More Discrete Distributions

Keegan Korthauer
Department of Statistics
UW Madison

COMMON DISCRETE DISTRIBUTIONS

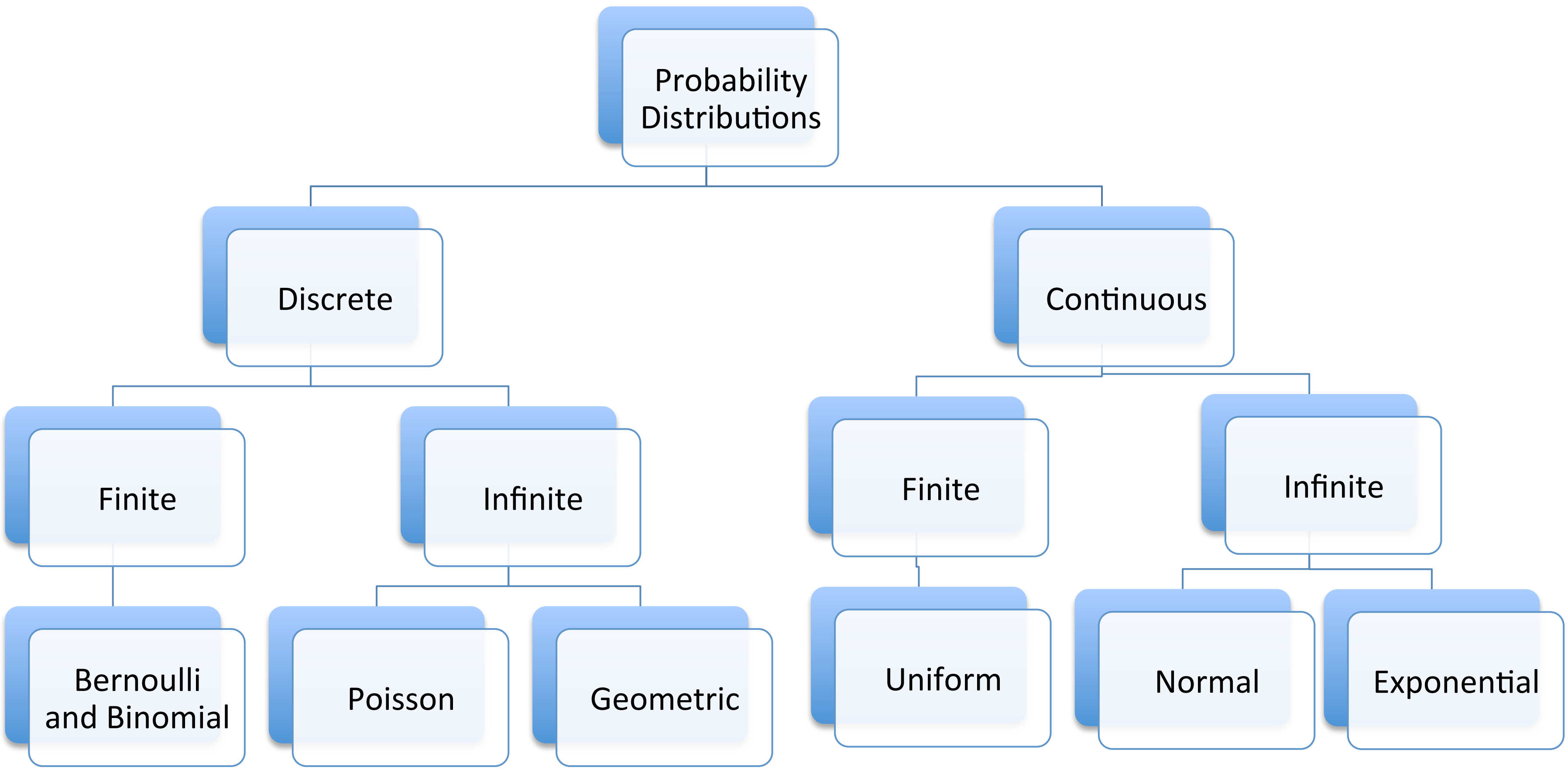
Bernoulli

Binomial

Poisson

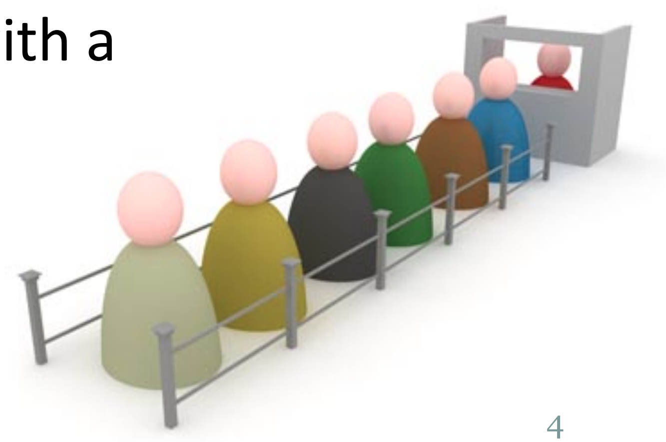
Geometric

Some Common Distributions



Another Discrete Distribution

- In binomial experiments we are interested in how many trials will be 'successes' (can range from 0 to n)
- What if we are interested in the number of occurrences of a certain event within a certain time period (can range from 0 to infinity)?
- Example: the number of arrivals in a queue within a 30 minute period
- We can represent this random variable with a **Poisson** distribution
- The Poisson distribution is characterized by a rate parameter λ



X ~ Poisson(λ)

- Probability mass function

$$P(X = x) = \begin{cases} e^{-\lambda} \dfrac{\lambda^x}{x!} & \text{if x is a non-negative integer} \\ 0 & \text{otherwise} \end{cases}$$

- Mean and variance

$$\mu_X = \lambda$$

$$\sigma_X^2 = \lambda$$

Poisson Distribution – Illustration

- http://www.math.uah.edu/stat/applets/SpecialCalculator.html (choose 'Poisson' from the dropdown menu)
- Shape of the distribution (PMF) is determined by the rate parameter λ
  - Note that the range of x is all non-negative integers, even though the demo only shows a finite range of x-values

Example - Call Center

- Suppose the number of calls to a credit card company calling center per minute (X) follows a Poisson distribution with rate parameter 2
- Find the mean and standard deviation of X

mean = 2, standard deviation = sqrt(2)

- Find P(X > 2)

= 0.3233

Poisson Approximation to the Binomial

- When n is very large and p is small, the Poisson distribution approximates the binomial distribution with $\lambda = np$
- Example: Let X be a binomial random variable with n=10,000 and p=0.0001

$$P(X=3)=\binom{10{,}000}{3}0.0001^{3}(0.9999)^{9997}=0.06131$$

Let Y be a Poisson RV with $\lambda = np = 1$

$$P(Y=3)=e^{-1}\frac{1^{3}}{3!}=0.06131$$

