Journal of the Chinese Institute of Industrial Engineers

ISSN: 1017-0669 (Print) 2151-7606 (Online) Journal homepage: http://www.tandfonline.com/loi/tjci20

A BAYESIAN APPROACH EMPLOYING GENERALIZED DIRICHLET PRIORS IN PREDICTING MICROCHIP YIELDS

Tzu-Tsung Wong

A BAYESIAN APPROACH EMPLOYING GENERALIZED DIRICHLET PRIORS IN PREDICTING MICROCHIP YIELDS

Tzu-Tsung Wong*
Institute of Information Management
National Cheng Kung University
1, Ta-Sheuh Road, Tainan City 701, Taiwan, ROC

ABSTRACT

In the production model studied by Jewell and Chou, since some of the sorting probabilities for different categories of microelectronic chips tend to be positively correlated, a Dirichlet distribution is an inappropriate prior for that model. Jewell and Chou therefore propose an approximation approach to predict coproduct yields. Since a generalized Dirichlet distribution allows variables to be positively correlated, a Bayesian method by assuming generalized Dirichlet priors is presented to calculate the probabilities of future yields in this paper. We consider not only the mean values, but also either the variances or the covariances of the sorting probabilities to construct generalized Dirichlet priors. The numerical results indicate that the generalized Dirichlet distribution should be a reasonable prior, and the computation in forecasting coproduct output is relatively straightforward with respect to the approximation approach.

Keywords: Bayesian analysis, conjugate, correlation, generalized Dirichlet distribution

1. INTRODUCTION

In a production process of microelectronic chips, an automatic sorting device is used to divide chips into categories by size. The probability that a chip falls in a given category is assumed stable during a production process, hence the distribution of the output given the sorting probabilities of the categories will follow a multinomial distribution. Since engineers may adjust the parameters of a production process to improve the yields for some categories to meet market requirements, the sorting probability of each category generally cannot be determined accurately. Thus, incorporating prior experience to predict the yield of each category should be of interest for production planning. This Bayesian approach is widely used in estimating probabilities [5, 9].

Note that the sum of all sorting probabilities equals one, and that any sorting probability is nonnegative. In handling compositional data with unit-sum and nonnegativity constraints, analysts usually select a Dirichlet distribution to be the prior density in Bayesian analysis [2, 4, 11]. However, in the production model for microchips proposed by Jewell and Chou [6], since some of the sorting probabilities are significantly positively correlated, the Dirichlet distribution becomes an inappropriate prior. Without a proper prior, a Bayesian approach becomes infeasible, hence they use the prior means and the covariances of the sorting probabilities to construct an approximate linearized approach to forecast mean yields of chips. In that approach, forecasting variances or higher moments of the yields could be problematic.

Generalized Dirichlet distributions are first derived by Connor and Mosimann [3], and Lochner [8] shows that the variables in a generalized Dirichlet random vector can be positively correlated. This distribution has been employed in several applications [1, 7]. Since a generalized Dirichlet random vector allows variables to be positively correlated, we will assume that the sorting probabilities have a generalized Dirichlet prior to construct a Bayesian approach for the production problem proposed by Jewell and Chou [6] in this paper. As you will see, our approach is not only relatively straightforward and simple in computation, but also applicable at any time during a production process.

A brief introduction of the generalized Dirichlet distribution in Bayesian analysis is

* Corresponding author:tzutsung@mail.ncku.edu.tw

presented in section 2. The advantages of assuming a generalized Dirichlet prior instead of a Dirichlet prior are pointed out in section 3. In section 4, we will assume that the sorting probabilities have a generalized Dirichlet prior and propose a Bayesian approach to predict the yields of microelectronic chips. The conclusions of this paper are summarized in section 5.

2. GENERALIZED DIRICHLET DISTRIBUTION

A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ is said to have a k-variate generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$ if the joint density of $\mathbf{X}$ is

$$f(\mathbf{x}) = \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} x_i^{\alpha_i - 1}(1 - x_1 - \ldots - x_i)^{\gamma_i}$$

for $x_1+x_2+...+x_k \le 1$ and $x_j \ge 0$ for $j = 1, 2, ..., k$, where $\gamma_j = \beta_j - \alpha_{j+1} - \beta_{j+1}$ for $j = 1, 2, ..., k-1$, and $\gamma_j = \beta_j - 1$. When $\beta_j = \alpha_{j+1} + \beta_{j+1}$ for $j = 1, 2, ..., k-1$, the generalized Dirichlet distribution reduces to a k-variate Dirichlet distribution $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_k)$. Thus, the generalized Dirichlet distribution includes the Dirichlet distribution as a special case.

