Review of Conditional Expectation and Conditional Variance

For simplicity, I will limit myself to probability experiments with a finite number of outcomes. For random variables that are continuous one needs more complicated measure theory for a rigorous treatment.

1 Probability Experiment

Denote the result of an experiment by one of the outcomes in **the sample space** $\Omega = \{\omega_1, \ldots, \omega_k\}$.

- For example, if the experiment is to choose one person at random from a population of size N with a particular disease, then the result of the experiment is $\Omega = \{A_1, \ldots, A_N\}$ where the different A's uniquely identify the individuals in the population.

- If the experiment is to sample n individuals from the population then the outcomes would be all possible n-tuple combinations of these N individuals; for example $\Omega = \{(A_{i_1}, \ldots, A_{i_n}),\ \text{for all } i_1, \ldots, i_n = 1, \ldots, N\}$. With replacement there are $k = N^n$th combinations; without replacement there are $k = N \times (N-1) \times \ldots \times (N-n+1)$ combinations of outcomes if order of subjects in the sample is important, and $k = \begin{pmatrix} N \\ n \end{pmatrix}$ combinations of outcomes if order is not important.

Denote by $p(\omega)$ the probability of outcome ω occurring, where $\sum_{\omega \in \Omega} p(\omega) = 1$.

2 Random variable

A random variable, usually denoted by a capital Roman letter such as $X, Y, \ldots$ is a function that assigns a number to each outcome in the sample space.

- For example, in the experiment where we sample one individual from the population

- $X(\omega)$= survival time for person ω
- $Y(\omega)$= blood pressure for person ω
- $Z(\omega)$= height of person ω

The **probability distribution** of a random variable X is just a list of all different possible values that X can take together with the corresponding probabilities; i.e. $\{(x, P(X = x)),\text{ for all possible } x\}$, where $P(X = x) = \sum_{\omega:X(\omega)=x} p(\omega)$.

The **mean** or **expectation** of X is

$$E(X) = \sum_{\omega\in\Omega} X(\omega)p(\omega) = \sum_x xP(X = x),$$

and the **variance** of X is

$$var(X) = \sum_{\omega\in\Omega}\{X(\omega) - E(X)\}^2 p(\omega) = \sum_x \{x - E(X)\}^2 P(X = x)$$

$$= E\{X - E(X)\}^2 = E(X^2) - \{E(X)\}^2.$$

3 Conditional Expectation

Suppose we have two random variables X and Y defined for the same probability experiment, then we denote the conditional expectation of X, conditional on knowing that $Y = y$, by $E(X|Y = y)$ and this is computed as

$$E(X|Y = y) = \sum_{\omega:Y(\omega)=y} X(\omega)\frac{p(\omega)}{P(Y = y)}.$$

The conditional expectation of X given Y, denoted by $E(X|Y)$ is itself a random variable which assigns the value $E(X|Y = y)$ to every outcome ω for which $Y(\omega) = y$. Specifically, we note that $E(X|Y)$ is a function of Y.

Since $E(X|Y)$ is itself a random variable, it also has an expectation given by $E\{E(X|Y)\}$. By the definition of expectation this equals

$$E\{E(X|Y)\} = \sum_{\omega\in\Omega} E(X|Y)(\omega)p(\omega).$$

By rearranging this sum, first within the partition $\{\omega : Y(\omega) = y\}$, and then across the partitions for different values of y, we get

$$E\{E(X|Y)\} = \sum_y \left\{ \frac{\sum_{\omega:Y(\omega)=y} X(\omega)p(\omega)}{P(Y=y)} \right\} P(Y=y)$$

$$= \sum_{\omega \in \Omega} X(\omega)p(\omega) = E(X).$$

Thus we have proved the very important result that

$$E\{E(X|Y)\} = E(X).$$

4 Conditional Variance

There is also a very important relationship involving conditional variance. Just like conditional expectation. the conditional variance of X given Y, denoted as $var(X|Y)$, is a random variable, which assigns the value $var(X|Y=y)$ to each outcome ω, where $Y(\omega) = y$, and

$$var(X|Y=y) = E[\{X - E(X|Y=y)\}^2|Y=y] = \sum_{\omega:Y(\omega)=y} \{X(\omega) - E(X|Y=y)\}^2 \frac{p(\omega)}{p(Y=y)}.$$

Equivalently,

$$var(X|Y=y) = E(X^2|Y=y) - \{E(X|Y=y)\}^2.$$

It turns out that the variance of a random variable X equals

$$var(X) = E\{var(X|Y)\} + var\{E(X|Y)\}.$$

This follows because

$$E\{var(X|Y)\} = E[E(X^2|Y) - \{E(X|Y)\}^2] = E(X^2) - E[\{E(X|Y)\}^2], \tag{1}$$

and

$$var\{E(X|Y)\} = E[\{E(X|Y)\}^2] - [E\{E(X|Y)\}]^2 = E[\{E(X|Y)\}^2] - \{E(X)\}^2. \tag{2}$$

Adding (1) and (2) together yields

$$E\{var(X|Y)\} + var\{E(X|Y)\} = E(X^2) - \{E(X)\}^2 = var(X),$$

as desired.

If we think of partitioning the sample space into regions $\{\omega : Y(\omega) = y\}$ for different values of y, then the formula above can be interpreted in words as

> "the variance of X is equal to the mean of the within partition variances of X plus the variance of the within partition means of X". This kind of partitioning of variances is often carried out in ANOVA models.