Chapter 3 Part 2

04/27/2021

Remainder of this Chapter

- This chapter is designed so as a good reference for later chapters.
- We may or may not need all of these distributions
- As such, we will define most of them but not all

Geometric Random Variable

- Suppose we are to perform independent, identical Bernoulli trials until the first success.
- If we wish to model Y, the number of failures before the first success, we can consider the following pmf:

$$P(Y = y) = (1 - p)^y p \quad \text{for} \quad y = 0, 1, \ldots, \infty. \tag{1}$$

Geometric Random Variable

- We can think about this function as modeling the probability of y failures, then 1 success.
- In this case, Y follows a **geometric distribution** with $\mathrm{E}(Y) = \frac{1-p}{p}$ and $\mathrm{SD}(Y) = \sqrt{\frac{1-p}{p^2}}$.

Geometric distributions with $p = 0.3,\ 0.5$ and 0.7.

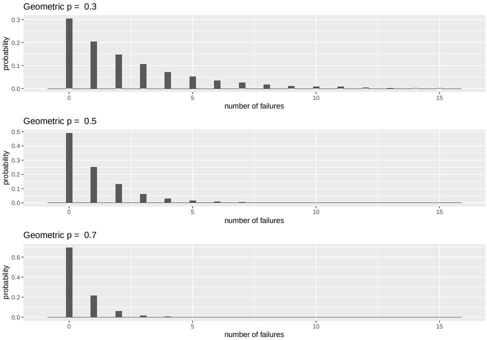

Figure 1: Geometric Plots

- Notice that as p increases, the range of plausible values decreases and means shift towards 0.

Geometric Distribution

- Video Example, click Click HERE
- The function `dgeom(y, p)` will output the probability of y failures before the first success where $Y \sim \mathrm{Geometric}(p)$.

Negative Binomial Random Variable

- What if we were to carry out multiple independent and identical Bernoulli trails until the r^{th} success occurs?
- If we model Y, the number of failures before the r^{th} success, then Y follows a **negative binomial distribution** where

$$P(Y = y) = \binom{y + r - 1}{r - 1}(1 - p)^{y}(p)^{r} \quad \text{for} \quad y = 0, 1, \ldots, \infty. \tag{2}$$

Negative Binomial Random Variable

- If $Y \sim \text{Negative Binomial}(r, p)$ then $\mathsf{E}(Y) = \frac{r(1-p)}{p}$ and $\mathsf{SD}(Y) = \sqrt{\frac{r(1-p)}{p^2}}$.

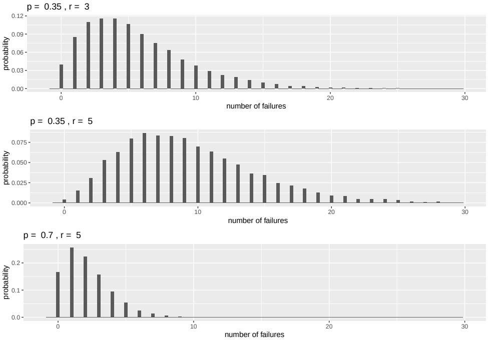

Figure 2: Negative Binomial

- Notice how centers shift right as r increases, and left as p

Negative Binomial Random Variable

- Note that if we set $r = 1$, then

$$P(Y = y) = \binom{y}{0}(1-p)^y p$$
$$= (1-p)^y p \quad \text{for} \quad y = 0, 1, \ldots, \infty,$$

which is the probability mass function of a geometric random variable! - Thus, a geometric random variable is, in fact, a special case of a negative binomial random variable.

- While negative binomial random variables typically are expressed as above using binomial coefficients (expressions such as $\binom{x}{y}$), we can generalize our definition to allow non-integer values of r.
- R function `dnbinom(y, r, p)` for the probability of y failures before the r^{th} success given probability p.

Negative Binomial Video

Click HERE

Hypergeometric Random Variable

- In all previous random variables, we considered a Bernoulli process, where the probability of a success remained constant across all trials.
- What if this probability is dynamic?
- The **hypergeometric random variable** helps us address some of these situations. Specifically, what if we wanted to select n items *without replacement* from a collection of N objects, m of which are considered successes?
- In that case, the probability of selecting a "success" depends on the previous selections. If we model Y, the number of successes after n selections, Y follows a **hypergeometric distribution** where

$$P(Y = y) = \frac{\binom{m}{y}\binom{N-m}{n-y}}{\binom{N}{n}} \quad \text{for} \quad y = 0, 1, \ldots, \min(m, n). \tag{3}$$

Hypergeometric Random Variable

- If Y follows a hypergeometric distribution and we define $p = m/N$, then $\mathrm{E}(Y) = np$ and $\mathrm{SD}(Y) = \sqrt{np(1-p)\frac{N-n}{N-1}}$.

