Duration of a game: using generating functions

Suppose we wish to calculate the mean (average value) of the total number of points played in a game. Using a standard formula for calculating the mean value of a discrete distribution, this c n be calculated by $\mu = E(X) = \sum_x xf(x)$. Similarly the variance (standard deviation squared) of the total number of points played in a game can be calculated by $\sigma^2 = E(X^2) - E(X)^2 = \sum_x (x)^2 f(x) - (\sum_x xf(x))^2$; which is recognized as a measure of the dispersion of a set of data from its mean. Both the mean and the variance contain important information to describe the shape of the distribution and these characteristics could be used to compare one distribution to another. For example, comparing the mean and variance of a tiebreak set to an advantage set to identify why `long' matches can occur. However, if a distribution is not symmetric (as typically occurs in a game and an advantage set) the mean and variance do not `adequately' describe the shape of the distribution. Two other characteristics that are used to describe the distribution and measure risk are skewness and kurtosis. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the centre point. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. In order to calculate the skewness and kurtosis, it becomes convenient to work with generating functions, as the four characteristics (mean, variance, skewness and kurtosis) can readily be obtained.

The expectation of a random variable X is calculated by $E(X) = \sum_x xf(x)$.

The expectation of the n^{th} power of a random variable X is calculated by $E(X^n) = \sum_x x^n f(x)$. This is also known as the n^{th} moment of the random variable X and represented by m_{nx}, such that: $m_{nx} = E(X^n)$ for n=1,2,3,4,…

The moment generating function for a random variable X is defined by $M_X(t) = E(e^{tX})$
$= \sum_x e^{tx} f(x)$.

The n^{th} moment for a random variable X can be recovered from the moment generating function by differentiating n times and setting t=0. Thus $E(X^n) = M_X^{(n)}(0) = d^n/dt^n M_X(0)$

The moment generating function for the total number of points played in a game from the outset, $M_X(t)$, becomes:

$\sum_x e^{tx} f(x)$
$= e^{4t}(p^4+q^4) + e^{5t}(4p^4q + 4q^4p) + e^{6t}(10p^4q^2+10q^4p^2)+(N(3,3)(1-N(1,1))e^{8t})/(1-N(1,1)e^{2t})$

where p represents the probability of the server winning a point, q=1-p and $N(g,h)$ be the probability of reaching a point score (g,h) in a game from the outset where g and h represent the projected point score for the server and receiver respectively.

The cumulant generating function for the total number of points played in a game from the outset, $K_X(t)$ becomes $log_e(M_X(t))$.

The first derivative of the cumulant generating function evaluated at t = 0, $K^{(1)}{}_X(0)$, is equivalent to the mean of the total number of points played in a game, $\mu(X)$. It follows that:

$\mu(X)= 4(pq(6p^2q^2-1)-1)/(1-2pq)$

The second derivative of the cumulant generating function evaluated at t = 0, $K^{(2)}{}_X(0)$, is equivalent to the variance of the total number of points played in a game, $\sigma^2(X)$. It follows that:

$\sigma^2(X)=4pq(1-pq(1-12pq(3-pq(5+12p^2q^2))))/(1-2pq)^2$

Let $\gamma_1(X)$ represent the coefficient of skewness of the total number of points played in a game.

The third derivative of the cumulant generating function evaluated at t = 0, becomes $K^{(3)}{}_X(0)$. The coefficient of skewness of the total number of points played in a game can be calculated by:

$\gamma_1(X)= K^{(3)}{}_X(0)/ K^{(2)}{}_X(0)^{3/2}$

Let $\gamma_2(X)$ represent the coefficient of excess kurtosis of the total number of points played in a game. In general, excess kurtosis = kurtosis - 3, so the normal distribution has an excess kurtosis of 0, and therefore a kurtosis of 3.

The fourth derivative of the cumulant generating function evaluated at t = 0, becomes $K^{(4)}{}_X(0)$. The coefficient of excess kurtosis of the total number of points played in a game can be calculated by:

$\gamma_2(X)= K^{(4)}{}_X(0)/ K^{(2)}{}_X(0)^2$

Let $\sigma(X)$ represent the standard deviation of the total number of points played in a game. Let $c_v(X)$ represent the coefficient of variation of the total number of points played in a game. It follows that $\sigma(X) =\sqrt{\sigma^2(X)}$ and $c_v(X)= \sigma(X)/\mu(X)$

Table 1 represents $\mu(X)$, $\sigma(X)$, $c_v(X)$, $\gamma_1(X)$ and $\gamma_2(X)$ for different values of p. The calculations were performed using *Mathematica*. For example when p=0.60, $\mu(X)$=6.48 and $\sigma(X)$=2.59. Notice how $\gamma_1(X)$ and $\gamma_2(X)$ increase for increasing values of p, whereas $\sigma(X)$ decreases for increasing values of p. This is evidence to show that the standard deviation is insufficient information for measuring risk.

p	$\mu(X)$	$\sigma(X)$	$c_v(X)$	$\gamma_1(X)$	$\gamma_2(X)$
0.50	6.75	2.77	0.41	2.16	6.95
0.55	6.68	2.73	0.41	2.17	7.01
0.60	6.48	2.59	0.40	2.20	7.21
0.65	6.19	2.37	0.38	2.25	7.59
0.70	5.83	2.10	0.36	2.34	8.25
0.75	5.45	1.78	0.33	2.46	9.27
0.80	5.09	1.44	0.28	2.61	10.71

Table 1: The parameters of the distributions of the number of points played in a game for different values of p