HAZARD, SURVIVAL AND DURATION

Consider a sequence of t independent Bernoulli trials, where the probability of the event is p and the probability of the non-event is $1-p$. If we consider the event to be the elimination of the player, then its absence over n trials can be described as their survival. The probability of this survival throughout t trials will be given by the binomial mass function $b(x;t,p)$ when $x=0$:

$$b(0,t,p)=(1-p)^t. \tag{1}$$

A similar result prevails in the case of the non-occurrence of such an eliminating event over a continuous finite period of time; and it is straightforward to convert the model of survival or elimination through Bernoulli trials into similar model in terms of events that are distributed randomly in time. This might be acheived by depicting a Poisson process in continuous time as a limiting case of a binomial process. However, matters are simplified by taking, as the point of departure, the special case of the binomial given under (1).

We may begin by considering that of each trial occupies a single unit of time. The probability of the occurrence of the eliminating event within a single period can be denoted by p, as before, and we may impose the assumption that the probability of the occurrence of two such events within the same interval is zero or, at least, that it is vanishingly small.

Such an assumption usually makes sense only if the time intervals are very short. Therefore, model can be improved by subdividing the intervals to give

$$\left\{\left(1-\frac{p}{n}\right)^n\right\}^t \simeq (1-p)^t, \tag{2}$$

where the approximation comes from taking the first two terms of the binomial expansion

$$\left(1-\frac{p}{n}\right)^n = 1-p+\frac{(n-1)}{2n}p^2-\frac{(n-1)(n-2)}{n^2 3!}p^3+\cdots. \tag{3}$$

In the limit, when the number of subdivisions increases indefinitely, there is

$$\lim(n\rightarrow\infty)\left(1-\frac{p}{n}\right)^n = e^{-p}. \tag{4}$$

Therefore, the probability of survival in the period $[0,t]$ is given by

$$S(t)=e^{-pt}, \tag{5}$$

which is the continuous-time analogue of equation (1).

The probability of being eliminated during the period $[0,t]$ is given by

$$1-S(t)=F(t)=\int_0^t f(s)ds = 1-e^{-pt}, \tag{6}$$

which is a cumulative distribution function. The corresponding density function, defined over the set of times at which elimination might occur, is just

$$f(t) = pe^{-pt}, \tag{7}$$

which is the derivative of the cumulative function $F(t)$ in respect of t. This is the so-called exponential waiting time function; and it is an instance of a duration distribution. A particular characteristic of this distribution is that the time already spent in waiting has no statistical effect upon future waiting times.

It is clear, from the foregoing derivation, that the value of p is an index of the hazard of being eliminated. The hazard of elimination in the period $[t, t + \Delta_t]$ is given by

$$H(t, t + \Delta_t) = P(t < \tau < t + \Delta_t | \tau > t) \simeq \frac{f(t)}{S(t)}\Delta_t = \frac{f(t)}{1 - F(t)}\Delta_t. \tag{8}$$

This is the probability of the occurrence of the eliminating event in period $[t, t + \Delta_t]$ given survival, i.e. the non-occurrence of the event, over the period $[0, t]$.

It is manifest that, under the present assumptions, $H(t, t + \Delta_t) = p\Delta_t$; which is to say that the hazard rate p is constant through time. However, one can envisage hazards that vary over time. For example, the hazard of human death is, typically, high at birth; and, thereafter, it declines before rising again in old age. It might also rise temporarily in youth before falling to its lowest level in early middle age.

To accommodate hazards that vary with time, we may define a hazard function by

$$p(t) = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt}\ln\{1 - F(t)\}. \tag{9}$$

The final equality is obtained via the function-of-a-function rule, using the fact that $d\ln(x)/dx = 1/x$.

There is clearly a one-to one correspondence between the hazard function and the cumulative function $F(t)$, which gives the probability of elimination within the period $[0, t]$; and it follows quite readily that the inverse mapping from the hazard function to the cumulative distribution is given by

$$F(t) = 1 - \exp\left\{ -\int_0^t p(s)ds \right\}. \tag{10}$$

In many applied contexts, when the exponential waiting time model is inappropriate, it is easier to specify the form of the hazard function than to specify directly the survival function $S(t)$ or the duration distribution $f(t)$.

In econometrics, hazard functions have been used in explaining such matters as the duration of strikes and the length of time spend by individuals in unemployment. A considerably variety of hazard functions have been proposed. Often such functions incorporate explanatory variables which are deemed to affect the hazard.