Counting statistics of random events: A tutorial

This tutorial presents derivations of some results from the theory of the statistics of random events that are studied in counting experiments.

The statistics of counting random events falls under the rubric of what statisticians call "point processes". A point process in time is one where a particular type of event is classified only according to a single point on a time line. Some examples of phenomena that could generate time points are the births of babies in a particular hospital, the passage of cars past a counting gate on a toll bridge, or the arrival of particles from outer space at a detector. Such point processes may be further categorized as "Poisson processes" if the occurrence of any particular point is completely independent of the occurrence of any other point, as is the case in our cosmic ray experiment. If the criterion of independence is not satisfied, the point process may be more generally known as a "renewal process", a name which is derived from the study of part failures and replacements in industrial contexts (e.g., a new light bulb is not installed until the old one burns out) [4].

This section presents in a tutorial fashion the derivation of some important formulas in the theory of Poisson processes. These results are neither new nor little known, especially among statisticians. The goal is to show derivations that a student with the mathematical background of first-year calculus can follow and believe. These are "physicist's proofs" where a plausible argument is chosen over rigorous demonstration. Suggested exercises are given to help the student cement his or her understanding.

The raw data for our statistical investigations comes in the form of a random pulse train, as illustrated in Fig. 1. The pulse width T_w is important only insofar as it determines the maximum rate of pulses that may be represented by the pulse train, since pulses which occur more frequently than $1/T_w$ cannot be resolved. The occurrence time of each pulse, denoted as t_1, t_2, etc., as measured from some (arbitrary) starting time, is marked at the leading edge of each pulse.

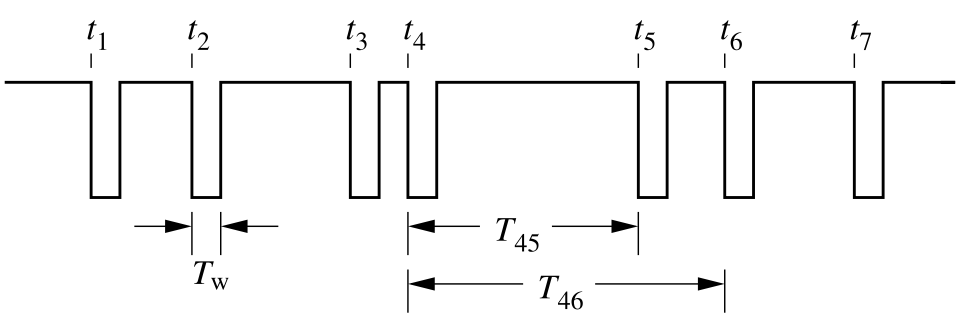

Figure 1: A random pulse train, as might be seen on an oscilloscope at a particular instant. This example shows negative-going ("fast NIM") pulses typical of those used in nuclear counting experiments.

If the pulse train is being created by the detection of decays from a long-lived radioactive nuclide or by cosmic rays, experiments show that although the time between specific pulses T_{ij} is completely unpredictable, the time intervals do converge to a well-defined average τ in the following sense: If one counts very many pulses (i.e., thousands) N over a long time interval T, then $\tau \approx T/N$. More specifically, if one estimates τ by counting N pulses over number of trials, the variation of the calculated τ's from these trials will cluster about each other with a fractional standard deviation of $1/\sqrt{N}$. (Later we will strengthen this assertion.) We can define an average *rate* of pulses r by the relation $r = 1/\tau$.

We are interested in two problems:

1. What is the *distribution of intervals* between successive pulses in a given pulse train? More generally, what is the distribution of *scaled intervals*: intervals between two pulses separated by an arbitrary number of pulses in between?
2. What is the *distribution of counts* within a succession of fixed-length periods?

