496

496

PREDICTION INTERVALS WITH THE DIRICHLET PRIOR*

by

Gregory Campbell

and

Myles Hollander

Department of Statistics
Division of Mathematical Sciences
Purdue University
Mimeograph Series #496

June 1977

*Gregory Campbell is Assistant Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907. Myles Hollander is Professor, Department of Statistics, Florida State University, Tallahassee, FL 32306. The research was sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grants AFOSR-74-2581B and AFOSR-76-3109.

Prediction Intervals with the Dirichlet Prior

by

Gregory Campbell and Myles Hollander

ABSTRACT

Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_N$ be consecutive samples from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ (the real line $\mathcal{R}$ with the Borel σ-field $\mathcal{B}$) with parameter α. Typically, prediction intervals employ the previous observations $X_1,\ldots,X_n$ in order to predict a specified function of the future sample $Y_1,\ldots,Y_N$. Here one- and two-sided prediction intervals for at least k of N future observations are developed for the situation in which, in addition to the previous sample, there is prior information available. The information is specified via the parameter α of the Dirichlet process.

1. INTRODUCTION

Let $X_1,\ldots,X_n$ be a random sample of size n from a distribution function F. Let $Y_1,\ldots,Y_N$ be a second random sample of size N from the same distribution function F and let $g(Y_1,\ldots,Y_N)$ be some function of these random variables. Then, if $L_1(X_1,\ldots,X_n)$ and $L_2(X_1,\ldots,X_n)$ are statistics based on the initial sample, $[L_1,L_2]$ is said to be a 100γ percent prediction interval for $g(Y_1,\ldots,Y_N)$ if

$$\Pr\{L_1(X_1,\ldots,X_n) \le g(Y_1,\ldots,Y_N) \le L_2(X_1,\ldots,X_n)\} = \gamma.$$

Key words: Prediction intervals; Dirichlet process; Bayesian nonparametric methods; Coverage property.

Parametric prediction intervals have been considered by many authors, including Proschan (1953), Chew (1966), and Hahn (1969, 1970a, 1970b, 1972). Wilks (1942, 1962) introduced nonparametric prediction intervals for the case in which F is an unknown continuous distribution function and one is interested in intervals to contain at least k of N future observations. Fligner and Wolfe (1976) have approached nonparametric prediction intervals via a sample analogue to the probability integral transformation and to a coverage property (see Section 4). In particular, they have reviewed the results of Wilks, developed additional prediction intervals, and generalized prediction intervals to the case of an unknown discontinuous distribution function. A Bayesian approach to prediction intervals is presented in Guttman (1970).

This paper combines nonparametric and Bayesian approaches to develop intervals which allow the use of both prior information and the data of the initial sample, without requiring strong parametric assumptions. Our Bayesian nonparametric prediction intervals are derived using Ferguson's (1973) Dirichlet process prior on the space of distribution functions. The Dirichlet process is introduced in Section 2. Section 3 presents the construction of one-sided Bayesian nonparametric prediction intervals for at least k of N future observations. The possibility of a coverage property for a sample from a Dirichlet process is investigated in Section 4. Section 4 also contains some useful results concerning the distribution of the order statistics from a Dirichlet sample. The two-sided prediction interval problem with prior information in the form of a Dirichlet process prior is solved in Section 5. The final section contains an example which illustrates the procedure of constructing Bayesian nonparametric prediction intervals, and discusses the implementation of such prediction intervals.

Let $Z_1,\dots,Z_k$ be independent gamma random variables with shape parameter $\alpha_i \geq 0$ and scale parameter 1, i=1,...,k. Define $Y_i = Z_i / \sum_{j=1}^{k} Z_j$. If $\sum_{i=1}^{k} \alpha_i > 0$, then $(Y_1,\dots,Y_k)$ is said to have a *Dirichlet distribution* with parameter $(\alpha_1,\dots,\alpha_k)$. If all the α_i are strictly positive, the distribution of $(Y_1,\dots,Y_{k-1})$ is absolutely continuous with density

$$f(y_1,\dots,y_k) = \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)} \left(\prod_{i=1}^{k-1} y_i^{\alpha_i - 1}\right)\left(1-\sum_{i=1}^{k-1} y_i\right)^{\alpha_k - 1} I_S(y_1,\dots,y_{k-1}),$$

where S denotes the simplex $y_i \geq 0$ for i=1,...,k-1 and $\sum_{i=1}^{k-1} y_i \leq 1$. The Dirichlet distribution is also called the multi-beta, in that for k=2, it reduces to the beta distribution.