Let Z_j for $j = 1, 2, \ldots, k$ be independent beta random variables with parameters α_j and β_j, and let $X_1 = Z_1$ and $X_j = Z_j \prod_{i=1}^{j-1}(1 - Z_i)$ for $j = 2, 3, ..., k$. Then by Connor and Mosimann [3], random vector $\mathbf{X} = (X_1, X_2, ..., X_k)$ has a k-variate generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$. Wong [13] shows that the general moment function of this generalized Dirichlet distribution is

$$E(X_1^{r_1} X_2^{r_2} \ldots X_k^{r_k}) = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j + r_j)\Gamma(\beta_j + \delta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j + \beta_j + r_j + \delta_j)},$$

where $\delta_j = r_{j+1}+r_{j+2}+...+r_k$ for $j = 1, 2, ..., k-1$ and $\delta_k = 0$. Let $X_{k+1} = 1 - X_1 - ... - X_k$. Then we have

$$E(X_j) = E\left[Z_j \prod_{i=1}^{j-1}(1 - Z_i)\right] = \frac{\alpha_j}{\alpha_j + \beta_j} \prod_{i=1}^{j-1} \frac{\beta_i}{\alpha_i + \beta_i},$$

$$\mathrm{Var}(X_j) = \frac{\alpha_j(\alpha_j + 1)}{(\alpha_j + \beta_j)(\alpha_j + \beta_j + 1)} \prod_{i=1}^{j-1} \frac{\beta_i(\beta_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} - E(X_j)^2$$

for $j = 1, 2, \ldots, k$ and

$$E(X_{k+1}) = E\left[\prod_{i=1}^{k}(1 - Z_i)\right] = \prod_{i=1}^{k} \frac{\beta_i}{\alpha_i + \beta_i},$$

$$\mathrm{Var}(X_{k+1}) = \prod_{i=1}^{k} \frac{\beta_i(\beta_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} - E(X_{k+1})^2.$$

Connor and Mosimann [3] also show that

$$\mathrm{Cov}(X_1, X_j) = -\frac{E(X_j)}{E(1 - X_1)} \mathrm{Var}(X_1)$$

for $j = 2, 3, \ldots, k+1$,

$$\mathrm{Cov}(X_j, X_{j+1}) = E(Z_{j+1})E[Z_j(1 - Z_j)] \prod_{i=1}^{j-1} E[(1 - Z_i)^2] - E(X_j)E(X_{j+1})$$

for $j > 1$, and

$$\mathrm{Cov}(X_j, X_m) = \left[\frac{E(Z_m)}{E(Z_{j+1})}\right]\left[\prod_{i=j+1}^{m-1} E(1 - Z_i)\right] \mathrm{Cov}(X_j, X_{j+1})$$

for $m > j > 1$. Thus, X_1 is always negatively correlated with all other random variables. However, Lochner [8] shows that $\mathrm{Cov}(X_j, X_m)$ can be positive for $j, m > 1$. If there exists some $m > j$ such that X_j and X_m are positively (negatively) correlated, then X_j and X_i will be positively (negatively) correlated for all $i > j$.

Let X_j be the probability for a trial turning to be outcome j for $j = 1, 2, \ldots, k+1$, and let y_j be the number of trials turning to be outcome j in M trials given $\mathbf{x} = (x_1, x_2, ..., x_k)$. Suppose that the joint prior of $\mathbf{X}$ is a generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$, and that the likelihood function of $\mathbf{y}|\mathbf{x}$ (denoted $L(\mathbf{y}|\mathbf{x})$) follows a multinomial distribution, where $\mathbf{y} = \{y_1, y_2, ..., y_{k+1}\}$. Wong [13] shows that the joint posterior of $\mathbf{x}|\mathbf{y}$ is a generalized Dirichlet distribution $GD_k(\alpha_1', \alpha_2', ..., \alpha_k'; \beta_1', \beta_2', ..., \beta_k')$, where $\alpha_j' = \alpha_j + y_j$ and $\beta_j' = \beta_j + y_{j+1} + y_{j+2} + ... + y_{k+1}$ for $j = 1, 2, ..., k$. This means that the generalized Dirichlet distribution is conjugate to the multinomial sampling.

