ij} - \pi_i \pi_j)$. □

REMARKS.

1. It follows under the assumptions of Proposition 4 that

$$P_{\hat{a}}(\text{match}) \doteq 1 - e^{-\lambda};$$

for a probability of approximately 1/2 we need:

$$k = 1.2 \sqrt{\frac{A(A+1)}{\sum a_i(a_i+1)}}.$$

This reduces to the cases treated above when all a_i 's are equal to a constant c .

2. Consider our class example with $n = 364 = 7 \times 52$ for convenience. We construct a 2 parameter family of Dirichlet priors writing $a_i = A\pi_i$, with $\pi_1 + \pi_2 + \cdots + \pi_n = 1$. Assign weekdays parameter $\pi_i = a$, weekends $\pi_i = \gamma a$, with $260a + 104\gamma a = 1$. Here γ is the parameter 'ratio of weekends to weekdays', (roughly we said $\gamma \doteq .7$) and A measures the strength of prior conviction. From Proposition 4, the chance of a match is approximately

$$1 - e^{-\lambda}, \text{ with } \lambda = \frac{\binom{k}{2}}{(A+1)} \{a^2 A(260 + 104\gamma^2) + 1\}, \quad a = \frac{1}{260 + 104\gamma}.$$

We give some values to show the impact of varying A and γ when $n = 364$. The following table shows the value of k needed to have probability approximately 1/2 of a match:

A^γ	.5	.7	1
1	2.2	2.2	2.2
364	16.1	16.3	16.4
728	18.4	18.6	18.8
∞	22.2	22.4	22.6

Small values of A force strong concentrations of the prior on the lumped values. The value $A = 364$ is roughly comparable to a uniform prior.

Here believable variations in γ make small difference to the conclusions. The main change comes from varying A . We conclude that for the birthday problem, believable uncertainties only change classical conclusions by a small amount.

3. We have proved a Poisson approximation for more general priors. If n and k are large, with

$$P \left\{ \binom{k}{2} \sum_{i=1}^n p_i^2 \leq \lambda \right\} \implies F(\lambda) \text{ weakly,}$$

then the number of matches is approximately an F mixture of Poisson (λ) variables. In the examples above, F is a point mass at λ . We have also used tree like priors which have F continuous. These results are attractive in showing that the only feature of the prior that matters is the length of the vector $(p_1, p_2, \dots, p_n) : \sum p_i^2$. Efron (1993) has developed some technology for assigning meaningful priors to one function of many parameters.

4. A different approach to the proof of proposition 2.4 and numerical computation, uses the conditional representation (2.3). From here, Le Cam's method (see Holst(1978)) may be used, perhaps combined with Edgeworth corrections to prove limit theorems or get numerical approximations. These approaches remain useful when some $a_i < 1$. Then the negative binomial is no longer log-concave, so the proof based on Barbour, Holst and Janson's version of Stein's method breaks down.
5. There is another approach to using Stein's method which is also useful for completely general a_i . This is based on the use of an exchangeable pair as in Stein (1986). This was our first approach to proof. Since the same argument can be used in the coupon collector's problem, we detail it here. For $\widetilde{K} = (K_1, \dots, K_n)$ with a Dirichlet/multinomial law from parameter \widetilde{a} , we construct \widetilde{K}' so that the pair $(\widetilde{K}, \widetilde{K}')$ is exchangeable. The argument is simple: regard \widetilde{K} as a replacement of k balls in n boxes, choose a ball uniformly at random and delete it, giving \hat{k} , then add a ball to box i with probability $(a_i + \hat{k}_i)/(k + A - 1)$. It is easy to check that this works, and then the techniques of Stein (1986) are in force.
6. We cannot leave off discussion of the birthday problem without remarking that the classical computation underlying Proposition 1 yields the strongest finite form of de Finetti's basic representation theorem; see Diaconis and Freedman (1981).

3. The Coupon Collector's Problem

We begin by reviewing the classical analysis of the coupon collector's problem. This is followed by a closed form answer for the Bayesian analysis

under a uniform prior and approximate analyses under a general Dirichlet prior.

3.1 *Classical coupons.* Laplace (1812) introduced the original coupon collector's problem: if k balls are dropped uniformly and independently into n boxes, the chance that all boxes are covered is

$$P(\text{cover}) = \sum_{j=0}^n (-1)^j \binom{n}{j} (1 - j/n)^k$$

Useful approximations to this expression were given by Von Mises (1939), see also Erdős and Renyi (1961) who show that if

$$k \sim n \log n + \theta n, \quad -\infty < \theta < \infty$$

$$\text{then} \quad P(\text{cover}) \sim e^{-e^{-\theta}}. \quad (3.1)$$

For example, $P(\text{cover}) \doteq 1/2$ for $\theta = 0.366$. When $n = 365$ this gives $k = 2287$ or as Feller (1968) puts it: in a village of 2300 inhabitants it is about even odds that every birthday is covered. For $P(\text{cover}) = .99$, $\theta = 4.60$. When $n = 365$, this gives $k = 3832$.

