Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 907–913
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Bayesian nonparametric inference of stochastically ordered distributions, with Pólya trees and Bernstein polynomials ☆

George Karabatsos[a], Stephen G. Walker[b,*]

[a]University of Illinois-Chicago, 1040 W. Harrison Street (MC 147), Chicago, IL 60607, USA
[b]Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury CT2 7NZ, UK

Received 13 June 2006; received in revised form 1 October 2006; accepted 19 December 2006
Available online 19 January 2007

Abstract

We introduce approaches to performing Bayesian nonparametric statistical inference for distribution functions exhibiting a stochastic ordering. We consider Pólya tree prior distributions, and Bernstein polynomial prior distributions, and each prior provides an appealing and simple way of introducing the stochastic order.

Keywords: Stochastic order; Bayesian nonparametrics; Polya tree prior; Bernstein polynomial prior

1. Introduction

In statistics, stochastic order is a fundamental concept of ordering distributions (e.g., Lehmann and Rojo, 1992; Lehmann, 1955). A univariate distribution F_2 is said to dominate distribution F_1 in the stochastic order, denoted by $F_1 \leqslant_{st} F_2$, whenever $F_1(t) \geqslant F_2(t)$ holds for all t, and of course, this concept can be extended to the ordering of two or more univariate distributions, $F_1, \ldots, F_K$. For at least three reasons, it is often of interest to implement nonparametric estimation of a set of distribution functions $F_1, \ldots, F_K$ under stochastic order constraints. First, a nonparametric approach allows the statistician to circumvent unrealistic assumptions about the distributional form of the K distributions. Second, stochastic order constraints provide a natural way for the statistician to incorporate his prior beliefs about the ordering of distributions. Finally, the incorporation of order constraints, when appropriate, can improve the efficiency in the estimation of parameters (e.g., Robertson et al., 1988).

Much research has dealt with the development of nonparametric approaches to finding the optimal point estimate of a set of distributions $F_1, \ldots, F_K$ subject to stochastic order constraints. Such methods either involve maximizing the empirical likelihood subject to the stochastic order constraint (Brunk et al.,

☆This research is supported by National Science Foundation Grant SES-0242030 in the program of Methodology, Measurement, and Statistics. The second author is financed by an EPSRC Advanced Research Fellowship. The list of the authors' names are presented in alphabetical order.
*Corresponding author. Tel.: +44 01227 823800; fax: +44 01227 827932.
E-mail addresses: georgek@uic.edu (G. Karabatsos), S.G.Walker@kent.ac.uk (S.G. Walker).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2006.12.012

1966; Robertson and Wright, 1974; Dykstra, 1982; Feltz and Dykstra, 1985; Lee, 1987; Dykstra and Feltz, 1989; El Barmi and Dykstra, 1994; Dykstra et al., 1996; Dardanoni and Forcina, 1998; Hoff, 2000), swapping values between pairs of empirical c.d.f.s that violate the stochastic order (Lo, 1987), or minimizing the squared distance to the K empirical c.d.f.s subject to the stochastic order constraint (Gangnon and King, 2002). Most of the recent research has dealt with the estimation of the posterior distribution of $F_1, \ldots, F_K$ under the constraints of a stochastic order (Arjas and Gasbarra, 1994, 1996; Evans et al., 1997; Hoff et al., 2001; Hoff, 2003; Gelfand and Kottas, 2001), through the use of Dirichlet process priors.

In this paper we show how a Bayesian nonparametric approach can be implemented in a simple way, using either Pólya tree prior distributions (Ferguson, 1974), or Bernstein polynomial prior distributions (Petrone 1999a,b). It is also well known that both types support to absolutely continuous distributions.

Moreover, with the Pólya tree prior, inference is remarkably straightforward to implement. In this case, the posterior distribution retains a useful conjugacy property which it exhibits in the traditional situation of inference on a single random distribution function. With the single Pólya tree posterior distribution, the Bayes estimate, that is the posterior mean, can be found without recourse to simulation. Conditional on the Pólya tree prior, the posterior mean is based on the product of means of beta distributions, and conditional on the Bernstein polynomial prior, the posterior mean is based on a weighted sum of means of beta distributions. In the two-dimensional constrained case, as described in this paper, we only need to extend this to find the means of beta distributions under an order constraint. While numerical methods can be used, it is most convenient to find these means using simulation.

