Chapter 4

Nonparametric Bayes Estimation

4.1 Nonparametric Bayes estimates, based on Dirichlet processes

For a long time there were a lot of unsuccessful efforts directed toward the solution of many nonparametric problems with the help of the Bayes approach. This can be explained mainly by difficulties a researcher encounters, when he attempts to find a suitable prior distribution, determined on a sample space. Such a distribution in nonparametric problems is chosen in the form of a set of probability distributions on the given sample space. The first work in this field where some progress has been achieved belongs to Ferguson [80]. Ferguson formulated the requirements which must be imposed on a prior distribution:

1) The support of a prior distribution must be large with respect to some suitable topology of a space of probability distributions, defined on the sample space;
2) The posterior distribution under the given sample of observations from the real distribution of probabilities must have as simple a form as possible.

These properties are contradictory, bearing in mind the fact that each of them can be found from each other. In the work [80] a class of prior distributions was proposed, named Dirichlet processes, which not only possess the first property but also satisfy the second one. Exactly such a choice was offered, because a prior distribution of a random probability measure appears to be also a Dirichlet process. Another argument in favor of using the Dirichlet distribution in practical applications is explained by the fact that this distribution is a good approximation of many parametric probability distributions. Special attention is paid to this question in the works by Dalal [51] and Hall [52]. This distribution appears also in problems dealing with order statistics [266]. In the Bayes parametric theory it is used as a conjugate with the sampling likelihood kernel for the parameters of a multinomial distribution [63]. Below we give the definition of the Dirichlet distribution and formulate (from
a practical point-of-view) some of its important properties.

### 4.1.1 Definition of the Dirichlet process

Denote by $\Gamma(\alpha, \beta)$ a gamma probability distribution with the shape parameter $\alpha \geq 0$ and scalar parameter $\beta > 0$. For $\alpha > 0$ this distribution has the probability density

$$F(z; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{-z/\beta} z^{\alpha-1} I_{[0,\infty)}(z),$$

(4.1)

where $I_{S}(z)$ is the indicator function of the set $S$ identically equal to unity for all $z \in S$ and equal to zero otherwise.

Let $z_1, z_2, \ldots, z_k$ be independent random variables, $z_j \sim \Gamma(\alpha_j, 1)$ where $\alpha_j \geq 0$ for all $j$ and $\alpha_j > 0$ for some $j$. The Dirichlet distribution with the parameters $\alpha_1, \alpha_2, \ldots, \alpha_K$, denoted further on by $D(\alpha_1, \alpha_2, \ldots, \alpha_K)$, is determined as the distribution of the variables $Y_1, Y_2, \ldots, Y_K$, defined in the following way:

$$Y_j = \frac{z_j}{\sum_{i=1}^{k} z_i}, \quad j = 1, 2, \ldots, k.$$ 

Note that if some $\alpha_j = 0$, then the corresponding $Y_j$ also degenerates to zero. Provided that $\alpha_j > 0$ for all $j$, $(k-1)$-dimensional distribution of the variables $Y_1, \ldots, Y_{K-1}$ is absolutely continuous with probability density

$$f(y_1, \ldots, y_{K}; \alpha_1, \ldots, \alpha_K) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_K)} \prod_{j=1}^{k-1} y_j^{\alpha_j-1} \left(1 - \sum_{j=1}^{k-1} y_j\right)^{\alpha_K-1} I_{S}(y_1, \ldots, y_{k-1}),$$

(4.2)

where $S$ is the following set:

$$\left\{ (y_1, \ldots, y_{k-1}) : y_j \geq 0, \sum_{j=1}^{k-1} y_j \leq 1 \right\}.$$ 

For $k = 2$, the expression (4.2) is transformed into the beta probability distribution which will be denoted by $Be(\alpha_1, \alpha_2)$.

The use of the Dirichlet distribution is based on the following properties:

**Property 4.1.** If $(Y_1, \ldots, Y_K) \sim D$ and $r_1, r_2, \ldots, r_\ell$ are some integer numbers, satisfying the inequality $0 < r_1 < r_2 < \cdots < r_\ell = k$, then

$$\left( \sum_{i=1}^{r_1} Y_i, \sum_{i=r_1+1}^{r_2} Y_i, \ldots, \sum_{i=r_{\ell-1}+1}^{r_\ell} Y_i \right) \sim D \left( \sum_{i=1}^{r_1} \alpha_i, \sum_{i=r_1+1}^{r_2} \alpha_i, \ldots, \sum_{i=r_{\ell-1}+1}^{r_\ell} \alpha_i \right).$$

**Property 4.2.** If a prior probability distribution of the variables $Y_1, \ldots, Y_K$ is $D(\alpha_1, \ldots, \alpha_K)$ and if $P\{X = j \mid Y_1, \ldots, Y_K\} = Y_j$ is almost surely for $j = 1, 2, \ldots, k$, then the posterior probability distribution of the variables $Y_1, \ldots, Y_K$ for $X = j$ is the Dirichlet distribution $D\left(\alpha_1^{(j)}, \ldots, \alpha_K^{(j)}\right)$, where $\alpha_i^{(j)} = \alpha_i$ as $i \neq j$ and $\alpha_i^{(j+1)} = \alpha_i + 1$ as $i = j$. 
