From last time: common discrete RVs

**Bernoulli**$(p)$:

\[ P\{X = 1\} = p, \quad P\{X = 0\} = 1 - p \]

**binom**$(n, p)$:

\[ P\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n \]

**geometric**$(p)$:

\[ P\{X = k\} = (1 - p)^{k-1} p \quad \text{for } k = 1, 2, \ldots \]

**hypergeometric**$(n, N, m)$:

\[ P\{X = k\} = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \quad \text{for } \max(n-(N-m), 0) \leq k \leq \min(n, m) \]
Suppose an urn contains $m$ white and $N - m$ black balls, and we sample $n$ of them without replacement. Let $X$ be the number of white balls in our sample. Then $X$ is hypergeometric($n, N, m$).
Proposition (Ross E.g. 4.8j or 7.2g)

If $X$ is hypergeometric($n, N, m$), then

$$E[X] = \frac{nm}{N}$$

and

$$\text{Var}(X) = np(1 - p)\left(1 - \frac{n - 1}{N - 1}\right), \quad \text{where } p = \frac{m}{N}.$$ 

See Ross E.g. 4.8j for the calculation of the variance and some other things.
Today we will meet some results of a new flavour:

*Approximating one random variable by another.*

Reasons why this can be important:

- Sometimes it shows that two different experiments are likely to give similar results.
- Sometimes it lets us approximate a complicated PMF by another for which calculations are easier.
- Sometimes it lets us ignore certain details of a RV: for example, by reducing the number of parameters (such as the $n$ and the $p$ in $\text{binom}(n, p)$) that we need to specify.
The first example is quite simple.

**Proposition (Ross pp153-4)**

Let $X$ be \textit{hypergeometric}(n, N, m), and let $p = m/N$ (the fraction of white balls in the urn).

If $m \gg n$ and $N - m \gg n$ (that is, the numbers of white and black balls are both much bigger than the size of the sample), then

$$P\{X = k\} \approx \binom{n}{k} p^k (1 - p)^{n-k}.$$ 

\textbf{THUS:} $X$ is approximately \textit{binom}(n, p).
We see this effect in the expectation and variance as well. If $X$ is hypergeometric($n, N, m$), $p = m/N$ and $m, N - m \gg n$, then

$$E[X] = \frac{nm}{N} = np$$

and

$$\text{Var}(X) = np(1 - p)\left(1 - \frac{n - 1}{N - 1}\right) \approx np(1 - p).$$

That is, they roughly match the expectation and variance of binom($n, p$).
This is a rigorous statement of a fairly obvious phenomenon:

Suppose we sample \( n \) balls from an urn which contains \( m \) white and \( N - m \) black balls. If the number of balls we sample is much smaller than the total number of balls of either colour, then sampling with and without replacement are roughly equivalent.

Intuitively this is because, when we sample \textit{with} replacement, we’re still very unlikely to pick the same ball twice.
The rest of today will be given to another approximation, less obvious but more powerful: the Poisson approximation.
Here’s a picture to get us started:
These are (parts of) the PMFs of three different binomials, with parameter choices \((10, \frac{3}{10})\), \((20, \frac{3}{20})\) and \((100, \frac{3}{100})\).

They are chosen to have the same expectation:

\[
10 \times \frac{3}{10} = 20 \times \frac{3}{20} = 100 \times \frac{3}{100} = 3.
\]

But we can see that in fact their PMFs look very similar, not just their expectations.
Suppose we fix the expectation of a binom$(n, p)$ to be some value $\lambda > 0$ (so $\lambda = np$), and then assume $n$ is very large, and hence $p = \lambda/n$ is very small. It turns out that in this extreme, binomial RVs start to look like another fixed RV ‘in the limit’.

