NOTES ON THE BINOMIAL AND POISSON DISTRIBUTIONS

The binomial distribution deals with events where each trial yields only one of two possible outcomes, such as "heads" and "tails". If

p=probability of a "success" on each trial, and

q=probability of a "failure" on each trial (p+q=1), then

the probability of getting k successes in n trials is given by

$$P(k;n,p) = \binom{n}{k} p^k \, q^{(n-k)}, \tag{1}$$

where $$\binom{n}{k} = \left(\frac{n!}{k!\,(n-k)!}\right).$$

(Note that there are both capital P's and lower case p's. You will have to keep them straight!)

The average number of successes in a large number of trials is "m", where

$$m = np. \tag{2}$$

Note that in this simple case (the simple binomial), we assume that p is constant for each trial. It is assumed, furthermore, that the outcome of any one trial is independent of the outcome of any other.

It is useful to be able to predict P(k;n,p) in order to compare the expected outcome of an experiment with the actual outcome. To determine P, it is necessary to know n and p (see equation 1). But n and p cannot be known without knowing the extent to which the measured number of successes varies over a number of *sets* of trials. A single set of trials can be run by having a number of coins flipped at once. Let's say that 15 students each flipped a coin once, and 9 of them got heads. That represents one set of trials, with an n of 15. By repeating this process N times, we have created N sets of n trials. For the binomial distribution, the variance (σ^2) about the mean for N sets of n trials, is given by

$$\sigma^2 = npq = np\,(1-p) = m\,(1-p). \tag{3}$$

Solving for p yields $$p = \left(1 - \frac{\sigma^2}{m}\right). \tag{4}$$

We now have p in terms of two measurable variables, m and σ^2 . Remember that m is simply the mean number of "successes" per set of n trials and σ^2 is the variance about this mean for N sets of trials. Once p is known, n can be determined (from equation 2).

A similar approach utilizes the coefficient of variation ($CV=\frac{\sigma}{m}$), which is given as

$$CV = \sqrt{\frac{1}{m} - \frac{1}{n}}. \tag{5}$$

Solving for n yields

$$n = \frac{m}{1 - \frac{\sigma^2}{m}}, \tag{6}$$

which is simply a rearrangement of equations (5) and (3). Once n is known, p is calculated. The use of equations (4) and (6) represent two different but non-independent means of measuring p and n.

The process of transmitter release may be analogous to that of flipping coins. Each time an action potential invades a nerve terminal, an integral number of transmitter quanta are released. The actual number

released fluctuates from trial to trial in a stochastic manner. The variable "n" represents the "number of quanta available for release" or the number of "release sites", while "p" represents the probability that any one of those quanta will be released (or the probability that release will occur from one of those sites). The variable "m" represents the average quantal content, the number of quanta released per stimulus, as determined over a large number of trials (stimuli). The variable σ^2 represents the variance about the mean, m, in the sampled population of quanta released over the N trials. Note now that the number of sets of trials, N, is equated with the number of stimuli. Each trial represents the sampling of a number of "quanta available for release". The total number of sampled events in the experiment is actually nN, but we have divided them into N sets, determining an outcome for each, in order to measure σ^2.

Only in rare instances is it possible to determine m and σ^2 directly when studying synaptic transmission. It simply isn't possible to count quanta under ordinary conditions. Instead, one measures a postsynaptic response, which is more-or-less continually graded. There is a finite amount of noise in the system, and the unit potentials, which sum to produce the overall response, have a variance of their own. Thus, if the unit potential has a mean size of 0.5 mV (if we are measureing a potential change) and a coefficient of variation (standard deviation divided by mean) of 0.25, then one cannot know whether a 2.5 mV response is an average "5", a large "4" or a small "6". It is still possible to estimate m from a number N of trials by dividing the mean response of the epsp (or epsc) by the mean unit response (usually equated with the spontaneously occurring signals). If the responses are measured as voltage deflections, then one may have to correct for non-linear summation before analyzing the data further (what is “non-linear summation?”). p (lower case!) can then be calculated using an expression similar to equation (5), but having terms for the variance of the evoked response as well for the spontaneous response. Thus

$$p = 1 - \frac{S^2}{m\gamma^2} + \frac{s^2}{\gamma^2}, \tag{7}$$

where S^2 = variance of the evoked response, γ = mean size of the spontaneous response, s^2 = variance of the spontaneous responses, and m = average quantal content (the mean size of the evoked response divided by the mean size of the spontaneous response).

