

AD-A101 688

FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/6 12/1
CONVERGENCE OF DIRICHLET MEASURES AND THE INTERPRETATION OF THE--ETC(U)
JUN 81 J SETHURAMAN, R C TIWARI DAAG29-79-C-0158

UNCLASSIFIED

FSU-TR-M583

ARO-16713.16-M

NL

1 of 1
AD A
017690



END
DATE
FILMED
8-81
DTIC

LEVEL

ARO 16713.16-M

12 14

AD A101688

9

Convergence of Dirichlet Measures and the Interpretation of Their Parameter

by

10

Jayaram/Sethuraman and Ram C./Tiwari
The Florida State University

12 14

11

FSU - 15-M
TH - D - 51 - A

June 1981

FSU Technical Report No. M583
USARO Technical Report No. D-51

DTIC
ELECT
JUL 21 1981

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

*Research supported by the U.S. Army Research Office under Grant No. DAAG-29-79-C-0158. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

†Now at Allahabad University, India.

AMS Classification Numbers: 60B10, 60C10.

KEY WORDS: Non-parametric Bayesian analysis, Dirichlet measures.

DTIC FILE COPY

81 7 20 111

4/0 277

JCB

Convergence of Dirichlet Measures and
the Interpretation of Their Parameter*

by

Jayaram Sethuraman and Ram C. Tiwari†
The Florida State University

1. Introduction. The form of the Bayes estimate of the population mean with respect to a Dirichlet prior with parameter α has given rise to the interpretation that $\alpha(X)$ is the prior sample size. Furthermore, if $\alpha(X)$ is made to tend to zero, then the Bayes estimate mathematically converges to the classical estimator, namely the sample mean. This has further given rise to the general feeling that allowing $\alpha(X)$ to become small not only makes the 'prior sample size' small but also that it corresponds to no prior information. By investigating the limits of prior distributions as the parameter α tends to various values, we show that it is misleading to think of $\alpha(X)$ as the prior sample size and the smallness of $\alpha(X)$ as no prior information. In fact very small values of $\alpha(X)$ actually mean that we have very definite information concerning the unknown true distribution.

*Research supported by the U.S. Army Research Office under Grant No. DAAG 29-79-C-0158. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

†Now at Allahabad University, India.

AMS Classification Numbers: 60B10, 60C10.

KEY WORDS: Non-parametric Bayesian analysis, Dirichlet measures.

2. The Dirichlet measure. Let (X, \mathcal{A}) be separable metric space endowed with the corresponding Borel σ -field. Let \mathcal{P} and \mathcal{M} be the class of probability measures and finite measures (countably additive) on (X, \mathcal{A}) . The natural σ -field, $\sigma(\mathcal{P})$, on \mathcal{P} is the smallest σ -field in \mathcal{P} such that the function $P \mapsto P(A)$ is measurable for each A in \mathcal{A} . There is also the notion of weak convergence in both \mathcal{P} and \mathcal{M} , namely, $\alpha_r \xrightarrow{w} \alpha$ if and only if $\int g d\alpha_r \rightarrow \int g d\alpha$ for all bounded continuous functions on X . Under this convergence \mathcal{P} becomes a separable complete metric space (Prohorov [4]) and the σ -field $\sigma(\mathcal{P})$ above is the Borel σ -field in \mathcal{P} . To each non-zero measure α in \mathcal{M} , we denote by $\bar{\alpha}$ the corresponding normalized measure, namely $\bar{\alpha}(A) = \alpha(A)/\alpha(X)$, $A \in \mathcal{A}$.