Modern approaches to studying the coupon collector's problem introduce a random variable W , the number of empty cells when k balls are dropped into n boxes. One observes

$$E(W) = n(1 - 1/n)^k$$

and then shows that if n and $k \rightarrow \infty$ in such a way that $E(W) \rightarrow \lambda$, then

$$P(W = j) \rightarrow e^{-\lambda} \lambda^j / j!. \quad (3.2)$$

From this $P(\text{cover}) = P(W = 0) = e^{-\lambda}$. Writing $k = n \log n + \theta n$ gives $\lambda = e^{-\theta}$, so (3.2) implies (3.1).

A rigorous version of the Poisson limit theorem with good error terms is in Barbour, Holst and Janson (1992). See also Rosen (1970) for the case of varying probabilities and Aldous (1992) for many variations and applications.

3.2 *Coupons under a uniform prior.*

PROPOSITION 3.5 *Under a uniform prior on Δ_n*

$$P(\text{cover}) = \binom{k-1}{n-1} / \binom{n+k-1}{n-1}. \quad (3.3)$$

If n and k are large in such a way that $k/n^2 \rightarrow \theta > 0$ then

$$P(\text{cover}) = e^{-1/\theta} (1 + O(1/n)). \tag{3.4}$$

PROOF. Under the uniform prior, all configurations are equally likely with probability given by the denominator in (3.3). The numerator counts the number of configurations which cover.

The limiting approximation in (3.4) is easily proved by simplifying the ratio, exponentiating and expanding $\log(1 - x)$ as in Proposition 1 above. \square

REMARKS. Using a uniform prior radically changes things: the classical result (3.1) shows a sharply changing probability in a neighbourhood of $n \log n$. The uniform prior gives a gradually changing probability in a neighbourhood of n^2 . For example, $P_u(\text{cover}) = 1/2$ for $\theta = 1.44$. When $n = 365$ this will need $k = 191,844$. For $P_u(\text{cover}) = .99$, $\theta = 99.5$ and $n = 365$ gives $k = 13,255,888$.

It is not hard to understand these results: under a uniform prior, the minimum p_i is of order $1/n^2$, so it takes k of order n^2 to have a good chance of dropping in at least one ball in each box.

3.3 *General Dirichlet prior.* The cover probability will be derived from a Poisson approximation due to Barbour, Holst and Jansen (1992,page 129). Let $\bar{a} = (a_1, a_2, \dots, a_n)$ with $a_i > 0$, $A = a_1 + a_2 + \dots + a_n$. Let W be the number of empty cells under a Dirichlet multinomial allocation. Then for any $B \subset \{0, 1, 2, \dots\}$, they show

$$|P\{W \in B\} - Po_\lambda(B)| \leq (1 - e^{-\lambda})(1 - \sigma^2/\lambda) \tag{3.5}$$

with $\lambda = E(W)$, $\sigma^2 = var(W)$. These are explicitly given by letting $\lambda = \sum p_i$ and $\sigma^2 = \sum_i p_i(1 - p_i) + \sum_{i \neq k} (p_{ik} - p_i p_k)$ and writing:

$$p_i = \frac{\binom{-(A-a_i)}{k}}{\binom{-A}{k}}, \quad p_{ik} = \frac{\binom{-(A-a_i-a_k)}{k}}{\binom{-A}{k}}, \quad i \neq k.$$

The inequality (3.5) shows that the Poisson approximation holds provided σ^2 is close to λ . Note that W is a sum of negatively correlated indicators so that $\sigma^2 < \lambda$.

As an example, take the symmetric case where $a_i = c$, $1 \leq i \leq n$. For c a fixed integer ,

$$p_i = p = \frac{(cn - 1) \dots (cn - c)}{(cn + k - 1) \dots (cn + k - c)}, \quad \lambda = np.$$

Now np has a finite limit if and only if $k \sim \theta n^{(c+1)/c}$. Notice that this agrees with the results of the previous section when $c = 1$. The same asymptotics hold for any fixed $c > 0$ using (3.5). We may conclude:

COROLLARY 3.1 *Under a Dirichlet (c, c, \dots) allocation with $c > 0$ fixed*

$$P_c(\text{cover}) \sim e^{-(c/\theta)^c} \quad \text{for } k \sim \theta n^{(c+1)/c} \quad (3.6)$$

REMARKS.

1. We see that the parameter c enters in a complex fashion and clearly makes a tremendous difference.
2. When $n = 365$, we compute k_c needed to have $P(\text{cover}) \doteq 1/2$:

c	1	2	3	4	5	10	20	∞
k_c	191,844	16,000	8,841	6,994	4,555	3176	2685	2297

We do not have a reasonable conjecture for the Bayesian probability of covering under more general priors (as in Remark 4 in Section 2).

4. The Matching Problem

We treat the matching problem and show that the classical computation has a natural Bayesian interpretation under a uniform prior. We then work with informative priors, first by Monte Carlo and then by analytic methods.