In order to determine whether release obeys binomial statistics, one must calculate both p and n. One can then generate an expected distribution of responses for N trials (from equation 1) and can determine whether the actual distribution of responses is adequately fit by the expected distribution (using the χ^2 test, for example). This can be done directly if individual quanta can be counted or indirectly if only the size of response is measured. If individual quanta cannot be counted, then an expected distribution of responses must be generated (see Boyd and Martin, 1956). If the fit is good, then we can say that release obeys binomial statistics. Unfortunately, this does not mean that the assumptions made in setting up equation (1) are correct. It turns out from Monte Carlo simulations that a good fit is often obtained even when p is not the same for all "available quanta", a condition known as **non-uniformity** or when the values of "p" change over time ("**non-stationarity**"). If release is non-uniform, then usually p will be overestimated and n will be underestimated. This is because estimates will preferentially reflect the more "active " sites and will raise the estimate for p (and lower it for n). If release is non-stationary, then usually p will be underestimated and n will be overestimated. This is because a temporal variation results in an increase in the overall variance, which will increase estimates of n (see equation 6).

One of the difficulties of applying the binomial distribution is that both n and p must be determined before an expected distribution of events can be generated. As we have seen, it isn't always easy to do this and know that the estimates are accurate. However, the equation defining the binomial distribution (1) is greatly simplified if n >> k (the number of successes), which will be true if p << 1. Equation 1, in expanded form, is

$$p(k;n,p) = \left(\frac{n(n-1)(n-2)\cdots(n-k+1)(n-k)(n-k-1)\cdots(1)}{k!\,(n-k)!}\right) p^k\, q^{(n-k)} \tag{8}$$

The terms in the numerator, up to (n-k+1), are $\cong n^k$ (since n>>k), and the terms in the remainder of the numerator are = (n-k)! Therefore, if p « 1,

$$P(k;n,p) = \frac{n^k}{k!} p^k q^{n-k} \quad (9)$$

Rearranging and substituting yields

$$P(k;n,p) = \frac{m^k}{k!}\left(1 - \frac{m}{n}\right)^n. \quad (10)$$

Furthermore, $\lim_{n\to\infty}\left(1 - \frac{m}{n}\right)^n = e^{-m}$ (a Taylor series). (11)

Therefore, $p(k;n,p) = \frac{m^k}{k!} e^{-m}$. (12)

Equation 12 is the Poisson equation. In order to determine whether a distribution of responses is well fitted by the Poisson law, one need only determine m, the mean quantal content, and use m to determine P(0), P(1), etc. For example, the probability that a stimulus will lead to zero successes (k=0) is

$$P(0) = \frac{m^0}{0!} e^{-m} = e^{-m} \quad (13)$$

The best method of measuring m is to count the quanta. If this is not possible, then one must divide the mean response by the unit response. This is the so-called "direct" method of determing m, although you can appreciate that it is far from direct. This single, experimentally determined number can then be used to generate the entire distribution of expected responses. If the Poisson distribution holds in a particular experimental situation (i.e., if p <<1), then one can calculate the mean in one of two ways, the "failures method" and the "coefficient of variation" method. The failures method uses equation 13, which, yields

$$-m = \ln P(0).$$

But since $P(0) = \frac{N_o}{N}$, by rearrangement

$$m = \ln \frac{N}{No} \quad (14)$$

where N is the number of trials (remember, this is the number of stimuli and should not be confused with n, the number of "available quanta" or "sites"), and N_o the number of trials resulting in no released quanta (i.e., the number of "failures").

The second method of estimating m is the coefficient of variation method. In the binomial case

$\sigma^2 = m(1-p)$, from equation (4). But since, in the Poisson, $p \ll 1$

$$m = \sigma^2 \quad (15a)$$

This is often expressed as $$m = \frac{1}{(CV)^2} \quad (15b)$$

where $CV = \sigma/m$. Once m is estimated either by the method of failures or the coefficient of variation method, it can then be compared to the "directly" estimated value of m. A more complete analysis would be to use the "direct" estimate of m to calculate the entire distribution of expected responses, including the number of failures., using equation (12), and then to compare that distribution with the actual one.