ON BEHAVIORS OF MEANS OF DISTRIBUTIONS WITH DIRICHLET PROCESSES

Rep. Fac. Sci. Kagoshima Univ., (Math., Phys. & Chem.),
No. 13, p. 41–45, 1980

ON BEHAVIORS OF MEANS OF DISTRIBUTIONS WITH DIRICHLET PROCESSES

By

Hajime YAMATO*

(Received September 30, 1980)

Abstract

The behavior of the mean of a distribution is discussed when the distribution is a Dirichlet process. The mean is symmetrically distributed if the parameter of the Dirichlet process is symmetric. The moment of the mean is evaluated for any order in case it exists. Under certain conditions the mean has approximately the distribution associated with the parameter of the Dirichlet process.

1. Introduction and Summary

As a prior distribution against a distribution in a nonparametric Bayesian statistical problem, Ferguson (1973) intrduced the Dirichlet process and applied it to many problems. The author (1977a, b) applied the Dirichlet process to estimation of estimable parameters. It will be valuable to know the behavior of the parameter involved in a statistical problem in which the Dirichlet process is assumed to be a prior distribution against a distribution. For the quantile, its distribution function is given in 5 (d) of Ferguson (1973). We shall discuss the behavior of the mean of a distribution which is a Dirichlet process.

Let $\boldsymbol{R}$ be the real line and $\boldsymbol{B}$ be the σ-field of Boreal subsets of $\boldsymbol{R}$. We denote a distribution on $(\boldsymbol{R}, \boldsymbol{B})$ by P and its mean by $\mu(P)$. Let α be a σ-additive non-null finite measure on $(\boldsymbol{R}, \boldsymbol{B})$ and we denote the probability measure $\alpha(\cdot)/\alpha(\boldsymbol{R})$ by $Q(\cdot)$. We assume that $\boldsymbol{P}$ is a Dirichlet process on $(\boldsymbol{R}, \boldsymbol{B})$ with parameter α and show the following results.

In the section 2, it is shown that if the measure α is symmetric about a constant ξ and $\int_{\boldsymbol{R}}|x|d\alpha(x)$ is finite then $\mu(\boldsymbol{P})$ is distributed symmetrically about ξ. By the symmetry of the measure α about ξ, we mean that $\alpha(B)=\alpha(T^{-1}(B))$ for any $B\in\boldsymbol{B}$ and the transformation $T(x)=2\xi-x(x\in\boldsymbol{R})$.

In the section 3, the moment of $\mu(\boldsymbol{P})$ is evaluated for any order in case it exists. It is seen that if there exists the moment of the distribution Q for any order and Q is the unique distribution having these monents then $\mu(\boldsymbol{P})$ has approximately the distribution Q for a small $\alpha(\boldsymbol{R})$.

* Department of Mathematics, Faculty of Science, Kagoshima University.
This research was partly supported by the Grant-in-Aid for Scientific Research Project No. 564075 from the Ministry of Education.

Preparatively in case the measurable space is $(\boldsymbol{R}, \boldsymbol{B})$, we quote the the definition of the Dirichlet process and its properties from Ferguson (1973) and Yamato (1977a, b).

DEFINITION (Ferguson). Let α be a non-null finite measure on $(\boldsymbol{R}, \boldsymbol{B})$. We say $\boldsymbol{P}$ is a Dirichlet process on $(\boldsymbol{R}, \boldsymbol{B})$ with parameter α if for every $k=1, 2, \cdots$, and measurable partition $(B_1 \cdots, B_k)$ of $\boldsymbol{R}$, the distribution of $(\boldsymbol{P}(B_k), \cdots, \boldsymbol{P}(B_k))$ is Dirichlet, $D(\alpha(B_1), \cdots, \alpha(B_k))$.

Hereafter it is birefly denoted by $\boldsymbol{P} \in \boldsymbol{D}(\alpha)$ that $\boldsymbol{P}$ is a Dirichlet process on $(\boldsymbol{R}, \boldsymbol{B})$ with parameter α.

LEMMA 1 (Ferguson). Let $\boldsymbol{P} \in \boldsymbol{D}(\alpha)$. If α is σ-additive, then so is $\boldsymbol{P}$ in the sense that for a fixed decreasing sequence of measurable sets $A_n \downarrow \phi$, we have $\boldsymbol{P}(A_n) \to 0$ with probability one.

In what follows, we assume that α is σ-additive and E denotes the expectation with respect to the Dirichlet process $\boldsymbol{P} \in \boldsymbol{D}(\alpha)$.