By Dirichlet integral, it can be shown that when a priori is a generalized Dirichlet distribution $GD_k(\alpha_1, \alpha_2, ..., \alpha_k; \beta_1, \beta_2, ..., \beta_k)$, the probability mass function of its prior predictive distribution will be

$$p(\mathbf{y}) = \int L(\mathbf{y} \mid \mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j')\Gamma(\beta_j')}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j' + \beta_j')}.$$

Similarly, when the likelihood function follows a multinomial distribution, the probability mass function of its posterior predictive distribution will be

$$p(\mathbf{w} \mid \mathbf{y}) = \prod_{j=1}^{k} \frac{\Gamma(\alpha_j^{'} + \beta_j^{'})\Gamma(\alpha_j^{''})\Gamma(\beta_j^{''})}{\Gamma(\alpha_j^{'})\Gamma(\beta_j^{'})\Gamma(\alpha_j^{''} + \beta_j^{''})},$$

where $\mathbf{w} = (w_1, w_2, ..., w_{k+1})$, $\alpha_j'' = \alpha_j'+w_j$, and $\beta_j'' = \beta_j'+w_{j+1}+w_{j+2}+...+w_{k+1}$ for j = 1, 2, ..., k.

For the production model introduced in section 1, let y_j be the number of chips produced for category j. We are interested in evaluating the probability of a specific data set $\mathbf{y} = \{y_1, y_2, ..., y_{k+1}\}$ due to market requirements given that the production capacity in some fixed period (e.g., a month) is $M = y_1+y_2+...+y_{k+1}$. Since $p(\mathbf{y}|M) = p(\mathbf{y},M)/p(M) = p(\mathbf{y})/p(M) \propto p(\mathbf{y})$, we have

$$p(\mathbf{y} \mid M) = \frac{M!}{y_1!y_2!...y_{k+1}!} \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \beta_j)\Gamma(\alpha_j^{'})\Gamma(\beta_j^{'})}{\Gamma(\alpha_j)\Gamma(\beta_j)\Gamma(\alpha_j^{'} + \beta_j^{'})}.$$

Alternatively, when we have produced y_j chips for category j for j = 1, 2, ..., k+1, we may also be interested in evaluating the probability of a data set $\mathbf{w} = \{w_1, w_2, ..., w_{k+1}\}$ such that the production capacity $M = \sum_{j=1}^{k+1}(w_j + y_j)$. In this case, we can use the posterior predictive distribution to calculate

$$p(\mathbf{w} \mid M, \mathbf{y}) = \frac{Q!}{w_1!w_2!...w_{k+1}!} \prod_{j=1}^{k} \frac{\Gamma(\alpha_j^{'} + \beta_j^{'})\Gamma(\alpha_j^{''})\Gamma(\beta_j^{''})}{\Gamma(\alpha_j^{'})\Gamma(\beta_j^{'})\Gamma(\alpha_j^{''} + \beta_j^{''})},$$

where $Q = w_1+w_2+...+w_{k+1}$.

3. COMPARISON

When random vector $\mathbf{X} = (X_1, X_2, ..., X_k)$ has a k-variate Dirichlet distribution $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$, Wilks [12] shows that

$$E(X_j) = \frac{\alpha_j}{\alpha},$$
$$Var(X_j) = \frac{\alpha_j(\alpha_j + 1)}{\alpha(\alpha+1)} - \left(\frac{\alpha_j}{\alpha}\right)^2$$

for j = 1, 2, ..., k+1, where $\alpha = \alpha_1+\alpha_2+...+\alpha_{k+1}$. For any $j \neq m$, the covariance between X_j and X_m is

$$Cov(X_j, X_m) = \frac{\alpha_j\alpha_m}{\alpha(\alpha+1)} - E(X_j)E(X_m)$$
$$= -\frac{1}{\alpha+1}E(X_j)E(X_m).$$

Hence, the variables in a Dirichlet random vector are all negatively correlated. Note that the number of parameters in a k-variate Dirichlet distribution is k+1. In constructing a Dirichlet prior, if the mean probabilities of the variables have been considered, there is only one degree of freedom (by selecting the value for α) that can be used to adjust the spread of the distribution. This implies that a Dirichlet prior is inappropriate in specifying the correlations among variables [10]. Another restriction of the Dirichlet distribution is that variables with the same mean must have the same variance, as shown in the following lemma.

Lemma 1 Let random vector $\mathbf{X} = (X_1, X_2, ..., X_k)$ have a k-variate Dirichlet distribution $D_k(\alpha_1, \alpha_2, ..., \alpha_k; \alpha_{k+1})$. Then

1) $E(X_j)/E(X_m) = Cov(X_i,X_j)/Cov(X_i,X_m)$ for any $i \neq j, m$.
2) If $E(X_j)/E(X_m) = b$ for some $b \geq 1$, we will have $1 \leq Var(X_j)/Var(X_m) \leq b$.

