L E C T U R E 17

Example 19.1. Suppose that a parameter Θ, takes on values $\theta_1 = 1$, $\theta_2 = 10$, and $\theta_3 = 20$. The distribution of X is discrete and depends on Θ as shown in the following table:

$$\begin{pmatrix} & \theta_1 & \theta_2 & \theta_3 \\ \\ x_1 & 0.1 & 0.2 & 0.4 \\ x_2 & 0.1 & 0.2 & 0.2 \\ x_3 & 0.2 & 0.2 & 0.2 \\ x_4 & 0.6 & 0.4 & 0.2 \end{pmatrix}$$

Assume a prior distribution of Θ is

$$p(\theta_1) = 0.5, \qquad p(\theta_2) = 0.25, \qquad p(\theta_3) = 0.25$$

a) Suppose that x_2 is observed. What is the posterior distribution of Θ?

b) What is the Bayes estimate under squared error loss in this case?

c) What is the Bayes estimate for the loss function $L(\theta, \widehat{\theta}) = |\theta - \widehat{\theta}|$?

Solution. Using Bayes formula we get:

$$P(\theta/x_2) = \frac{P(\theta) \cdot P(x_2/\theta)}{P(x_2)}.$$

Prior distribution for Θ is the following:

Θ :	1	10	20
$P(\Theta)$:	0.5	0.25	0.25

Therefore, we can find the posterior distribution for parameter θ:

$$P(\theta_i/x_2) = \frac{P(\theta_i \cap x_2)}{P(x_2)}, \quad i = 1, 2, 3,$$

and using the Total probability formula we obtain:

$$P(x_2) = 0.1 \cdot 0.5 + 0.2 \cdot 0.25 + 0.2 \cdot 0.25 = 0.15$$

$$P(\theta_1/x_2) = \frac{0.5 \cdot 0.1}{0.15} = \frac{5}{15} = \frac{1}{3},$$

$$P(\theta_2/x_2) = \frac{0.2 \cdot 0.25}{0.15} = \frac{5}{15} = \frac{1}{3},$$

and

$$P(\theta_3/x_2) = \frac{0.25 \cdot 0.2}{0.15} = \frac{5}{15} = \frac{1}{3},$$

Posterior distribution for parameter θ is

$\Theta:$	1	10	20
$P(\Theta/x_2):$	1/3	1/3	1/3

b) Bayes estimator is posterior expectation:

$$\frac{31}{3} = 10.(3) \approx 10.3333.$$

c) Bayes estimator is the posterior median, that is 10.

Corollary 19.1. If η_1, η_2, ..., η_n are independent and Γ random variables with $(\alpha_1, 1)$ and $(\alpha_2, 1)$,...$(\alpha_n, 1)$ then $\eta_1 + \eta_2 + ...\eta_n$ is Γ distribution with $\left(\sum\limits_{i=1}^{n} \alpha_i, 1\right)$ parameters, that is, density function of the sum $\sum_{i=1}^{n} \eta_i$ has the following form:

$$f_{\eta_1+\eta_2+...+\eta_n}(x) = \frac{1}{\Gamma(\alpha_1 + \alpha_2 + ... + \alpha_n)}\, x^{\alpha_1+\alpha_2+...+\alpha_n-1}\, e^{-x}, \qquad \text{for} \quad x > 0,$$

and 0 for $x \leq 0$.

Lemma 19.2. The χ^2-distribution has the additive property that if η_1 and η_2 are independent χ^2 random variables with n_1 and n_2 degrees of freedom, respectively, then $\eta_1 + \eta_2$ is χ^2 with $n_1 + n_2$ degrees of freedom, that is, density function of the sum $\eta_1 + \eta_2$ has the following form:

$$f_{\eta_1+\eta_2}(x) = \frac{1}{2^{(n_1+n_2)/2}\, \Gamma\left(\dfrac{n_1+n_2}{2}\right)}\, x^{\frac{n_1+n_2}{2}-1}\, e^{-x/2}, \qquad \text{for} \quad x > 0,$$

and 0 for $x \leq 0$.