To make these problems clear, we remind the reader of a few basic definitions. A *distribution* is the collected result of many *trials* of a particular experiment. Each trial produces a value of a *random variable*. For example, in the distribution of counts, one trial consists of adding up the number of counts that happen in a particular time period. This number is the value of the random variable $N(t)$, where t denotes the time length of the period. The collection of numbers $N(t_i)$, where i is an index denoting each trial period, is the distribution. Since $N(t)$ can only take on non-negative integer values, the count distribution is *discrete*. The random variable in the interval distribution is the time between pulses T_{ij}. Since it can take on any nonzero value, the interval distribution is *continuous*. Distributions are often presented as *histograms* (vertical bar graphs), in which the horizontal axis represents the value (or range of values, in the case of a continuous distribution) of the random variable and the height of the bars represents how often that value occurs.

The theoretical problem is the derivation of probability distribution functions: functions that give the likelihood of finding a particular value or range of values of the random variable. We will address the theoretical problem by making a basic assumption about our pulse train and invoking two rules. The rules are from the theory of probabilities [1]. Assume that two events of a similar type A and B follow from the same proximate cause. The probabilities of these events $P(A)$ and $P(B)$ are subject to these two rules:

> **I.** *The rule of compound probabilities.* If the two events A and B are mutually exclusive, then the probability of having either event $P(A \,\text{or}\, B)$ is given by the sum $P(A) + P(B)$.
>
> **II.** *The rule of independent probabilities.* If the two events A and B are independent of each other, then the probability of having both events occur $P(A \,\text{and}\, B)$ is given by the product $P(A) \times P(B)$.

Simple examples of these two rules are illustrated by considering rolls of dice. A die has six numbered sides, and when it is thrown any one number will be equally likely to land face up: the probability of any number is 1/6. If we ask, "What is the probability to roll an even number?", we are invoking rule **I**, since the roll of any number excludes the roll of any other number. Thus,

$$P(\text{even}) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}\,.$$

If we ask, "What is the probability to roll two 1s in a row?", we are invoking rule **II**, since each successive roll of the die is independent of any previous roll. Thus

$$P(1\,\&\,1) = P(1) \times P(1) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}\,.$$

If we ask, "What is the probability of rolling only one 1 in two rolls of the die?", we want to know the probability of a 1 on the first roll *and* not 1 on the second *or* a 1 on the second roll *and* not on the first. Thus, since successive rolls of the dice are independent events and the two possible desired outcomes are mutually exclusive,

$$\begin{aligned} P(\text{only one 1 in 2 rolls}) &= P(1;\text{1st}) \times P^*(1;\text{2nd}) \;+\; P^*(1;\text{1st}) \times P(1;\text{2nd}) \\ &= \frac{1}{6}\times\frac{5}{6} \;+\; \frac{5}{6}\times\frac{1}{6} \\ &= \frac{10}{36}\,, \end{aligned}$$

where P^* is the compliment of P, that is, it is the probability of *not* obtaining the specified outcome. By convention, probability distributions are normalized: the random variable *must* take on one of its possible values, and since any value X either does or does not occur, $P(X) + P^*(X) = 1$, by rule **I**.

The basic assumption we make about our data is

> Any infinitesimally small interval of time dt is equally likely to contain a pulse. The probability of finding a pulse in dt is given by $r\,dt$.

For this assumption to make sense, $r\,dt$ must be much smaller than one, and the probability of finding two pulses in dt must be negligible. As long as r is a constant, these restrictions can be assured by making dt small enough.

We will first apply the rules we have stated to a real problem in particle physics: random coincidences.

Rate of random coincidences

The use of coincidence counting is pervasive in elementary particle experiments for one big reason: noise. For example, if you set up a scintillator paddle to count the passage of cosmic-ray muons, *most* of the counts you record will be due to non-muon events such as electrical noise, low-level background radiation, etc. Fortunately, muons have fairly high energy, so that if they pass through a thin scintillator paddle they will continue out the other side and can be detected by a second paddle nearby. Such an event produces two pulses, one in each detector, at (nearly) the same time. By using a logic gate, one can force the counting electronics to record only those pulses which occur in coincidence. A schematic of this setup is shown in Fig. 2. In order for the coincidence technique to work profitably, the rate of random coincidences must be appreciably lower than the rate of "true" coincidences. We can use our results to predict this rate.