The following expression for the $r_1,\dots,r_\ell^{\text{th}}$ moment of the distribution of $(Y_1,\dots,Y_k)$, for $\ell \leq k$ and r_i a nonnegative integer, will be useful in the sequel:

$$E(Y_1^{r_1}\dots Y_\ell^{r_\ell}) = \frac{\Gamma(\alpha_1+r_1)\dots\Gamma(\alpha_\ell+r_\ell)\Gamma(\alpha)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_\ell)\Gamma(\alpha+r)}, \tag{2.1}$$

where $\alpha = \sum_{i=1}^{k} \alpha_i$ and $r = \sum_{j=1}^{\ell} r_j$. (For a proof of this result and a more complete treatment of the Dirichlet distribution see Wilks (1962). For further background on the Dirichlet distribution and its generalizations, see, for example, Connor and Mosimann (1969) and Good (1965).) Let $y^{[k]}$ denote the ascending factorial $y(y+1)\dots(y+k-1)$ with $y^{[0]}\equiv 1$. Then the right-hand side of (2.1) can be rewritten as $\alpha_1^{[r_1]}\dots\alpha_\ell^{[r_\ell]}/\alpha^{[r]}$.

The Dirichlet process on the real line can now be defined. Let α be a nonnegative measure on the real line $\mathcal{R}$ with Borel σ-field $\mathcal{B}$. Then P is a *Dirichlet process* on $(\mathcal{R},\mathcal{B})$ with parameter α if, for every m=1,2,..., and every measurable partition $B_1,\dots,B_m$ of $\mathcal{R}$, $(P(B_1),\dots,P(B_m))$ has a Dirichlet distribution with parameter $(\alpha(B_1),\dots,\alpha(B_m))$. This process gives rise to a probability on the set of distribution functions, as shown in the landmark paper of Ferguson (1973). By a sample from the process, it will be understood that a distribution function F is chosen by this probability and then a random sample obtained from F. (See Ferguson (1973) and Berk and Savage (1977) for a more rigorous mathematical treatment.) The tractability of Ferguson's approach lies in part in following result (Theorem 1 of Ferguson, 1973). The posterior distribution of the Dirichlet process P with parameter α, given a sample $X_1,\ldots,X_r$ from P, is again a Dirichlet process with as a parameter the updated measure $\alpha + \sum_{i=1}^{r} \delta_{X_i}$, where δ_z is the measure which concentrates all its mass of one at the point z.

For the purposes of this paper, F is taken to be a *random* distribution function from Ferguson's Dirichlet process prior. Given F, the first sample $X_1,\ldots,X_n$ is a random sample from F. The second sample $Y_1,\ldots,Y_N$ is then a sample from the conditional Dirichlet process, given $X_1,\ldots,X_n$. One wishes to predict a specified function of the second sample. In particular, several prediction intervals are obtained to contain at least q of the N future observations.

3. ONE-SIDED PREDICTION INTERVALS WITH THE DIRICHLET PRIOR

In this section 100γ percent prediction intervals of the form (x,∞) are found for at least q of N future observations. Let

$$R(x) = \Pr\{x < \text{at least } q \text{ of the } N\ Y\text{'s} < \infty\}. \tag{3.1}$$

Note that R(x) is decreasing in x. The problem is to find x_0 such that $R(x_0) = \gamma$, for then (x_0,∞) is the desired interval.

Unlike the nonparametric prediction intervals of Wilks (1942, 1962) and Fligner and Wolfe (1976), it is possible, using the Dirichlet process prior, to form prediction intervals for the case of no initial sample of X's (i.e., n=0). Call this problem the "no data" problem. This problem is first solved and then extended in a natural way to obtain the solution of the "data" problem (n > 0).