2. Bayesian nonparametric models of stochastic order

2.1. Model based on the Pólya tree prior

We now describe a procedure for constructing two random distribution functions on $[0,1]$ with the order constraint. So let $Y_{\varepsilon 0}$, where ε is a sequence from $\{0,1\}$, generate F_1 and $Z_{\varepsilon 0}$ generate F_2 according to the description given for single Pólya tree prior distributions. See Ferguson (1974) and Lavine (1992) for further details. So, for example, $F_1(\frac{1}{2}) = Y_0$, $F_1(\frac{1}{4}) = Y_0 Y_{00}$ and $F_1(\frac{3}{4}) = Y_0 + Y_1 Y_{10}$ and so on. Here $Y_1 = 1 - Y_0$. In general, if $B_{\varepsilon(m)}$ is one of the dyadic intervals at level m, $m = 1, 2, \ldots$, for which there will be 2^m intervals, then we can characterize such an interval by $\varepsilon(m) = \varepsilon_1 \ldots \varepsilon_m$ where each $\varepsilon_j \in \{0,1\}$. Then the mass assigned to the set $B_{\varepsilon(m)}$ is given by

$$F(B_{\varepsilon(m)}) = \prod_{j=1}^{m} Y_{\varepsilon_1 \ldots \varepsilon_j}$$

and note that for all ε, it is that $Y_{\varepsilon 0} + Y_{\varepsilon 1} = Y_\varepsilon$. The Y's are generated as independent beta distributions, that is

$$(Y_{\varepsilon 0}, Y_{\varepsilon 1}) \sim \text{beta}(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1}),$$

where each $\alpha > 0$. A canonical choice of parameter, which provides absolutely continuous distribution functions, with respect to the Lebesgue measure, is given by $\alpha_{\varepsilon(m)} = cm^2$ for all $\varepsilon(m)$ characterizing an interval at level m.

To generate the appropriate order between the two Pólya tree random distribution functions, one only needs to ensure that $Y_{\varepsilon 0} \geqslant Z_{\varepsilon 0}$ for all ε.

Theorem 1. *If $Y_{\varepsilon 0} \geqslant Z_{\varepsilon 0}$ for all ε, then $F_1(t_\varepsilon) \geqslant F_2(t_\varepsilon)$ for all ε where t_ε is the right limit of B_ε.*

Proof. We prove this by an inductive argument. Assume that this hypothesis is true for level m and let $\varepsilon = \varepsilon_1 \ldots \varepsilon_m$, with ε^* defining the interval to the left of ε. Then

$$\begin{aligned} F_1(t_{\varepsilon 0}) &= F_1(t_{\varepsilon^*}) + Y_{\varepsilon 0}\{F_1(t_\varepsilon) - F_1(t_{\varepsilon^*})\} \\ &\geqslant F_1(t_{\varepsilon^*}) + Z_{\varepsilon 0}\{F_1(t_\varepsilon) - F_1(t_{\varepsilon^*})\} \\ &= (1 - Z_{\varepsilon 0})F_1(t_{\varepsilon^*}) + Z_{\varepsilon 0}F_1(t_\varepsilon) \\ &\geqslant (1 - Z_{\varepsilon 0})F_2(t_{\varepsilon^*}) + Z_{\varepsilon 0}F_2(t_\varepsilon) \\ &= F_2(t_{\varepsilon 0}). \end{aligned}$$

We also have

$$F_1(t_{\varepsilon 1}) = F_1(t_\varepsilon) \geqslant F_2(t_\varepsilon) = F_2(t_{\varepsilon 1}).$$

Clearly the hypothesis is true for $m = 1$ since we have $F_1(\frac{1}{2}) \geqslant F_2(\frac{1}{2})$. To illustrate for $m = 2$, it is easy to see that $F_1(\frac{1}{4}) \geqslant F_2(\frac{1}{4})$. Also,

$$\begin{aligned} F_1(3/4) &= Y_0 + (1 - Y_0)Y_{10} = Y_0(1 - Y_{10}) + Y_{10} \geqslant Z_0(1 - Y_{10}) + Y_{10} \\ &= Z_0 + Y_{10}(1 - Z_0) \geqslant Z_0 + Z_{10}(1 - Z_0) \\ &= F_2(3/4). \qquad \square \end{aligned}$$

The $\{t_\varepsilon\}$ generate the dyadic points, that is the rationals, in $(0, 1)$ and hence from the absolute continuity of the random distribution functions, it follows that $F_1(t) \geqslant F_2(t)$ for all $t \in (0, 1)$.

