Probability and Statistics (continued)

1 Example: Two People with the same Birthday

Suppose there are $N$ people in a room. What is the probability that 2 or more have a birthday on the same day? Assume that the probability, $P_0$, that a person has a birthday on a particular day is a constant and equal to $1/365$. First, let us find the probability that 2 or more people in the room have a birthday on a particular day, say today. So a success is defined as having a birthday today. Then its probability distribution is

$$
\phi(j, N, P_0) = \frac{P_0^j(1 - P_0)^{N-j}N!}{j!(N-j)!}
$$

And the probability of 2 or more such successes is

$$
P_{2+} = 1 - \phi(0) - \phi(1)
$$

or

$$
P_{2+} = 1 - (1 - P_0)^N - P_0(1 - P_0)^{N-1}N
$$

In the above equation we used the results that

$$
\phi(0) = (1 - P_0)^N \quad (1)
$$

and

$$
\phi(1) = P_0(1 - P_0)^{N-1}N/(N-1)! = P_0(1 - P_0)^{N-1}N
$$

Now we step through all 365 days of the year, and determine, for each day, if there are two people with a birthday on that day. Now set success’ be having 2 or more birthdays on the same day. The probability of success' in 1 trial is $P_{2+}$. We want $P'$, the probability of one or more success’ in 365 trials. So

$$
P' = 1 - \phi'(0)
$$

and using equation 1,

$$
P' = 1 - (1 - P_{2+})^{365}
$$

For what $N$ does this probability exceed 0.5?

Properties of Probability Distributions
2 Normalization

Let the probability of any result be 1. Then

$$\sum_{j=0}^{\infty} \phi(j) = 1$$

If you have an un-normalized distribution, say $h(j)$, which gives the number of occurrences of $j$ in particular set of numbers, it can be made into a normalized distribution, or normalized by finding the sum

$$N = \sum_{j=0}^{\infty} h(j)$$

and then $\phi(j) = h(j)/N$ will be normalized.

3 Expectation Value

The expectation value of some quantity $f$ which depends on $j$ is defined as

$$< f >= \sum_{j=0}^{\infty} f(j) \phi(j)$$

4 Moments

The $n$th moment of a distribution is the expectation value of of $j^n$, so

$$< j^n >= \sum_{j=0}^{\infty} j^n \phi(j)$$

A distribution is characterized by its moments. In general, different distributions have different moments. A formula for a particular moment of a particular distribution may not be correct for some other distribution.

4.1 1st Moment

If $n = 1$, the 1st moment is the mean or average of $j$. That is,

$$\bar{j} = < j >= \sum_{j=0}^{\infty} j \phi(j)$$  \hspace{1cm} (2)

For example, consider a set of $N$ test grades between 0 and 100. If $n(j)$ is the number of students who got grade $j$, then

$$\phi(j) = n(j)/N$$
So Equation 2 becomes
\[ \bar{j} = \frac{1}{N} \sum_{j=0}^{100} j \ n(j) \]

or
\[ \bar{j} = \frac{1}{N} \sum \text{(all grades)} \]

which is the usual definition of an average. It can be shown that for the binomial distribution,
\[ < j > = Np \]

4.2 Moments about the Mean

A moment about the mean is defined as
\[ < (j - < j >)^n > = \sum_{j=0}^{\infty} (j - < j >)^n \phi(j) \]

For \( n = 2 \), the conventional notation is
\[ < (j - < j >)^2 > = \sigma^2 \]

The quantity \( \sigma^2 \) is called the variance, and \( \sigma \) is the standard deviation. So
\[ \sigma^2 = \sum_{j=0}^{\infty} (j - < j >)^2 \phi(j) \]

It can be shown that, for the binomial distribution,
\[ \sigma = \sqrt{Np(1 - p)} \]

For the two binomial distributions in the figure, the \( \sigma \)'s are 2.24 and 2.95. This illustrates how the standard deviation is a measure of the “width” of the distribution.