Proof. By the expression for the covariance between any two variables in a Dirichlet random vector, the first part of this lemma follows. To show the second part of this lemma, when $E(X_j)/E(X_m) = b$, we have $\alpha_j = b\alpha_m$. The ratio $Var(X_j)/Var(X_m)$ can be expressed as

$$\frac{Var(X_j)}{Var(X_m)} = \frac{E(X_j^2) - E^2(X_j)}{E(X_m^2) - E^2(X_m)} = \frac{E(X_j)(\alpha - \alpha_j)}{E(X_m)(\alpha - \alpha_m)}$$
$$= b\frac{\alpha - \alpha_j}{\alpha - \alpha_m}.$$

Since $b \geq 1$ implies $\alpha_j \geq \alpha_m$, we have $(\alpha-\alpha_j)/(\alpha-\alpha_m) \leq 1$, hence $Var(X_j)/Var(X_m) \leq b$. We also have

$$Var(X_j) - Var(X_m) = \frac{\alpha_j(\alpha - \alpha_j) - \alpha_m(\alpha - \alpha_m)}{\alpha^2(\alpha+1)}$$
$$= \frac{(\alpha_j - \alpha_m)(\alpha - \alpha_j - \alpha_m)}{\alpha^2(\alpha+1)} \geq 0. \blacklozenge$$

Lemma 1 shows that in a Dirichlet random vector, if $E(X_j)/E(X_m) = b \geq 1$, then $Cov(X_i,X_j)/Cov(X_i,X_m)$ must exactly equal to b for any $i \neq j, m$. In addition, when the ratio b is greater than or equal to 1, the ratio for the variances of X_j and X_m must be between 1 and b. This implies that when we use the mean probabilities to solve for the parameters of a Dirichlet distribution, we also set rigorous constraints for the variances and covariances of the variables. These constraints can make our model unrealistic.

Since the number of parameters in a k-variate generalized Dirichlet distribution is 2k when analysts have considered the means of the k variables, the remaining degrees of freedom that can be used to adjust the distribution are k. Analysts can choose k variances and/or covariances to construct a generalized Dirichlet prior. This means that the generalized Dirichlet distribution is a more realistic prior than the Dirichlet distribution. In addition, by the expressions given in section 2 for the generalized Dirichlet distribution, variables with the same mean need not have the same variance. This result increases the applicability of the generalized Dirichlet distribution. If analysts only know the means of variables, please refer to Wong [13] in constructing a generalized Dirichlet prior. The following methods use either k variances or k consecutive covariances (i.e., $Cov(X_j,X_{j+1})$ for j = 1, 2, ..., k) together with k means to construct a generalized Dirichlet prior.

When the means and the variances of variables X_1 through X_{k+1} are known, we can have the first and second moments of each variable. Since $X_1 = Z_1$, we can use the first and second moments of X_1 to set two linear equations $E(X_1) = \alpha_1/(\alpha_1+\beta_1)$ and $E(X_1^2)/E(X_1) = (\alpha_1+1)/(\alpha_1+\beta_1+1)$ to solve for α_1 and β_1. When the values of α_1 and β_1 are known, since $X_2 = Z_2(1–Z_1)$, we can use the first and second moments of X_2 to set two linear equations for α_2 and β_2, and so on.

Now, suppose that we have the $E(X_j)$ and $Cov(X_j,X_{j+1})$ for j = 1, 2, …, k. Since $X_1 = Z_1$ and $X_j = Z_j \prod_{i=1}^{j-1}(1-Z_i)$ for j = 2, 3, ..., k, we can use the k mean values to solve for the values of the $E(Z_i)$ for i = 1, 2, ..., k. Note that the covariance between X_j and X_{j+1} can be expressed as

$$\begin{aligned} Cov(X_j, X_{j+1}) &= E(X_j X_{j+1}) - E(X_j)E(X_{j+1}) \\ &= E(Z_{j+1})E[Z_j(1-Z_j)]\prod_{i=1}^{j-1}E(Z_i) \\ &\quad - E(X_j)E(X_{j+1}) \\ \Rightarrow E[Z_j(1-Z_j)] &= \frac{Cov(X_j, X_{j+1}) + E(X_j)E(X_{j+1})}{E(Z_{j+1})\prod_{i=1}^{j-1}E(Z_i)} \\ \Rightarrow \frac{\beta_j}{\alpha_j+\beta_j+1} &= \frac{Cov(X_j, X_{j+1}) + E(X_j)E(X_{j+1})}{E(Z_{j+1})\prod_{i=1}^{j}E(Z_i)}, \end{aligned}$$

because $E[Z_j(1–Z_j)] = \beta_j E(Z_j)/(\alpha_j+\beta_j+1)$. The above equation can be combined with the equation for $E(Z_j)$ to solve for α_j and β_j. Again, all equations for the parameters of the generalized Dirichlet distribution are linear.