4.1 *A classical and Bayesian matching.* The matching problem asks for the number of fixed points in a uniformly distributed random permutation. It was stated and solved by Montmort (1708). In modern notation, if W is the number of fixed points of a random permutation of n objects, then for any $B \subset \{0, 1, 2, \dots, \}$

$$|P(W \in B) - P_{o_1}(B)| \leq 2^n / (n+1)!. \quad (4.1)$$

There is a “match” if $W \geq 1$. It follows that:

$$P(\text{match}) = P(W \geq 1) \sim 1 - 1/e \quad (4.2)$$

For a history of the matching problem and many variations (decks of cards with repeated values, almost matches etc.) see Takacs (1980).

The matching problem can be stated in several scenarios: a hat checker returning hats having lost the checks, a teacher returning homeworks in a

random manner, mixed up letters put into envelopes at random. In these variations, the uniform distribution is suspect.

Consider demonstrating the matching problem with a borrowed deck of cards. The cards are turned up one at a time calling ‘ace of spades’, ‘two of spades’, etc. in a fixed order. If the deck is believed well shuffled, then a Bayesian will have a uniform prior over all $n!$ orders and the classical computation described above is also the Bayesian computation. Suppose next that the first few cards turned over are hearts. You may well think that the cards were used in a game where the final arrangement clumps the suits. Half a dozen other patterns could easily provoke a similar reaction. These reactions are not compatible with a uniform prior. In the next few sections we explore alternatives to the uniform prior.

4.2 *Clumping*. One natural alternative to uniformity is that the cards of like suit or value tend to be close together. For π a permutation of n objects, consider the statistic:

$$T(\pi) = \#\{i : |\pi_{i+1} - \pi_i| = 1, 1 \leq i \leq n, \pi_{n+1} = \pi_1\} \tag{4.3}$$

This measures the number of cards adjacent in the original ordered deck which are still adjacent. A variety of other clumping statistics will be discussed below.

A nonuniform distribution on permutations which permits varying probabilities for such clumps may be specified by putting an exponential family through T :

$$P_\theta(\pi) = Ze^{-\theta T(\pi)}, 0 < \theta < \infty. \tag{4.4}$$

Here θ is a parameter and Z is a normalizing constant. Note first that $\theta = 0$ gives the uniform distribution. For $\theta > 0$, the distribution P_θ concentrates on permutations with large values of T . As $\theta \rightarrow \infty$, P_θ concentrates on $2n$ specific permutations: the cyclic shifts of the identity and their reversals. To calibrate θ , note that under P_0 , T has an approximate Poisson(2) distribution. It follows that, for fixed θ and large n , under P_θ , T is approximately Poisson($2e^\theta$). Thus, if you think “there are about k adjacent values”, then $\theta = \log(k/2)$. Of course, a prior could also be used for θ .

The following table shows the probability of a match with prior P_θ when $n = 52$ for various values of θ . We note that the prior hardly changes things. As θ tends to infinity $P_\theta(\text{match})$ tends to $1/2$.

θ	0	0.5	1	2	5	10	20	100	200
$E(T)$	2.0	3.2	5.1	11.8	34.5	34.2	31.3	31.3	32.6
P_θ (Match)	.64	.62	.61	.62	.45	.46	.56	.52	.51

The foregoing table is based on an implementation of the Metropolis algorithm for sampling from P_θ . There is one subtlety here which may be of independent interest. A naive version of the Metropolis algorithm uses a base chain. This will here be defined by specifying a set \mathcal{S} of permutations (in this case all transpositions). The walk takes place on the set of all $n!$ permutations. If the process is currently at the permutation π , choose $s \in \mathcal{S}$ uniformly and compute $\sigma = s\pi$. If $T(\sigma) \geq T(\pi)$, the process moves to σ . If $T(\sigma) < T(\pi)$, flip a coin with probability of heads $P_\theta(\sigma)/P_\theta(\pi)$. If the coin comes up head, the process moves to σ . If the coin comes up tail, the process stays at π . This describes the Metropolis algorithm. Hammersley and Handscomb (1965) or Diaconis and Saloff-Coste (1998) give further details.

To get the right answer from this simulation procedure it is important to realize that the statistic T has some invariance properties. It takes its maximum value n at the identity permutation and at all cyclic shifts and their reversals. Thus P_θ has $2n$ ‘peaks’.

If one used the Metropolis algorithm based on local changes, (e.g., \mathcal{S} is all transpositions), it takes time of order e^n to escape from a peak. The statistic of interest ‘match/no match’ is not invariant under the symmetries of T , one could easily report a wrong answer. Indeed, since $e^3 \doteq 20$ and $2^{10} \doteq 10^3$, then $e^{52} \doteq e \times 20^{17} \doteq e \times 128 \times 10^{20} \doteq 3 \times 10^{22}$ so that in this case an order of 10^{22} steps would have been necessary just to escape from one peak.

