Lecture 23: Oct 18

Last time

- Common Discrete Distributions (Chapter 3)

Today

- Common Discrete Distributions (Chapter 3)
- Common Continuous Distributions

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

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:

$$
\begin{aligned}
E(X) &= 1/p \\
Var(X) &= (1-p)/p^2
\end{aligned}
$$

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
$$f(y) = \binom{r+y-1}{y} p^r q^y, \quad y = 0, 1, 2, \ldots$$
- 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

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

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 $\lambda = r(1-p)/p$.

Common continuous distributions

Uniform Distribution A random variable X having a pdf

$$f(x) = \begin{cases} 1 & \text{for } 0 < x \leqslant 1 \\ 0 & otherwise \end{cases}$$

is said to have a *uniform distribution* over the interval $(0, 1)$.

The cdf is:

$$F(y) = \int_{-\infty}^{y} f(x)dx = \begin{cases} 0 & \text{for } y \leqslant 0 \\ y & \text{for } 0 \leqslant y \leqslant 1 \\ 1 & \text{for } y > 1 \end{cases}$$

- Unifrom; $Y \sim U[a,b]$
- sample space: $[a,b]$
- pdf:

$$f(y) = \begin{cases} \frac{1}{b-a} & \text{for } a < y \leqslant b \\ 0 & otherwise \end{cases}$$

- cdf:

$$F(y) = \int_{-\infty}^{y} f(x)dx = \begin{cases} 0 & \text{for } y \leqslant a \\ \frac{y-a}{b-a} & \text{for } a \leqslant y \leqslant b \\ 1 & \text{for } y > b \end{cases}$$

- moments:

$$E(Y) = (a+b)/2$$
$$Var(Y) = \frac{(b-a)^2}{12}$$

Notes

- The uniform extends to the continuous case the idea of equally likely outcomes.
- If $Y \sim U[0,1]$, then $a + (b-a)Y \sim U[a,b]$

Exponential Distribution Denoted $X \sim Exp(\lambda)$:

- sample space: $x \geqslant 0$
- pdf:

$$f(x) = \begin{cases} \lambda e^{-\lambda y} & \text{for } y \geqslant 0 \\ 0 & otherwise \end{cases}$$

- cdf:

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geqslant 0 \\ 0 & \text{for } x < 0 \end{cases}$$

- moments:

$$E(X) = 1/\lambda$$
$$Var(X) = 1/\lambda^2$$
$$M_X(t) = \lambda/(\lambda - t), \quad t < \lambda$$

Interpretation The exponential can be derived as the waiting time between Poisson events. Suppose that the number of events in a unit interval of time follows a Poisson(λ) distribution. Then, let Y be the time until the first event.

$$\Pr(Y > t) = \Pr(0 \text{ events in } [0,t])$$

and the number of events in $[0,t]$ follows a Poisson distribution with parameter λt. Therefore,

$$\Pr(Y > t) = e^{-\lambda t}.$$

The cdf of Y is

$$F(t) = 1 - \Pr(Y > t) = 1 - e^{-\lambda t}$$

and hence the density is $f(t) = \lambda e^{-\lambda t}$.

Alternative parameterization Many books write the density as

$$f(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta} & \text{for } y \geqslant 0 \\ 0 & otherwise \end{cases}$$

so that $E(Y) = \theta$ and $Var(Y) = \theta^2$. In this case $\theta = 1/\lambda$ is called the *mean parameter*, while $\lambda = 1/\theta$ is called the *rate parameter*.

Memoryless property The exponential has a memoryless property, just like the geometric.

$$\Pr(Y > s + t | Y > t) = \Pr(Y > s)$$

Same interpretation as the geometric for continuous time:

- The probability of an event in a time interval depends only on the length of the interval, not the absolute time of the interval.
- The underlying Poisson process is stationary: the rate λ is constant. (In the geometric case, the probability, p of getting an event in every discrete time unit is constant).

Shifted exponential Let $X \sim Exp(\lambda)$ and $Y = X + v, v \in \mathbb{R}$. Then, Y has the *shifted exponential distribution* with pdf:

$$f(y) = \begin{cases} \lambda e^{-(y-v)\lambda} & \text{for } y \geqslant v \\ 0 & otherwise \end{cases}$$

Interpretation:

- $v > 0$: Event is delayed
- $v < 0$: The news of the event is delayed

Does the shifted expoenential maintain the memoryless property?

Double exponential The *double exponential distribution* is formed by reflecting an exponential distribution around zero. It has pdf:

$$f(x) = \frac{1}{2} \lambda e^{-\lambda |x|}, \quad x \in \mathbb{R}$$

Suppose X has the above distribution with $\lambda = 1$. Now let $Y = \sigma X + \mu, \mu \in \mathbb{R}$ (shifting) and $\sigma > 0$ (scaling). Then Y has the *Laplace distribution* with pdf:

$$f_Y(y) = \frac{1}{2\sigma} \exp\left(-\frac{|y - \mu|}{\sigma} \right)$$

with moments

$$EY = \mu, \quad Var(Y) = 2\sigma^2$$

The Laplace distribution provides an alternative to the normal for centered data with fatter tails but all finite moments.