Week 9 Recap

Monday October 16: Poisson approximation for binomials distributions

We discussed earlier how the Binomial $\mathrm{Bin}(n,p)$ can be approximated by a normal distribution when $np(1-p)$ is large enough. In this section we discuss what to do when $np(1-p)$ is small. Since n is an integer (and, most likely, a large integer), in order for $np(1-p)$ to be small, it must be that p or $q=1-p$ is small.
By exchanging p and q if necessary, we can assume that p is small, so small that $np(1-p)\approx np$ is also relatively small. It will be useful to think of p is given by

$$p=\lambda/n,\ \ \text{equivalently}\ \ \lambda=np.$$

As a concrete example, we can think of the probability of wining a daily jackpot at a lottery. Players buy daily tickets and we let X be the number of times a player wins in one calendar year (results from different weeks are independent). Clearly
X is a binomial $\mathrm{Bin}(365,p)$ random variable where p is the probability of winning which is extremely small (in this case, $\lambda=365\times p$ itself is very small).

Approximation of a binomial using a Poisson distribution Fix $\lambda>0$. Let n be such that $\lambda/n<1$. Let $S_n\sim\mathcal{B}(n,\lambda/n)$. Then

$$\lim_{n\to\infty}P(S_n=k)=e^{-\lambda}\frac{\lambda^k}{k!}.$$

Proof: We want to show that $P(S_n=k)\approx e^{-\lambda}\frac{\lambda^k}{k!}$. Because S_n is a $\mathrm{Bin}(n,\lambda/n)\$$ random variable,

$$P(S_n=k)=(\lambda/n)^k(1-\lambda/n)^{n-k}\times\frac{n(n-1)\dots(n-k+1)}{k!}.$$

Let us rearrange the right-hand side, putting together the terms which do not depend on n, that is,

$$P(S_n=k)=\frac{\lambda^k}{k!}(1-\lambda/n)^n(1-\lambda/n)^{-k}\frac{n(n-1)\dots(n-k+1)}{n^k}.$$

Obviously, we want to keep the first fraction. In high school, you learned (perhaps with minimal understanding then) that $(1-\lambda/n)^n$ tends to $e^{-\lambda}$ when n tends to $+\infty$. We can understand why this is true thanks to our good calculus training by writing $(1-\lambda/n)^n = e^{n\log(1-\lambda/n)}$ and

$$\begin{aligned}\lim_{n\to\infty} n\log(1-\lambda/n) &= -\lambda \lim_{h\to 0}(-1/\lambda h)(\log(1-\lambda h)-\log(1))\\ &= -\lambda \log'(1) = -\lambda.\end{aligned}$$

It follows from this and the continuity of the exponential function that $(1-\lambda/n)^n$ tends indeed to $e^{-\lambda}$.

Note: Whether or not this is a good proof of the desired limit depends in part on how the function $x \mapsto e^x$ is defined. Here, I am assuming that the base e exponential function $x \mapsto e^x$ is defined as the inverse function of $\log(x) = \int_1^x \frac{dx}{x}$. It is possible to define e^x as the limit of $(1+x/n)^n$. If that is the original definition, then there is nothing to prove!

The third factor, $(1-\lambda/n)^{-k}$, tends to 1 when n tends to infinity. Finally, observe that the last fraction $\frac{n(n-1)\dots(n-k+1)}{n^k}$ has the same number, k, of factors in its numerator and denominator. Using this fact, we rewrite this fraction in the form

$$\frac{n(n-1)\dots(n-k+1)}{n^k} = 1\times(1-1/n)\dots(1-(k-1)/n).$$

Each of the k factors in this product tends to 1 when n tends to infinity so the entire product tends to 1. Putting all these computations together shows that, for each fixed k,

$$\lim_{n\to\infty} P(S_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

This is the desired result.

The following is a much more powerful statement which gives a useful bound for the error made by replacing a binomial by a Poisson with the same mean.
Observe that the bound is valid for the approximation of the probability of an arbitrary event A.

Approximation of a binomial using a Poisson distribution. Suppose that $S \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathcal{P}(np)$ (i.e., $\lambda = np$). Then, for any subset A of $\{0,1,2,\dots\}$,

$$|P(S\in A) - P(Y\in A)| \le np^2 = \lambda^2/n.$$

This is a special case of the following more general statement.