For fixed x, let I_x, J_x, and K_x denote the random variables for the number of Y's that are less than, equal to, and greater than x, respectively. In the "no data" problem, $Y_1,\ldots,Y_N$ is merely a sample from a Dirichlet process with parameter α. For notational convenience, the subscript x for I,J, and K is suppressed.

Theorem 1: For $Y_1,\ldots,Y_N$ a sample from a Dirichlet process with parameter α,

$$\Pr\{(I,J,K) = (i,j,k)\} = \binom{N}{i,j,k} \alpha(-\infty,x)^{[i]}\alpha(\{x\})^{[j]}\alpha(x,\infty)^{[k]}/\alpha(R)^{[N]}. \quad (3.2)$$

Proof: For distribution function F given, a multinomial argument yields

$$\Pr\{(I,J,K) = (i,j,k)|F\} = \binom{N}{i,j,k} F(x^-)^i[F(x)-F(x^-)]^j[1-F(x)]^k. \quad (3.3)$$

Integration of both sides of (3.3) with respect to the probability Q_α on the set of distribution function gives

$$\Pr\{(I,J,K) = (i,j,k)\} = \binom{N}{i,j,k} \int F(x^-)^i[F(x)-F(x^-)]^j[1-F(x)]^k dQ_\alpha(F).$$

Then, by definition of the Dirichlet process, $(F(x^-),F(x)-F(x^-),1-F(x))$ has a Dirichlet distribution. Application of the $i,j,k^{\underline{th}}$ moment of this Dirichlet distribution yields the right-hand-side of (3.2), completing the proof.||

The random variables $(I_1,\ldots,I_k)$ are said to have a *Dirichlet compound multinomial* distribution (see Johnson and Kotz, 1969, p.309) with parameters $N,\alpha_1,\ldots,\alpha_k$ if, for non-negative integers $i_1,\ldots,i_k$ such that $\sum_{j=1}^{k} i_j = N$,

$$\Pr\{I_1=i_1,\ldots,I_k=i_k\} = \frac{N!}{[\sum_{j=1}^{k} i_j]^N} \prod_{j=1}^{k} \frac{\alpha_j^{[i_j]}}{i_j!}.$$

The Dirichlet compound multinomial results (as the name indicates) by placing a Dirichlet distribution on the parameters of a multinomial distribution. It is clear that the distribution of (I,J,K), given by (3.2), is Dirichlet compound multinomial with parameters $N,\alpha(-\infty,x),\alpha(\{x\}),\alpha(x,\infty)$.

The one-sided prediction interval problem is find x_0 such that $R(x_0) = \gamma$. This equation can be rewritten as

$$\sum_{k=q}^{N} \Pr\{\text{exactly } k \text{ of the } N \text{ future } Y \text{ observations} > x_0\} = \gamma.$$

Now, for the "no sample" problem,

$$\Pr\{\text{exactly } k \text{ of } N \text{ future observations} > x\} = P\{K=k\}.$$

Since the distribution of (I,J,K) is Dirichlet compound multinomial, the distribution of K has what is called a beta compound binomial distribution or a Pólya-Eggenberger distribution (see Johnson and Kotz, 1969, p.229). It follows that

$$\Pr\{K=k\} = \binom{N}{k} \alpha(-\infty,x]^{[N-k]}\alpha(x,\infty)^{[k]}/\alpha(\mathbb{R})^{[N]}.$$

Therefore, the solution is sought for the following equation in x:

$$\sum_{k=q}^{N} \binom{N}{k} \alpha(-\infty,x]^{[N-k]}\alpha(x,\infty)^{[k]}/\alpha(\mathbb{R})^{[N]} = \gamma. \qquad (3.4)$$

The monotonicity of $R(x)$ from the definition ensures that, for $0 < \gamma < 1$, there is either a solution x_0 to equation (3.4) or there exists an x_1 such that $R(x_1) < \gamma \leq R(x_1^-)$. If the Dirichlet parameter α is a nonatomic measure, so that $\alpha(-\infty,t)$ is a continuous function in t, then the left-hand-side of (3.4) is continuous. Further, since $R(x)$ ranges from 1 to 0, in such a case a solution exists (it may not be unique). In the second case, if $R(x_1) < \gamma \leq R(x_1^-)$, the closed interval $[x_1,\infty)$ is a prediction interval for at least q of N future observations with prediction coefficient at least γ.