LEMMA 2 (Ferguson). Let $\boldsymbol{P} \in \boldsymbol{D}(\alpha)$ and g be a measurable real-valued function defined on $(\boldsymbol{R}, \boldsymbol{B})$. If $\int_{\boldsymbol{R}} |g(x)| d\alpha(x) < \infty$, then $\int_{\boldsymbol{R}} |(g(x)| d\boldsymbol{P}(x) < \infty$ with probability one and

$$E \int_{\boldsymbol{R}} g(x)\, d\boldsymbol{P}(x) = \int_{\boldsymbol{R}} g(x)\, dQ(x)\,.$$

LEMMA 3 (Yamato). Let $\boldsymbol{P} \in \boldsymbol{D}(\alpha)$ and let $g(x, y)$ be a measurable real-valued function defined on $(\boldsymbol{R}^2, \boldsymbol{B}^2)$ and symmetric in x, y. If $\int_{\boldsymbol{R}^2} |g(x, y)| d\alpha(x)\, d\alpha(y) < 8$ and $\int_{\boldsymbol{R}} |g(x, x)| d\alpha(x) < \infty$, then

$$E \iint g(x,y)\, d\boldsymbol{P}(x)\, d\boldsymbol{P}(y)$$
$$= \frac{\alpha(R)}{\alpha(\boldsymbol{R})+1} \int_{\boldsymbol{R}^2} g(x,y)\, dQ(x)\, dQ(y) + \frac{1}{\alpha(\boldsymbol{R})+1} \int_{\boldsymbol{R}} g(x,x)\, dQ(x)\,.$$

LEMMA 4 (Yamato). Let $\boldsymbol{P} \in \boldsymbol{D}(\alpha)$ and let $g(x_1, \cdots, x_k)$ be a measurable real-valued function defined on the k-fold product of the measurable space $(\boldsymbol{R}, \boldsymbol{B})$, $(\boldsymbol{R}^k, \boldsymbol{B}^k)$, and symmetric in $x_1, \cdots, x_k$. Then we have

$$E \int_{\boldsymbol{R}^k} g(x_1, \cdots, x_k)\, \Pi_{i=1}^k d\boldsymbol{P}(x_i)$$

$$= \Sigma^* \frac{k!\, [\alpha(\boldsymbol{R})]^{\Sigma_{i=1}^k m_i}}{\Pi_{i=1}^k [i^{m_i}(m_i!)]\, [\alpha(\boldsymbol{R})]^{(k)}} \int_{\boldsymbol{R}^{\Sigma_{i=1}^k m_i}} g(x_{11}, \cdots, x_{1m_1}, x_{21}, x_{21}, \cdots,$$

$$x_{2m_2}, x_{2m_2}, \cdots, x_{k1}, \cdots, x_{k1}, \cdots, x_{km_k}, \cdots, x_{km_k})\, \Pi_{i=1}^k \Pi_{j=1}^{m_i} dQ(x_{ij})\,,$$

under the condition that all the integrals of the right-hand side exist. Where (i) Σ^* denotes the summation over all combinations $(m_1, m_2, \cdots, m_k)$ of k nonnegative integers with $\Sigma_{i=1}^k i m_i = k$. (ii) In the arguments of the right-hand side $x_{11}, \cdots, x_{1m_1}$, $x_{21}, x_{21}, \cdots, x_{2m_2}, x_{2m_2}, \cdots, x_{k1} \cdots, x_{k1} \cdots, x_{km_k} \cdots, x_{km_k}$ the x_{is} appears at i times. (iii) For any real a, $a^{(0)}=1$ and $a^{(j)}=a(a+1)\cdots(a+j+1)$ for integer $j=1, 2, \cdots$.

2. Symmetry

THEOREM. Let $\boldsymbol{P}\in\boldsymbol{D}(\alpha)$ with a σ-additive non-null finite measure α. If the measure α is symmetric about a constant ξ and $\int_{\boldsymbol{R}}|x|d\alpha(x)$ is finite, then the mean $\mu(\boldsymbol{P})=\int_{\boldsymbol{R}}xd\boldsymbol{P}(x)$ is distributed symmetrically about ξ.