Several hypergeometric distributions

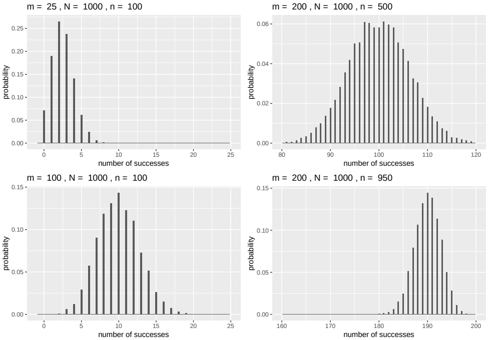

Figure 3: Hypergeometric

E(X)=n*p
p = m/N
E(X) = n*m/N

- On the left, N and n are held constant. As $m \to N/2$, the distribution becomes more and more symmetric.
- On the right, m and N are held constant. Both distributions are displayed on the same scale. We can see that as $n \to N$

Hypergeometric Random Variable in R

- If we wish to calculate probabilities through R, `dhyper(y, m, N-m, n)` gives $P(Y = y)$ given n draws without replacement from m successes and $N - m$ failures.

Hypergeometric Video

Click HERE

Poisson Random Variable

- Sometimes, random variables are based on a **Poisson process**.
- In a Poisson process, we are counting the number of events per unit of time or space and the number of events depends only on the length or size of the interval.
- We can then model Y, the number of events in one of these sections with the **Poisson distribution**, where

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad \text{for} \quad y = 0, 1, \dots, \infty, \tag{4}$$

where λ is the mean or expected count in the unit of time or space of interest. - This probability mass function has $\mathrm{E}(Y) = \lambda$ and $\mathrm{SD}(Y) = \sqrt{\lambda}$.

Poisson Distribution Graphs

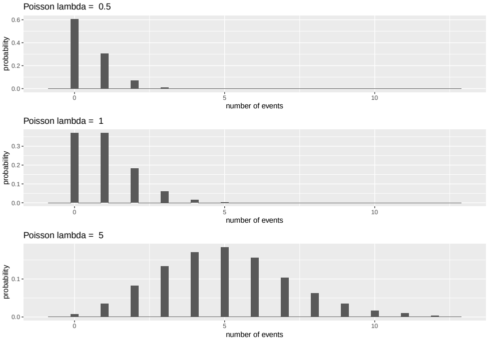

Figure 4: (ref:multPois)

- Notice how distributions become more symmetric as λ increases.
- If we wish to use R, `dpois(y, lambda)` outputs the probability of y events given λ.

Poisson Video Example

Click HERE

Continuous Random Variables

- A continuous random variable can take on an uncountably infinite number of values.
- With continuous random variables, we define probabilities using **probability density functions** (pdfs).
- Probabilities are calculated by computing the area under the density curve over the interval of interest. So, given a pdf, $f(y)$, we can compute

$$P(a \leq Y \leq b) = \int_a^b f(y)dy.$$

Continuous Random Variables

This hints at a few properties of continuous random variables:

- $\int_{-\infty}^{\infty} f(y)dy = 1$.
- For any value y, $P(Y = y) = \int_{y}^{y} f(y)dy = 0$.
- Because of the above property, $P(y < Y) = P(y \leq Y)$. We will typically use the first notation rather than the second, but both are equally valid.

Exponential Random Variable

- Suppose we have a Poisson process with rate λ, and we wish to model the wait time Y until the first event.
- We could model Y using an **exponential distribution**, where

$$f(y) = \lambda e^{-\lambda y} \quad \text{for} \quad y > 0, \tag{5}$$

- $E(Y) = 1/\lambda$ and $SD(Y) = 1/\lambda$.

Exponential Distribution

Exponential distributions with $\lambda = 0.5, 1$, and 5.

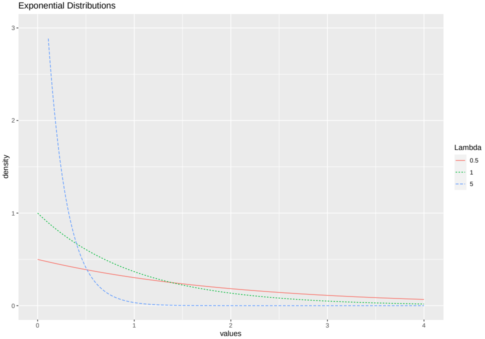

Figure 5: Exponential Distribution

- As λ increases, E(Y) tends towards 0, and distributions "die off" quicker.