For large values of α_j, distribution (19.3) is concentrated far from the borders of the simplex, which corresponds to more or less uniform discrete distributions of vector p. On the other hand, for small values α_j, the Dirichlet distribution is concentrated near the borders of the simplex, which corresponds to extremely non-uniform distributions of p, which are some large p_j and the rest are small. In particular, if all α_j equal to some small value α, then, from symmetry, all p_j will have the same mathematical expectation $1/n$, but probability that at least one of p_j will be much greater than the average; for which j or which value p_j will be great - only a matter of chance.

It is difficult to use directly formula (19.3), because the linear dependence of the components p_j. It has long been known that it is much more convenient to describe the Dirichlet distribution in terms of independent gamma quantities.

Let $Y_1, Y_2, ..., Y_n$ be independent, positive random variables, and Y_j have the following density function:

$$g_\alpha(y) = \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}, \quad y > 0, \tag{19.4}$$

where $\alpha = \alpha_j$. Suppose that $Y = Y_1 + Y_2 + ... + Y_n$. It is not difficult to verify that a vector p with components

$$p_j = \frac{Y_j}{Y} \tag{19.5}$$

has distribution $D(\alpha_1, \alpha_2, ..., \alpha_n)$ and does not depend on Y. The proof consists in a direct calculation using the change of variables carried out using the function acting from R^n to R^n and given by the formula

$$(Y_1, Y_2, ..., Y_n) \mapsto (Y, p_1, p_2, ..., p_{n-1}).$$

As a consequence of this calculation, we obtain that the random variable Y also has a distribution (19.4) with the parameter

$$\alpha = \alpha_1 + \alpha_2 + ... + \alpha_n.$$

Proof. From probability theory we know the following statement.

Lemma 19.3. If η_1 and η_2 are 2 independent random variables with $f_1(x)$ and $f_2(x)$ density functions. The density function of the variable $\dfrac{\eta_1}{\eta_2}$ has the following form:

$$f_{\eta_1/\eta_2}(x) = \int_0^{+\infty} y f_2(y) f_1(xy)\, dy - \int_{-\infty}^{0} y f_2(y) f_1(xy)\, dy.$$

Exercise 19.1. Calculate the density function of variable $\dfrac{\eta_1}{\eta_2}$, if both η_1 and η_2 are standard normal random variables.

Since $f_2(y) = 0$ if $y \leq 0$, in our case we have

$$f_{\eta_1/\eta_2}(x) = \int_0^{+\infty} y f_2(y) f_1(xy)\, dy =$$

$$= \frac{1}{\Gamma\left(\sum_{j=1}^{n} \alpha_j\right)} \frac{1}{\Gamma(\alpha_i)} \int_0^{+\infty} y \cdot y^{\sum \alpha_j - 1}\, e^{-y}\, (xy)^{\alpha_i - 1} e^{-xy} dy =$$

$$= \frac{x^{\alpha_i - 1}}{\Gamma\left(\sum_{j=1}^{n} \alpha_j\right)} \frac{1}{\Gamma(\alpha_i)} \int_0^{+\infty} y^{\sum \alpha_j + \alpha_i - 1}\, e^{-y(1+x)}\, dy =$$

Make the change of variable $y(1+x) = t$ $dy = \dfrac{dt}{1+x}$, we obtain

$$\frac{x^{\alpha_i}}{\Gamma\left(\sum_{j=1}^{n} \alpha_j\right) \Gamma(\alpha_i)\, (1+x)^{\sum \alpha_j + \alpha_i}} \int_0^{+\infty} t^{\sum \alpha_j + \alpha_i - 1}\, e^{-t}\, dt =$$

$$= \frac{x^{\alpha_i - 1}}{(1+x)^{\sum \alpha_j + \alpha_i}} \frac{\Gamma\left(\sum_{j=1}^{n} \alpha_j + \alpha_i\right)}{\Gamma\left(\sum_{j=1}^{n} \alpha_j\right) \Gamma(\alpha_i)}.$$

It is the Beta distribution $B\left(\alpha_i, \sum_{j=1}^{n} \alpha_j + \alpha_i\right)$.

