

## Lecture 24: Oct 26

Last time

- Common Discrete Distributions (Chapter 3)

Today

- Negative binomial distribution
- Common Continuous Distributions

**Negative 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

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

Letting  $r \rightarrow \infty$ , we get

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

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

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

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

The cdf is:

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

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

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

- cdf:

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

- 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:  $x \geq 0$
- pdf:

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

- cdf:

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

- 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  $\text{Poisson}(\lambda)$  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  $\lambda t$ . 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

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

so that  $E(Y) = \theta$  and  $\text{Var}(Y) = \theta^2$ . 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  $\lambda$  is constant. (In the geometric case, the probability,  $p$  of getting an event in every discrete time unit is constant).

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

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

Interpretation:

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

Does the shifted exponential 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.