For any two variables X_j and X_m in a Dirichlet random vector, it can be shown that X_j and $X_j/(1–X_m)$ are independent. Although the generalized Dirichlet distribution is more realistic than the Dirichlet distribution, there still exist conditional independencies among the variables in a generalized Dirichlet random vector. When random vector **X** has a generalized Dirichlet distribution, for any j < m, we have

$$\frac{X_m}{1-\sum_{i=1}^{j}X_i} = \frac{Z_m\prod_{i=1}^{m-1}(1-Z_i)}{\prod_{i=1}^{j}(1-Z_i)} = Z_m\prod_{i=j+1}^{m-1}(1-Z_i).$$

Since the Z_i for i = 1, 2, ..., k are independent, $X_m/(1-\Sigma_{i=1}^{j}X_i)$ is independent of X_j for any m > j.

4. BAYESIAN APPROACH

Let X_j be the probability for a chip to be in category j. In the production model proposed by Jewell and Chou [6], microelectronic chips are divided into 4 categories by size. By historical data, the mean probabilities are $E(X_1) = 0.1479$, $E(X_2) = 0.5096$, $E(X_3) = 0.1261$, and $E(X_4) = 0.2164$, and their covariance matrix is as follows:

$$T_0 = \begin{bmatrix} +0.00101 & -0.00153 & -0.00005 & +0.00057 \\ -0.00153 & +0.02886 & -0.01553 & -0.01180 \\ -0.00005 & -0.01553 & +0.01078 & +0.00480 \\ +0.00057 & -0.01180 & +0.00480 & +0.00643 \end{bmatrix}.$$

Let ρ_{ij} be the correlation coefficient between X_i and X_j. Then both $\rho_{14} = 0.224$ and $\rho_{34} = 0.576$ are significantly positive, and $\rho_{13} = -0.014$ is very close to zero.

As illustrated by Jewell and Chou [6], the Dirichlet distribution constructed by considering $E(X_1)$, $E(X_2)$, $E(X_3)$, and $Var(X_2)$ is $D_3(1.1326, 3.9033, 0.9659; 1.6577)$. In this case, the covariance matrix is:

$$T_1 = \begin{bmatrix} +0.01455 & -0.00870 & -0.00215 & +0.00370 \\ -0.00870 & +0.02886 & -0.00742 & -0.01274 \\ -0.00215 & -0.00742 & +0.01273 & -0.00315 \\ -0.00370 & -0.01274 & -0.00315 & +0.01958 \end{bmatrix},$$

and we have $\rho_{13} = -0.158$, $\rho_{14} = -0.219$, and $\rho_{34} = -0.200$ that are all significantly negative. Thus, it is inappropriate to assume that the sorting probabilities X_1 through X_3 have a Dirichlet distribution.

Suppose that $\mathbf{X} = (X_1, X_2, X_3)$ has a 3-variate generalized Dirichlet distribution $GD_3(\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2, \beta_3)$. By considering the $E(X_j)$ and the $Var(X_j)$ for j = 1, 2, 3, the generalized Dirichlet distribution is $GD_3(18.32503, 3.06981, 1.442; 105.57645, 2.06321, 2.47838)$, and its covariance matrix is:

$$T_2 = \begin{bmatrix} +0.00101 & -0.00060 & -0.00015 & -0.00026 \\ -0.00060 & +0.02886 & -0.01040 & -0.01785 \\ -0.00015 & -0.01040 & +0.01078 & -0.00023 \\ -0.00026 & -0.01785 & -0.00023 & +0.01834 \end{bmatrix}.$$

In this case, we have ρ_{13} = -0.045, ρ_{14} = -0.060, and ρ_{34} = -0.016 that are relatively insignificantly negative.

Although both $Cov(X_1,X_4)$ and $Cov(X_3,X_4)$ become negatively correlated in matrix T_2, correlation coefficients ρ_{14} and ρ_{34} in matrix T_2 are close to zero. Let $Cov_T(X_j,X_m)$ represent $Cov(X_j,X_m)$ in matrix T. Then we have $\frac{Cov_{T_0}(X_1,X_2)}{Cov_{T_2}(X_1,X_2)} = 2.54$, $\frac{Cov_{T_0}(X_1,X_3)}{Cov_{T_2}(X_1,X_3)} = 0.31$, $\frac{Cov_{T_0}(X_2,X_3)}{Cov_{T_2}(X_2,X_3)} = 1.49$, and $\frac{Cov_{T_0}(X_2,X_4)}{Cov_{T_2}(X_2,X_4)} = 0.66$ that are all between 1/3 and 3. Based on these observations, the values in matrix T_2 are neither significantly overrated nor significantly underrated. Hence, matrix T_2 should be an acceptable approximation of matrix T_0.