In order to obtain a conjugate prior distribution, one can simply take a joint density for $(y_{\varepsilon 0}, z_{\varepsilon 0})$ as

$$f(y_{\varepsilon 0}, z_{\varepsilon 0}) \propto y_{\varepsilon 0}^{\alpha_{\varepsilon 0}-1}(1 - y_{\varepsilon 0})^{\beta_{\varepsilon 0}-1} \times z_{\varepsilon 0}^{\gamma_{\varepsilon 0}-1}(1 - z_{\varepsilon 0})^{\delta_{\varepsilon 0}-1}\mathbf{1}(y_{\varepsilon 0} \geqslant z_{\varepsilon 0}).$$

In other words, we have the joint density function as the product of the two univariate densities joined with the constraint $\mathbf{1}(y_{\varepsilon 0} \geqslant z_{\varepsilon 0})$. This seems a suitable and minimal way to impose the order constraint, without it seems leading to further dependent structures that are undesirable. The updated model is simply the above with revised parameters

$$\alpha'_{\varepsilon 0} = \alpha_{\varepsilon 0} + n_{\varepsilon 0},$$

where $n_{\varepsilon 0}$ is the number of observations from F_1 which are in the interval characterized by $\varepsilon\mathbf{0}$. The update of the other parameters follows straightforwardly as well.

The model can be extended to K ordered distributions without difficulty. We have

$$F_1(t) \geqslant F_2(t) \geqslant \cdots \geqslant F_K(t)$$

simply by extending the bivariate density to

$$f(y_{\varepsilon 0}^{(1)}, \ldots, y_{\varepsilon 0}^{(K)}) = \prod_{k=1}^{K} (y_{\varepsilon 0}^{(l)})^{\alpha_{\varepsilon 0}^{(k)}-1}(1 - y_{\varepsilon 0}^{(l)})^{\beta_{\varepsilon 0}^{(k)}-1}\mathbf{1}(y_{\varepsilon 0}^{(1)} \geqslant \cdots \geqslant y_{\varepsilon 0}^{(K)}).$$

Conjugacy in this case is also maintained.

The posterior distribution is easily available; however, it is not easy to make statistical inference from the posterior without using simulation techniques. Fortunately, there is a very easy sampling technique using elementary, but numerous Gibbs samplers.

The idea, in the bivariate case, is to sample from the joint density of $(y_{\varepsilon 0}, z_{\varepsilon 0})$, independently of all other joint distributions with different ε. This can be done with an independent Gibbs sampler for each ε. In the bivariate case we sample from $f(y_{\varepsilon 0}|z_{\varepsilon 0})$ and then from $f(z_{\varepsilon 0}|y_{\varepsilon 0})$ in the usual way. Both of these densities are truncated beta distributions. In the K-dimensional model, independent Gibbs samplers for each ε can be run where the conditional densities are again truncated beta distributions.

Rather than sample truncated beta distributions, an alternative approach relies on the introduction of latent variables. So consider the joint density, for ε, the subscript of which will be suppressed, given by

$$f(y; z; u; v) \propto y^{a-1}\mathbf{1}(u < (1 - y)^{b-1})z^{c-1}\mathbf{1}(v < (1 - z)^{d-1})\mathbf{1}(z < y).$$

A Gibbs sampler now follows by sampling from the conditional densities of u and v, which are both obvious uniform distributions. Now the conditional distributions of z and y are simple beta distributions truncated to intervals, and hence can be sampled using the inverse c.d.f. technique.

