Poisson Distribution

1 Derivation

Consider an event that occurs at random times. We want $p(N|\mu)$ where N is the number of occurrences, and μ is the average number of occurrences in a give time interval t.
λ = probability per unit time of event occurring
$\Delta p = \lambda \Delta t$ = probability of occurrence in small time Δt.
Let $t = M\Delta t$. Since t is finite and eventually we'll take $\Delta t \to 0$, M is large and eventually we'll take $M \to \infty$. For simplicity, M is an integer.
Let $p_0(t)$ be the probability that no event occurs in t.

$$p_0(t) \;=\; (1-\lambda\Delta t)^M \;=\; \left(1-\frac{\lambda t}{M}\right)^M$$

$$p_0(t) \;=\; 1+M\left(-\frac{\lambda t}{M}\right)+\frac{M(M-1)}{2!}\left(-\frac{\lambda t}{M}\right)^2+\frac{M(M-1)(M-2)}{3!}\left(-\frac{\lambda t}{M}\right)^3+....$$

but

$$\lim_{M\to\infty} M(M-1) = M^2 \quad\text{and}\quad \lim_{M\to\infty} M(M-1)(M-2) = M^3 \quad\text{etc.}$$

therefore

$$\lim_{M\to\infty} p_0(t) \;=\; 1+(-\lambda t)+\frac{(-\lambda t)^2}{2!}+\frac{(-\lambda t)^3}{3!}+... = e^{-\lambda t}$$

$$\boxed{p_0(t) = e^{-\lambda t}}$$

Let $p_1(t)$ be the probability that one event happens between $t = 0$ and t. Then $p_1(t+dt)$ is the probability that no event happens before t multiplied by the probability that it happens in dt plus the probability that one event happens before t multiplied by the probability that no other event happens in dt. This can be written as

$$p_1(t+dt) = p_0(t)\lambda dt + p_1(t)(1-\lambda dt) = e^{-\lambda t}dt + p_1(t)(1-\lambda dt)$$

$$\frac{dp_1(t)}{dt} = \lambda e^{-\lambda t} - \lambda p_1(t)$$

The solution of this simple differential equation is

$$\boxed{p_1(t) = \lambda t e^{-\lambda t}}$$

Next we will show by induction that $p_N(t) = (\lambda t)^N e^{-\lambda t}/N!$. This works for $N = 0$ and $N = 1$. Let's assume that it works for N and show that it also works for $N + 1$.
Using the same reasoning that led us to a differential equation for p_1 in terms of p_0, we can write

$$\frac{dp_{N+1}(t)}{dt} = x - \lambda\ p_{N+1}(t) + \lambda\ p_N(t)$$

Let's see if our assumed solution works. With the assumed solution we have

$$\frac{dp_{N+1}(t)}{dt} \quad = \quad - \frac{\lambda(\lambda t)^{N+1} e^{-\lambda t}}{(N+1)!} \quad + \quad \frac{\lambda(\lambda t)^N e^{-\lambda t}}{(N)!} \quad = \quad -\lambda p_{N+1}(t) \ + \ \lambda p_N(t)$$

Yes, it works. Identifying $\mu = \lambda t$ we have

$$\boxed{p(N|\mu) = \frac{\mu^N e^{-\mu}}{N!}}$$

2 Mean

Clearly $< N >= \mu$. It is easy to verify it:

$$< N >= \sum_{n=0}^{n=\infty} n\ p(n|\mu) = e^{-\mu} \sum_{n=0}^{n=\infty} \frac{n\mu^n}{n!} = \mu e^{-\mu} \sum_{n=1}^{n=\infty} \frac{\mu^{n-1}}{(n-1)!} = \mu e^{-\mu} e^{+\mu} = \mu$$

3 Variance

$$< N^2 >= \sum_{n=0}^{n=\infty} n^2\ p(n|\mu) = \mu e^{-\mu} \sum_{n=1}^{n=\infty} \frac{n\mu^{n-1}}{(n-1)!}$$

$$< N^2 >= \mu e^{-\mu} \left(\sum_{n=1}^{n=\infty} \frac{(n-1)\mu^{n-1}}{(n-1)!} + \sum_{n=1}^{n=\infty} \frac{\mu^{n-1}}{(n-1)!} \right)$$

$$< N^2 >= \mu e^{-\mu} \left(\mu \sum_{n=2}^{n=\infty} \frac{\mu^{n-2}}{(n-2)!} + \sum_{n=1}^{n=\infty} \frac{\mu^{n-1}}{(n-1)!} \right) \ = \ \mu e^{-\mu}\ (\mu e^{\mu} + e^{\mu}) = \mu^2 + \mu$$

