Lecture 21: Oct 14

Last time

- Internal midterm evaluation open
- Moment generating function

Today

- Moment generating function
- Common Discrete Distributions (Chapter 3)

Theorem Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist.

1. If X and Y have **bounded support**, then $F_X(u) = F_Y(u)$ for all u if and only if $EX^r = EY^r$ for all integers $r = 0, 1, 2, \ldots$.
2. If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.

Theorem (Convergence of mgfs) Suppose $\{X_i, i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t), \quad \text{for all } t \text{ in a neighborhood of } 0,$$

and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x).$$

That is, *convergence*, for $|t| < h$, of mgfs to an mgf implies *convergence* of cdfs.

Poisson approximation One approximation that is usually taught in elementary statistics courses is that binomial probabilities can be approximated by Poisson probabilities. It is taught that the Poisson approximation is valid "when n is large and np is small", and rules of thumb are sometimes given.

The $Poisson(\lambda)$ pmf is given by

$$\Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

where λ is a positive constant. The approximation states that if $X \sim Binomial(n, p)$ and $Y \sim Poisson(\lambda)$, with $\lambda = np$, then

$$\Pr(X = x) \approx \Pr(Y = x)$$

for large n and small np. We now show that the mgf converge, lending credence to this approximation. Recall that

$$M_X(t) = [pe^t + (1-p)]^n.$$

For the $Poisson(\lambda)$ distribution, we can calculate (HW4, exercise 2.33)

$$M_Y(t) = e^{\lambda(e^t-1)},$$

and if we define $p = \lambda/n$, then $M_X(t) = [1 + (e^t - 1)\lambda/n]^n$ such that $M_X(t) \to M_Y(t)$ as $n \to \infty$.

Theorem For any constant a and b, the mgf of the random variable $aX + b$ is given by

$$M_{aX+b} = e^{bt} M_X(at).$$

Proof:

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

Hypergeometric Distribution Suppose a population of N entities is made up of two types: M of the first type and $N-M$ of the second type. Suppose we take a sample of size K. We wish to know X, the number in the sample of the first type. The probability mass function of X is given by:

$$f_X(x)=\frac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}}$$

for $x = \max(0, M-N+K), \dots, \min(M, K)$.

The sample space is defined so that all binomial coefficients are valid. We must have:

$$0 \leqslant x \leqslant K, \quad 0 \leqslant x \leqslant M, \quad 0 \leqslant K-x \leqslant N-M$$

Often $K < M$ and $K < N-M$ so the range becomes $0 \leqslant x \leqslant K$.

Hypergeometric vs Binomial   We can show that the limiting form of the hypergeometric pmf is the binomial pmf

$$\begin{aligned}
\Pr(s) &= \frac{\binom{M}{s}\binom{N-M}{n-s}}{\binom{N}{n}} \\
&= \frac{\frac{M!}{s!(M-s)!}\frac{(N-M)!}{(n-s)!(N-M-n+s)!}}{\frac{N!}{n!(N-n)!}} \\
&= \frac{\frac{n!}{s!(n-s)!}\frac{M!}{(M-s)!}\frac{(N-M)!}{(N-M-n+s)!}}{\frac{N!}{(N-n)!}}
\end{aligned}$$

Note

$$\begin{aligned}
\frac{M!}{(M-s)!} &= \frac{M(M-1)(M-2)\dots(M-s)!}{(M-s)!} \\
&= M^s\left[1(1-\frac{1}{M})\dots(1-\frac{s-1}{M})\right] \\
\frac{N!}{(N-n)!} &= N^n\left[1(1-\frac{1}{N})\dots(1-\frac{n-1}{N})\right] \\
\frac{(N-M)!}{[(N-M)-(n-s)]!} &= (N-M)^{n-s}\left[1(1-\frac{1}{N-M})\dots(1-\frac{n-s-1}{N-M})\right]
\end{aligned}$$

Letting $N \to \infty, M \to \infty, \frac{M}{N} \to p$, we have

$$\begin{aligned}
\Pr(s) &= \frac{\binom{M}{s}\binom{N-M}{n-s}}{\binom{N}{s}} \\
&\approx \binom{n}{s}\frac{M^s(N-M)^{n-s}}{N^n} \\
&= \binom{n}{s}\left(\frac{M}{N}\right)^s\left(1-\frac{M}{N}\right)^{n-s} \\
&\to \binom{n}{s}p^s(1-p)^{n-s}
\end{aligned}$$

In summary, we have

Hypergeometric	$\to$	Binomial	$\to$	Poisson
$N \to \infty$		$n \to \infty$		$\lambda = np$
$M \to \infty$		$p \to 0$		
$\frac{M}{N} \to p$		$np \to \lambda$		

Geometric Distribution Consider a series of iid Bernoulli Trials with p = probability of success in each trial. Define a random variable X representing the number of trials until first success. Note X includes the trial at which the success occurs (one parameterization). Then, X has a geometric distribution.

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

$$f(x) = \Pr(X = x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \dots \\ 0 & otherwise \end{cases}$$

- cdf:

$$F(x) = \Pr(X \leqslant x) = 1 - (1-p)^x$$

- Moments:

$$E(X) = 1/p$$
$$Var(X) = (1-p)/p^2$$

Memoryless property. Suppose $k > i$, then

$$\Pr(X > k | X > i) = \Pr(X > k - i)$$

Proof:

$$\begin{aligned} \Pr(X > k | X > i) &= \frac{\Pr(X > k)}{\Pr(X > i)} = \frac{(1-p)^k}{(1-p)^i} \\ &= (1-p)^{k-i} = \Pr(X > k - i) \end{aligned}$$

Example Suppose X is number of years you live, and X follows a geometric distribution, then

$$\begin{aligned} \Pr(\text{survive two more years}) &= \Pr(X > \text{current age} + 2 | X > \text{current age}) \\ &= \Pr(X > 2) \end{aligned}$$

This model is clearly too simple for human populations (since we do age).

Negative Binomial Distribution Still in the context of iid Bernoulli trials, define a random variable corresponding to the number of trials required to have s successes. We say $X \sim Negbin(s, p)$.

- Sample space: $\{s, (s+1), \dots\}$
- pmf: for $x = s, s+1, s+2, \dots$

$$\begin{aligned} f(x) &= \binom{x-1}{s-1} p^{s-1} q^{x-s} \cdot p \\ &= \binom{x-1}{s-1} p^s q^{x-s} \end{aligned}$$

- cdf: no closed form
- Expectation: $EX = s/p$.
- Variance: $Var(X) = s(1-p)/p^2$

Notes

- Why the name? See Casella & Berger p.95.
- $X \sim Negbin(1, p)$ is the same as $X \sim Geometric(p)$
- $Negbin(n, p)$ is the same as the sum of n $Geometric(p)$ random variables

Other parameterizations The negative binomial distribution is sometimes defined in terms of the random variable Y = number of failures before the rth success. Then

- Sample space: $\{0, 1, 2, \dots\}$
- pmf
$$f(y) = \binom{r+y-1}{y} p^r q^y, \quad y = 0, 1, 2, \dots$$
- cdf: no closed form
- Expectation: $EY = r(1-p)/p$
- Variance: $Var(Y) = r(1-p)/p^2$

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

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

Letting $s \to \infty$, we get

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

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