STSCI 4780
Relationships between distributions

PRELIMINARY

Tom Loredo, CCAPS & SDS, Cornell University

© 2020-05-05

Sum of Bernoulli 0/1's is binomial

FFSSSSFSSSFS ($n = 8$ successes in $N = 12$ trials)

Bernoulli process

$$\begin{aligned} p(D|\alpha, M) &= p(\text{failure}|\alpha, M) \times p(\text{failure}|\alpha, M) \times \cdots \\ &= \alpha^n (1-\alpha)^{N-n} \\ &= \mathcal{L}(\alpha) \end{aligned}$$

Binomial distribution

Let $\mathcal{S}$ = a sequence of flips with n heads.

$$\begin{aligned} p(n|\alpha, M) &= \sum_{\mathcal{S}} p(\mathcal{S}|\alpha, M)\, p(n|\mathcal{S}, \alpha, M) \\ &= \alpha^n (1-\alpha)^{N-n} C_{n,N} \end{aligned}$$

$$C_{n,N} \equiv \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Sum of same-α binomial counts is binomial

N_1 trials with n_1 successes, followed by N_2 trials with n_2 successes, gives total successes $n = n_1 + n_2$ in $N = N_1 + N_2$ trials

- The total corresponds to the successes from N trials, for which

$$p(n|\alpha, N) = \binom{N}{n} \alpha^n (1-\alpha)^{N-n}$$

- Show it explicitly using the LTP:

$$\begin{aligned}
p(n|\alpha, N) &= \sum_{n_1} \sum_{n_2} p(n_1|\alpha, N_1) p(n_2|\alpha, N_2) p(n|n_1, n_2, \ldots) \\
&= \sum_{n_1} p(n_1|\alpha, N_1) p(n_2 = n - n_1|\alpha, N_2) \\
&= \sum_{n_1} \binom{N_1}{n_1} \binom{N_2}{n - n_1} \alpha^n (1-\alpha)^{N-n} \\
&= \binom{N}{n} \alpha^n (1-\alpha)^{N-n}
\end{aligned}$$

by a binomial coefficient identity

Sum of Poisson counts is Poisson

n_1 counts in an interval with $r_1 T_1 = \lambda$, n_2 counts in an interval with $r_2 T_2 = \mu$, total counts $n = n_1 + n_2$:

$$\begin{aligned}
p(n|\lambda,\mu) &= \sum_{n_1}\sum_{n_2} p(n_1|\lambda)p(n_2|\mu)p(n|n_1,n_2,\ldots) \\
&= \sum_{n_1} p(n_1|\lambda)p(n_2 = n - n_1|\mu) \\
&= \sum_{n_1} \frac{\lambda^{n_1}}{n_1!}e^{-\lambda}\frac{\mu^{n-n_1}}{(n-n_1)!}e^{-\mu} \\
&= \frac{e^{-(\lambda+\mu)}}{n!}\sum_{n_1}\frac{n!}{n_1!(n-n_1)!}\lambda^{n_1}\mu^{n-n_1} \\
&= \frac{(\lambda+\mu)^n}{n!}e^{-(\lambda+\mu)}
\end{aligned}$$

Poisson with summed rate $\times$ interval

Sum of normally distributed values is normal

Suppose $x \sim \mathrm{Norm}(0, \sigma^2)$ and $y \sim \mathrm{Norm}(0, \sigma^2)$;
what is the PDF for $s = x + y$?

$$p(s) \propto \exp\left[-\frac{s^2}{4\sigma^2}\right]$$

I.e., normal with $\sigma_s^2 = \sigma_x^2 + \sigma_y^2 = 2\sigma^2$

When the distribution for a sum is in the same family as the distributions for the components, the family is *infinitely divisible*

Poisson for large expected counts

Recall the Poisson distribution for n counts from a process with rate r over an interval T:

$$p(n|r, M) = \frac{(rT)^n}{n!} e^{-rT} = \frac{\mu^n}{n!} e^{-\mu} \quad \text{for } \mu \equiv rT$$

Expectation value of n:

$$\begin{aligned}
\mathbb{E}(n) &\equiv \sum_{n=0}^{\infty} n p(n|r, I) \\
&= e^{-\mu} \sum_{n=1}^{\infty} \frac{\mu^n}{(n-1)!} \\
&= \mu\, e^{-\mu} \sum_{m=0}^{\infty} \frac{\mu^m}{m!} \quad \text{for } m = n - 1 \\
&= \mu
\end{aligned}$$