Since $\sigma^2 =< N^2 > - < N >^2$ we get

$$\boxed{\sigma^2 = \mu}$$

4 Sum of Poisson variables

Let a and b be independent Poisson random variables with $< a >= a_0$ and $< b >= b_0$. The random variable $c = a + b$ is also a Poisson variable with $c_0 \equiv < c >= a_0 + b_0$. Proof:

$$p(c = k) = \sum_{i=0}^{k} p(b = k - i) p(a = i) = e^{-(a_0+b_0)} \sum_{i=0}^{k} \frac{b_0^{k-i} a_0^i}{(k-i)! i!}$$

Multiply top and bottom by $k!$:

$$p(c = k) = \frac{e^{-(a_0+b_0)}}{k!} \sum_{i=0}^{k} \binom{k}{i} b_0^{k-i} a_0^i = \frac{e^{-(a_0+b_0)}}{k!} (a_0 + b_0)^k = \frac{c_0^k \ e^{-c_0}}{k!}$$

5 Limit of large N

For large N use Stirling formula $N! \approx \sqrt{2\pi N} e^{-N} N^N$. Then

$$p(N|\mu) = \frac{\mu^N e^{-\mu}}{N!} \approx \frac{1}{\sqrt{2\pi N}} \left(\frac{\mu}{N}\right)^N e^{N-\mu}$$

Write $N = \mu(1 + \delta)$ with $\mu >> 1$ and $\delta << 1$, i.e., large N and large μ and the two are not too different from each other. Then $\mu/N = 1/(1 + \delta)$, $N - \mu = \delta\mu$ and $\sqrt{N} = \sqrt{\mu}\sqrt{1 + \delta}$. Substituting we get

$$p(N|\mu) \approx \frac{e^{\delta\mu}}{\sqrt{2\pi\mu}} (1 + \delta)^{-\mu(1+\delta)-0.5} = \frac{1}{\alpha} \ \frac{e^{\delta\mu}}{\sqrt{2\pi\mu}}$$

Where $\alpha = (1 + \delta)^{\mu(1+\delta)+0.5}$.

$$\log\alpha = \left(\mu(1 + \delta) + \frac{1}{2}\right) \log(1 + \delta)$$

Since δ is small, expand $\log(1 + \delta) \approx \delta - \delta^2/2$. Then to order $\mu\delta^2$

$$\log\alpha \approx \mu\delta + \frac{1}{2}\mu\delta^2 \quad \rightarrow \quad \frac{1}{\alpha} \approx e^{-\delta\mu} e^{-\delta^2\mu/2}$$

Plugging $1/\alpha$ back in the equation for $p(N|\mu)$:

$$p(N|\mu) \approx \frac{\exp(-\delta^2\mu)}{\sqrt{2\pi\mu}}$$

But since $N = \mu(1 + \delta)$, $\delta^2\mu = (N - \mu)^2/\mu$. So, finally

$$\boxed{p(N|\mu) \approx \frac{1}{\sqrt{2\pi\mu}} \ e^{-(N-\mu)^2/(2\mu)}}$$

A gaussian of mean μ and standard deviation $\sqrt{\mu}$.

6 Weighted events

In HEP we often have to deal with samples of weighted events, e.g., when we make a histogram where we add two or more Monte Carlo components with different weights. The prescription is that the uncertainty in each bin is $\sigma^2 = \sum w_i^2$ where the sum is over all events and w_i is the weight of event i. This prescription can be implemented using the "Sumw2" prescription in ROOT histograms.
Lets justify this for two samples N_1 and N_2 with weights w_1 and w_2 and means μ_1 and μ_2. The argument easily generalizes to more samples, even to a sample where every weight is different, as can happen in some event generators that output weighted events.

Since we are interested in $N = w_1 N_1 + w_2 N_2$ the estimate of the mean of N is $w_1\mu_1 + w_2\mu_2 = w_1 N_1 + w_2 N_2$.
Let's now estimate the variance of N. The variance of $w_1 N_1$ is

$$\mathrm{Var}(w_1N_1) =< (w_1N_1)^2 > - < w_1N_1 >^2 = w_1^2 < N_1^2 > -w_1^2 < N_1 >^2$$

$$\mathrm{Var}(w_1N_1) = w_1^2(< N_1^2 > - < N_1 >^2 = w_1^2 \mathrm{Var}(N_1) = w_1^2 N_1 = \sum_i w_i^2$$

were $w_i = w_1$ is the weight of the $i - th$ event in the 1st sample. The variance $\mathrm{Var}(w_2N_2)$ is given by a similar sum over the 2nd sample. Then since N_1 and N_2 are independent $\mathrm{Var}(N) = \mathrm{Var}(w_1N_1) + \mathrm{Var}(w_2N_2)$, $\mathrm{Var}(N) = \sum w_i^2$ where now the sum is over all events.
Note that the pdf for N is not a Poisson anymore. In the limit of large number of events in each sample it is approximately Gaussian.