EXPECTATIONS OF FUNCTIONS OF SAMPLES FROM DISTRIBUTIONS CHOSEN FROM DIRICHLET PROCESSES

Rep. Fac. Sci., Kagoshima Univ., (Math., Phys. & Chem.),
No 17, p.1–8, 1984

EXPECTATIONS OF FUNCTIONS OF SAMPLES FROM DISTRIBUTIONS CHOSEN FROM DIRICHLET PROCESSES

By

Hajime Yamato*

(Received September 7, 1984)

Abstract

For samples from distributions chosen from Dirichlet processes, we evaluate expectations of their functions. By making use of this result, we derive some properties of the samples and evaluate expectations of random functionals of Dirichlet processes.

1. Introduction

Ferguson [2] introduces the Dirichlet process as a prior distribution for Bayesian nonparametric inference. It is well-known that a distribution chosen from a Dirichlet process is discrete with probability one. It has a positive probability that some observations of a sample from a distribution chosen from a Dirichlet process are equal, even if parameter is nonatomic (see Antoniak [1], p. 1160). We shall consider a function of a sample from a distribution chosen from a Dirichlet process and give its expectation, from which we shall derive some properties of a sample and evaluate expectation of a random functional of a Dirichlet process.

The author assumes that readers are familiar with the Dirichlet process. For the definition of a Dirichlet process see Ferguson [2]. Let $\boldsymbol{X}$ be a set and let $\boldsymbol{A}$ be a σ-field of subsets of $\boldsymbol{X}$. Let α be a nonnull finite measure on $(\boldsymbol{X},\boldsymbol{A})$. $Q(\cdot)$ denotes a distribution $\alpha(\cdot)/\alpha(\boldsymbol{X})$ and M denotes $\alpha(\boldsymbol{X})$. We list some properties of a Dirichlet process.

Lemma 1.1(Ferguson [2]). *Let $\boldsymbol{P}$ be a Dirichlet process on $(\boldsymbol{X},\boldsymbol{A})$ with parameter α and let $\boldsymbol{X}$ be a sample of size* 1 *from $\boldsymbol{P}$. Then for $A\in\boldsymbol{A}$*

$$P(X\in A)=Q(A).$$

Let $X_1, \cdots, X_n$ be a sample of size n from a distribution $\boldsymbol{P}$ chosen from a Dirichlet process. Then, as stated in Korwar and Hollander [3], we can view the observations $X_1, \cdots, X_n$ as being obtained equentially as follows : Let X_1 be a sample of size 1 from $\boldsymbol{P}$; having obtained X_1, let X_2 be a sample of size 1 from the conditional distribution $\boldsymbol{P}$ given X_1 ; and so on until $X_1, \cdots, X_n$ are obtained. Thus we have the following lemma, which is essentially similar to the statement of Zehnwirth [5], p. 16.

* Department of Mathematics, Faculty of Science, Kagoshima University, Kagoshima 890, Japan

Lemma 1. 2. *Let $\boldsymbol{P}$ be a Dirichlet process on $(\boldsymbol{X},\boldsymbol{A})$ with parameter α and let $X_1, \cdots, X_n$ be a sample of size n from $\boldsymbol{P}$. Then we can view as follows : X_1 has the distribution Q and for $k=1, \cdots, n-1$, the conditional distribution of X_{k+1} given $X_1, \cdots, X_k$ is the distribution $(MQ(\cdot)+\sum_1^k \delta_{X_j}(\cdot))/(M+k)$, where for $x \in X$, δ_x denotes the measure on $(\boldsymbol{X},\boldsymbol{A})$ giving the mass one to the point x.*

In Section 2, we evaluate expectation of a function of a sample, $X_1, \cdots, X_n$, from a distribution chosen from a Dirichlet process, $\mathrm{E}h(X_1, \cdots, X_n)$, for a measurable function h under certain conditions. In Section 3, we shall give some properties of a sample, which yields Proposition 3 of Antoniak [3] as a special case. Furthermore we shall give the conditional distribution of a sample, which yields Theorem 2.5 of Korwar and Hollander [3] as its corollary. Finally we evaluate expectation of a random functional of a Dirichlet process. The evaluation is essentially as same as Lemma 5 of Yamato [4].