**Proposition (Ross p136)**

Fix $\lambda > 0$. Let $X$ be binom$(n, p)$ for a very large value of $n$ and with $p = \lambda/n$. Then

$$P\{X = k\} \approx e^{-\lambda} \frac{\lambda^k}{k!}.$$. 
For each \( \lambda \), the function of \( k \) which shows up in this theorem gives us a new PMF:

**Definition**

Let \( \lambda > 0 \). A discrete RV \( Y \) is **Poisson with parameter** \( \lambda \) (or ‘\( \text{Poi}(\lambda) \)’) if its possible values are \( 0, 1, 2, \ldots \), and

\[
P\{Y = k\} = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for} \ k = 0, 1, 2, \ldots
\]
First checks: if $Y$ is Poi($\lambda$), then

$$\sum_{k=0}^{\infty} P\{ Y = k \} = 1 \quad \text{(as it should be)},$$

$$E[Y] = \lambda \quad (= \text{limit of } E[\text{binom}(n, \lambda/n)] \text{ as } n \to \infty)$$

and

$$\text{Var}(Y) = \lambda \quad (= \text{limit of } \text{Var}(\text{binom}(n, \lambda/n)) \text{ as } n \to \infty).$$

(See Ross pp137-8.)
The Poisson approximation is valid in many cases of interest. A Poisson distribution is a natural assumption for:

- The number of misprints on a page of a book.
- The number of people in a town who live to be 100.
- The number of wrong telephone numbers that are dialed in a day.
- The number of packets of dog biscuits sold by a pet store each day.
- …

In all these cases, the RV is really counting how many things actually happen from a very *large* number of independent trials, each with very *low* probability.
Example (Ross E.g. 4.7a)

Suppose the number of typos on a single page of a book has a Poisson distribution with parameter $1/2$. Find the probability that there is at least one error on a given page.

Example (Ross E.g. 4.7b)

The probability that a screw produced by a particular machine will be defective is 10%. Find/approximate the probability that in a packet of 10 there will be at most one defective screw.
The Poisson approximation is extremely important because of the following:

*To specify the PMF of a binom(n, p) RV, you need two parameters, but to specify the PMF of a Poi(λ), you need only one.*

Very often, we have an RV which we believe is binomial, but we don’t know exactly the number of trials (n) or the success probability (p). But as long as we have data which tells us the expectation (\(= np\)), we can still use the Poisson approximation!
Example (Ross 4.7c)

We have a large block of radioactive material. We have measured that on average it emits 3.2 $\alpha$-particles per second. Approximate the probability that at most two $\alpha$-particles are emitted during a given one-second interval.

NOTE: The $\alpha$-particles are being emitted by the radioactive atoms. There’s a huge number of atoms, and to a good approximation they all emit atoms within a given time-interval independently. But we don’t know how many atoms there are, nor the probability of emission by a given atom per second. We only know the average emission rate. So there’s not enough information to model this using a binomial, but there is enough for the Poisson approximation!
Poisson RVs frequently appear when modeling how many ‘event’s (beware: not our usual use of this word in this course!) of a certain kind occur during an interval of time.

Here, an ‘event’ could be:

- the emission of an $\alpha$-particle,
- an earthquake,
- a customer walking into a store.

These are all ‘event’s that occur at ‘random times’.
The precise setting for this use of Poisson RVs is the following. Events occur at random moments. Fix $\lambda > 0$, and suppose we know that:

- If $h$ is very short, then the probability of an ‘event’ during an interval of length $h$ is approximately $\lambda h$ (up to an error which is much smaller than $h$).
- If $h$ is very short, then the probability of two or more ‘event’s during an interval of length $h$ is negligibly small.
- For any time-intervals $T_1, \ldots, T_n$ which don’t overlap, the numbers of ‘event’s that occur in these time intervals are independent.

Then (see Ross, p144):

The number of events that occur in a time-interval of length $t$ is $\text{Poi}(\lambda t)$.
Example (Ross E.g. 4.7e)

Suppose that California has two (noticeable) earthquakes per week on average.

(a) Find the probability of at least three earthquakes in the next two weeks.

(b) Let $T$ be the RV which gives the time until the next earthquake. Find the CDF of $T$.

QUESTION: Is this really a good model for earthquakes?