Approximation by a Poisson distribution Let $(X_i)_1^n$ be independent Bernoulli random variables with parameters p_i. Let Y_i be independent Poisson random variables with parameters p_i. Let $X = \sum_1^n X_i$ and $Y = \sum_1^n Y_i$. Then

$$|P(X \in A) - P(Y \in A)| \leq \sum_1^n p_i^2.$$

The first statement is a consequence of the second because a binomial (n, p) can be viewed as a sum of n independent Bernoulli p and a Poisson np can be view as the sum of n independent Poisson p. This last fact can be seen as follows. Let U, V be independent Poisson with respective parameters u, v defined on a same probability space (Ω, P), and let $Z = U + V$. Then

$$\begin{aligned} P(Z = k) &= \sum_{\ell=0}^{k} P(U = \ell, V = k - \ell) \\ &= \sum_0^k e^{-u} \frac{u^\ell}{\ell!} e^{-v} \frac{v^{k-\ell}}{(k-\ell)!} \\ &= e^{-(u+v)} \frac{(u+v)^k}{k!}. \end{aligned}$$

This shows that the sum of two independent Poisson random variables is a Poisson random variable (with parameter given by the sum of the two original parameters). Induction allows us to extend the statement to more than two independent Poisson random variables. It is now clear the first statement follows from the second by taking all the p_is equal to p.

Wednesday October 18:

Earlier, we have learned that the number X of successes in a long sequence of n repeated identical independent experiments with small probability of success $p = \lambda/n$ can be approximated by a Poisson random variable with parameter λ. This works well when $p^2 n$ is small. Examples of application are: The number of misprints on a page chosen at random in a document. The number of people visiting the post office on a given day, The number of students needing a makeup for the prelim.

Two examples: In real life, one is often interested in using such a Poisson approximation in complicated cases when the occurrences of what ever we are counting are not quite independent but we still expect the dependence to be weak enough for the same result to hold. We now describe such examples.

Example 1: Matchings Pick a permutation of $\{1,\dots,n\}$ at random and let X be the number of fixed points. We can write $X=\sum_1^n X_i$
where X_i is either 1 or 0 depending on whether or not card I is in position I (a fixed point in the permutation). This looks very much like counting the number of success in repeated experiment EXCEPT that these random variable X_i are not independent. However, if n is large and m is fixed, the random variables $X_1, X_2, \dots, X_m$ are almost independent in the sense that

$$\begin{aligned} P(X_1=1,\dots,X_m=1) &= \frac{1}{n}\frac{1}{n-1}\dots\frac{1}{n-m+1} \approx \left(\frac{1}{n}\right)^m \\ &= P(X_1=1)\dots P(X_m=1). \end{aligned}$$

If we ignore the lack of independence, we have n experiment, each has probability of success $1/n$, and we should be able to approximate
X by a Poisson random variable with parameter 1.

When we studied matchings using the inclusion-exclusion principle, we proved that this is correct by a direct computation. So

$$P(X=0)\approx 1/e, P(X=1)\approx 1/e, P(X=2)\approx 1/(2e), P(X=3)\approx 1/(6e), \text{ etc.}$$

Example 2: The birthday problem We want to approximate
$P(\text{no two people among } n \text{ have the same birthday})$ when we have a group of n individuals. We think of checking all the $n(n-1)/2$ different pairs of people, one by one. For $1\le i<j\le n$ let $X_{ij}=0$ or 1 depending on whether or not individual I and j have the same birthday. These are not quite independent but the dependence is weak (there is dependence only between pairs sharing one member). Here we have $n(n-1)/2$ repeated experiments (one for each pair) and $P(X_{ij}=1)=1/365$ so we should take $\lambda=\binom{n}{2}/365=n(n-1)/730.$ This gives
$P(\text{no two people among } n \text{ have the same birthday})\approx e^{-n(n-1)/730}$.

Checking how large n must be to be sure that this probability is larger than $1/2$, we find $n\ge 23$. But now we can solve the much more complicated case of approximating
$P(\text{no three people among } n \text{ have the same birthday}).$

Write down the argument that explains why
$P(\text{no three among n have the same birthday})\approx e^{-\frac{n(n-1)(n-2)}{799350}}$.

Events arising in time, a theoretical framework Assume we are interested in events that occur at various times in such a way that the following hypotheses are satisfied for some constant $\lambda > 0$:

- The probability that exactly ONE event occurs in any fixed time interval of length $s > 0$ is equal to $\lambda s + o(s)$ (read: little o of s; this describes a quantity that tends to zero faster than s). The $o(s)$ is the same for all intervals of length s.
- The probability that MORE than one event occur in any fixed time interval of length s is $o(s)$. The $o(s)$ is the same for all intervals of length s.
- For any finite family of disjoint intervals $I_1, \dots, I_j$ and integers $i_1, \dots, i_j$, the events $E_1, \dots, E_j$ that exactly i_k events occur in the interval I_k, $1 \leq k \leq j$, are independent.

Poisson distribution for rare events: Under the assumptions stated above, the number $N(t)$ of events occurring in the interval $[0, t]$ (or any interval interval of length t) is a Poisson random variable with parameter λt.