Exponential Distribution

- To use R, `pexp(y, lambda)` outputs the probability $P(Y < y)$ given λ.

Gamma Random Variable

- Once again consider a Poisson process.
- When discussing exponential random variables, we modeled the wait time before one event occurred.
- If Y represents the wait time before r events occur in a Poisson process with rate λ, Y follows a **gamma distribution** where

$$f(y) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y} \quad \text{for} \quad y > 0. \tag{6}$$

- If $Y \sim \mathrm{Gamma}(r, \lambda)$ then $\mathsf{E}(Y) = r/\lambda$ and $\mathsf{SD}(Y) = \sqrt{r/\lambda^2}$.
- Note: $\Gamma(r) = (r-1)!$ (There is more to it)

Gamma Distribution

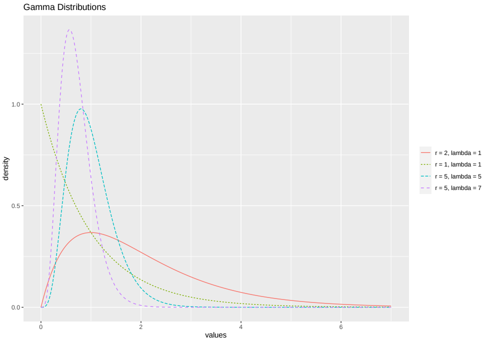

Figure 6: (ref:multGamma)

- Observe that means increase as r increases, but decrease as λ increases.

Gamma vs Others

- Note that if we let $r = 1$, we have the following pdf,

$$
\begin{aligned}
f(y) &= \frac{\lambda}{\Gamma(1)} y^{1-1} e^{-\lambda y} \\
&= \lambda e^{-\lambda y} \quad \text{for} \quad y > 0,
\end{aligned}
$$

an exponential distribution. - Just as how the geometric distribution was a special case of the negative binomial, exponential distributions are in fact a special case of gamma distributions!

- Just like negative binomial, the pdf of a gamma distribution is defined for all real, non-negative r.
- In R, `pgamma(y, r, lambda)` outputs the probability $P(Y < y)$ given r and λ.

Beta Random Variable

- So far, all of our continuous variables have had no upper bound.
- If we want to limit our possible values to a smaller interval, we may turn to a **beta random variable**.
- In fact, we often use beta random variables to model distributions of probabilities—bounded below by 0 and above by 1.
- The pdf is parameterized by two values, α and β ($\alpha, \beta > 0$). We can describe a beta random variable by the following pdf:

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \quad \text{for} \quad 0 < y < 1. \tag{7}$$

- If $Y \sim \mathrm{Beta}(\alpha, \beta)$, then $\mathrm{E}(Y) = \alpha/(\alpha+\beta)$ and $\mathrm{SD}(Y) = \sqrt{\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$.

Beta Distribution

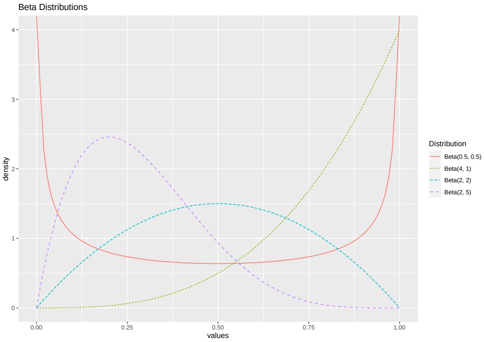

Figure 7: Beta Distribtion

- Note that when $\alpha = \beta$, distributions are symmetric.
- The distribution is left-skewed when $\alpha > \beta$ and right-skewed when $\beta > \alpha$.

Beta Distribution

- If $\alpha = \beta = 1$, then

$$f(y) = \frac{\Gamma(1)}{\Gamma(1)\Gamma(1)} y^0 (1-y)^0$$
$$= 1 \quad \text{for} \quad 0 < y < 1.$$

- This distribution is referred to as a **uniform distribution**.
- In R, `pbeta(y, alpha, beta)` yields $P(Y < y)$ assuming $Y \sim \mathrm{Beta}(\alpha, \beta)$.

Distributions Used in Testing

- We have spent most of this chapter discussing probability distributions that may come in handy when modeling.
- The following distributions, while rarely used in modeling, prove useful in hypothesis testing as certain commonly used test statistics follow these distributions.
- χ^2 distribution (requires a degree of freedom)
- Student t distribution
- F distribution (need 2 different degrees of freedom)
- Since we have used these In the past, we will leave their definitions to be referenced if needed

Table!

- Would not fit on a slide

Click HERE