2. Expectations of functions of samples

From Lemma 1.1 a sample of size 1, X_1, from a distribution $\boldsymbol{P}$ chosen from a Dirichlet process on $(\boldsymbol{X},\boldsymbol{A})$ with parameter α has the distribution Q. Therefore if the integral $\int_X h(x)dQ(x)$ exists for a real-valued measurable function h defined on $(\boldsymbol{X},\boldsymbol{A})$, then

$$\mathrm{E}h(X_1)=\int_X h(x)dQ(x). \tag{2.1}$$

$(\boldsymbol{X}^n,\boldsymbol{A}^n)$ denotes the n-fold product of measurable space $(\boldsymbol{X},\boldsymbol{A})$ for $n=2, 3, \cdots$. Let X_1, X_2 be a sample of size 2 from a distribution $\boldsymbol{P}$ chosen from a Dirichlet process on $(\boldsymbol{X},\boldsymbol{A})$ with parameter α. Let $h(x_1,x_2)$ be a real-valued measurable function defined on $(\boldsymbol{X}^2,\boldsymbol{A}^2)$ and symmetric in x_1, x_2. We suppose that the integrals $\int_{X^2} h(x_1, x_2)dQ(x_1)dQ(x_2)$ *and* $\int_X h(x_1, x_1)dQ(x_1)$ exist. Since by Lemma 1.2, given X_1, X_2 has the distribution $(MQ(\cdot)+\delta_{X_1}(\cdot))/(M+1)$,

$$\mathrm{E}[h(X_1, X_2)|X_1]=\{M\int_X h(X_1, x_2)dQ(x_2)+h(X_1, X_1)\}/(M+1),$$

where $\int_X h(x_1, x_2)dQ(x_2)$ exists and is integrable by Fubini's Theorem. Since X_1 has the distribution Q by Lemma 1.2 and there exists expectation of the right-hand side of the above equation,

$$\begin{aligned}&\mathrm{E}h(X_1, X_2)\\&=\{M\int_{X^2} h(x_1, x_2)dQ(x_1)dQ(x_2)+\int_X h(x_1, x_1)dQ(x_1)\}/(M+1)\end{aligned} \tag{2.2}$$

In general we have the following

Theorem 2.1. *Let $h(x_1, \cdots, x_n)$ be a real-valued measurable function defined on $(\boldsymbol{X}^n, \boldsymbol{A}^n)$ and symmetric in $x_1, \cdots, x_n$. Let $X_1, \cdots, X_n$ be a sample of size n from a distribution chosen from a Dirichlet process on $(\boldsymbol{X}, \boldsymbol{A})$ with parameter α. Then*

$$\mathrm{E}h(X_1, \cdots, X_n) \tag{2.3}$$
$$=\sum_{S(\sum im(i)=n)}\frac{n!M^{\sum m(i)}}{\Pi_{i=1}^{n} i^{m(i)}(m(i)!)M^{(n)}}\int_{X^{\sum m(i)}}h(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)},$$
$$x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1})\Pi_{i=1}^{n}\Pi_{j=1}^{m(i)}dQ(X_{ij})$$

provided all integrals of the right-hand side exist. Where $M^{(n)}=M(M+1)\cdots(M+n-1)$ *for a positive integer* n, $\sum_{S(\sum im(i)=n)}$ *denotes the summation over all sequences of* n *non-negative integers* $m(1), \cdots, m(n)$ *satisfying* $\sum_{i=1}^{n} im(i)=n$ *and in the arguments of the integrand of the right-hand side the number of* x_{ij} *is* i *for* $i=1, \cdots, n$ *and* $j=1, \cdots, m(i)$.

Proof. We shall prove Theorem by induction for a positive integer n. It is shown in (2.1) and (2.2) that Theorem holds for $n=1, 2$. We assume that Theorem holds for $n\geqq 2$ and show that Theorem holds for $n+1$.