Another way to prove this is to use the Laplace transform:

$$\int_0^{\infty} g_\alpha(y) e^{-\theta y}\, dy = \frac{1}{(1+\theta)^\alpha}, \quad \theta > -1, \tag{19.6}$$

which also shows that Gamma distribution $G(\alpha)$ defined by formula (19.4) is infinitely divisible and has a Levi-Khinchin representation

$$\frac{1}{(1+\theta)^\alpha} = \exp\{-\alpha \int_0^{\infty} (1 - e^{-\theta z}) z^{-1} e^{-z}\, dz\}. \tag{19.7}$$

Representation (19.7) corresponds to a subordinator, known as the Moran gamma process. This process is given by the parameters

$$\beta = 0, \quad \gamma(dz) = z^{-1} e^{-z} \, dz. \tag{19.8}$$

In this case, the increment $\varphi(t) - \varphi(s)$ has a distribution $G(t-s)$. Note, that

$$\gamma(0, \infty) = \int_0^\infty z^{-1} e^{-z} \, dz = \infty,$$

and therefore the jumps of the process φ are everywhere dense.

For $\alpha_1, \alpha_2, ..., \alpha_n > 0$ we set

$$t_0 = 0, \quad t_j = \alpha_1 + \alpha_2 + ... + \alpha_j, \quad 1 \le j \le n. \tag{19.9}$$

Then random variable $Y_j = \varphi(t_j) - \varphi(t_{j-1})$ has a distribution $G(\alpha_j)$, and all random variables Y_j are independent. Since

$$Y = Y_1 + Y_2 + ... + Y_n = \varphi(t_n),$$

we see that formula (19.1) in which

$$p_j = \frac{\varphi(t_j) - \varphi(t_{j-1})}{\varphi(t_n)} \tag{10}$$

sets a random vector from Δ_n with distribution $D(\alpha_1, \alpha_2, ..., \alpha_n)$.

The Poisson- Dirichlet distribution (continuation).

Just as the multinomial distribution is a multivariate extension of the binomial distribution, the Dirichlet distribution is a multivariate generalization of the beta distribution. If X is a k-dimensional vector and $X \sim$ Dirichlet $(\alpha_1, \alpha_2, ..., \alpha_k)$, then:

$$f(X) = \frac{\Gamma(\alpha_1 + ... + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} x_1^{\alpha_1 - 1} \cdots x_k^{\alpha_k - 1}.$$

Just as the beta distribution is a conjugate prior for the binomial distribution, the Dirichlet is a conjugate prior for the multinomial distribution. We can see this result clearly, if we combine a Dirichlet distribution as a prior with a multinomial distribution likelihood:

$$f(p_1...p_k|X) \propto f(X|p_1...p_k)f(p_1...p_k) \propto Multinomial(X|p_1...p_k)Dirichlet(p_1...p_k|\alpha_1...\alpha_k)$$

$$\propto Dirichlet(p_1...p_k|\alpha_1 + x_1, \alpha_2 + x_2, ..., \alpha_k + x_k) \propto p_1^{\alpha_1+x_1-1} \cdot p_2^{\alpha_2+x_2-1} \cdots p_k^{\alpha_k+x_k-1}.$$

Notice how here, as we discussed at the beginning of the section, the vector X in the original specification of the Dirichlet pdf has been changed to a vector p. In this specification, p is the random variable in the Dirichlet distribution, whereas $\alpha_1...\alpha_k$ are the parameters representing prior counts of outcomes in each of the k possible outcome categories. Also observe how the resulting Dirichlet posterior distribution looks just like the resulting beta posterior distribution, only with more possible outcomes.