In considering the $E(X_j)$ and the $Cov(X_j,X_{j+1})$ for j = 1, 2, 3 to construct the generalized Dirichlet distribution $GD_3(\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2, \beta_3)$, the significantly positive correlation ρ_{34} will still be positive while its value may change, because both $Var(X_3)$ and $Var(X_4)$ may change. Since X_1 is always negatively correlated with other variables in a generalized Dirichlet random vector, correlation coefficient ρ_{14} cannot be positive in a generalized Dirichlet approach. The generalized Dirichlet distribution generated from the $E(X_j)$ and the $Cov(X_j,X_{j+1})$ for j = 1, 2, 3 is GD_3(7.12406, 1.84934, 2.24663; 41.04404, 1.24293, 3.85544), and the covariance matrix of this distribution is:

$$T_3 = \begin{bmatrix} +0.00256 & -0.00153 & -0.00038 & -0.00065 \\ -0.00153 & +0.04372 & -0.01553 & -0.02665 \\ -0.00038 & -0.01553 & +0.01112 & +0.00480 \\ -0.00065 & -0.02665 & +0.00480 & +0.02251 \end{bmatrix}.$$

In this case, we have ρ_{13} = -0.071, ρ_{14} = -0.086, and ρ_{34} = 0.303. As expected, correlation coefficient ρ_{34} is still positive, but correlation coefficient ρ_{14} becomes negative. Both ρ_{13} and ρ_{14} are relatively insignificantly negative.

Note that all corresponding items in covariance matrices T_0 and T_3 have the same sign except $Cov(X_1,X_4)$. In addition, both ρ_{13} and ρ_{14} in T_3 are close to zero, and correlation ρ_{34} in T_3 is still significantly positive. Thus, covariance matrix T_3 should also be a reasonable approximation of covariance matrix T_0. However, except for the items in T_3 used to construct the distribution, the absolute value of each item in T_3 is larger than the absolute value of the corresponding item in T_0. This implies that an analyst with prior information represented by covariance matrix T_0 is more confident about the estimates of the X_j than an analyst with prior information represented by covariance matrix T_3. So, any forecasting result obtained from matrix T_3 will be conservative.

For the two generalized Dirichlet distributions constructed in this section, we can see that the sum of parameters α_1 and β_1 corresponding to X_1 is much larger than the sum of the parameters corresponding to any other variable. When we observe an event corresponding to X_1, the change on the mean probability of any variable will not be significant. However, when we observe an event corresponding to any variable other than X_1, the mean probability of any variable with an index greater than one will have a relatively significant change. This implies that in this production problem, the relative confidence levels about X_1 and other variables in **X** are significantly different. This again suggests that the sorting probabilities should not have a Dirichlet distribution.

Next, we will investigate the differences among the forecasting results by using either covariance matrix T_1, T_2, or T_3. Suppose that the production capacity before a specific time is $M = y_1+y_2+y_3+y_4 = 100$. We will fix two quantities in $\{y_1, y_2, y_3, y_4\}$ according to their mean probabilities to study the relationship between the other two variables. For instance, if we want to study the relationship between X_1 and X_2, we will calculate the production probabilities $P\{y_1, y_1+y_2 = 66, y_3 = 12| M = 100\}$ for y_1 = 0, 1, ..., 66. There are two positive covariances in the original covariance matrix T_0. Since $Cov(X_1,X_4)$ cannot be positive in a generalized Dirichlet approach, we will analyze the relationship between X_3 and X_4 to investigate whether the impact of positive correlations will appear in a figure for production probabilities. We will also analyze the relationship between X_1 and X_3 as a contrast case, since $Cov(X_1,X_3)$ in the four covariance matrices are all negative and close to zero except in matrix T_1. The numerical results of these two cases are shown in Figures 1 and 2.

In both figures, we can see that when random vector **X** has a Dirichlet distribution, the forecasting probability is not sensitive to the change of the horizontal axis. Note that the largest mean ratio in this problem is $E(X_2)/E(X_3)$ = 4.04. By part 2 of Lemma 1, the variance ratio of any two variables in **X** cannot be larger than 4.04 when **X** has a Dirichlet distribution. This explains why the curves for the Dirichlet distribution are flat in both figures. However, the curves for the two generalized Dirichlet distributions are both bell-shaped. This is consistent with the result that in a generalized Dirichlet random vector, variables with the same mean need not have the same variance.