Let $h(x_1, \cdots, x_{n+1})$ be a real-valued measurable function defined on $(\boldsymbol{X}^{n+1}, \boldsymbol{A}^{n+1})$ and symmetric in $x_1, \cdots, x_{n+1}$. We suppose existence of all integrals of the right-hand side of (2.3) for $n+1$ instead of n. By Lemma 1.2, given $X_1, \cdots, X_n$, the conditional distribution of X_{n+1} is $(MQ(\cdot)+\sum_{j=1}^{n}\delta_{X_j}(\cdot))/(M+n)$ and

$$\mathrm{E}[h(X_1, \cdots, X_n, X_{n+1})|X_1, \cdots, X_n] \tag{2.4}$$
$$=\{M\int_X h(X_1, \cdots X_n, x_{n+1})dQ(x_{n+1})+\sum_{j=1}^{n}h(X_1, \cdots, X_n, X_j)\}/(M+n),$$

where the integral $\int_X h(x_1, \cdots, x_n, x_{n+1})dQ(x_{n+1})$ exists and is integrable by Fubini's Theorem.

Since $\int_X h(x_1, \cdots, x_n, x_{n+1})dQ(x_{n+1})$ is symmetric in $x_1, \cdots, x_n$, the assumption yields

$$M\mathrm{E}\int_X h(X_1, \cdots, X_n, x_{n+1})dQ(x_{n+1})/(M+n) \tag{2.5}$$
$$=\sum_{S(\sum im(i)=n)}\frac{n!M^{\sum m(i)+1}}{\Pi_{i=1}^{n} i^{m(i)}(m(i)!)M^{(n+1)}}\int_{X^{\sum m(i)}}h(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots,$$
$$x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1}, x_{n+1})\Pi_{i=1}^{n}\Pi_{j=1}^{m(i)}dQ(x_{ij})dQ(x_{n+1}),$$

where all integrals of the right-hand side exist by the assumption.

Note that $g(x_1, \cdots, x_n)=\sum_{j=1}^{n}h(x_1, \cdots, x_n, x_j)$ is measurable function on $(\boldsymbol{X}^n, \boldsymbol{A}^n)$ and symmetric in $x_1, \cdots, x_n$, and $g(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1})=\sum^{*}h(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1}, x)$, where in the summation $\sum^{*}$ x takes $x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1}$. Therefore the assumption yields

$$\mathrm{E}\sum_{j=1}^{n}h(X_1, \cdots, X_n, X_j)/(M+n)$$
$$=\mathrm{E}g(X_1, \cdots, X_n)/(M+n) \tag{2.6}$$
$$=\sum_{S(\sum im(i)=n)}\frac{n!M^{\sum m(i)}}{\Pi_{i=1}^{n} i^{m(i)}(m(i)!)M^{(n+1)}}$$

$$\times\{\, m(1)\int_{X^{\sum m(i)}} h(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1}, x_{11})\Pi_{i=1}^{n}\Pi_{j=1}^{m(i)} dQ(x_{ij})$$

$$+2m(2)\int_{X^{\sum m(i)}} h(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1}, x_{21})\Pi_{i=1}^{n}\Pi_{j=1}^{m(i)} dQ(x_{ij})$$

$$\cdots \quad \cdots \quad \cdots$$

$$+(n-1)m(n-1)\int_{X^{\sum m(i)}} h(x_{11}, \cdots, x_{n-1,1}, \cdots, x_{n-1,1}, \cdots, x_{n-1,m(n-1)}, \cdots, x_{n-1,m(n-1)}, x_{n-1,1})\Pi_{i=1}^{n}\Pi_{j=1}^{m(i)} dQ(x_{ij})$$

$$+nm(m)\int_{X^{\sum m(i)}} h(x_{n1}, \cdots, x_{n1}, x_{n1}) dQ(x_{n1})\,\},$$

where if $m(n)\neq 0$ then $m(n)=1$ and $m(1)=\cdots=m(n-1)=0$.