We ran the Metropolis algorithm based on random transpositions coupled with cyclic shifts and reversals. After proposing a transposition and choosing to make that step, a random cyclic shift is made and with probability $1/2$, a reversal is made.

Each entry in the foregoing table was computed using 10,000 steps of a Monte Carlo simulation based on this refined version of the Metropolis algorithm. Figure 1 shows an example of one run for $\theta = 200$, the figure shows the evolution of both T and the number of fixed points. We see that the first 1000 steps or so do not suffice to stabilize the distribution of T . The simulations were started at random permutations.

The discussion above has been for the statistic T of (4.3). There are several other natural statistics, for example $W(\pi) = \sum_{i=1}^n |\pi(i+1) - \pi(i)|$. One could also consider statistics to identify clumping of cards of like value and suit. These can be combined in the form $\sum w_i P_{\theta_i, T_i}$ or using $\sum \theta_i T_i$ in the exponent. It seems plausible that any such complex mixture would lead to an approximate uniform distribution. We do not know of any theoretical justification for this belief.

The next two sections explore other models for non-uniform distributions.

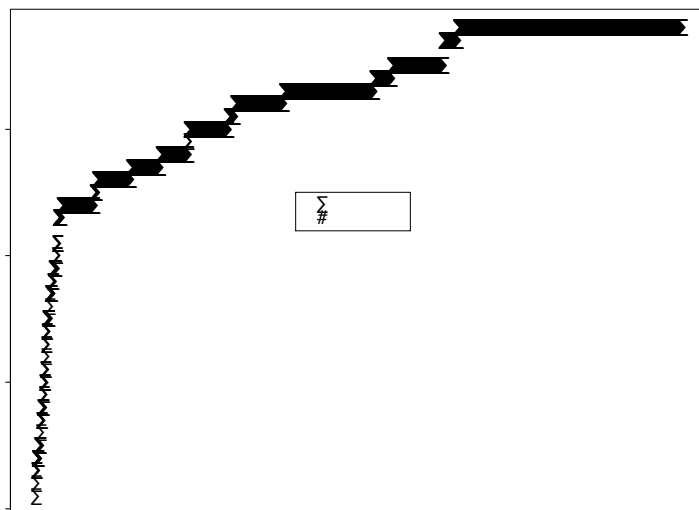


Figure 1: 10,000 Metropolis Steps

4.3 *Badly shuffled cards.* Suppose a deck of cards is initially in known order $1, 2, 3 \dots n$ and is riffle shuffled k times. We take each shuffle according to the Gilbert-Shannon-Reeds (GSR) distribution;

- The deck is cut in half according to a binomial distribution.
- Then the cards are interleaved by being dropped from the two hands alternately according to the following scheme:
 - If at this stage the left hand has L cards in it, and the right hand R cards,
 - the chance that the next card comes from the left hand will be $L/(L + R)$
 - the chance that the next card comes from the right will be $R/(L + R)$

This specifies a probability distribution over single shuffles which is actually uniform. Experiments reported in Diaconis (1988) show that the GSR distribution matches the way people ordinarily shuffle. See Bayer and Diaconis (1992) for more background.

Let P_k be the distribution on permutations after k independent GSR shuffles. Bayer and Diaconis (1992) show that as $k \rightarrow \infty$, P_k converges to uniform. This happens (roughly) for $k = (3/2) \log_2 n$. Let $W(\pi)$ denote the number of fixed points of the permutation π . The distribution of W under P_k is studied in Diaconis, McGrath and Pitman (1994) who prove:

THEOREM 4.1 *Under k GSR shuffles, the number of fixed points satisfies:*

$$E_k(W) = 1 + 2^{-k} + 2^{-2k} + \dots + 2^{-(n-1)k}. \quad (4.5)$$

For large n , and fixed k , W has an approximate negative binomial distribution, for $j = 0, 1, \dots$,

$$P(W = j) \sim \binom{j+k-1}{j} p^j (1-p)^k, \quad f = 2^k, p = 2^{-k}. \quad (4.6)$$

In particular, $P_k(\text{match}) \sim 1 - (1 - 2^{-k})^{2^k}$.

REMARKS. After $k = 1$ shuffle, $E_1(W) \sim 2$, $P_1(\text{match}) = 3/4$, as opposed to uniform where $P_\infty \sim 1 - 1/e$. Already at 2 shuffles: $P_2(\text{match}) \doteq .6836$, $P_\infty(\text{match}) \doteq .6321$. After only a few shuffles, the chances agree. Thus again, using a non-uniform prior on permutations only changes the answer slightly for the matching problem.

4.4 Cayley distance and Ewen's sampling formula. A useful family of non-uniform distributions on permutations is provided by Mallow's models:

$$Q_\theta = c(\theta) \theta^{d(\pi, \pi_0)}, \quad 0 < \theta \leq 1, \quad (4.7)$$

where π_0 is a fixed center (often taken to be the identity) and d is a metric on permutations. Thus, if $\theta = 1$, Q_θ is uniform. For $0 < \theta < 1$, Q_θ is largest at π_0 and falls off exponentially.