Having sampled a large number of $(Y_{\varepsilon_0}, Z_{\varepsilon_0})$ for each ε, one can estimate the means $(\bar{Y}_{\varepsilon_0}, \bar{Z}_{\varepsilon_0})$ and then

$$(\widehat{F}_1(B_{\varepsilon(m)}), \widehat{F}_2(B_{\varepsilon(m)})) = \left(\prod_{j=1}^{m} \bar{Y}_{\varepsilon_1 \ldots \varepsilon_j}, \prod_{j=1}^{m} \bar{Z}_{\varepsilon_1 \ldots \varepsilon_j} \right)$$

for all m and $\varepsilon_1 \ldots \varepsilon_m$.

2.2. Model based on the Bernstein polynomial prior

The Bernstein polynomial prior distribution, introduced by Petrone (1999a,b), assumes that a univariate random density function f (with sample space domain [0, 1]) has the representation:

$$f(x; J, w) = \sum_{j=1}^{J} w_{j,J} \mathrm{beta}(x; j, J - j + 1), \tag{1}$$

where the $\{w_{j,J}\}_{j=1}^{J}$ are weights, and J will be fixed. We will write

$$w_{j,J} = \lambda_{j,J} - \lambda_{j-1,J},$$

where the $\lambda_{j,J}$ are strictly increasing for $j = 1, \ldots, J$ and $\lambda_{0,J} = 0$ and $\lambda_{J,J} = 1$. In fact

$$\lambda_{j,J} = \sum_{l=1}^{j} w_{j,J}.$$

If we take a Dirichlet distribution for $(w_{1,J}, \ldots, w_{J,J})$, with parameters $(\alpha_{1,J}, \ldots, \alpha_{J,J})$, then let

$$p(\lambda_{1,J}, \ldots, \lambda_{J-1,J})$$

be the corresponding density function for $(\lambda_{1,J}, \ldots, \lambda_{J-1,J})$.

In our approach involving the inference of a stochastic ordering $F_1 \geqslant_{\mathrm{st}} \cdots \geqslant_{\mathrm{st}} F_K$ under the Bernstein polynomial prior, we enforce the stochastic order constraint $F_1(x) \geqslant \cdots \geqslant F_K(x)$ for all x, by enforcing the order constraint $\lambda_{j,J,1} \geqslant \lambda_{j,J,2} \geqslant \cdots \geqslant \lambda_{j,J,K}$ for a fixed J.

Therefore, we take, and without loss of generality we put $K = 2$, the joint density of $(\lambda_{1,J,1}, \ldots, \lambda_{J-1,J,1})$ and $(\lambda_{1,J,2}, \ldots, \lambda_{J-1,J,2})$ as

$$f(\lambda_{1,J,1}, \ldots, \lambda_{J-1,J,1}, \lambda_{1,J,2}, \ldots, \lambda_{J-1,J,2}) \propto p(\lambda_{1,J,1}, \ldots, \lambda_{J-1,J,1}) \times p(\lambda_{1,J,2}, \ldots, \lambda_{J-1,J,2}) \times \mathbf{1}(\lambda \in C),$$

where

$$C = \{\lambda_{j,J,1} > \lambda_{j,J,2}, j = 1, \ldots, J-1\}.$$

As with the approach involving the Pólya tree prior, the combination of product of Dirichlet distributions and indicator function provides a suitable and minimal way to impose the stochastic order constraint between F_1 and F_2, without it seems leading to further dependent structures that are undesirable. We have the following theorem which establishes the stochastic ordering:

Theorem 2. *If $\lambda \in C$ then $F_1(x) \geqslant F_2(x)$ for all x.*

Proof. Let $B_{j,J}(x)$ denote the distribution function of $\mathrm{beta}(x; j, J-j+1)$, that is

$$B_{j,J}(x) = \frac{\Gamma(J+1)}{\Gamma(j)\Gamma(J-j+1)} \int_0^x s^{j-1}(1-s)^{J-j}\, \mathrm{d}s$$

and it is easy to show that $B_{1,J}(x) \geqslant \cdots \geqslant B_{J-1,J}(x)$. In fact we have, using integration by parts on $B_{j,J}(x)$, with $j > 1$, that

$$B_{j,J}(x) = B_{j-1,J}(x) - \frac{\Gamma(J+1)}{\Gamma(j)\Gamma(J-j+1)} \frac{x^{j-1}(1-x)^{J-j+1}}{J-j+1}.$$