PROOF. Under the assumption, the mean $\mu(\boldsymbol{P})$ exists with probability one by the lemma 2. If we consider a transformation $T(x)=2\xi-x$ for $x\in\boldsymbol{R}$, then T is a measurable transformation from $(\boldsymbol{R}, \boldsymbol{B})$ to $(\boldsymbol{R}, \boldsymbol{B})$. We define a random probability measure $\boldsymbol{P}^*$ by $\boldsymbol{P}^*(B)=\boldsymbol{P}(T^{-1}(B))$ for any $B\in\boldsymbol{B}$. Then by the definition, for any measurable partition $(B_1\cdots, B_k)$ of $\boldsymbol{R}$, the distribution of $(\boldsymbol{P}^*(B_1),\cdots, \boldsymbol{P}^*(B_k))$ is Dirichlet, $D(\alpha(T^{-1}(B_1)),\cdots,\alpha(T^{-1}(B_k)))$, because $(T^{-1}(B_1), \cdots, T^{-1}(B_k))$ is also measruable partition of R. From the symmetry of α, for any measurable partition $(B_1,\cdots, B_k)$ of $\boldsymbol{R}$, the distribution of $(\boldsymbol{P}^*(B_1),\cdots,\boldsymbol{P}^*(B_k))$ is Dirichlet, $D(\alpha(B_1),\cdots, \alpha(B_k))$ and by the definition $\boldsymbol{P}^*$ is a Dirichlet process with parameter α. Thus $\boldsymbol{P}$, $\boldsymbol{P}^*\in\boldsymbol{D}(\alpha)$ and it follows that $\mu(\boldsymbol{P})-\xi$, $\mu(\boldsymbol{P}^*)-\xi$ are identically distributed.

From the lemma 1 $\boldsymbol{P}$, $\boldsymbol{P}^*$ are σ-additive w.p.l. (with probability one) and we have $\int_{\boldsymbol{R}}T(x)\,dP(x)=\int_{\boldsymbol{R}}td\boldsymbol{P}^*(t)$ w.p.1 (see, for example, Halmos (1966), p. 163) which yields $\int_{\boldsymbol{R}}xd\boldsymbol{P}^*(x)-\xi=\xi-\int_{\boldsymbol{R}}xd\boldsymbol{P}(x)$ w.p.l. Therefore $\mu(\boldsymbol{P})-\xi$ and $\xi-\mu(\boldsymbol{P})$ are identically distributed, which implies that $\mu(\boldsymbol{P})$ is distributed symmetrically about ξ.

3. Moments

The lemma 2 with $g(x)=x$ yields the well-known result that if $\boldsymbol{P}\in\boldsymbol{D}(\alpha)$ and there exists the mean of the distribution Q, $\mu(Q)$, then

$$E(\mu(\boldsymbol{P}))=\mu(Q)$$

(see 5(b) of Ferguson (1973)). From the lemmas 2,3 with $g(x)=x$ and $g(x, y)=xy$, we have easily the following

PROPOSITION 1. If $\boldsymbol{P}\in\boldsymbol{D}(\alpha)$ and $\int_{\boldsymbol{R}}x^2dQ(x)<\infty$, then

$$Var(\mu(\boldsymbol{P}))=\sigma_Q^2/(\alpha(\boldsymbol{R})+1),$$

where σ_Q^2 is the variance of the distribution Q.

For the k-th moment of $\mu(\boldsymbol{P})$, the lemma 4 with $g(x_1,\cdots, x_k)=x_1\cdots x_k$ yields the following

PROPOSITION 2. If $\boldsymbol{P}\in\boldsymbol{D}(\alpha)$ and there exists the k-th moment of the distribution Q, then

$$E(\mu(\boldsymbol{P})^k)$$

$$=\sum{}^*\frac{k!\,[\alpha(R)]^{\sum_{i=1}^k m_i}}{\prod_{i=1}^k[i^{m_i}(m_i\,!)]\,[\alpha(\boldsymbol{R})]^{(k)}}(\mu_1')^{m_1}(\mu_2')^{m_2}\cdots(\mu_k')^{m_k},$$

where k is a positive integer and μ_j' is the j-th moment of the distribution Q ($j=1, \cdots 2, k$).

Now we shall consider the limit of $E(\mu(\boldsymbol{P})^k)$ as $\alpha(\boldsymbol{R})$ tends to zero keeping Q fixed. In the above summation Σ^*, $\Sigma_{i=1}^k m_i \geq 2$ for $m_k=0$. Because if we assume $\Sigma_{i=1}^k m_i=1$ with $m_k=0$, then $\Sigma_{i=1}^k im_i \leq k-1<k$, which yields the contradiction. Therefore when we take the limit of the above equations as $\alpha(\boldsymbol{R})$ tends to zero keeping Q fixed, all terms vanish except for the one with $m_k=1$ and $m_1=\cdots m_{k-1}=0$. Thus we have the following

COROLLARY. $$\lim_{\alpha(\boldsymbol{R})\to 0} E(\mu(\boldsymbol{P})^k) = \mu_k' ,$$

where Q is fixed.