In this section we take $d = \text{Cayley's distance}$: $d = \text{minimum number of transpositions to bring } \pi \text{ to } \pi_0$.

See Diaconis (1988) or Marden (1995) for references and background. In particular it is proved in Diaconis (1988) that d is invariant under re-labelling and satisfies:

$$d(\pi, \pi_0) = n - \# \text{ cycles in } (\pi \pi_0^{-1})$$

Further, for Cayley distance it is known that the normalizing constant is given by

$$c(\theta) = \prod_{i=1}^n \frac{1}{1 + \theta(i-1)}.$$

PROPOSITION 4.6 *Under Mallow's model defined with the Cayley distance and with $\pi_0 = \mathbb{I}$, the number of fixed points satisfies:*

$$E_\theta(W) = \frac{n}{1 + \theta(n - 1)}. \tag{4.8}$$

For any $A \subset \{0, 1, 2, \dots, \}$,

$$|Q_\theta(W \in A) - P_{O_{1/\theta}}(A)| \leq \frac{1}{1 + \theta n} \left(\frac{1}{\theta} + \frac{n}{\frac{1}{\theta} + n - 1} \right). \tag{4.9}$$

PROOF. (4.8) will follow from the group theoretic discussion below. For (4.9), write

$$Q_\theta(\pi) = \prod \frac{1}{1 + \theta(i - 1)} \theta^{n - c(\pi)}, \text{ where } c(\pi) = \# \text{cycles in } \pi.$$

By changing from θ to $\tau = 1/\theta$, we can rewrite Q :

$$Q_\tau(\pi) = \tau^{c(\pi)} / \Gamma(\tau + n)$$

and retrieve a measure on permutations studied by Arratia, Barbour and Tavaré (1992, p.521). □

REMARKS. Arratia, Barbour and Tavaré (1992) and many other writers cited by them derive properties of many further functionals of π under Q_θ . One further property is given below.

If P is any probability on permutations, we may want to say “ P only holds roughly”. One way to do this is to convolve P with a measure such as Q_θ of (4.7). Here $Q_\theta * P(\pi) = \sum P(\pi\sigma^{-1})Q_\theta(\sigma)$ is close to $P(\pi)$ when θ is close to zero. The following result shows how the expected value in the matching problem changes under such a change in priors.

PROPOSITION 4.7 *Let P be a probability on the permutation group. Let Q_θ be defined by (4.7) with $\pi_0 = id$ and d as Cayley distance. Then, the expected number of fixed points W under $Q_\theta * P$ is:*

$$1 + \frac{(1 - \theta)}{(1 - \theta) + n\theta}(\mu - 1)$$

with μ the expected number of fixed points under P .

PROOF. Let $\rho(\pi)$ be the usual n -dimensional representation of π via permutation matrices. Then for any probability P , the expected number of fixed points is $\mu(P) = \text{Tr}(\hat{P}(\rho))$, where $\hat{P}(\rho) = \sum P(\pi)\rho(\pi)$. Now ρ can be decomposed as a direct sum $\rho = \rho_1 + \rho_2$ of a 1-dimensional trivial representation and an $n - 1$ dimensional irreducible representation (see Diaconis, 1988, chapter 2, for background).

The probability $Q_\theta(\pi) = Q_\theta(\sigma^{-1}\pi\sigma)$ is constant on conjugacy classes, so for any irreducible representation ρ_1 , $\hat{Q}(\rho_1) = c\mathbb{I}$ for a constant c . Thus, in an appropriate basis, $\hat{Q}(\rho)$ is a diagonal matrix with a one in the $(1, 1)$ entry and c in the $n - 1$ other diagonal entries. Further, in this basis, $\hat{P}(\rho)$ is block diagonal with a one in the $(1, 1)$ entry, zeros elsewhere in the first row and column, and some $(n - 1) \times (n - 1)$ block at the bottom right. By invariance of the trace under change of basis, the trace of this $(n - 1) \times (n - 1)$ block must be $\mu(P) - 1$.

Now $\widehat{Q * P}(\rho) = \hat{Q}(\rho)\hat{P}(\rho)$ has trace $1 + c(\mu - 1)$. Finally using the argument above with P replaced by Q_θ , $c = (\mu(Q) - 1)/(n - 1)$. Now $\mu(Q_\theta) = n/(1 + \theta(n - 1))$ is easy to see directly, for $\mu(Q_\theta)$ is $n \times Q_\theta(\pi(1) = 1)$ by invariance, and

$$\begin{aligned} Q_\theta(\pi(1) = 1) &= \prod_{i=1}^n \frac{1}{1 + \theta(i - 1)} \sum_{\pi: \pi(1)=1} \theta^{n-c(\pi)} \\ &= \prod_{i=1}^n \frac{1}{1 + \theta(i - 1)} \prod_{i=1}^{n-1} (1 + \theta(i - 1)) = \frac{1}{\theta(n - 1)}. \end{aligned}$$

Thus

$$\mu(Q_\theta) = \frac{n}{1 + \theta(n - 1)} \quad \text{and} \quad c = \frac{(1 - \theta)}{(1 - \theta) + n\theta}. \quad \square$$

From the analyses above, we see that the prior only makes a small difference in the analysis of the classical matching problem. Of course, there are differences; the rate of convergence to Poisson appears to be very different if we compare the $1/n$ rate in (4.9) to $2^n/(n + 1)!$.

