BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

`https://sakai.unc.edu/portal/site/bios660-bios672-3-credits`

Notes 11

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. Then, X has a geometric distribution.

sample space: $\{1, 2, ...\}$

pmf:

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

cdf:

$$F(x) = P\{(X \leq x)\} = 1 - (1-p)^x$$

Moments:

$$\text{E}(X) = 1/p$$
$$\text{Var}(X) = (1-p)/p^2$$

Memoryless property

Suppose $k > i$, then

$$P(X > k | X > i) = P(X > k - i)$$

Proof:

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

Example: Suppose X is # years you live

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

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

Note

Some texts use the geometric distribution to describe the distribution of the number of trials **before** the first success. In this case, the pdf changes to:

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

The moments also change. Be careful!

Example

In studies of human fertility, researchers are often interested in exposures that increase the time it takes for a couple to successfully conceive. If a woman has a $p = 25\%$ chance of conceiving in a particular menstrual cycle, what is the probability that she becomes pregnant on the 3rd cycle?

What is the expected number of cycles to pregnancy?

What is her probability of not becoming pregnant within 12 cycles? (this is the definition of clinical infertility)

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: $\{s, (s+1), ...\}$
pmf: for $x = s, s+1, s+2, \ldots,$

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

cdf: no closed form.
expectation: $\mathsf{E}(X) = s/p$
Variance: $\mathsf{Var}(X) = s(1-p)/p^2$

Notes

- Why the name? See C-B 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
- Note that the outcome $SFFSFSSFSS$ may have been generated from a $Binomial(10, p)$ or from a $NegBinom(6, p)$. Need to know the experimental design to compute probabilities.

Negative binomial sampling

Example: Suppose a proportion p of the population possess a certain characteristic (e.g., have a certain disease). How many people should we expect to sample in order to collect r people with that characteristic? (Suppose the population is big enough so that you can assume sampling with replacement)

How many should we sample in order to be 95% sure of getting r people with the characteristic?

Other parametrizations

As with the geometric, some books define a negative binomial random variable as the number of **failures before the s^{th} success**. This is equal to the previous definition minus s.

sample space: $\{0, 1, 2, \dots\}$

pmf:

$$f(x) = \binom{s+x-1}{x} p^s q^x, \qquad x = 0, 1, 2, \dots$$

cdf: no closed form.

expectation: $\mathsf{E}(X) = s(1-p)/p$

Variance: $\mathsf{Var}(X) = s(1-p)/p^2$

Negative binomial vs. Poisson

The negative binomial distribution is often good for modeling count data as an alternative to the Poisson.

In the previous parametrization, define

$$\lambda = \frac{s(1-p)}{p} \quad \Leftrightarrow \quad p = \frac{s}{s+\lambda}$$

Then we have

$$\mathsf{E}X = \lambda$$

$$\mathsf{Var}X = \frac{\lambda}{p} = \lambda\left(1+\frac{\lambda}{s}\right) = \lambda + \frac{\lambda^2}{s}$$

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.

cont.

In the previous parametrization, the pmf becomes

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

Letting $s \to \infty$ we get that

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

So for large s, the negative binomial can be approximated by a Poisson with parameter $\lambda = s(1-p)/p$.
We can see that also from convergence of the moments (Homework).

Uniform Distribution

A random variable X having a *pdf*

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

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

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

Uniform Distribution (cont.)

Uniform: $Y \sim U[a, b]$:

sample space $[a, b]$
pdf:

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

cdf:

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

moments

$$\begin{aligned} \mathrm{E}(Y) &= (a+b)/2 \\ \mathrm{Var}(Y) &= \frac{(b-a)^2}{12} \end{aligned}$$

Notes

- The uniform extends to the continuous case the idea of equally likely outcomes.
- If $Y \sim U[0,1]$, then $a + (b-a)Y \sim U[a,b]$.
- Useful for settings where individuals arrive at a destination, enter a study, etc. randomly over time.

Example: Suppose that during rush hour buses leave every 10 minutes, starting at 7:00am. Suppose you arrive at the bus stop according to a uniform distribution between 7am and 7:30am. What is the probability that you have to wait more than 5 minutes? What about 2 minutes?

Exponential Distribution 17 / 22

Exponential Distribution

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

sample space: $y \geq 0$
pdf:

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

cdf:

$$F(y) = \int_0^y \lambda e^{-\lambda y} dx = \begin{cases} 1 - e^{-\lambda y} & \text{for } \ y \geq 0 \\ 0 & \text{for } \ y < 0 \end{cases}$$

moments:

$$\begin{aligned} \text{E}(Y) &= 1/\lambda \\ \text{Var}(Y) &= 1/\lambda^2 \\ M_X(t) &= \lambda/(\lambda - t) \end{aligned}$$

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.

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

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

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

The *cdf* corresponding to Y is

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

and hence the density is

$$f(t) = \lambda e^{-\lambda t}$$

Notes

- Many books write the density as

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

  so that E$(Y) = \theta$ and Var$(Y) = \theta^2$

- The exponential has a *memorylessness property*, just like the geometric.

$$P(Y > s + t | Y > t) = P(Y > s)$$

  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, the prob. p of getting an event in every discrete time unit is constant.)

Shifted exponential

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

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

Double Exponential

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

$$f(x) = 0.5\lambda e^{-\lambda|x|}, \qquad x \in \mathbb{R}$$

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 *double exponential distribution* with pdf:

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

and moments

$$\text{E}Y = \mu, \qquad \text{Var}Y = 2\sigma^2$$

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