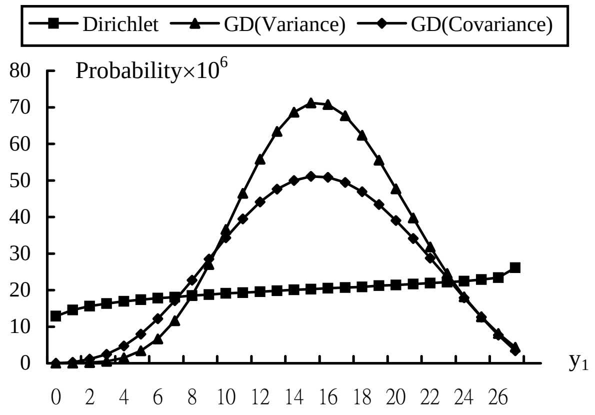

Figure 1. The forecasting probability of $P\{y_1+y_3 = 27, y_2 = 51| M=100\}$

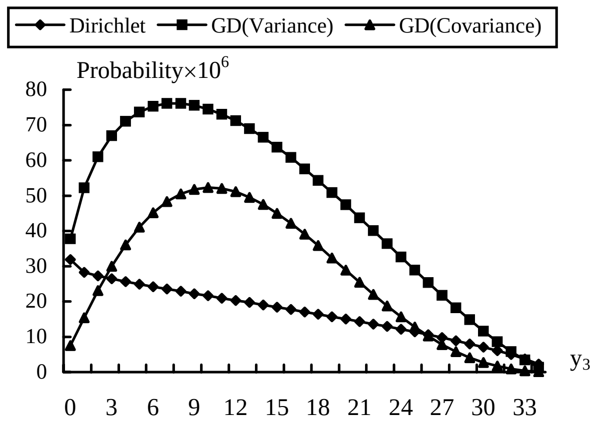

Figure 2. The forecasting probability of $P\{y_3+y_4 = 34, y_1 = 15|$

Consider only the curves for the generalized Dirichlet approaches. In Figure 1, since the mean value of X_1 is a little bit greater than the mean value of X_3, the value of y_1 with the largest forecasting probability slightly deviates from the center to the right. In Figure 2, since the mean value of X_3 is less than the mean value of X_4, the value of y_3 with the largest forecasting probability deviates from the center to the left. We examined other pairs of variables in **X** and found that all curves obey this rule. Thus, the direction of deviation in a forecasting probability curve is determined by the mean values of variables. Although $Cov(X_3,X_4)$ is negative in matrix T_2 and positive in matrix T_3, we are not able to determine whether the covariance of two variables will be positive or negative from the pattern of a forecasting probability curve.

By examining the variances in covariance matrices T_2 and T_3, we can see that each variance in T_2 is smaller than the corresponding variance in T_3. So, the bell-shaped curves generated from T_2 are steeper than the curves generated from T_3. Since the forecasting probability from matrix T_3 is generally too conservative, using covariance matrix T_2 to predict future yields should be a better choice in this case, even though $Cov(X_3,X_4)$ in matrix T_2 is slightly less than zero.

The most significant advantage of this generalized Dirichlet approach instead of the approximation approach proposed by Jewell and Chou [6] is that we can calculate any forecasting probability at any time in a production process. At the beginning of a production process, we can use the prior predictive expression of the generalized Dirichlet distribution to calculate any forecasting probability. If a production process has produced y_j chips for category j for j = 1, 2, 3, 4, we can use the posterior predictive expression to evaluate the probability of meeting market requirements. Since both expressions are in closed form, the computation of these probabilities of interest in a generalized Dirichlet approach will be tractable and efficient. No matter which predictive distribution will be used to estimate future yields, forecasting variances or higher moments of the yields is computationally tractable.

5. SUMMARY

When two variables in a random vector are significantly positively correlated, or variables with the same mean have significantly different variances, it is inappropriate to assume that the random vector has a Dirichlet distribution. Since some of the sorting probabilities in the production model for microelectronic chips are significantly positively correlated, and the sorting probabilities with similar mean values have significantly different variances, we assume that the sorting probabilities have a generalized Dirichlet prior (instead of a Dirichlet distribution) to predict the yield of each chip category in this paper. We consider not only the mean values of the sorting probabilities, but also either their variances or their covariances to construct generalized Dirichlet priors. The numerical results indicate that the generalized Dirichlet model should be an applicable approach for this realistic case.