In non-parametric Bayesian analysis, the 'true' probability measure P takes values in \mathcal{P} , is random and has a prior distribution. To facilitate the use of standard probability theory we must view P as a measurable map from some probability space (Ω, \mathcal{S}, Q) into $(\mathcal{P}, \sigma(\mathcal{P}))$ and the induced measure QP^{-1} becomes the prior distribution. For any non-zero measure α in \mathcal{M} , the Dirichlet prior measure D_α with parameter α , is defined as follows (Ferguson [3]): For any finite measurable partition (A_1, \dots, A_k) of X , the distribution of $(P(A_1), \dots, P(A_k))$ under D_α is the singular Dirichlet distribution $D(\alpha(A_1), \dots, \alpha(A_k))$ defined on the k -dimensional simplex as in Wilks [7] Section 7.7. Ferguson [3] used this definition and also an alternate definition (See Theorem 1 of Ferguson [3]), and derived many properties of Dirichlet priors and the corresponding Bayes estimates of population parameters. Blackwell [1] and Blackwell and MacQueen [2] have also given alternative definitions of the Dirichlet

Accession For	<input checked="" type="checkbox"/>
NTIS	<input type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<input type="checkbox"/>
By	<i>John S. ...</i>
Distribution/	
Availability Codes	
Dist	A
Special	

prior. We give below yet another definition of the Dirichlet prior which is more general than the previous ones since we will not have to assume that X is separable metric. Let α be a non-zero measure in M . Let (Ω, S, Q) be a probability space rich enough to support two independent sequences of i.i.d. random variables Y_1, Y_2, \dots and $\theta_1, \theta_2, \dots$, where Y_1 is X -valued and has distribution $\bar{\alpha}$ and θ_1 is real valued and has a Beta distribution with parameters 1 and $\alpha(X)$. Let $p_1 = \theta_1$, $p_2 = \theta_2(1-\theta_1)$, $p_3 = \theta_3(1-\theta_1)(1-\theta_2)$, For any y in X let δ_y stand for the degenerate probability measure at y . Define the measurable map P from (Ω, S) into $(P, \sigma(P))$ as follows:

$$P(A) = \sum_{j=1}^{\infty} p_j \delta_{Y_j}(A). \quad (1.1)$$

Then the induced distribution of P is the Dirichlet measure D_α with parameter α . The proof of this fact and that the standard properties of Dirichlet measures can be deduced from this will be given elsewhere, Sethuraman [5].

In the statistical problem of non-parametric Bayesian analysis we have a random variable P taking values in P and whose distribution is D_α . We also have a sample X_1, \dots, X_n , which are random variables taking values in X . Given P , these are i.i.d. with common distribution P . It is required to estimate a function $\phi(P)$, and the Bayes estimator $\hat{\phi}$ with respect to squared loss is given by

$$E(\phi(P) | X_1, \dots, X_n).$$

In particular, if $\phi(P) = \phi_g(P)$ where

$$\phi_g(P) = \int g(x)P(dx) \quad (1.2)$$

where g is a real valued measurable function on X with $\int g^2 d\alpha < \infty$, then the Bayes estimate is given by

$$\hat{\phi}_g = \frac{\alpha(X) \int g d\bar{\alpha} + n \int g dF_n}{\alpha(X) + n}, \quad (1.3)$$

where F_n is the empirical d.f. of X_1, \dots, X_n (Ferguson [3]). In this if we let $\alpha(X) \rightarrow 0$ we obtain the classical estimate $\int g dF_n$. Also the denominator in this estimate is $\alpha(X) + n$ which is $\alpha(X)$ plus the sample size. These facts have given rise to the interpretation that $\alpha(X)$ is the prior sample size and allowing $\alpha(X)$ to tend to zero corresponds to no prior information. In the next section we investigate what happens to Dirichlet measures when their parameters are allowed to converge to certain values. In section 4 we investigate what happens to Bayes estimates when the parameters of the corresponding Dirichlet priors are allowed to converge to the zero measure. From the results in these two sections it follows that small values of $\alpha(X)$ actually correspond to certain definitive information about P .