The solution to the prediction interval "data" problem is now considered. Thus, suppose that an initial sample $X_1,\ldots,X_n$ is observed from a Dirichlet process. The development for the "data" problem is immediate in that the Dirichlet process with parameter α is merely replaced by the Dirichlet process with updated parameter $\alpha' = \alpha + \sum_{i=1}^{n} \delta_{X_i}$ and one proceeds as in the "no data" problem. Thus, (I,J,K) given $(X_1,\ldots,X_n)$ has a Dirichlet compound multinomial distribution with parameters $N,\alpha'(-\infty,x),\alpha'(\{x\}),\alpha'(x,\infty)$. The prediction interval is obtained upon the solution of

$$\sum_{k=q}^{N} \binom{N}{k} \alpha'(-\infty,x]^{[N-k]}\alpha'(x,\infty)^{[k]}/\alpha'(\mathbb{R})^{[N]}=\gamma. \tag{3.5}$$

Here, α' is not nonatomic so either a solution x_0 exists or there exists an x_1 such that $[x_1,\infty)$ is a prediction interval for at least q of N future observations with prediction coefficient at least γ.

There are two special cases of note. When q=N, one obtains the one-sided upper prediction interval for all N future observations; when q=1, the interval is the one-sided upper prediction interval for the largest of N future observations.

4. INVESTIGATION OF THE COVERAGE PROPERTY FOR A DIRICHLET SAMPLE

The coverage property for a continuous distribution function F_0 with $Y_1,\ldots,Y_N$ a random sample from F_0 is as follows:

Coverage Property: If $Y_{(1)} \leq \ldots \leq Y_{(N)}$ denote the order statistics of the sample $Y_1,\ldots,Y_N$ from F_0, then, for integers p and q such that $0 \leq p < q \leq N+1$, the distribution of $F_0(Y_{(q)})-F_0(Y_{(p)})$ has the same distribution as $F_0(Y_{(q-p)})$ where, by convention, $F_0(Y_{(0)})=0$ and $F_0(Y_{(N+1)})=1$.

Fligner and Wolfe (1976) have extended the coverage property from the case of a continuous distribution function to that of the empirical distribution function F_n from the initial sample $X_1,\ldots,X_n$, also from F_0. In particular, they prove that the distribution of $F_n(Y_{(q)})-F_n(Y_{(p)})$ has the same distribution as $F_n(Y_{(q-p)})$.

A question of interest is whether the coverage property holds for $Y_1,\ldots,Y_n$ a sample from a Dirichlet process with parameter α. In particular, is it true that $\{\alpha(-\infty,Y_{(q)}]/\alpha(\mathbb{R})\}-\{\alpha(-\infty,Y_{(p)}]/\alpha(\mathbb{R})\}$ has the same distribution as $\alpha(-\infty,Y_{(q-p)}]/\alpha(\mathbb{R})$? If the coverage property were to hold, it would aid in constructing two-sided prediction intervals directly from one-sided intervals in that if $(Y_{(q-p)},\infty)$ were a one-sided 100γ percent prediction interval, then $(Y_{(p)},Y_{(q)}]$ would also be a 100γ percent prediction interval for fixed integers p and q with $0 \leq p < q \leq N+1$. In that event, one could employ the techniques derived in the preceding section.

