Lecture 22: Oct 18

Last time

- Moment Generating Function

Today

- Internal “evaluation” open where you can make anonymous suggestions
- Common Discrete Distribution

Common Discrete Distribution

Why parametric models?

- *Parametric models* or *distribution families* have a specific form but can change according to a fixed number of parameters.
- The objective is to model a population. Parametric models are often appropriate in common situations with similar mechanisms.
- Parametric models have many known and useful properties and are easy to work with. When fitting a population, only a few parameters need to be estimated: *parametric inference*.
- Sometimes one does not want to make parametric assumptions and would rather work with non-parametric models. But non-parametric models can be infinite dimensional.
- In this course, we emphasize parametric models.

Discrete uniform X has the discrete unifrom$(1, N)$ distribution if X is equally likely to be one of $\{1, 2, \ldots, N\}$.

- Sample space: $\{1, 2, \ldots, N\}$
- pmf:

$$f_X(x) = \frac{1}{N}, \quad x = 1, 2, \ldots, N$$

- cdf:

$$F_X(x) = \Pr(X \leqslant x) = \begin{cases} 0 & x < 0 \\ \lfloor x \rfloor / N & 0 \leqslant x < N \\ 1 & N \leqslant x \end{cases}$$

- moments:

$$EX = \frac{N+1}{2}$$

Bernoulli Distribution Consider an experiment where outcomes are binary (say, Success or Failure) and the probability of success is p. Define the following random variable

$$Y = \begin{cases} 1 & \text{outcome is success} \\ 0 & \text{outcome is failure} \end{cases}$$

Then, Y has a Bernoulli Distribution.

- Sample space: $\{0, 1\}$.
- pmf: $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. We can write this as:

$$f(y) = \Pr(Y = y) = \begin{cases} p^y(1-p)^{1-y} & y = 0, 1 \\ 0 & othersie \end{cases}$$

- what are the cdf, mean and variance?

Binomial Distribution A $Binomial(n, p)$ random variable X is defined as the number of successes in n i.i.d. (independent, identically distributed) Bernoulli trials, each with probability p of success:

$$X = \sum_{i=1}^{n} Y_i, \quad Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} Bernoulli(p)$$

- Sample space: $\{0, 1, \ldots, n\}$
- pmf:

$$f_X(s) = \begin{cases} \binom{n}{s} p^x (1-p)^{n-s} & s = 0, 1, \ldots, n \\ 0 & otherwise \end{cases}$$

- cdf:

$$F_X(x) = \sum_{s=0}^{x} \binom{n}{s} p^s (1-p)^{n-s} \quad \text{(no closed form)}$$

Poisson Distribution The Poisson distribution was derived by the French mathematician Poisson in 1837 as a limiting version of the binomial distribution. The Poisson distribution is often used to model the number of occurrences in a given time interval. One of the basic assumptions on which the Poisson distribution is built is that, for small time intervals, the probability of an arrival is proportional to the length of waiting time. This makes it a reasonable model for situations such as waiting for a bus, waiting for customers to arrive in a bank.

The Poisson distribution has a single parameter λ, sometimes called the intensity parameter. A Poisson random variable X, takes values in the nonnegative integers with pmf

$$\Pr(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

To see that $\sum_{x=0}^{\infty} P(X = x|\lambda) = 1$, recall the Taylor series expansion of $e^{\lambda} = \sum\limits_{i=0}^{\infty} \frac{\lambda^i}{i!}$. Thus

$$\sum_{x=0}^{\infty} \Pr(X = x|\lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

What is the mean and variance of X?

$$\begin{aligned} EX &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} \\ &= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \\ &= \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} \\ &= \lambda \end{aligned}$$

Similarly

$$\begin{aligned} EX^2 &= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{(x-1)!} \\ &= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} + \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-2)!} \\ &= \lambda + \lambda^2 \end{aligned}$$

So that

$$Var(X) = EX^2 - (EX)^2 = \lambda$$

- Sample space: $\{0, 1, \ldots\}$
- pmf: $\Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$
- cdf: $F_X(x) = \sum_{s=0}^{x} \frac{e^{-\lambda}\lambda^s}{s!}$

Negarive binomial vs. Poisson The negative binomial distribution is often good for modeling count data as an alternative to the Poisson. In the previous parameterization, define

$$\lambda = \frac{r(1-p)}{p} \iff p = \frac{r}{r+\lambda}$$

Then we have

$$EX = \lambda$$

$$Var(X) = \frac{\lambda}{p} = \lambda(1 + \frac{\lambda}{r}) = \lambda + \frac{\lambda^2}{r}$$

For the Poisson we had that the variance equals the mean.
For the negative binomial, the variance is equal to the mean plus a quadratic term. Thus the negative binomial can capture overdispersion in count data.
In the previous parameterization, the pmf becomes

$$\begin{aligned} f(y) &= \binom{r+y-1}{y} p^r q^y = \frac{(r+y-1)!}{y!(r-1)!} \left(\frac{r}{r+\lambda}\right)^r \left(\frac{\lambda}{r+\lambda}\right)^y \\ &= \frac{\lambda^y}{y!} \frac{r(r+1)\dots(r+y-1)}{(r+\lambda)^y} \left(1 + \frac{\lambda}{r}\right)^{-r} \end{aligned}$$

Letting $r \to \infty$, we get

$$f(x) \to \frac{\lambda^x}{x!} e^{-\lambda}$$

So for large r, the negative binomial can be approximated by a Poisson with parameter $\lambda = r(1-p)/p$.