3. Convergence of Dirichlet measures. In this section we study the convergence of Dirichlet measures as their parameter is allowed to converge in appropriate ways. Since $(P, \sigma(P))$ is a separable complete metric space endowed with its Borel σ -field, we can talk about the usual weak convergence of probability measures on $(P, \sigma(P))$ and of Dirichlet measures, in particular.

THEOREM 3.1. Let $\{\alpha_r\}$ be a sequence of measures in M and let the sequence of normalized measures $\{\bar{\alpha}_r\}$ be tight. Then the sequence $\{D_{\alpha_r}\}$ of Dirichlet measures is tight.

PROOF. Fix $\epsilon > 0$. There exists a sequence of compact sets K_d in X such that

$$\sup \bar{\alpha}_r(K_d^c) \leq 6\epsilon/d^3\pi^2, \quad (3.1)$$

$d = 1, 2, \dots$. Let

$$M_d = \{P: P(K_d^c) \leq 1/d\}, \quad (3.2)$$

$d = 1, 2, \dots$ and let

$$M = \bigcap_d M_d. \quad (3.3)$$

Then clearly M is a compact subset of P in the weak topology. Now, by the Chebysheff inequality

$$D_{\alpha_r}(M_d^c) \leq d E_{D_{\alpha_r}}(P(K_d^c)) = d \bar{\alpha}_r(K_d^c) \leq 6\epsilon/\pi^2 d^2 \quad (3.4)$$

and

$$D_{\alpha_r}(M^c) \leq \sum_d 6\epsilon/\pi^2 d^2 = \epsilon, \text{ for all } r. \quad (3.5)$$

This proves that $\{D_{\alpha_r}\}$ is tight. \square

THEOREM 3.2. Let $\{\alpha_r\}$ be a sequence of measures in M such that

$$\sup_A |\alpha_r(A) - \alpha_0(A)| \rightarrow 0 \quad (3.6)$$

where α_0 is a non-zero measure in M . Then D_{α_r} converges to D_{α_0} weakly.

PROOF. The proof of this result rests heavily on the constructive definition of the Dirichlet measure in (1.1) and the following result which is proved in Sethuraman [6].

Let $\{\beta_r\}$ be a sequence of probability measures on an arbitrary measurable space (Y, \mathcal{B}) and let

$$\sup_B |\beta_r(B) - \beta_0(B)| \rightarrow 0, \quad (3.7)$$

where β_0 is a probability measure on (Y, \mathcal{B}) . Then there exists a sequence of Y -valued random variables $\{Y_r\}_0^\infty$ with marginal distributions $\{\beta_r\}_0^\infty$ such that

$$\text{Prob. } \{Y_r \neq Y_0\} \rightarrow 0 \text{ as } r \rightarrow \infty. \quad (3.8)$$

From (1.1) and the above result, we can find independent sequences of i.i.d. random variables $\{Y_j^r\}, \{\theta_j^r\}, r = 0, 1, 2, \dots$ such that the distribution of Y_1^r is $\bar{\alpha}_r$, the distribution of θ_1^r is Beta with parameters 1 and $\alpha_r(X)$, $r = 0, 1, \dots$, and

$$\text{Prob. } (Y_j^r \neq Y_j^0) \rightarrow 0 \quad (3.9)$$

and

$$\text{Prob. } (\theta_j^r \neq \theta_j^0) \rightarrow 0 \text{ as } r \rightarrow \infty, j = 1, 2, \dots \quad (3.10)$$

Furthermore, if $p_1^r = \theta_1^r, p_j^r = \theta_j^r(1-\theta_{j-1}^r) \dots (1-\theta_1^r)$ for $j \geq 1$, and

$$P^r(A) = \sum_{j=1}^{\infty} p_j^r \delta_{Y_j^r}(A), \quad (3.11)$$

then the distribution of P^r is the Dirichlet measure D_{α_r} , $r = 0, 1, \dots$