Similarly, $\mathrm{Var}(n) = \mathbb{E}(n^2) - \mu^2 = \mu$, so the standard deviation $\sigma_n = \mu^{1/2}$

For $\mu \gg 1$, expect $n \approx \mu \pm \mu^{1/2}$ so

$$\delta \equiv \frac{n - \mu}{\mu} \ll 1$$

In terms of δ, we can write $n = \mu(\delta + 1)$

Look for the leading order term in $\log p(n)$; use two approximations, Stirling's approximation:

$$\log(n!) \approx n \log(n) - n$$

and a Taylor expansion for the logarithm:

$$\log(1 + x) \approx x - \frac{x^2}{2} + \cdots$$

$$\begin{aligned}
\log p(n) &\approx n \log \mu - n \log n + n - \mu \\
&= -n \log \frac{n}{\mu} + n - \mu \\
&= -\mu(\delta + 1) \log(1 + \delta) + \mu(\delta + 1) - \mu \\
&\approx -\mu(\delta + 1) \left(\delta + \frac{\delta^2}{2} \right) + \mu\delta \\
&\approx -\mu\delta^2 - \mu\delta + \frac{\mu\delta^2}{2} + \mu\delta \\
&= -\frac{\mu\delta^2}{2}
\end{aligned}$$

So

$$p(n|\mu) \propto \exp\left(-\frac{(n - \mu)^2}{2\mu} \right)$$

a *Gaussian distribution* (evaluated at integer values of n)

Binomial for rare events

How many Cornell students share your birthday?

- $N \approx 24,000 \gg 1$
- $\alpha \approx \frac{1}{365} \ll 1$
- Expected number $\mu \equiv \mathbb{E}(n) = \alpha N \approx 66 \gg 1$

1000 bacteria are mixed in a liter of water. How many are in a 0.1 ml sample?

- $N = 1000 \gg 1$
- $\alpha = 10^{-4}$
- Expected number $\mu \equiv \mathbb{E}(n) = \alpha N = 0.1 \ll 1$

Seek an approximation for $p(n| \ldots)$ for small α, but not necessarily small μ (or n): A rare event can happen many times in a very large sample

Recall the binomial sampling distribution for n successes in N trials, given success probability α:

$$p(n|\alpha, M) = \frac{N!}{n!(N-n)!}\alpha^n(1-\alpha)^{N-n}$$

Expected number of successes $\mu \equiv \mathbb{E}(n) = \alpha N$

Recursion relation:

$$\begin{aligned}
\frac{p(n)}{p(n-1)} &= \frac{N!}{n!(N-n)!}\frac{(n-1)!(N-n+1)!}{N!}\frac{\alpha}{1-\alpha} \\
&= \frac{N-n+1}{n}\frac{\alpha}{1-\alpha}
\end{aligned}$$

Consider the limit where $N \to \infty$ and $\alpha \to 0$, but with $\mu = \alpha N$ fixed and not necessarily small (but $\mu \ll N$); focus on $n \sim \mu$ so $n \ll N$ as well:

$$\frac{p(n)}{p(n-1)} \approx \frac{N\alpha}{n} = \frac{\mu}{n}$$

In that same limit, writing α in terms of μ and N,

$$p(0) = (1-\alpha)^N = \left(1 - \frac{\mu}{N}\right)^N \approx e^{-\mu}$$

Now evaluate $p(n)$ using the recurrence relation:

$$\begin{aligned} p(1) &= \frac{\mu}{1} \times p(0) = \mu e^{-\mu} \\ p(2) &= \frac{\mu}{2} \times p(1) = \frac{\mu^2}{2} e^{-\mu} \\ p(n) &= \frac{\mu}{n} \times p(n-1) = \frac{\mu^n}{n!} e^{-\mu} = \textit{Poisson} \end{aligned}$$

The *Poisson limit theorem* or *law of rare events* (events rare in *proportion*, though possibly numerous)

Binomial for large N: de Moivre-Laplace theorem

For N Bernoulli outcomes with probability α and n successes, note that

$$\mathbb{E}(n) = \alpha N; \quad \mathrm{Var}(n) = \alpha(1-\alpha)N$$

Define

$$\Delta = \frac{n - \alpha N}{\sqrt{\alpha(1-\alpha)N}}$$

Then for large N,

$$p(n|\alpha, N) \propto \exp\left(-\frac{\Delta^2}{2}\right)$$

This is a special case of a more general sum-of-outcomes result, the *central limit theorem* (CLT)