Hence,

$$\begin{aligned}F_1(x) &= \sum_{j=1}^{J}(\lambda_{j,J,1}-\lambda_{j-1,J,1})\,B_{j,J}(x)\\ &= B_{J,J}(x)+\sum_{j=1}^{J-1}\lambda_{j,J,1}\{B_{j,J}(x)-B_{j+1,J}(x)\}\\ &\geqslant B_{J,J}(x)+\sum_{j=1}^{J-1}\lambda_{j,J,2}\{B_{j,J}(x)-B_{j+1,J}(x)\}\\ &= F_2(x). \qquad \square\end{aligned}$$

There is a straightforward Gibbs sampler for estimating this model. The start is to introduce the latent variable $s \in \{1,\ldots,J\}$ such that it has a joint density with the data as

$$f(s,x|w) = w_{s,J}\,\mathrm{beta}(x; s, J-s+1).$$

Hence,

$$\mathrm{P}(s=j|x,w) \propto w_{j,J}\,\frac{(x/(1-x))^j}{\Gamma(j)\,\Gamma(J-j+1)}.$$

Introducing such a variable for each data point, $\{s_{i,k}\}$, for $k=1,2$, the other full conditional density to provide for the Gibbs sampler is that for

$$(w_{1,J,1},\ldots,w_{J-1,J,1},w_{1,J,2},\ldots,w_{J-1,J,2})$$

which is given by

$$\mathrm{Dir}(\alpha_{1,J,1}+n_{1,1},\ldots,\alpha_{J,J,1}+n_{J,1})\times\mathrm{Dir}(\alpha_{1,J,2}+n_{1,2},\ldots,\alpha_{J,J,2}+n_{J,2})$$
$$\times\,\mathbf{1}\left(\sum_{l=1}^{j}w_{l,J,1}>\sum_{l=1}^{j}w_{l,J,2}\;\forall j=1,\ldots,J-1\right),$$

where

$$n_{j,k} = \#\{s_{i,k}=j\}.$$

This joint density has full conditional densities which not only are simple to find, but also are simple to sample, being truncated beta distributions.

3. Application

To provide a straightforward illustration of our methods to estimating distribution functions under the stochastic order, we analyze a data set presented in Belin and Rubin (1990). The data were collected from a psychology experiment, where each of 17 subjects had their reactions times (in milliseconds) measured 30 times. In total, the data consist of two sets of reaction times, one set consisting of 180 reaction-times of 6 schizophrenics (S), and the other set consisting of 330 reaction times of 11 nonschizophrenics (N).

Psychological research has suggested that schizophrenics have a higher reaction time due to attentional deficit and general motor reflex retardation. With this prior knowledge, it was of interest to estimate the posterior distribution of the two distributions of reaction times ($F_\mathrm{S}, F_\mathrm{N}$) under the stochastic order constraint $F_\mathrm{S} \geqslant_{\mathrm{st}} F_\mathrm{N}$.

We performed this estimation using the approach involving Pólya tree priors (Section 2.1), and using the approach involving Bernstein polynomial priors (Section 2.2). The largest reaction time was 1714 ms and so the data were divided by 1715 to put it onto the interval [0, 1]. This enabled the implementation of these two different approaches (as they each assume a [0, 1] sample space), though the results will be reported in terms of milliseconds (by a simple transformation back from [0, 1]).

For the Pólya tree approach, the level of the tree was taken to 6, so there are $2^6 = 64$ partitions of the interval $[0, 1]$. All the parameters of the beta distributions were set to 1.5. A Gibbs sampler was run on each of the $(Y_{\varepsilon 0}, Z_{\varepsilon 0})$ and each was run for 50,000 iterations. For the Bernstein polynomial approach, we specified $J = 100$ and we assumed a noninformative Dirichlet prior by setting $\alpha_{j,J,k} = 0.01$ for $k = 1, 2$. Here a Gibbs sampler was run for 10,000 iterations.