The matching problem is often given in terms of letters in envelopes or hats returned from a cloakroom. Bayesian analyses of the hat/cloakroom version depend on the situation. For instance, if the cloakroom attendant did not keep tickets, the chances are there would be special hats that would be remembered, thus favouring permutations with a certain number of fixed points (the specially noticeable hats). The secretary might remember certain of the letters, because of known clients, or postal codes and so on.

5. Final Remarks

We have collected here a brief review of some other Bayesian versions of standard probability problems. There are several recent textbooks which survey the active development of the Bayesian approach to statistics. Berger (1995), Bernardo and Smith (1995), Robert (1993) and Schervish (1995) develop axioms and tools available. We restrict our attention to our theme and briefly review Bayesian versions of standard probability problems.

5.1 *Laws of large numbers and de Finetti's theorem.* Let X_1, X_2, \dots be an infinite sequence of binary exchangeable random variables. Let $S_n = X_1 + X_2 + \dots + X_n$, de Finetti's law of large numbers asserts that

$$S_n/n \text{ converges almost surely to a limit } \theta \in [0, 1]. \quad (5.1)$$

This θ is a random variable whose law we will call μ . The law μ of (5.1) uniquely determines the law of $\{X_i\}_{i=1}^\infty$ through de Finetti's representation:

$$P\{X_1 = e_1, \dots, X_n = e_n\} = \int_0^1 x^{S_n} (1-x)^{n-S_n} \mu(dx)$$

where $S_n = e_1 + e_2 + \dots + e_n$ and $\{e_i\}_{i=1}^\infty$ is any binary sequence.

The literature on de Finetti's theorem and its extensions is reviewed in Aldous (1985), Diaconis (1988), and Diaconis and Freedman (1984) and Lauritzen (1988).

5.2 *Central limit theorem.* Let X_1, X_2, \dots be an infinite sequence of binary exchangeable random variables. The central limit theorem states that, for θ the limit of (5.1)

$$P\left\{n^{-1/2}(S_n - n\theta) \leq x\right\} \longrightarrow \int_0^1 \Phi((p(1-p))^{1/2}x) \mu(dp). \quad (5.2)$$

Thus, as expected, one has fluctuations of order $n^{1/2}$ about the random mean $n\theta$; the shape of the fluctuations being a scaled mixture of Gaussians, with the mixing measure determined by μ .

Again, this theorem admits sweeping generalizations giving a Bayesian version of the general central limit problem. Here mixtures of normals are replaced by mixtures of general infinitely divisible laws. Perhaps the best recent results, and a survey of previous work appears in Fortini, Ladelli and Regazzini (1996).

5.3 *Comparisons.* It can be confusing to compare conclusions from the classical and Bayesian versions of the central limit theorem. Classically,

if one flips a fair coin n times, one expects Gaussian fluctuations of size $n^{1/2}$ about $n/2$. The Bayesian result (5.2) looks much more complex. The difference is this:

the Bayesian analysis is not assuming the coin is “fair” but rather building on the fact that the long term frequency is “unknown”.

Consider a classical statistician contemplating repeated flips of a drawing pin (thumb tack), what predictions could be made about the next n flips? The question doesn’t make sense classically, one would have to get some data and try things out.

With substantial data the Bayesian and frequentist predictions agree: the data swamps the prior and the posterior is basically point mass at the observed frequency of successes. With moderate amounts of data the prior still comes in: Diaconis and Freedman (1988) is a recent reference giving bibliographical pointers and bounds for this classical work of Laplace (1812). Beckett and Diaconis (1994) examine real flips of real thumb tacks and show that the basic exchangeability assumption is open to question.

Consider a Bayesian analysis of repeated trials, with a uniform prior. The number of successes in the next n trials is uniform on $\{0, 1, 2, \dots, n\}$. The central limit theorem says it has roughly $n^{1/2}$ fluctuations about $n\theta$, where θ is uniform on $[0, 1]$. This is not useful. Of course, with a tightly peaked prior, the situation would be different. Further the Bayesian central limit theorems are useful in justifying the use of Normal models in applications.

5.4 Some further examples. There is much further work to be done incorporating uncertainty and prior knowledge into basic probability calculations.

Our interest in this area was stimulated by Scarsini (1988). His paper studies optimal play in a game of heads and tails where the player has an exchangeable prior for the outcomes. Chamberlain and Rothschild (1981) study the question “What’s the chance that my vote matters in an election with $2n$ other people?”. In a classical analysis, with each voter assumed to vote for one of two sides with probability $1/2$, the chance that one vote breaks a tie is around $(\pi n)^{-1/2}$. If a uniform prior is used the chance is only $1/n$.