Let the rate of random pulses recorded by detector A be be r_A and the rate of random pulses recorded by detector B be r_B. Each time a pulse from either A or B enters the gate, the gate is triggered: it "opens" for a short time interval T called the *resolving time*, meaning that if a pulse from the other detector occurs during T, then a coincidence pulse is produced at the output. Since r_A and r_B are the average rates of random pulses, the rate of coincidences will also be random (in the absence of any true coincidence events). Let R_2 be this rate; more specifically, we let $R_2\,dt$ be the probability that a random coincidence event between 2 detectors will occur within the infinitesimal time interval dt.

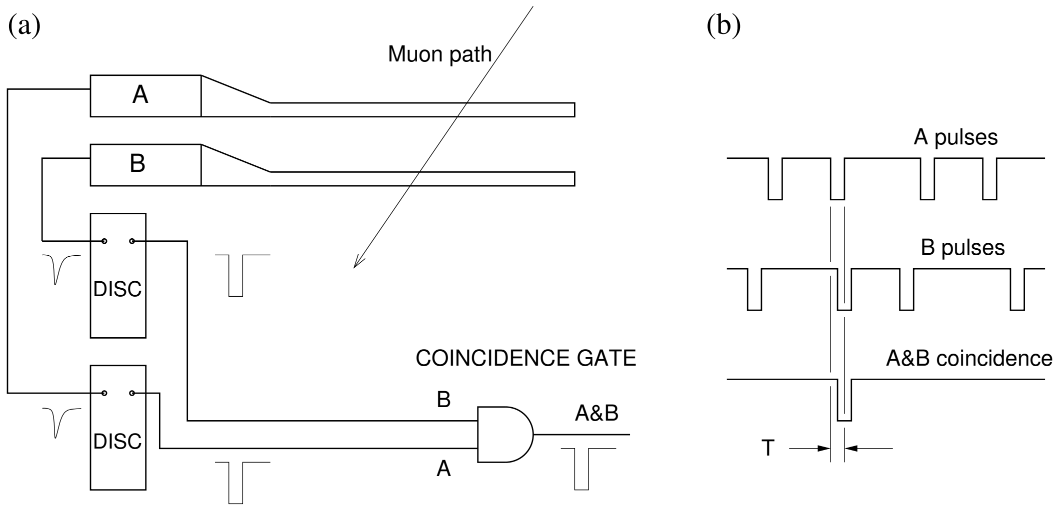

Figure 2: (a) A two-paddle coincidence setup. Variable height detector pulses are discriminated and turned into digital pulses. The coincidence gate produces a pulse when two input pulses overlap within a gate time T. (b) Example of two pulse trains showing one coincidence.

By rules **I** and **II** $R_2\,dt$ is given by the following construction:

$$R_2\,dt = \begin{pmatrix}\text{Probability of the gate}\\ \text{being triggered by}\\ \text{detector } A \text{ in } dt.\end{pmatrix} \times \begin{pmatrix}\text{Probability of detector}\\ B \text{ delivering a pulse}\\ \text{within resolving time}\\ T \text{ of the trigger time.}\end{pmatrix}$$

$$+ \begin{pmatrix}\text{Probability of the gate}\\ \text{being triggered by}\\ \text{detector } B \text{ in } dt.\end{pmatrix} \times \begin{pmatrix}\text{Probability of detector}\\ A \text{ delivering a pulse}\\ \text{within resolving time}\\ T \text{ of the trigger time.}\end{pmatrix} .$$

In symbolic form, this equation reads

$$R_2\,dt = r_A\,dt \times r_B T + r_B\,dt \times r_A T \;.$$

Typically the resolving time T is much shorter than the average rates r_A and r_B so that $rT \ll 1$. In this case, we may approximate the probability of finding a pulse in the finite time T by using the infinitesimal probability: $r\,dt \approx rT$. So the rate of 2-fold random coincidences R_2 is given by

$$R_2 = 2 r_A r_B T \;. \tag{1}$$

Exercise 1 *Show that if you added a third paddle C to the setup, the rate of random 3-fold coincidences R_3 would be equal to $3r_A r_B r_C T^2$.*