However, the coverage property does not hold for samples from a Dirichlet process. It suffices to demonstrate this for the case N=2, p=1, and q=2 by comparison of the mean of $\alpha(-\infty,Y_{(2)}]-\alpha(-\infty,Y_{(1)}] = \alpha(Y_{(1)},Y_{(2)}]$ and the mean of $\alpha(-\infty,Y_{(1)}]$. If the coverage property were true, then, in particular, $E\alpha(-\infty,Y_{(1)}] = E\alpha(Y_{(1)},Y_{(2)}]$ or, equivalently,

$$2E\alpha(-\infty,Y_{(1)}] = E\alpha(-\infty,Y_{(2)}]. \tag{4.1}$$

Theorem 2 below, which gives the distribution of the r^{th} order statistic of a sample of size from a Dirichlet process, will be used to show that inequality (4.1) *does not* hold. Since the Dirichlet process places all its mass on discrete distribution functions (see, for example, Ferguson (1973), Blackwell (1973), and Berk and Savage (1977)), there can be ties in the samples from Dirichlet processes. Nonetheless, one can order the random variables from a sample of size n from a Dirichlet process and derive the distribution of the order statistics.

Theorem 2: For $1 \leq r \leq n$, the distribution F_r of the r^{th} order statistic of a sample of size n from a Dirichlet process with parameter α is given by

$$F_r(x) = \sum_{i=r}^{n} \binom{n}{i} \alpha(-\infty,x]^{[i]}\alpha(x,\infty)^{[n-i]}/\alpha(\mathbb{R})^{[n]}. \tag{4.2}$$

Proof: Suppose F is a known distribution function with $X_1,\ldots,X_n$ the random sample from F. Then the distribution of $X_{(r)}$, the r^{th} order statistic is:

$$\Pr\{X_{(r)} \leq x|F\} = \sum_{i=r}^{n} \binom{n}{i} F(x)^i[1-F(x)]^{n-i}. \tag{4.3}$$

If, in fact, F is a random distribution function from a Dirichlet process, then by definition, for x fixed, F(x) has a beta distribution with parameters $\alpha(-\infty,x]$ and $\alpha(x,\infty)$. Then integrating both sides of (4.3) over F, one obtains

$$F_r(x) = \Pr\{X_{(r)} \le x\} = \sum_{i=r}^{n} \binom{n}{i} \int F(x)^i [1-F(x)]^{n-i} dQ_\alpha(F)$$

$$= \sum_{i=r}^{n} \binom{n}{i} \alpha(-\infty,x]^{[i]} \alpha(x,\infty)^{[n-i]} / \alpha(\mathbb{R})^{[n]}.$$

The final line above follows by the moments of the beta (Dirichlet) distribution.||

It is a simple matter to also derive the joint distribution of the r^{th} and s^{th} order statistics ($r < s$).

Theorem 3: If $X_1,\ldots,X_n$ is a sample of size n from a Dirichlet process with parameter α, the joint distribution of the r^{th} order statistics $X_{(r)}$ and the s^{th} order statistic $X_{(s)}$, for $1 \le r < s \le n$, is given by:

$$F_{r,s}(x,y) = \sum_{i=r}^{n} \sum_{j=\max(0,s-i)}^{n-i} \binom{n}{i,j,n-i-j} \alpha(-\infty,x]^{[i]} \alpha(x,y]^{[j]} \cdot \alpha(y,\infty)^{[n-i-j]} / \alpha(\mathbb{R})^{[n]}. \quad (x<y) \tag{4.4}$$

Proof: Given the distribution function F, the joint distribution of $X_{(r)}$ and $X_{(s)}$ is, for $x < y$:

$$\Pr\{X_{(r)} \le x, X_{(s)} \le y\} = \sum_{i=r}^{n} \sum_{j=\max(0,s-i)}^{n-i} \binom{n}{i,j,n-i-j} F(x)^i [F(y)-F(x)]^j [1-F(y)]^{n-i-j}. \tag{4.5}$$

Integrating both sides of (4.5) with respect to F, using the definition of the Dirichlet process for the partition $(-\infty,x]$, $(x,y]$, (y,∞), and employing the moments of the Dirichlet distribution completes the proof.||