A set $(x_{11}, \cdots, x_{1m'(1)}, x_{21}, x_{21}, \cdots, x_{2m'(2)}, x_{2m'(2)}, \cdots, x_{n+1,1}, \cdots, x_{n+1,1})$ with $\sum_{i=1}^{n+1} im'(i)=n+1$ can be obtained from some of sets $(x_{11}, \cdots, x_{1m(1)}, x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}, \cdots, x_{n1}, \cdots, x_{n1})$ with $\sum_{i=1}^{n} im(i)=n$ by the following ways. {1} : A new variable enters and $m(1)=m'(1)-1$, $m(2)=m'(2)$, $\cdots$, $m(n)=m'(n)$, $m'(n+1)=0$, $\sum_{i=1}^{n} m(i)+1=\sum_{i=1}^{n} m'(i)$, which is seen in arguments of the integrand of the right-hand side of (2.5). {2} : One of $x_{11}, \cdots, x_{1m(1)}$ enters again and $m(1)=m'(1)+1$, $m(2)=m'(2)-1$, $m(3)=m'(3)$, $\cdots$, $m(n)=m'(n)$, $m'(n+1)=0$, $\sum_{i=1}^{n} m(i)=\sum_{i=1}^{n+1} m'(i)$, which is seen in arguments of the integrand of the first integral of the right-hand side of (2.6). {3} : One of $x_{21}, x_{21}, \cdots, x_{2m(2)}, x_{2m(2)}$ enters again and $m(1)=m'(1)$, $m(2)=m'(2)+1$, $m(3)=m'(3)-1$, $m(4)=m'(4)$, $\cdots$, $m(n)=m'(n)$, $m'(n+1)=0$, $\sum_{i=1}^{n} m(i)=\sum_{i=1}^{n+1} m'(i)$, which is seen in arguments of the integrand of the second integral.

$\cdots \quad \cdots \quad \cdots$

{$n-1$} : One of $x_{n-1,1}, \cdots, x_{n-1,1}, \cdots, x_{n-1,m(n-1)}, \cdots, x_{n-1,m(n-1)}$ enters again and $m(1)=m'(1)$, $\cdots$, $m(n-2)=m'(n-2)$, $m(n-1)=m'(n-1)+1$, $m(n)=m'(n)-1$, $m'(n+1)=0$, $\sum_{i=1}^{n} m(i)=\sum_{i=1}^{n+1} m'(i)$. For $n\geqq 2$, if $m(n-1)\neq 0$ then $m(n)=0$. This is seen in arguments of the integrand of the third intergral of the right-hand side of (2.6).

{n} : One of $x_{n1}, \cdots, x_{n1}$, whose number is n, enters again and $m(1)=m'(1)=0$, $\cdots$, $m(n-1)=m'(n-1)=0$, $m(n)=m'(n)+1$, $m'(n)=0$, $m'(n+1)=1$, $\sum_{i=1}^{n} m(i)=\sum_{i=1}^{n+1} m'(i)$, which is seen in arguments of the integrand of the last integral.

Therefore from (2.4), (2.5), (2.6) we have

$$\mathbf{E}h(\mathbf{X}_1, \cdots, X_{n+1})$$

$$=\{\, M\mathbf{E}\int_X h(X_1, \cdots, X_n, x_{n+1}) dQ(x_{n+1})+\mathbf{E}\sum_{j=1}^{n} h(X_1, \cdots, X_n, X_j)\,\}/(M+1)$$

$$=\sum_{S(\sum im'(i)=n+1)}\int_{X^{\sum m'(i)}} h(x_{11}, \cdots, x_{1m'(1)}, x_{21}, x_{21}, \cdots, x_{2m'(2)}, x_{2m'(2)}, \cdots, x_{n+1,1}, \cdots, x_{n+1,1})\Pi_{i=1}^{n+1}\Pi_{j=1}^{m'(i)} dQ(x_{ij})$$

$$\times\frac{n!M^{\sum m'(i)}}{\Pi_{i=1}^{n+1} i^{m'(i)}(m'(i)!)M^{(n+1)}}\{\, m'(1)+\frac{2m'(2)}{m'(1)+1}\times(m'(1)+1)+$$

$$+\frac{3m'(3)}{2(m'(2)+1)}\times 2(m'(2)+1)+\cdots+\frac{nm'(n)}{(n-1)(m'(n-1)+1)}\times(n-1)(m'(n-1)+1)$$
$$+\frac{(n+1)m'(n+1)}{n(m'(n)+1)}\times n(m'(n)+1)\}$$
$$=\sum_{S(\sum im'(i)=n+1)}\frac{(n+1)!M^{\sum m'(i)}}{\Pi_{i=1}^{n+1}i^{m'(i)}(m'(i)!)M^{(n+1)}}\int_X{}^{\sum m'(i)}h(x_{11},\ \cdots,\ x_{1m'(1)},\ x_{21},\ x_{21},\ \cdots,$$
$$x_{2m'(2)},\ x_{2m'(2)},\ \cdots,\ x_{n+1,1},\ \cdots,\ x_{n+1,1})\Pi_{i=1}^{n+1}\Pi_{j=1}^{m'(i)}dQ(x_{ij}),$$

where $\sum_{S(\sum im'(i)=n+1)}$ denotes the summation over all sequences of $n+1$ nonnegative integers $m'(1),\ \cdots,\ m'(n+1)$ satisfying $\sum_{i=1}^{n+1}im'(i)=n+1$ and in arguments of the integrand the number of x_{ij} is i for $i=1,\ \cdots,\ n+1$ and $j=1,\ \cdots,\ m'(i)$. Thus the theorem is proved.