From (3.11) it can be easily shown that, for any integer m ,

$$\begin{aligned} \sup_A |P^r(A) - P^0(A)| &\leq \sum_{j=1}^m |p_j^r - p_j^0| + \sum_{j=1}^m I(Y_j^r \neq Y_j^0) \\ &+ 2 \prod_{j=1}^m (1 - \theta_j^0) + \prod_{j=1}^m (1 - \theta_j^r). \end{aligned} \quad (3.12)$$

From the construction above and (3.8), (3.9) and (3.12) and by first

choosing m appropriately and then allowing r to tend to ∞ that

$\sup_A |P^r(A) - P^0(A)| \rightarrow 0$ in probability which is a stronger assertion than made in the theorem, namely that $D_{\alpha_r} \rightarrow D_{\alpha_0}$ weakly. \square

THEOREM 3.3. Let $\{\alpha_r\}$ be a sequence of measures in M such that

$$\alpha_r(X) \rightarrow 0 \text{ and } \sup_A |\bar{\alpha}_r(A) - \bar{\alpha}_0(A)| \rightarrow 0 \text{ as } r \rightarrow \infty, \quad (3.13)$$

where $\bar{\alpha}_0$ is a probability measure in \mathcal{P} . Then the measures D_{α_r} converge to a random degenerate measure δ_{Y^0} where Y^0 has distribution $\bar{\alpha}_0$.

PROOF. As before we can construct independent sequences of i.i.d. random variables $\{Y_j^r\}$ and $\{\theta_j^r\}$, and an independent random variable Y^0 , such that Y_1^r has distribution $\bar{\alpha}_r$, Y^0 has distribution $\bar{\alpha}_0$, the distribution of θ_1^r is Beta with parameters 1 and $\alpha_r(X)$, $r = 1, 2, \dots$, and

$$\text{Prob. } (Y_1^r \neq Y^0) \rightarrow 0 \text{ as } r \rightarrow \infty. \quad (3.14)$$

Furthermore, if $p_1^r = \theta_1^r$, $p_j^r = \theta_j^r(1 - \theta_{j-1}^r) \dots (1 - \theta_1^r)$, for $j \geq 1$, and

$$P^r(A) = \sum_{j=1}^{\infty} P_j^r \delta_{Y_j^r}(A), \quad (3.15)$$

then the distribution of P^r is the Dirichlet measure with parameter α_r ,
 $r = 1, 2, \dots$

From (3.15), it is easily seen that

$$\sup_A |P^r(A) - \delta_{Y_0^r}(A)| \leq I(Y_1^r \neq Y_0^r) + 2(1-p_1^r). \quad (3.16)$$

From (3.14) and the fact that $\alpha_r(X) \rightarrow 0$, it follows that

$\sup_A |P^r(A) - \delta_{Y_0^r}(A)| \rightarrow 0$ in probability which again is stronger than the
 assertion of the theorem. \square

From Theorem 3.2 it is clear that allowing $\alpha_r(X)$ to tend to zero does not correspond to no information on P . In fact if $\alpha_r(X) \rightarrow 0$ and the normalized measure $\bar{\alpha}_r$ converges in the strong sense of (3.13) to a probability measure $\bar{\alpha}_0$, then the information about P is that it is a probability measure concentrated at a particular point in X which is chosen at random according to $\bar{\alpha}_0$. This is definitely very strong information about P and most probably not of the type any statistician would be willing to make.

4. Convergence of Bayes estimates. In this section we are mainly interested in the limits of Bayes estimates of various function $\phi(P)$ as $\alpha(X) \rightarrow 0$. We will therefore make the following assumption throughout this section:

$$\alpha_r(X) \rightarrow 0 \text{ and } \sup_A |\bar{\alpha}_r(A) - \bar{\alpha}_0(A)| \rightarrow 0, \quad (4.1)$$

where $\bar{\alpha}_0$ is a probability measure in P . We will also be mainly concerned

with a special class of functions $\phi(P)$ as defined below. Let g be a permutation invariant measurable function from X^k into R^1 such that