By an application of Theorem 2, the distributions, of the first and second order statistics, for the case N=2, are

$$F_1(x) = [\{2\alpha(-\infty,x]\alpha(x,\infty)\} + \alpha(-\infty,x]^{[2]}]/\alpha(\mathbb{R})^{[2]},$$

$$F_2(x) = \alpha(-\infty,x]^{[2]}/\alpha(\mathbb{R})^{[2]}.$$

It suffices to consider the special case of $\alpha(-\infty,x] = x$ for $x \in [0,1]$ with $\alpha([0,1]) = 1$ and $\alpha(\mathcal{R}-[0,1]) = 0$. Then,

$$E\alpha(-\infty,X_{(1)}] = E(X_{(1)}) = \int_0^1 x dF_1(x)$$

$$= \int_0^1 (1-F_1(x))dx = \int_0^1 \{1-x(1-x)-\tfrac{1}{2}x(x+1)\}dx = 5/12.$$

In a similar fashion,

$$E\alpha(-\infty,X_{(2)}] = E(X_{(2)}) = \int_0^1 x dF_2(x)$$

$$= \int_0^1 (1-F_2(x))dx = \int_0^1 \{1 - \tfrac{1}{2}x(x+1)\}dx = 7/12.$$

Thus equation (4.1) does not hold for this special case. Therefore, the coverage property is not valid for a sample from a Dirichlet process.

5. TWO-SIDED PREDICTION INTERVALS WITH THE DIRICHLET PRIOR

The problem of generating two-sided 100γ percent prediction intervals of the form (x,y), for $x < y$, to contain at least q of N future observations from a Dirichlet process, requires more notational development. Let I, J, K, L, and M (all dependent on x and/or y with the notational dependences suppressed) be random variables for the number of $Y_1,\ldots,Y_N$ that are less than x, equal to x, between x and y, equal to y, and greater than y, respectively. (Note that I, J, and K have been redefined and should not be confused with their use in Section 3.)

Theorem 4: If $X_1,\ldots,X_n$ is a sample from a Dirichlet process P (say) with parameter α and $Y_1,\ldots,Y_N$ is a second sample from the conditional process P given $X_1,\ldots,X_n$, then for x and y with $x < y$:

$$\Pr\{(I,J,K,L,M) = (i,j,k,\ell,m)|X_1,\ldots,X_n\}$$

$$= \binom{N}{i,j,k,\ell,m}\alpha'(-\infty,x)^{[i]}\alpha'(\{x\})^{[j]}\alpha'(x,y)^{[k]}\alpha'(\{y\})^{[\ell]} \cdot\alpha'(y,\infty)^{[m]}/\alpha'(\mathcal{R})^{[N]}, \tag{5.1}$$

where $\alpha' = \alpha + \sum_{i=1}^{n} \delta_{X_i}$.

Proof: The conditional probability distribution of (I,J,K,L,M) given $X_1,\ldots,X_n$ and F is obtained by a multinomial argument. Integration over F and application of the mean of the Dirichlet distribution for $(F(x^-),\ F(x) - F(x^-),\ F(y^-) - F(x),\ F(y) - F(y^-),\ 1 - F(y))$ yields (5.1).||

The distribution of (I,J,K,L,M) given $X_1,\ldots,X_n$ is Dirichlet compound multinomial with parameters N, $\alpha'(-\infty,x), \alpha'(\{x\}),\ \alpha'(x,y), \alpha'(\{y\}),\ \alpha'(y,\infty)$. Note that if n = 0 and x = y so that K = 0 and J and L are combined, Theorem 1 is obtained as a special case.

For $x < y$, define

$$R(x,y) = \Pr\{\text{at least } q \text{ of the Y's are in the interval } (x,y)\}$$

$$= \sum_{p=q}^{N} \Pr\{\text{exactly } p \text{ of the Y's are in the interval } (x,y)\}.$$

Note that for x fixed, R(x,y) is increasing in y and that for y fixed, R(x,y) is decreasing in x. The prediction interval problem is to find (x_0,y_0) such that $R(x_0,y_0) = \gamma$. However, from Theorem 4 and the fact that the marginals of the Dirichlet compound multinomial are beta compound binomial, K has a beta compound binomial with parameters N, $\alpha'(x,y), \alpha'(\mathcal{R}-(x,y))$. Thus,

$$R(x,y) = \sum_{p=q}^{N} \Pr\{K=p\} = \sum_{p=q}^{N} \binom{N}{p} \alpha'(x,y)^{[p]} \alpha'(\mathcal{R}-(x,y))^{[N-p]} / \alpha'(\mathcal{R})^{[N]} . \quad (5.2)$$