DasGupta (1995) studies how the odds change in a game of craps if the chances for the die are not assumed known. Armero and Bararri (1996) studies features of the basic single server queue under an exponential model with a prior on the parameters.

L.J. Savage (1973) and Engel (1992) give Bayesian developments of problems like Buffon’s needle using subjectivist versions of Poincaré’s early efforts

on the “method of arbitrary functions”. This work shows that for some problems essentially any prior will lead to the same final answer. This in turn agrees with the classical distribution. This is typically a case for problems with sensitive dependence on initial conditions.

5.5 *Further problems.* This paper has only scratched the surface of a rich terrain. There are dozens of juicy examples in Feller’s book (not to mention volume 2). Sometimes priors matter, sometimes not. We have used conjugate priors. Sometimes only one feature of the prior matters. It seems worthwhile to make a catalog of examples to help us understand Bayesian robustness.

Acknowledgments. This work had its origin in conversations with Marco Scarsini. Distance has prevented us from working this out jointly. A non-technical version of this paper (Diaconis and Holmes, 1996) was presented at the workshop in honour of Dick Jeffreys. We thank Eugenio Regazzini, Dominico Constantini, Donato Cifarelli and Maria Carla Galovotti for their help. Abner Shimoni and Sandy Zabell raised interesting questions incorporated herein. Anirban DasGupta made wonderfully useful comments on the original manuscript.

References

- ALDOUS, D. (1985). *Exchangeability and Related Topics* (École d’Été de Probabilités de Saint-Flour, XIII — 1983). Lecture Notes in Mathematics **1117**, Springer Verlag, New York.
- ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer Verlag, New York.
- ARMERO, C. and BAYARRI, M.J. (1996). Bayesian questions and answers in queues. In *Bayesian Statistics Vol. 5* (Alicante, 1994), J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds., Oxford University Press, Oxford, 3-23.
- ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method, *Ann. Probab.*, **17**, 9-25.
- ARRATIA, R., BARBOUR, A.D. and TAVARÉ, S. (1992). Poisson process approximations for the Ewens sampling formula, *Ann. Appl. Probab.*, **2**, 519-535.
- BARBOUR, A., HOLST, L. and JANSEN, S. (1992). *Poisson Approximations*. Cambridge University Press, Cambridge.
- BAYER, D. and DIACONIS, P. (1992). Trailing the dovetail shuffle to its lair, *Ann. Appl. Probab.*, **2**, 294-313.
- BAYES T. (1764). An essay towards solving a problem in the doctrine of chances, *Philos. Trans. Roy. Soc. London* (for 1763), **53**, 370-418.
- BECKETT, L. and DIACONIS, P. (1994). Spectral Analysis for Discrete Longitudinal Data, *Adv. Appl. Math.*, **103**, 107-128.
- BERNARDO, J.M. and SMITH, A.F.M. (1994). *Bayesian Theory*. Wiley, New York.

- CHAMBERLAIN, G. and ROTHSCHILD, M. (1981). A note on the probability of casting a decisive vote. *J. Econom. Theory*, **25**, 152-162.
- DASGUPTA, A. (1995). Winning at craps with a random die, Tech. Rep. #95-24, Dept. of Statistics, Purdue University.
- DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS Lecture Notes and Monograph Series **11**, Institute of Mathematical Statistics, Hayward, CA.
- DIACONIS, P. (1988). Recent progress on de Finetti's notions of exchangeability. In *Bayesian Statistics Vol. 3* (Valencia 1987), J.-M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., Oxford University Press, Oxford, 111-125.
- DIACONIS, P. and EFRON, B. (1987). Probabilistic-geometric theorems arising from the analysis of contingency tables. In *Contributions to the Theory and Applications of Statistics: A volume in honour of Herbert Solomon*, A.E. Gelfand, ed., Academic Press, New York, 103-125.
- DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences, *Ann. Probab.*, **8**, 745-764.
- DIACONIS, P. and FREEDMAN, D. (1984). Partial exchangeability and sufficiency. In *Statistics: Applications and New Directions* (Proc. Indian Stat. Inst. Golden Jubilee International Conference on Statistics, Calcutta, 1981), J.K. Ghosh and J. Roy, eds. Indian Statistical Institute, Calcutta, 205-236.
- DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory, *Ann. Inst. Henri Poincaré*, **23**, 397-423.
- DIACONIS, P. and FREEDMAN, D. (1990). On the uniform consistency of Bayes estimates for multinomial probabilities, *Ann. Statist.*, **18**, 1317-1327.
- DIACONIS, P. and HOLMES, S. (1996). Are there still things to do in Bayesian Statistics? In *Erkenntnis* (special issue in honour of R.C. Jeffrey), **45**, 145-158.
- DIACONIS, P., MCGRATH, M. and PITMAN, J. (1994). Riffle shuffles, cycles and descents, *Combinatorica*, **66**, 379-386.
- DIACONIS, P. and MOSTELLER, F. (1989). Methods for studying coincidences *J. Amer. Statist. Assoc.*, **84**, 853-861.
- DIACONIS, P. and SALOFF-COSTE, L. (1998). What do we know about the Metropolis algorithm?, *Jour. Comput. Sys. Sci.*, **57**, 20-36.
- EFRON, B., (1993). Bayes and likelihood calculations from confidence intervals, *Biometrika*, **80**, 3-26.
- ENGEL, E.M., (1992). *A Road to Randomness in Physical Systems*. Springer Lecture Notes in Statistics **71**, Springer Verlag, New York.
- ERDÖS, P. and RÉNYI, A. (1961). On a classical problem of probability theory, *Mat. Kutato Int. Köz.*, **6**, 215-220.
- FELLER W. (1968). *An Introduction to Probability Theory and its Applications, Vol. I*, Wiley, New York.
- FORTINI, S., LADELLI, L. and REGAZZINI, E., (1996). A Central limit theorem for partially exchangeable random variables. *Theor. Probab. Appl.*, **41**, 353-374.
- GOOD, I.J., (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables, *Ann. Statist.*, **5**, 1159-1189.
- GOOD, I.J., (1979). Bayes' billiard table argument extended to multinomials, *Journ. Statist. Comput. Simulation*, **9**, 161-163.