$$\int |g(x_1, \dots, x_1, x_2, \dots, x_2, \dots, x_m, \dots, x_m)| d\bar{\alpha}(x_1) \dots d\bar{\alpha}(x_m) < \infty \quad (4.2)$$

for all possible combinations of arguments $(x_1, \dots, x_1, x_2, \dots, x_2, \dots, x_m, \dots, x_m)$ from all distinct ($m = k$) to all identical ($m = 1$). When the function g vanishes whenever any two coordinates are equal, condition (4.2) reduces to the simple condition

$$\int |g(x_1, \dots, x_k)| d\bar{\alpha}(x_1) \dots d\bar{\alpha}(x_k) < \infty. \quad (4.3)$$

Define the parametric function

$$\phi_g(P) = \int g(x_1, \dots, x_m) dP(x_1) \dots dP(x_k) \quad (4.4)$$

for all those P 's for which it exists. Let P have D_α as the prior distribution and let (X_1, \dots, X_n) be a sample from P . Under further assumptions concerning the second moment of g under $\bar{\alpha}^k$, the Bayes estimate (with respect to squared error loss) of $\phi_g(P)$ based on the sample is

$$\hat{\phi}_{g,\alpha}^n = E_{D_\alpha}(\phi_g(P) | X_1, \dots, X_n), \quad (4.5)$$

and based on no sample is

$$\hat{\phi}_{g,\alpha}^0 = E_{D_\alpha}(\phi_g(P)). \quad (4.6)$$

Since the conditional distribution of P given (X_1, \dots, X_n) is $D_{\alpha+nF_n}$, where F_n is the empirical distribution function of (X_1, \dots, X_n) , we have

$$\hat{\phi}_{g,\alpha}^n = \hat{\phi}_{g,\alpha+nF_n}^0. \quad (4.7)$$

Suppose that we substitute $\alpha = \alpha_r$ where $\{\alpha_r\}$ satisfies (4.1). From the results of section 3 we know that

$$D_{\alpha_r} \rightarrow \delta_{Y^0} \text{ weakly,} \quad (4.8)$$

and

$$D_{\alpha_r + nF_n} \rightarrow D_{nF_n} \quad (4.9)$$

as $r \rightarrow \infty$. The main result of this section pertains to the convergence of the Bayes estimates $\hat{\phi}_{g, \alpha_r}^0$ and $\hat{\phi}_{g, \alpha_r + nF_n}^0$.

THEOREM 4.1. Let condition (4.1) hold. Let g be a continuous function from X^k into R^1 . Let $g(x_1, \dots, x_1, x_2, \dots, x_2, \dots, x_m, \dots, x_m)$ be uniformly integrable with respect to $\bar{\alpha}_r^m$, for all combinations of arguments $(x_1, \dots, x_1, x_2, \dots, x_2, \dots, x_m, \dots, x_m)$ from all distinct to all identical. Then

$$\hat{\phi}_{g, \alpha_r}^0 \rightarrow \int g(x, \dots, x) d\bar{\alpha}_0(x) \quad (4.10)$$

and

$$\hat{\phi}_{g, \alpha_r + nF_n}^0 \rightarrow \hat{\phi}_{g, nF_n}^0 = E_{D_{nF_n}}(g(Z_1, \dots, Z_k)) \quad (4.11)$$

where (Z_1, \dots, Z_k) is a sample from P where P has the distribution D_{nF_n} .