By applying 4.30 of Kendall and Stuart (1969) to the above corollary it is seen that if there exists the moment of the distribution Q for any order and Q is the unique distribution having these moments then the distribution of $\mu(\boldsymbol{P})$ converges to Q as $\alpha(\boldsymbol{R})$ tends to zero keeping Q fixed. Thus under the same condition $\mu(\boldsymbol{P})$ has approximately distribution Q for a small $\alpha(\boldsymbol{R})$.

At last we consider the moment of the posterior distribution for low order. Let $X_1,\cdots, X_n$ be a smaple of size n from a distribution $\boldsymbol{P}$ with $\boldsymbol{P}\in\boldsymbol{D}(\alpha)$. We shall put $q_n=\alpha(\boldsymbol{R})/(\alpha(\boldsymbol{R})+n)$.

Then the posterior distribution of $\boldsymbol{P}$ given $X_1,\cdots, X_n$ is a Dirichlet process $\boldsymbol{D}(\alpha+\Sigma_{i=1}^n \delta_{x_i})$, where δ_x denotes the unit measure on $(\boldsymbol{R}, \boldsymbol{B})$ concentrated at the point x. If there exists the mean of the distribution Q, then

$$E[\mu(\boldsymbol{P})\,|\,X_1, \cdots, X_n] = q_n\mu(Q)+(1-q_n)\bar{X} ,$$

where $\bar{X}=\Sigma_{i=1}^n X_i/n$. (See Ferguson (1973).)

Since the posterior distribution of $\boldsymbol{P}$ given $X_1,\cdots, X_n$ is a Dirichlet process $\boldsymbol{D}(\alpha+\Sigma_{i=1}^n \delta_{x_i})$, under the condition $\int_{\boldsymbol{R}} x^2 d\alpha(x)<\infty$ the proposition 1 yields

$$Var[\mu(\boldsymbol{P})\,|\,X_1, \cdots, X_n] = \sigma_{\hat{p}_n}^2/(\alpha(R)+n+1) ,$$

where $\hat{p}_n(\cdot)=q_nQ(\cdot)+(1-q_n)P_n(\cdot)$ with the empirical distribution P_n based on the sample $X_1\cdots, X_n$. We have

$$\sigma_{\hat{p}_n}^2 = q_n \int_{\boldsymbol{R}} x^2\, dQ(x)+(1-q_n)\int_{\boldsymbol{R}} x^2\, dp_n(x)-[q_n\mu(Q)+(1-q_n)\bar{X}]^2$$
$$= q_n\sigma_Q^2+(1-q_n)s_n^2+q_n(1-q_n)\,[\mu(Q)-\bar{X}]^2 ,$$

where $s_n^2=\Sigma_{i=1}^n (X_i-\bar{X})^2/n$. Thus we have the following

PROPOSITION 3. If $X_1\cdots, X_n$ is a sample from a distribution $\boldsymbol{P}$ with $\boldsymbol{P}\in\boldsymbol{D}(\alpha)$ and $\int_{\boldsymbol{R}} x^2 d\alpha(x)<\infty$, then

$$Var[\mu(\boldsymbol{P})\,|\,X_1, \cdots, X_n] = \{q_n\sigma_Q^2+(1-q_n)\,s_n^2+q_n(1-q_n)\,[\mu(Q)-\bar{X}]^2\}/(\alpha(\boldsymbol{R})+n+1) .$$

References

[1] Ferguson, T.S. (1973). *A Bayesian analysis of some nonparametric problems.* Ann. Statist. **1** 209–230.

[2] Halmos, P.R. (1966). Measure Theory. Van Nostrand, New York.

[3] Kendall. M.G. and Stuart, A. (1969). The Advanced Theory of Statistics, Vol. 1. Charles Griffin, London.

[4] Yamato, H. (1977a). *Relations between limiting Bayes estimates and the U-statistics for estimable parameters of degrees 2 and 3.* Commun. Statist. -Theory Methods. **6** 55–66.

[5] Yamato, H. (1977b). *Relations between limiting Bayes estimates and U-statistics for estimable parameters.* J. Japan Statist. Soc. **7** 57–66.