

Learning the parameters of the Multinomial Distribution

C. Flinn

January 2006

For simplicity, assume that there are 3 possible wages, and the values of possible wage offers are known to be $w_i = i$, $i = 1, 2, 3$. Then the inferential problem is to learn the values of the probabilities attached to each w_i , π_i , where $\pi_i > 0$, $i = 1, 2, 3$; $\sum_i \pi_i = 1$.

Each agent begins with a prior on the values of the vector π , which we denote by $F(\pi)$. Now imagine the agent has drawn N times, and has observed the value w_i a total of N_i times. Denote the sample information by $R = (N_1, N_2, N_3)$. Given values of the parameters π , the probability of the sample is given by

$$L(R; \pi) = \pi_1^{N_1} \pi_2^{N_2} \pi_3^{N_3}.$$

We think of the actual population value of the parameters as being the realization of a random variable π , and this is how we think of the prior. Then we should think of the likelihood $L(\pi)$ as being a conditional likelihood of the sample given the parameter realization π . Then the joint distribution

$$L(R|\pi)f(\pi|A)$$

combines the data and prior information, where $f(\pi|A)$ summarizes the prior information and depends on parameters A , specified a priori by the decision-maker. The posterior distribution of π , after observing the sample information R , is

$$\tilde{f}(\pi|R, A) = \frac{L(R|\pi)f(\pi|A)}{\int L(R|\pi)f(\pi|A)d\pi}. \quad (1)$$

Using the posterior density $\tilde{f}(\pi|R, A)$, agents want to determine the probability that the next observation is w_i , $i = 1, 2, 3$, which of course is a random variable. Rothschild defines the expected value of this probability, computed with respect to the posterior, by λ_i . Then to compute the marginal posterior density of π_1 , for example,

$$\tilde{f}_1(\pi_1|R, A) = \int \int \frac{L(R|\pi)f(\pi|A)}{\int L(R|\pi)f(\pi|A)d\pi} d\pi_2 d\pi_3.$$

Then we have

$$\begin{aligned}
\lambda_1(R, A) &= \int \pi_1 \tilde{f}_1(\pi_1|R, A) d\pi_1 \\
&= \int \pi_1 \left[\int \int \frac{L(R|\pi) f(\pi|A)}{\int L(R|\pi) f(\pi|A) d\pi} d\pi_2 d\pi_3 \right] d\pi_1 \\
&= \frac{\int \int \int \pi_1 L(R|\pi) f(\pi|A) d\pi_1 d\pi_2 d\pi_3}{\int L(R|\pi) f(\pi|A) d\pi}
\end{aligned}$$

after exchanging orders of integration. This is the expression given in that appears in equation (8) of Rothschild.

There is some discussion of the use of the Dirichlet distribution as the prior, F . The Dirichlet is a conjugate prior for the multinomial distribution, in that if f belongs to the Dirichlet family, then so does the posterior density \tilde{f} . For a multinomial with 3 parameters (for our example), the Dirichlet is characterized by 3 parameters as well. In terms of updating properties, the Dirichlet parameters $A = (a_1 \ a_2 \ a_3)$ can be considered as previous observations of the sampling process. That is, we can consider the effective number of observations of wage i as being equal to $a_i + N_i$. In terms of learning, this makes it clear that as $N \rightarrow \infty$ the data dominates the prior and estimates of the π realization converge to the maximum likelihood estimates (e.g., $\hat{\pi}_1 = N_1/N$).