Distribution of intervals

Now, let's return to our main questions. We first calculate the probability that there will be *no* pulse in a finite time interval T. Since we assume that a pulse may occur at any instant with equal

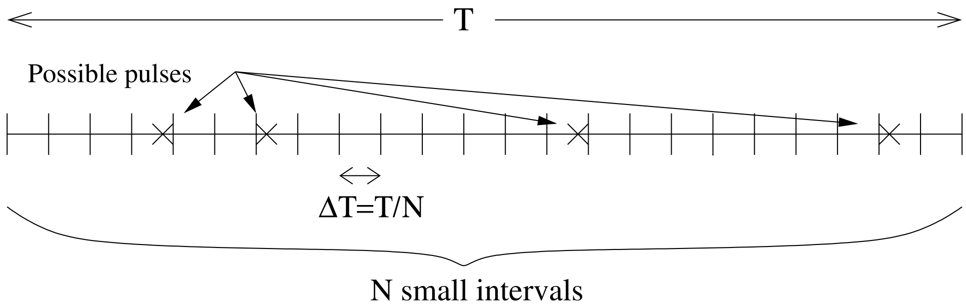

Figure 3: Construction used in calculating the probability of zero pulses in a finite time T.

probability, it does not matter where along the time line we start our clock. Since our assumption involves the infinitesimal interval dt, we divide T into small N small intervals ΔT, with an eye toward taking the limit of small ΔT. This construction is shown in Fig. 3. In any interval of length ΔT, when this interval becomes small, the probability of finding a pulse is $r\Delta T$, so by Rule **1**, the probability of not finding a pulse is $1 - r\Delta T$. Since the probability of finding a pulse in any small interval is independent of the probability of finding a pulse in any other interval, the probability of not finding a pulse in T, $P_0(T)$ is, by rule **II**, the product of the probabilities for all of the N small intervals:

$$P_0(T) \approx (1 - r\Delta T)^N \ . \tag{2}$$

If we write $\Delta T = T/N$, the limit of infinitesimal ΔT is found by letting N get large, and in this limit, the approximation tends to equality. Thus,

$$P_0(T) = \lim_{N\to\infty} \left(1 - \frac{rT}{N}\right)^N = e^{-rT} \ . \tag{3}$$

The final step in this unsurprising result can be proved by taking the logarithm of both sides of the equation, since $\ln(1+x) \approx x$ for small x.

We now use the result of Eq. (3) to derive the probability density function of 1-pulse intervals, $I_1(t)$. We define this continuous probability density function as follows: $I_1(t)\,dt$ is the probability of finding a 1-pulse interval of length between t and $t + dt$. This means the probability of finding a pulse during the infinitesimal time interval dt after a time interval of length t following a previous pulse during which there have been no pulses. This situation is illustrated by Fig. 4.

If we pick a pulse in the train, and assign its time as $t = 0$, then we obtain $I_1(t)$ by the following construction from rule **II**:

$$\begin{aligned} I_1(t)\,dt &= \left(\begin{array}{l}\text{Probability of finding}\\ \text{zero pulses between 0}\\ \text{and } t\end{array}\right) \times \left(\begin{array}{l}\text{Probability of finding}\\ \text{a pulse between } t \text{ and}\\ t + dt\end{array}\right) \\ &= P_0(t) \times r\,dt \\ &= re^{-rt}\,dt \ . \end{aligned}$$

Thus

$$I_1(t) = re^{-rt} \ . \tag{4}$$

This result is known as the exponential distribution and has a couple of interesting features. First, the condition that each infinitesimal interval dt is equally likely to contain a pulse leads to the