PROOF. The easiest way to prove this result is to use the representation (1.1) for the random probability measure P with a Dirichlet distribution. The uniform integrability conditions on g with respect to α_r immediately show that $\phi_g(P^r)$ is uniformly integrable with respect to D_{α_r} since it is the convex combination of uniformly integrable functions as given below:

$$\phi_g(P^r) = \sum_{(j_1, \dots, j_k)} p_{j_1}^r \dots p_{j_k}^r g(Y_{j_1}^r, \dots, Y_{j_k}^r),$$

where Y_1^r, \dots are i.i.d. with common distribution $\bar{\alpha}_r$. This fact and (4.8) and (4.9) establish the results (4.10) and (4.11) of the theorem. \square

The results of this theorem generalize those of Ferguson [3] Section 5b and 5e and Yamato [8], [9]. Also when $g(x_1, \dots, x_k)$ is such that it vanishes whenever two coordinates are equal, it is easy to see that

$$\hat{\phi}_{g, nF_n}^{\circ} = \frac{n^{(k)}}{n^k} U_{g, n}$$

where $U_{g, n}$ is the usual U statistic based on g and the sample (X_1, \dots, X_n) . This result is also contained in Yamato [8], [9].

5. Acknowledgements. I wish to thank Professor Roger Berger and Wai Chan for their valuable assistance in the preparation of this paper.

REFERENCES

- [1] Blackwell, D. (1973) Discreteness of Ferguson selections. Ann. Statist. 1, 356-358.
- [2] Blackwell, D. and MacQueen, J.B. (1973) Ferguson distributions via Pólya urn schemes. Ann. Statist. 1, 353-355.
- [3] Ferguson, T.S. (1973) A Bayesian analysis of some non-parametric problems. Ann. Statist. 1, 209-230.
- [4] Prohorov, Yu. V. (1956) Convergence of random processes and limit theorems in probability theory. Theor. Prob. and its Appl. (English trans.) 1, 157-214.
- [5] Sethuraman, J. (1978) On a new construction of the Dirichlet measure. Unpublished.
- [6] Sethuraman, J. (1981) Some extensions of the Skorohod representation theorem. In preparation.
- [7] Wilks, S.S. Mathematical Statistics. John Wiley & Sons, N.Y.
- [8] Yamato, H. (1977a) Relation between limiting Bayes estimates and the U-statistics for estimable parameters of degree 2 and 3. Comm. Statist. - Theor. Meth. A6(1), 55-56.
- [9] Yamato, H. (1977b) Relations between limiting Bayes estimates and U-statistics for estimable parameters. Jour. Japanese Statistical Soc. 7, 57-66.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M-583 USARO No. D-51	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Convergence of Dirichlet Measures and the Interpretation of Their Parameter	5. TYPE OF REPORT & PERIOD COVERED Technical Report	6. PERFORMING ORG. REPORT NUMBER
		8. CONTRACT OR GRANT NUMBER(s) DAAG 29-79-C-0158
7. AUTHOR(s) Jayaram Sethuraman and Ram C. Tiwari	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 60B10, 60C10	12. REPORT DATE June, 1981
9. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University Department of Statistics Tallahassee, FL 32306		13. NUMBER OF PAGES 12
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709	15. SECURITY CLASS. (of this report)	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
	16. DISTRIBUTION STATEMENT (of this report) Approved for public release; distribution unlimited	17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)
18. SUPPLEMENTARY NOTES	THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSIDERED AS AN OFFICIAL STATEMENT OF THE ARMY RESEARCH OFFICE, RESEARCH TRIANGLE PARK, NC, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.	
19. KEY WORDS Nonparametric Bayesian analysis, Dirichlet measures		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The form of the Bayes estimate of the population mean with respect to a Dirichlet prior with parameter α has given rise to the interpretation that $\alpha(X)$ is the prior sample size. Furthermore, if $\alpha(X)$ is made to tend to zero, then the Bayes estimate mathematically converges to the classical estimator, namely the sample mean. This has further given rise to the general feeling that allowing $\alpha(X)$ to become small not only makes the 'prior sample size' small but also that it corresponds to no prior information. By investigating the limits of prior distributions as the parameter α tends to various values, we show that it is misleading to think of $\alpha(X)$ as the prior sample size and the smallness of $\alpha(X)$ as no prior information. In fact very small values of $\alpha(X)$ actually mean that we have very definite information concerning the unknown true distribution.		