A trial-by-error solution to find (x_0,y_0) such that $R(x_0,y_0) = \gamma$ is one way of proceeding. The solution (if it exists) need not be unique and in fact an uncountably infinite number of pairs is possible. Note that as x or y is shifted, $\alpha'(x,y)$ may change, so that a computer in many cases is an invaluable aid in the determination of such prediction intervals for even small values of n and N.

It is clear that one could easily construct prediction intervals of the form [x,y], or (x,y] instead of (x,y). For example, for the interval [x,y], one employs the fact that J + K + L has a beta compound binomial distribution with parameters N, $\alpha'[x,y], \alpha'(\mathcal{R}-[x,y])$ and proceed as above.

In the event that $\alpha(\mathbb{R})$ is small, there may be no solution to $R(x,y) = \gamma$. In that case, one could find x_1 and y_1 such that $R(x_1,y_1) < \gamma \leq R(x_1^-,y_1^+)$. Then $[x_1,y_1]$ is a prediction interval for at least q of N future observations with prediction coefficient at least γ.

6. AN EXAMPLE

In this section, two-sided non-Bayesian nonparametric and Bayesian nonparametric (Dirichlet) prediction intervals for at least q of N future observations are illustrated using a numerical example originally introduced by Hahn (1970a). He gives the following data, on the readings of a new type of generator, recorded for five prototypes: 51.4, 49.5, 48.7, 49.3, 51.6. To illustrate our procedure we suppose that there is prior evidence (from past experience relating to a similar machine) which suggests that the underlying life distribution can be approximated by a normal distribution with a mean of 50 and a standard deviation of 1.25. Thus, to apply the two-sided Bayesian nonparametric prediction interval introduced in Section 5, we will set $\{\alpha(-\infty,x]/\alpha(\mathbb{R})\} = \Phi(\{x-50\}/1.25)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. We must also specify a value for $\alpha(\mathbb{R})$. This specification hinges on the degree of confidence or belief that one invests in this choice for the measure α. For this case, suppose we set $\alpha(\mathbb{R}) = 5$. Roughly speaking, this corresponds to a prior sample size of 5 observations. Since n also equals 5 here, the prior and the initial sample of size 5 are equally weighted in their contribution to the prediction interval. Rather than to construct the different prediction intervals (which may not be unique) for a fixed prediction coefficient, for simplification we let the prediction intervals $(X_{(1)},X_{(5)})$ and $[X_{(1)},X_{(5)}]$ be chosen and the prediction coefficients computed. (Note that any order statistics could have been chosen for the sake of comparison of Dirichlet and nonparametric prediction intervals, but that unlike the Dirichlet intervals, the nonparametric ones demand that only order statistics of the initial sample serve as endpoints.)

Consider the two-sided prediction interval for a single future observation (N=1). The non-Bayesian nonparametric prediction coefficient for the interval $(X_{(1)}, X_{(5)}) = (48.7, 51.6)$ based on the n=5 initial observations is as follows [see Wilks (1942) or Danziger and Davis (1964) for details]:

$$\Pr\{\text{exactly } N_0 \text{ of } N \text{ future observations fall in } (X_{(1)}, X_{(n)})\}$$
$$= n(n-1)(N-N_0+1)N!(N_0+n-2)!/\{N_0!(N+n)!\} \qquad (6.1)$$

Substituting into (6.1) with n=5 and $N=N_0=1$ yields the value 2/3 for the prediction coefficient. Contrast this with the Dirichlet prediction coefficient, for the same interval, as given by (5.2):

$$R(X_{(1)}, X_{(5)}) = \binom{1}{1}\alpha'(X_{(1)}, X_{(5)})^{[1]}/\alpha'(\mathcal{R})^{[1]}$$
$$= \{5(.7505) + 3\}/10 = .675.$$

However, if the interval is expanded to include the endpoints, the nonparametric prediction coefficient does not change, but the discreteness of the Dirichlet process causes an increase in the Dirichlet coefficient to $\{\alpha'[X_{(1)}, X_{(5)}]/\alpha'(\mathcal{R})\} = \{5(.7505)+5\}/10 = .875$.