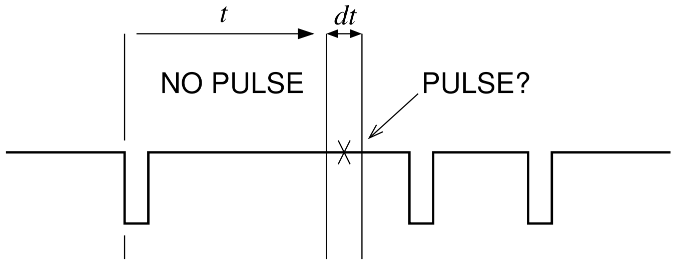

Figure 4: The construction leading to $I_1(t)\,dt$ asks how likely is a pulse to occur in dt after a previous pulse.

result that shorter intervals are much more common than longer ones. This is a vindication of the colloquial rule that "disasters come in threes": completely random events tend to cluster together in time. Perhaps more practical is the prediction that one may find the average rate r in a counting experiment by plotting the experimental interval distribution on a semi-log scale: both the slope and intercept will give r (after a some arithmetic).

Exercise 2 *(a) Show that $I_1(t)$ is normalized, that is, prove $\int_0^\infty I_1(t)\,dt = 1$. (b) Show that the average 1-pulse interval τ_1 is equal to $1/r$. The average interval is given by $\int_0^\infty tI_1(t)\,dt$.*

The situation for the distribution of 2-pulse intervals is a bit more complicated. We want the likelihood of finding an interval of a particular length t between two pulses that contains a single pulse anywhere inside at $t' < t$. One such possibility is shown in Fig. 5. The (differential) probability of finding an interval like this is given by rule **II**:

$$\begin{aligned} dP(t,t') &= \\ & \begin{pmatrix}\text{Probability of}\\ \text{finding 0}\\ \text{pulses between}\\ \text{0 and } t'\end{pmatrix} \times \begin{pmatrix}\text{Probability of}\\ \text{finding a pulse}\\ \text{between } t' \text{ and}\\ t'+dt'\end{pmatrix} \times \begin{pmatrix}\text{Probability of}\\ \text{finding 0}\\ \text{pulses between}\\ t' \text{ and } t\end{pmatrix} \times \begin{pmatrix}\text{Probability of}\\ \text{finding a pulse}\\ \text{between } t \text{ and}\\ t+dt\end{pmatrix} \\ &= P_0(t') \times r\,dt' \times P_0(t-t') \times r\,dt\ . \end{aligned}$$

This construction is nothing more than the probability of finding two 1-pulse intervals of particular lengths. But since the intermediate pulse can happen anywhere in the interval, by rule **I** we must sum the individual probabilities over all values of t' between 0 and t. Hence, we see that

$$\begin{aligned} I_2(t)\,dt &= \left(\int_0^t I_1(t')I_1(t-t')\,dt'\right)\,dt \\ &= \left(\int_0^t re^{-rt'}\,re^{-r(t-t')}\,dt'\right)\,dt \\ &= \left(r^2e^{rt}\int_0^t dt'\right)\,dt \\ &= r(rt)e^{-rt}\,dt\ . \end{aligned}$$

So the distribution function for 2-pulse intervals is

$$I_2(t) = r(rt)e^{-rt}\ . \tag{5}$$

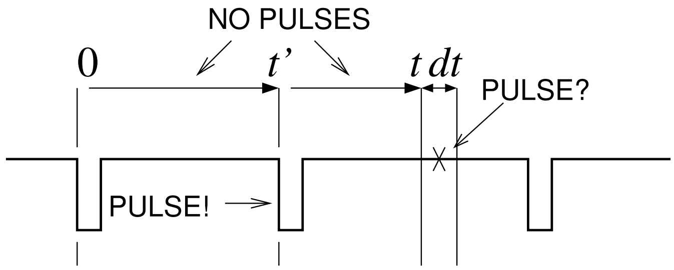

Figure 5: Construction used in calculating the probability a 2-pulse interval of length t.