We can rewrite Theorem 2.1 in the following form, which is seen useful later.

Corollary.

$$\mathbf{E}h(X_1,\ \cdots,\ X_n)$$
$$=\sum_{u=1}^{n}\sum_{S(\sum_1^u r(i)=n)}\frac{n!M^u}{\Pi_{i=1}^{n}(K_i(r(1),\ \cdots,\ r(u))!)\Pi_{i=1}^{u}r(i)M^{(n)}}$$
$$\int_{X^u}h(x_1,\ \cdots,\ x_1,\ \cdots,\ x_u,\ \cdots x_u)\Pi_{i=1}^{u}dQ(x_i),$$

where $\sum_{S(\sum_1^u r(i)=n)}$ *represents the summation over all sequences of* u *integers* $r(1),\ \cdots,\ r(u)$ *such that* $1\leq r(1)\leq\cdots\leq r(u)$ *and* $\sum_{i=1}^{u}r(i)=n$, $K_i(r(1),\ \cdots,\ r(u))$ *is the number of* j *such that* $r(j)=i$ $(j=1,\ \cdots,\ u)$ *for positive integers* u, i, $r(1),\ \cdots,\ r(u)$ *and in the arguments of the integrand of the right-hand side the number of* x_i *is* $r(i)$ *for* $i=1,\ \cdots,\ u$.

3. Applications

We consider a function h such that $h(x_1,\ x_2)=1$ if $x_1=x_2$ and $=0$ if $x_1\neq x_2$. Let X_1, X_2 be a sample of size 2 from a distribution chosen from a Dirichlet process on $(\boldsymbol{X},\ \boldsymbol{A})$ with parameter α. Then $\mathbf{E}h(X_1,\ X_2)=P(X_1=X_2)$ and by (2.2) we have

$$\mathbf{P}(X_1=X_2)=\{M\int_{x_1=x_2}dQ(x_1)dQ(x_2)+\int_X dQ(x_1)\}/(M+1) \tag{3.1}$$
$$=\{M\sum_{x\in D}Q^2(\{x\})+1\}/(M+1),$$

where D is a set of discontinuity points of the distribution Q, which is at most countable. In general, we consider a function h such that $h(x_1,\ \cdots,\ x_n)=1$ if $x_1=\cdots=x_n$ and $=0$ otherwise. Then by Theorem 2.1 we have the following

Proposition 3.1. *Let* $X_1,\ \cdots,\ X_n$ *be a sample of size* n *from a distribution chosen from a Dirichlet process on* $(\boldsymbol{X},\ \boldsymbol{A})$ *with parameter* α. *Then*

$$P(X_1=\cdots=X_n) \tag{3.2}$$
$$=\sum{}'_{S(\sum im(i)=n)}\{n!M^{\sum m(i)}\sum_{x\in D}Q^{\sum m(i)}(\{x\})/\Pi_{i=1}^{n}(m(i)!\,i^{m(i)})M^{(n)}\}+(n-1)!M/M^{(n)},$$

where the summation $\sum'_{S(\sum im(i)=n)}$ *is taken over all sequences of* n *nonnegative integers* $m(1),\ \cdots,$

$m(n)$ *satisfying* $\sum_1^n im(i)=n$, *except for* $m(1)=\cdots=m(n-1)=0$, $m(n)=1$.