The most significant advantage of our generalized Dirichlet model is that we always have a closed-form expression to predict the probabilities for different combinations of yields. Thus, the computation for a probability of interest will generally be relatively straightforward and simple with respect to the original approximation approach. In addition, since the generalized Dirichlet distribution is conjugate to the multinomial sampling, the posterior predictive distribution is also in closed form. When we already have some output from a production process before market requirements are satisfied, we can use the posterior predictive distribution of the generalized Dirichlet distribution to check whether the company will have enough products for all chip categories to meet the requirements in time. Thus, our approach can estimate the probability of meeting market requirements at any time in a production process.

ACKNOWLEDGEMENTS

This research is supported by the National Science Council in Taiwan under Grant No. NSC 89-2213-E-006-217.

REFERENCES

1. Bouguila N., D. Ziou and J. Vaillancourt, "Novel mixtures based on the Dirichlet distribution: Application to data and image classification," *Lecture Notes in Artificial Intelligence*, **2734**, 172-181, (2003).
2. Bouguila, N., D. Ziou and J. Vaillancourt, "Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application," *IEEE Transactions on Image Processing*, **13 (11)**, 1533-1543, (2004).
3. Connor, R. J. and J. E. Mosimann, "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution," *Journal of the American Statistical Association*, **64**, 194-206 (1969).
4. Gasbarra D, M. J. Sillanpaa and E. Arjas, "Backward simulation of ancestors of sampled individuals," *Theoretical Population Biology*, **67(2)**, 75-83, (2005).
5. Geweke, J., "Bayesian econometrics and forecasting," *Journal of Econometrics*, **100(1)**, 11-15, (2001).
6. Jewell, W. S and S. K. Chou, "Predicting Coproduct Yields in Microchip Fabrication," *Case Study in Bayesian Statistics* by C. Gatsonis, J. S. Hodges, R. E. Kass, and N. D. Singpurwalla, Springer-Verlag, New York, 351-361 (1993).
7. Lewy, P., "A generalized Dirichlet distribution accounting for singularities of the variables," *Biometrics*, **52(4)**, 1394-1409, (1996).
8. Lochner, R. H., "A Generalized Dirichlet Distribution in Bayesian Life Testing," *Journal of the Royal Statistical Society*, Series B, **37**, 103-113 (1975).
9. Mostaghimi, M., "Monetary policy, composite leading economic indicators and predicting the 2001 recession," *Journal of Forecasting*, **23(7)**, 463-477, (2004).
10. Thall, P. F. and H. G. Sung, Some Extensions and Applications of a Bayesian Strategy for Monitoring Multiple Outcomes in Clinical Trials, *Statistics in Medicine*, **17**, 1563-1580 (1998).
11. Vermeesch, P., "Statistical uncertainty associated with histograms in the Earth sciences," *Journal of Geophysical Research-Solid Earth*, **110(B2)**, Art. No. B02211 (2005).
12. Wilks, S. S., *Mathematical Statistics*, John Wiley, New York (1962).
13. Wong, T. T., "Generalized Dirichlet Distribution in Bayesian Analysis," *Applied Mathematics and Computation*, **97**, 165-181 (1998).

ABOUT THE AUTHORS

Tzu-Tsung Wong is an assistant professor in the Institute of Information Management at National Cheng Kung University, Taiwan, ROC. He received his Ph.D. degree majored in industrial engineering from the University of Wisconsin at Madison, USA. His research interests include Bayesian statistical analysis, naïve Bayesian classifiers, model trees for numeric prediction, and classification methods for microarray data.

用貝式統計分析來做晶片生產的預測

翁慈宗
成功大學資訊管理研究所
台南市大學路一號

摘要

在Jewell和Chou所討論的微電子晶片生產模式中，晶片需按照大小來分成四類，由於有些晶片類別產生的機率存在著正相關的關係，因此不適合假設各個晶片類別產生機率的連結分佈爲Dirichlet分佈，Jewell 和Chou便提出求近似解的方法，他們的方法很難用來計算變數二階以上的動率，而且用來預測生產量的計算較複雜，因此本研究假設各個晶片類別產生機率的連結分佈爲generalized Dirichlet分佈，然後用貝氏統計的方式來預測各類晶片的生產量。由於在generalized Dirichlet分佈中，允許各變數間爲正相關，而且對於相同期望值的變數，允許它們有不同的變異數，因此對本問題而言，generalized Dirichlet分佈是一個合適的多變量機率分佈。本研究所採用的貝氏統計分析方式，對於預測生產量的計算，比求近似解的方法簡單，可計算變數二階以上的動率，而且在使用的時機上較不受限制。

(聯絡人: tzutsung@mail.ncku.edu.tw)