Likewise, we may find the distribution for 3-pulse intervals $I_3(t)$ by multiplying $I_2(t')$ by $I_1(t-t')$ an integrating over all $t' < t$. This process leads to a general recursion relation

$$I_n(t) = \int_0^t I_{n-1}(t') I_1(t-t')\, dt' \ . \tag{6}$$

When applied to the exponential distribution the formula gives

$$I_n(t) = r\frac{(rt)^{(n-1)}e^{-rt}}{(n-1)!} \ . \tag{7}$$

Equation 7 is known as the (integer) gamma distribution or Erlang(ian) distribution (after the Danish engineer A. K. Erlang who used it to characterize telephone networks).

Exercise 3 *Convince yourself of the $(n-1)!$ factor in Eq. (7) by using $I_3(t)$ and $I_1(t)$ to show that*

$$I_4(t) = r\frac{(rt)^3 e^{-rt}}{1 \cdot 2 \cdot 3}$$

Exercise 4 *Show that the average n-pulse interval τ_n is equal to n/r, but that the most probable interval $T_{n,\mathrm{peak}}$ is given by $(n-1)/r$. The most probable interval is the interval at the peak of the distribution, and is found by evaluating $dI_n(t)/dt = 0$. Hint: $\int_0^\infty x^n e^{-x}\, dx = n!$.*

Graphs of $I_n(t)$ are shown in Fig. 6(a) for a few values of n. A couple of things are notable. First, the probability of finding zero length intervals for $n > 1$ is zero. This makes sense in that there must be at least one pulse within each such n-pulse interval, and even a single pulse takes time. More interestingly, the relative width of the peak of $I_n(t)$ shrinks as n increases. This is shown most clearly by scaling $I_n(t)$ so that the functions all have the same mean value $(1/r)$ while maintaining the normalized (unit) area, as is shown in Fig. 6(b). What this means is that for larger n, the collection of intervals becomes more uniform.

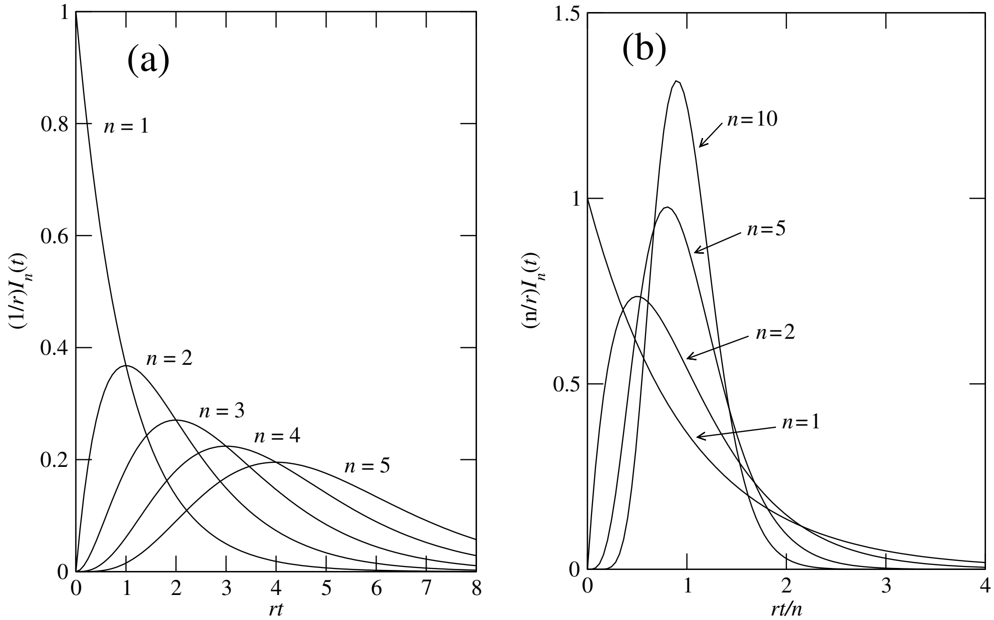

Figure 6: The n-pulse interval density functions, after Knoll [2, p. 98]. (a) $(1/r)I_n(t)$ for n from 1 to 5. Notice that the functions peak at $rt = n - 1$. (b) $(1/r)I_n(t)$ scaled by the transformations $t \to t/n$ and $(1/r)I_n(t) \to (n/r)I_n(t)$ to show that as n increases, the function becomes more sharply peaked around the mean value $\tau_n = n/r$.