Now we shall consider the case that α is nonatomic. For positive integers n, u, and a sequence of u positive integers $r(1)$, $\cdots$, $r(u)$ such that $1\leq r(1)\leq\cdots\leq r(u)$ and $\sum_{i=1}^u r(i)=n$, $R(r(1), \cdots, r(u))$ consists of points in $\boldsymbol{X}^n$ and is defined as follows; $(x_1, \cdots, x_n)\in R(r(1), \cdots, r(u))$ implies that $r(i)$ values of x are equal and different from the remainders for $i=1, \cdots, u$ and $(x_1', \cdots, x_n')$ belongs to $R(r(1), \cdots, r(u))$ for each permutation of $x_1, \cdots, x_n, x_1', \cdots, x_n'$. For $(x_1, x_2, \cdots, x_n)\in R(r(1), \cdots, r(u))$, we denote $x_1, x_2, \cdots, x_n$ by $y_1, \cdots, y_1, y_2, \cdots, y_2, \cdots, y_u, \cdots, y_u$ neglecting the order, where the number of $y_i's$ is $r(i)$ for $i=1, 2, \cdots, u$. If there are same values in $r(1), \cdots, r(u)$, then we define $y's$ as follows; Suppose that $r(k(1))=\cdots=r(k(j))=r$ for some $k(1)<\cdots<k(j)$ and $r(i)\neq r$ for $i\neq k(1), \cdots, k(j)$. If $x_{s(1)}=\cdots=x_{t(1)}$, $x_{s(2)}=\cdots=x_{t(2)}$, $\cdots$, $x_{s(j)}=\cdots=x_{t(j)}$ correspond to $y_{k(1)}, y_{k(2)}, \cdots, y_{k(j)}$, respectively, then $\min(s(1), \cdots, t(1))<\min(s(2), \cdots, t(2))<\cdots<\min(s(j), \cdots, t(j))$. For example, we consider the case of $n=5$, $u=3$, $r(1)=1$, $r(2)=r(3)=2$. For $(x_1, x_2, x_3, x_4, x_5)\in R(1, 2, 2)$ and $x_1\neq x_2=x_3\neq x_4=x_5$, $y_1=x_1$, $y_2=x_2$, $y_3=x_4$. For $(x_1, x_2, x_3, x_4, x_5)\in R(1, 2, 2)$ and $x_3\neq x_1=x_4\neq x_2=x_5$, $y_1=x_3$, $y_2=x_1$, $y_3=x_2$.

For a sample of size n, $X_1, \cdots, X_n$, such that $(X_1, \cdots, X_n)\in R(r(1), \cdots, r(u))$, we denote it by $Y_1, \cdots, Y_1, \cdots, Y_u, \cdots Y_u$, neglecting the order. $Y_1, \cdots, Y_u$ denotes the distinct observations in the sample. In case that there are same values in $r(1), \cdots, r(u)$, we define $Y's$ by the same method to $y's$.

Proposition 3.2. *We suppose that* α *is nonatomic. Then for positive integers* u, $r(1), \cdots, r(u)$ *satisfying* $1\leq r(1)\leq\cdots\leq r(u)$, $\sum_1^u r(i)=n$ *and any set* $A_i\in\boldsymbol{A}(i=1, \cdots, u)$,

$$\begin{aligned}&P(Y_i\in A_i(i=1, \cdots, u), (X_1, \cdots, X_n)\in R(r(1), \cdots, r(u))) \qquad (3.3)\\ &=n!M^u\Pi_{i=1}^u Q(A_i)/\Pi_{i=1}^n(K_i(r(1), \cdots, r(u))!)\Pi_{i=1}^u r(i)M^{(n)},\end{aligned}$$

Proof. We take a symmetric function h such that $h(x_1, \cdots, x_n)=1$ *if* $(x_1, \cdots, x_n)\in R(r(1), \cdots, r(u))$, $y_i\in A_i(i=1, \cdots, u)$ and $=0$ otherwise. Then $\mathbf{E}h(X_1, \cdots, X_n)=P(Y_i\in A_i(i=1, \cdots, u), (X_1, \cdots, X_n)\in R(r(1), \cdots, r(u)))$. Thus by nothing that α is nonatomic, we have the proposition from Corollary of Theorem 2.1.

If we take $\mathbf{A}_i=\boldsymbol{X}$ for $i=1, \cdots, u$ in Proposition 3.2, then we have the following corollary, which is essentially as same as Proposition 3 of Antoniak [1].

Corollary. *If* α *is nonatomic, then*

$$\begin{aligned}&P((X_1, \cdots, X_n)\in R(r(1), \cdots, r(u))) \qquad (3.4)\\ &=n!M^u/\Pi_{i=1}^n(K_i(r(1), \cdots, r(u))!)\Pi_{i=1}^u r(i)M^{(n)}.\end{aligned}$$

Theorem 3.1. *We suppose that* α *is nonatomic. Given* $(X_1, \cdots, X_n)\in R(r(1), \cdots, r(u))$, $Y_1, \cdots, Y_u$ *are independent and identically identically distributed with the distribution* Q.