Fig. 1 presents the estimated posterior means c.d.f.s of F_S and F_N, under the stochastic order constraint $F_S \geqslant_{st} F_N$, using the approach involving Pólya tree priors. Fig. 2 presents the estimated posterior means under the Bernstein polynomial priors. We find that the posterior means for both approaches to be quite similar.

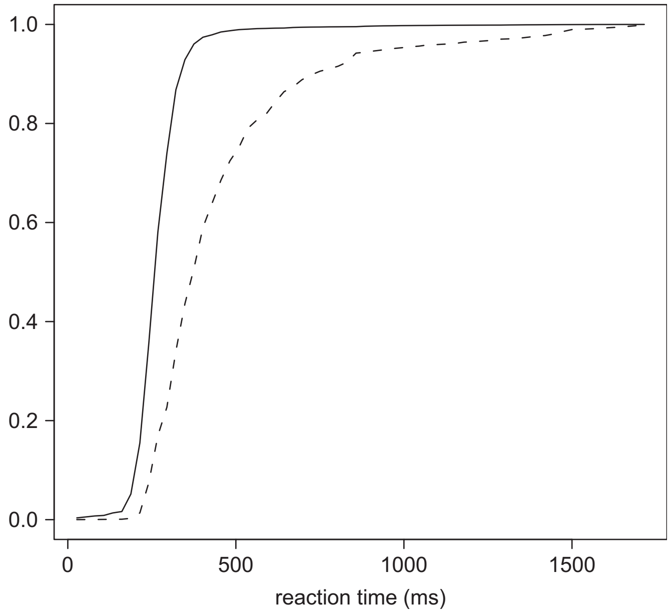

Fig. 1. The estimated posterior mean of the c.d.f.s for the nonschizophrenic group (solid line) and the schizophrenic group (dotted line), using the Pólya tree approach to stochastic ordering.

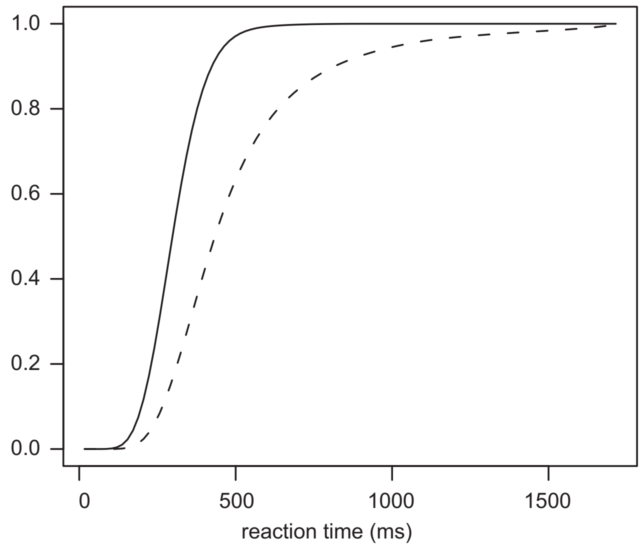

Fig. 2. The estimated posterior mean of the c.d.f.s for the nonschizophrenic group (solid line) and the schizophrenic group (dotted line), using the Bernstein polynomial approach to stochastic ordering.

4. Discussion

For both the Pólya tree prior and the Bernstein polynomial prior, we have developed a general approach for estimating the posterior distribution of a set of continuous distribution functions $F_1, \ldots, F_K$ subject to the constraints of a stochastic order. All the methods are straightforward to implement, since they only require generating samples from many independent truncated (order-constrained) beta densities.

Our methods can be extended in a straightforward manner to address situations where it is of interest to estimate a set of distributions under a partial stochastic order, such as $F_1(t) \geqslant F_2(t), F_3(t) \geqslant F_4(t)$ for all t, where no order is specified between F_2 and F_3. Other straightforward modifications also permit the estimation of a pair of distribution functions under a partial stochastic order of the form $F_1(t) \geqslant F_2(t)$ for some t in an interval $\mathscr{X}' \subseteq \Re$. In either case, only an appropriate subset of beta densities would be subject to order constraints, instead of all of the beta densities.