Exercise 5 *Evaluate the variance* σ_n^2 *of* $I_n(t)$*. The variance is given by the formula*

$$\sigma_n^2 = \int_0^\infty (t^2 - \tau_n^2) I_n(t)\, dt \ .$$

Then show that the fractional standard deviation in τ_n*,* σ_n/τ_n *is equal to* $1/\sqrt{n}$*.*

Distribution of counts

Up to now, we have been framing our investigation thusly, "Given a fixed number of counts, how are the intervals that contain exactly this number distributed?" In other words, we are using the count number as a *parameter* in our problem, and letting the time be the *variable*. But frequently, our experimental situation is different: the independent parameter is the time period, and the dependent variable is the number of counts that occur within that time period. This is a related, but not identical question. Most obviously, as stated in the introduction, a time interval can take on any value, but a count number can only take integer values; the distribution of interval lengths is a *continuous* distribution, whereas the distribution of counts is a *discrete* distribution.

The relationship between the two distributions can be stated plainly: A period of a given length containing q counts, T_q, has *fewer* than n counts if and only if the interval containing n counts, T_n, is *longer* than T_q. That is,

$$q < n \ \text{ if and only if } \ T_n > T_q \ .$$

This suggests the following interpretation of what has been derived so far. The probability of finding an n-count interval of between length T and $T + dt$ is the same as the probability of finding

Table 1: A comparison of the two types of distributions that are contained in the "Poisson" formula. Note $\sigma_{_E}^2 = \tau_n^2/n$ and $\sigma_{_P}^2 = \overline{n}$.

	Erlang	Poisson
Type:	Continuous	Discrete
Variable:	t	n
Mean:	$\tau_n = n/r$	$\overline{n} = rt$
Variance:	$\sigma_{_E}^2 = n/r^2$	$\sigma_{_P}^2 = rt$

exactly $n-1$ counts in an interval of length T times the probability of finding one more count in T to $T+dt$. Stated in terms of our formulas:

$$I_n(T)\,dt = \frac{(rT)^{(n-1)}e^{-rT}}{(n-1)!} \times r\,dt = P(n-1; rT) \times r\,dt\ . \tag{8}$$

Thus, we identify $P(n-1;rT)$ as none other than the Poisson distribution giving the probability of finding $n-1$ events given that the average number of events is rT. This fact is used in the derivations of the interval distribution function given by Knoll [2] and Melissinos [3]. Most derivations of the Poisson distribution are based on taking a particular limit of the discrete binomial distribution. The derivation given here follows a more direct route from the basic rules of probability as applied to random pulse trains.

As is well known, the Poisson distribution is a distribution where the mean is equal to the variance; only one parameter is needed to define the whole distribution. This is not true of the Erlang distribution; for a given rate constant r and count number n, the mean interval time is n/r but the variance is n/r^2. A comparison of the two distributions given by our formula is given in Table 1.

References

[1] Wannier, Gregory H., *Statistical Physics,* (Dover, Mineola NY, 1987), p. 29.

[2] Knoll, Glenn F., *Radiation Detection and Measurement*, 2nd edition, (John Wiley, New York, 1989), pp. 96–99.

[3] Melissinos, Adrian C., and Jim Napolitano, *Experiments in Modern Physics*, 2nd edition (Academic Press, New York, 2003), pp. 401–404, 470–473.

[4] Cox, D. R., *Renewal Theory,* (Meuthen and Co. Science Paperbacks, Butler & Tanner Ltd., 1967).

Prepared D. Pengra

`counting_stats_tutorial_b.tex -- Updated 29 April 2008`