Proof. For any $A_i\in\boldsymbol{A}(i=1, \cdots, u)$, by Proposition 3.2 and its corollary we have

$$\mathrm{P}(Y_i \in A_i (i=1, \cdots, u) | (X_1, \cdots, X_n) \in R(r(1), \cdots, r(u))) = \Pi_{i=1}^{u} Q(A_i). \tag{3.5}$$

Note that the conditional probability given by (3.5) depends on a positive integer u and is constant for all sequences of u positive integers $r(1), \cdots, r(u)$ satisfying $1 \leq r(1) \leq \cdots \leq r(u)$ *and* $\sum_{i=1}^{u} r(i) = n$ with fixed u and n. Thus

$$\mathrm{P}(Y_i \in A_i (i=1, \cdots, u) | \cup^* \{(X_1, \cdots, X_n) \in R(r(1), \cdots, r(u))\}) = \Pi_{i=1}^{u} Q(A_i),$$

where $\cup^*$ is the union over all sequences of u positive integers $r(1), \cdots, r(u)$ suth that $1 \leq r(1) \leq \cdots \leq r(u)$, $\sum_{i=1}^{u} r(i) = n$ with fixed u, n. The event $\cup^* \{(X_1, \cdots, X_n) \in R(r(1), \cdots, r(u))\}$ denotes that the number of distinct observations in the sample $X_1, \cdots, X_n$ is u. Therefore we have the following corollary, which is Theorem 2.5 of Korwar and Hollander [3].

Corollary (Korwar and Hollander [3]). *Given the number of distinct observations in the sample,* u, $Y_1, \cdots, Y_u$ *are independent and identically distributed with the distribution* Q.

Finally, by the use of Corollary of Theorem 2.1 we shall prove the following proposition 3.3, which is essentially as same as Lemma 5 of Yamato [4].

Proposition 3.3. *Let* $h(x_1, \cdots, x_n)$ *be a real-valued measurable function defined on* $(\boldsymbol{X}^n, \boldsymbol{A}^n)$ *and symmetric in* $x_1, \cdots, x_n$. *Let* $\boldsymbol{P}$ *be a Dirichlet process on* $(\boldsymbol{X}, \boldsymbol{A})$ *with parameter* α. *Then*

$$\mathrm{E}\int_{X^n} h(x_1, \cdots, x_n) \Pi_{i=1}^{n} d\boldsymbol{P}(x_i) = \sum_{u=1}^{n} \sum_{S(\sum_1^u r(i)=n)} \frac{n! M^u}{\Pi_{i=1}^{n} (K_i(r(1), \cdots, r(u))!) \Pi_{i=1}^{u} r(i) M^{(n)}} \int_{X^u} h(x_1, \cdots, x_1, \cdots, x_u, \cdots, x_u) \Pi_{i=1}^{u} dQ(x_i), \tag{3.6}$$

provided all integrals of the right-hand side exist.

Proof. Let $X_1, \cdots, X_n$ be a sample of size n from a distribution $\boldsymbol{P}$. Since given $\boldsymbol{P}$, $X_1, \cdots, X_n$ are independent and identically distributed with the distribution $\boldsymbol{P}$,

$$\int_{X^n} h(x_1, \cdots, x_n) \Pi_{i=1}^{n} d\boldsymbol{P}(x_i) = \mathrm{E}[h(X_1, \cdots, X_n) | \boldsymbol{P}].$$

Taking expectation of the both sides of the above equation and applying Corollary of Theorem 2.1, we get the desired result.

References

[1] C.E. Antoniak, Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Ann. Statist.* 2 (1974) 1152-1174.

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

[3] R.M. Korwar and M. Hollander, Contributions to the theory of Dirichlet processes, *Ann. Probability* 1 (1973) 705-7111.

[4] H. Yamato, Relations between limiting Bayes estimates and U-statistics for estimable parameters, *J. Japan Statist. Soc.* 7 (1977) 57-66.
[5] B. Zehnwirth, Credibility and the Dirichlet process, *Scand. Actuarical J.* (1979) 13-23.