1. Conditioning of random variables

Let $X_1, \dots, X_{k+\ell}$ be random variables on a common probability space. Let $f(t_1, \dots, t_{k+\ell})$ be the pmf of $(X_1, \dots, X_{k+\ell})$ and let $g(t_1, \dots, t_\ell)$ be the pmf of $(X_{k+1}, \dots, X_{k+\ell})$ (of course we can compute g from f by summing over the first k indices). Then, for any $s_1, \dots, s_\ell$ such that $\mathbf{P}\{X_{k+1} = s_1, \dots, X_{k+\ell} = s_\ell\} > 0$, we can define

(1)

$$h_{s_1,\dots,s_\ell}(t_1, \dots, t_k) = \mathbf{P}\{X_1 = t_1, \dots, X_k = t_k \;\Big|\; X_{k+1} = s_1, \dots, X_{k+\ell} = s_\ell\} = \frac{f(t_1, \dots, t_k, s_1, \dots, s_\ell)}{g(s_1, \dots, s_\ell)}.$$

It is easy to see that $h_{s_1,\dots,s_\ell}(\cdot)$ is a pmf on $\mathbb{R}^k$. It is called the conditional pmf of $(X_1, \dots, X_k)$ given that $X_{k+1} = s_1, \dots, X_{k+\ell} = s_\ell$.

Its interpretation is as follows. Originally, we had random observables $X_1, \dots, X_k$ which had a certain joint pmf. Then, we observe the values of the random variables $X_{k+1}, \dots, X_{k+\ell}$, say they turn out to be $s_1, \dots, s_\ell$, respectively. Then, we update the distribution (or pmf) of $X_1, \dots, X_k$ according to the above recipe. The conditional pmf is the new function $h_{s_1,\dots,s_\ell}$.

Exercise 1. Let $(X_1, \dots, X_{n-1})$ be a random vector with multinomial distribution with parameters $r, n, p_1, \dots, p_n$. Let $k < n-1$. Given that $X_{k+1} = s_1, \dots, X_{n-1} = s_{n-k+1}$, show that the conditional distribution of $(X_1, \dots, X_k)$ is multinomial with parameters $r', n', q_1, \dots, q_{k+1}$, where $r' = r - (s_1 + \dots + s_{n-k+1})$, $n' = k+1$, $q_j = p_j/(p_1 + \dots + p_k + p_n)$ for $j \le k$ and $q_{k+1} = p_n/(p_1 + \dots + p_k + p_n)$.

This looks complicated, but is utterly obvious if you think in terms of assigning r balls into n urns by putting each ball into the urns with probabilities $p_1, \dots, p_n$ and letting X_j denote the number of balls that end up in the j^{th} urn.

Conditional densities: Now, suppose that $X_1, \dots, X_{k+\ell}$ have joint density $f(t_1, \dots, t_{k+\ell})$ and let $g(s_1, \dots, s_\ell)$ be the density of $(X_{k+1}, \dots, X_{k+\ell})$. Then, we define the conditional density of $(X_1, \dots, X_k)$ given $X_{k+1} = s_1, \dots, X_{k+\ell} = s_\ell$ as

$$h_{s_1,\dots,s_\ell}(t_1, \dots, t_k) = \frac{f(t_1, \dots, t_k, s_1, \dots, s_\ell)}{g(s_1, \dots, s_\ell)}. \tag{2}$$

This is well-defined whenever $g(s_1, \dots, s_\ell) > 0$.

Remark 2. Note the difference between (1) and (2). In the latter, we have left out the middle term because $\mathbf{P}\{X_{k+1} = s_1, \dots, X_{k+\ell} = s_\ell\} = 0$. In (1), the definition of pmf comes from the definition of conditional probability of events, but in (2) this is not so. We simply define the conditional density by analogy with the case of conditional pmf. This is similar to the difference between interpretation of pmf ($f(t)$ is actually the probability of an event) and pdf ($f(t)$ is not the probability of an event, but the density of probability near t).

Example 3. Let (X, Y) have bivariate normal density $f(x, y) = \frac{\sqrt{ab-c^2}}{2\pi} e^{-\frac{1}{2}(ax^2+by^2+2cxy)}$ (so we assume $a > 0, b > 0, ab - c^2 > 0$). We can show that the marginal distribution of Y is $N(0, \frac{a}{ab-c^2})$, i.e., it has density $g(y) = \frac{\sqrt{ab-c^2}}{\sqrt{2\pi a}} e^{-\frac{ab-c^2}{2a}y^2}$. Hence, the conditional density of X given $Y = y$ is

$$h_y(x) = \frac{f(x,y)}{g(y)} = \frac{\sqrt{a}}{\sqrt{2\pi}} e^{-\frac{a}{2}(x+\frac{c}{a}y)^2}.$$

Thus, the conditional distribution of X given $Y = y$ is $N(-\frac{cy}{a}, \frac{1}{a})$. Compare this with marginal (unconditional) distribution of X which is $N(0, \frac{b}{ab-c^2})$.

In the special case when $c = 0$, we see that for any value of y, the conditional distribution of X given $Y = y$ is the same as the unconditional distribution of X. What does this mean? It is just another way of saying that X and Y are independent! Indeed, when $c = 0$, the joint density $f(x, y)$ splits into a product of two functions, one of x alone and one of y alone.

Exercise 4. Let (X, Y) have joint density $f(x, y)$. Let the marginal densities of X and Y be $g(x)$ and $h(y)$, respectively. Let $h_x(y)$ be the conditional density of Y given $X = x$.

(1) If X and Y are independent, show that for any x, we have $h_x(y) = h(y)$ for all y.

(2) If $h_x(y) = h(y)$ for all y and for all x, show that X and Y are independent.

Analogous statements hold for the case of pmf as well.