References

Arjas, E., Gasbarra, D., 1994. Nonparametric Bayesian inference from right-censored survival data, using the Gibbs sampler. Statist. Sinica 4, 505–524.

Arjas, E., Gasbarra, D., 1996. Bayesian inference of survival probabilities under stochastic ordering constraints. J. Amer. Statist. Assoc. 91, 1101–1109.

Belin, T., Rubin, D., 1990. Analysis of a finite mixture model with variance components. Proceedings of the American Statistical Association, Social Statistics Section, pp. 211–215.

Brunk, H., Franck, W., Hanson, D., Hogg, R., 1966. Maximum likelihood estimation of the distributions of two stochastically ordered random variables. J. Amer. Statist. Assoc. 61, 1067–1080.

Dardanoni, V., Forcina, A., 1998. A unified approach to likelihood inference on stochastic orderings in a non-parametric context. J. Amer. Statist. Assoc. 93, 1112–1123.

Dykstra, R., 1982. Maximum likelihood estimation of the survival functions of stochastically ordered random variables. J. Amer. Statist. Assoc. 77, 621–628.

Dykstra, R., Feltz, C., 1989. Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual. Biometrika 76, 331–341.

Dykstra, R., Lee, C., Yan, X., 1996. Multinomial estimation procedures for two stochastically ordered distributions. Statist. Probab. Lett. 30, 353–361.

El Barmi, H., Dykstra, R., 1994. Restricted multinomial maximum likelihood estimation based upon Fechnel duality. Statist. Probab. Lett. 21, 121–130.

Evans, M., Gilula, Z., G.I., Swartz, T., 1997. Bayesian analysis of stochastically ordered distributions of categorical variables. J. Amer. Statist. Assoc. 92, 208–214.

Feltz, C., Dykstra, R., 1985. Maximum likelihood estimation of the survival functions of N stochastically ordered random variables. J. Amer. Statist. Assoc. 80, 1012–1019.

Ferguson, T., 1974. Prior distributions on spaces of probability measures. Ann. Statist. 2, 615–629.

Gangnon, R., King, W., 2002. Minimum distance estimation of the distribution functions of stochastically ordered random variables. Appl. Statist. 51, 485–502.

Gelfand, A., Kottas, A., 2001. Nonparametric Bayesian modeling for stochastic order. Ann. Inst. Statist. Math. 53, 865–876.

Hoff, P., 2000. Constrained nonparametric maximum likelihood via mixtures. J. Comput. Graphical Statist. 9, 633–641.

Hoff, P., 2003. Bayesian methods for partial stochastic orderings. Biometrika 90, 303–317.

Hoff, P., Halberg, R., Shedlovsky, A., Dove, W., Newton, M., 2001. Identifying carriers of a genetic modifier using nonparametric Bayes methods. In: Gatsonis, C., Kass, R., Carlin, B., Carriquiry, A., Gelman, A., Verdinelli, I., West, M. (Eds.), Case Studies In Bayesian Statistics. Springer, New York.

Lavine, M., 1992. Some aspects of Pólya tree distributions for statistical modeling. Ann. Statist. 20, 1222–1235.

Lee, C., 1987. Maximum likelihood estimates for stochastically ordered multinomial populations with fixed and random zeroes. In: MacNeill, I., Umphrey, G. (Eds.), Foundations of Statistical Inference. D. Reidel, Dordrecht.

Lehmann, E., 1955. Ordered families of distributions. Ann. Math. Statist. 26, 399–419.

Lehmann, E., Rojo, J., 1992. Invariant directional orderings. Ann. Statist. 20, 2100–2110.

Lo, S., 1987. Estimation of distribution functions under order restrictions. Statist. Decisions 5, 251–262.

Petrone, S., 1999a. Bayesian density estimation using Bernstein polynomials. Canadian J. Statist. 27, 105–126.

Petrone, S., 1999b. Random Bernstein polynomials. Scand. J. Statist. 26, 373–393.

Robertson, T., Wright, F., 1974. On the maximum likelihood estimation of stochastically ordered random variates. Ann. Statist. 2, 528–534.

Robertson, T., Wright, F., Dykstra, R., 1988. Order Restricted Statistical Inference. Wiley, Chichester.