To illustrate the crucial nature of the choice of $\alpha(\mathcal{R})$, suppose $\alpha(\mathcal{R}) = 20$. Then the Dirichlet prediction coefficient of (48.7, 51.6) is $(20(.7505)+3)/25 = .720$. The limit as $\alpha(\mathcal{R})$ tends to infinity can also be easily computed. As $\alpha(\mathcal{R})$ increases, greater confidence is placed on the prior at the expense of the initial sample. In this case that is reflected by the result that in the limit the prediction coefficient for (48.7, 51.6) (and also for [48.7, 51.6]) is .7505. This value is of course $\Pr(48.7 < X < 51.6)$, where X is normal with mean 50 and standard deviation 1.25.

Note that the nonparametric and Dirichlet prediction coefficients also do not agree as $\alpha(\mathcal{R})$ tends to zero (corresponding to less and less confidence in the prior). In our example, the nonparametric coefficient for (48.7, 51.6) remains 2/3, whereas the Dirichlet coefficient approaches .6 as $\alpha(\mathcal{R})$ tends to zero.

REFERENCES

Berk, R. H. and Savage, I. R. (1977), "Dirichlet Processes Produce Discrete Measures: An Elementary Proof," Contributions to Statistics-Jaroslav Hajek Memorial Volume, Prague: North Holland Publishing Co. (to appear)

Blackwell, D. (1973), "Discreteness of Ferguson Selections", Annals of Statistics, 1, 353-355.

Chew, V. (1966), "Confidence, Prediction and Tolerance Regions for the Multivariate Normal Distribution", Journal of the American Statistical Association 61, 605-617.

Connor, R. J. and Mosimann, J. E. (1969), "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution," Journal of the American Statistical Association, 64, 194-206.

Danziger, L. and Davis, S. A. (1964), "Tables of Distribution-Free Tolerance Intervals," Annals of Mathematical Statistics, 35, 1361-1365.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," Annals of Statistics, 1, 209-230.

Fligner, M. A. and Wolfe, D. A. (1976), "Some Applications of Sample Analogues to the Probability Integral Transformation and a Coverage Property", The American Statistician, 30, 78-85.

Good, I. J. (1965), The Estimation of Probabilities, Cambridge, Mass.: The MIT Press.

Guttman, I. (1970), Statistical Tolerances Regions: Classical and Bayesian, London: Charles Griffin and Company Limited.

Hahn, G. J. (1969), "Factors for Calculating Two-Sided Prediction Intervals for Samples from a Normal Distribution", Journal of the American Statistical Association, 64, 878-888.

Hahn, G. J. (1970a), "Statistical Intervals for a Normal Population, Part I. Tables, Examples, and Applications," Journal of Quality Technology, 2, 115-121.

Hahn, G. J. (1970b), "Additional Factors for Calculating Two-Sided Prediction Intervals for Samples from a Normal Distribution", Journal of the American Statistical Association, 65, 1668-1676.

Hahn, G. J. (1972), "Simultaneous Prediction Intervals to Contain the Standard Deviation or Ranges of Future Samples from a Normal Distribution", Journal of the American Statistical Association, 67, 938-942.

Johnson, N. L. and Kotz, S. (1969), Distributions in Statistics: Discrete Distributions, New York: John Wiley and Sons.

Proschan, F. (1953), "Confidence and Tolerance Intervals for the Normal Distribution", Journal of the American Statistical Association, 48, 550-564.

Wilks, S. S. (1942), "Statistical Prediction with Special Reference to the Problem of Tolerance Limits," Annals of Mathematical Statistics, 13, 400-409.

Wilks, S. S. (1962), Mathematical Statistics, New York: John Wiley and Sons.