5 Progeny of the Binomial Distribution

Given the binomial distribution
\[ \phi(j, N, P_0) = \frac{P_0^j (1 - P_0)^{N-j} N!}{j! (N-j)!} \]

let \( N \to \infty \), and \( P \to 0 \) in such a way that the average (the first moment, \( NP \)) stays finite. Let the average now be denoted by \( a \). Then it can be shown that the resulting distribution, which is the probability of \( j \) when the average is \( a \), is
\[ \phi(j, a) = \frac{a^j e^{-a}}{j!} \]

This is called the Poisson distribution. Note that the allowed range of \( j \) is 0 to \( \infty \).
Figure 1: Binomial distributions $P(j,20,5)$, and $P(j,80,125)$ have standard deviations 2.24 and 2.95 respectively.

5.1 Example

Cosmic ray muons randomly strike a detector and make counts in a counter at a rate of 2.34 per second. In 2.00 seconds, what is the probability of $j$ counts?

The average number in 2 seconds, is $a = 4.68$ using this in Equation 3 yields the probability distribution shown in this graph:

Note that the probability of 0 counts is not 0. It is $\phi(0) = e^{-a} = .00928$.

5.2 Moments of the Poisson Distribution

It can be shown that

1. 

$$\bar{j} = \sum_{j=0}^{\infty} j \frac{a^j e^{-a}}{j!} = a$$

as expected.

2. 

$$\sigma^2 = \sum_{j=0}^{\infty} (j - a)^2 \frac{a^j e^{-a}}{j!} = a$$
So that the standard deviation of the Poisson distribution is

\[ \sigma = \sqrt{\lambda} \]

This is a very useful result. If a particular quantity has a Poisson distribution, for example counts, \( n_k \) in the \( k \)th channel of a multi-channel analyzer, the average may be estimated as the number of counts, and so the standard deviation is

\[ \sigma = \sqrt{n_k} \]

This is a good estimate of the uncertainty in the number \( n_k \).

### 5.3 Example: Do two runs agree?

In Run 1 we get 10 events; in Run 2 we get 12 events. Has the rate changed? In other words, are the results of the two runs statistically consistent?

Since \( \sigma_1 = \sqrt{10} = 3.16 \), 12 is within 1 \( \sigma \) of this. So there is no disagreement. We are now getting into new subject; that of agreement or disagreement, and confidence limits.

### 5.4 Example: You get nothing.

In a search experiment, we find 0 events. What limit can be put on \( \lambda \), the true average number of events? We define the 90% confidence limit such that the probability of getting
your value, or less, is 10%. Here the value is 0. So with a Poisson probability distribution,

\[ \phi(0) = e^{-a} = 0.1 \]

This means that \( a < 2.3 \). That is, with \( a < 2.3 \) the probability of your result is < 10%.

### 5.5 Example: Accidental Coincidences

Random events have an average rate \( R \). Given a time interval \( \Delta t \), what is the probability of getting 2 or more events in this interval? What is the rate of such coincidences?

The probability of 2 or more in \( \Delta t \) is

\[ \phi(\geq 2) = 1 - \phi(0) - \phi(1) \]

Using the Poisson distribution with the average \( a = R \Delta t \)

\[ \phi(\geq 2) = 1 - e^{-R \Delta t} - R \Delta t e^{-R \Delta t} \]

Making a Taylor expansion of the exponential (assuming \( a \) is small)

\[ \phi(\geq 2) = 1 - (1 - R \Delta t) - R \Delta t (1 - R \Delta t) \]

\[ \phi(\geq 2) = R^2 (\Delta t)^2 \]

The rate of accidental coincidences is

\[ R_{accid} = (\# \text{ Delta } t \text{ intervals/unit time})(\text{Prob of } 2 \text{ in } \Delta t) \]

\[ = (1/\Delta t)R^2 (\Delta t)^2 = R^2 \Delta t \]

Note that many electronic coincidence circuits which have pulse inputs report a coincidence whenever two or more of the input pulses have any time overlap. So in this case the time interval to use in the above equation is twice the pulse width.