- HOLST, L. (1979). Two conditional limit theorems with applications in sampling theory, *Ann. Statist.*, **7**, 551-557.
- HOLST, L. (1981). Some conditional limit theorems in exponential families, *Ann. Probab.*, **9**, 818-830.
- HOLST, L. (1986). On birthday, collectors', occupancy and other classical urn problems, *Internat. Statist. Rev.*, **54**, 15-27.
- JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications, *Ann. Statist.*, **11**, 286-295.
- JOHNSON, N. and KOTZ, S. (1977). *Urn Models and their Applications*, Wiley, New York.
- KOLCHIN, V., SEVAST'YANOV, B. and CHISTYAKOV, V. (1978). *Random Allocations*, Winston, Washington D.C.
- LAPLACE, P., (1812). *Theorie Analytique des Probabilités*, Courcier, Paris, (3rd ed. 1820 with supplements).
- LAURITZEN, S. (1988). *Extremal Families and Systems of Sufficient Statistics*, Springer Lecture Notes in Statistics **49**, Springer Verlag, New York.
- MARDEN, J.I. (1995). *Analysing and Modelling Rank Data*. Chapman and Hall, London.
- ROBERT, C.P. (1994). *The Bayesian Choice: a decision-theoretic motivation*, Springer Verlag, New York.
- ROSÉN, B. (1970). On the coupon collector's waiting time, *Ann. Math. Statist.*, **41**, 1952-1969.
- SAVAGE, L.J. (1973). Probability in science: a personalistic account. In *Logic, Methodology and Philosophy of Science. IV* (Bucharest, 1971), P. Suppes, L. Henkin, A. Joja and Gr.C. Moisil, North-Holland, Amsterdam-London, 417-428. Reprinted in *The Writings of Leonard Jimmie Savage: a memorial selection*, Amer. Statist. Assoc., Washington, 666-677.
- SCARSINI, M. (1988). On a simple sequential decision problem, *Statist. Neerlandica*, **42**, 29-35.
- SCHERVISH, M.J. (1995). *Theory of Statistics*, Springer Verlag, New York.
- STEIN, C. (1986). *Approximate Computation of Expectations*, IMS Lecture Notes and Monographs, Institute of Mathematical Statistics, Hayward, California.
- STEIN, C. (1990). Applications of Newton's Identities to a General Birthday Problem and to the Poisson Binomial Distribution. Tech. Report, Dept. of Statistics, Stanford University.
- TAKÁCS, L. (1980). The problem of coincidences, *Arch. Hist. Exact. Sci.*, **21**, 229-244.
- VON MISES, R. (1939). Über aufteilungen und besetzungswahrscheinlichkeiten, *Rev. Fac. Sci. Istanbul*, **4**, 145-163, Reprinted in *Selected Papers of R. Von Mises, Vol. 2*, Amer. Math. Soc., Providence, RI, 313-331.
- WILKS, S.S. (1962). *Mathematical Statistics*. Wiley, New York.
- ZABELL, S. (1982). W.E. Johnson's "sufficientness" postulate, *Ann. Statist.*, **10**, 1091-1099.

PERSI DIACONIS
 STATISTICS DEPARTMENT
 SEQUOIA HALL, 390 SERRA MALL
 STANFORD UNIVERSITY
 STANFORD, CA 94305-4065, USA

SUSAN HOLMES
 STATISTICS DEPARTMENT
 SEQUOIA HALL, 390 SERRA MALL
 STANFORD UNIVERSITY
 STANFORD, CA 94305-4065, USA
 AND INRA, BIOMÉTRIE, FRANCE
 E-mail: susan@stat.stanford.edu