Poisson Approximation to the Binomial

- Mean of binomial is np, so it makes sense that the mean of Poisson is λ
- Variance of binomial is $np(1-p) \approx np$ for very small p, so it makes sense that the variance of Poisson is also λ

- **Rule of thumb: when $n \geq 100$ and $np \leq 10$, the Poisson(np) distribution will closely approximate the binomial(n,p) distribution**

Example – Exercise 4.3.5

A sensor network consists of a large number of microprocessors spread out over an area, connected to each other and to a base station.

In a certain network the probability that a message will fail to reach the base station is 0.005 and there are 1000 messages sent per day.

- What is the probability that exactly 3 messages fail to reach the base station?

  $P(X=3) = 0.14037$

- What is the probability that fewer than 994 messages reach the base station?

  $P(X>6) = 0.2378$

When λ is Unknown

- So far, we've dealt with examples where we have knowledge of the value of the rate parameter λ
- In practice experiments are performed to estimate a rate λ that represents the mean number of events that occur in a time period (or distance of some sort)
- How can we estimate it?
  - Count the number of events X that occur in *t* units of time/space
  - Then X ~ Poisson(λt)
  - Our best guess for λ:

$$\hat{\lambda} = \frac{X}{t}$$

How 'good' is our estimate?

- In other words, how much bias and uncertainty is there in our estimate?

- Bias = $\mu_{\hat{\lambda}} - \lambda = E(X/t) - \lambda = \dfrac{\lambda t}{t} - \lambda = 0$

- Uncertainty =

In practice, substitute $\hat{\lambda}$ for λ

$$\sigma_{\hat{\lambda}} = \sigma_{X/t} = \frac{\sigma_X}{t} = \frac{\sqrt{\lambda t}}{t} = \sqrt{\frac{\lambda}{t}}$$

Example 4.25 – Radioactive Decay

- A certain mass of a radioactive substance emits alpha particles at a mean rate of λ particles per second
- A physicist counts 1594 emissions in 100 seconds
- Estimate λ and find the uncertainty in the estimate

$$\hat{\lambda} = X / t = 1594 / 100 = 15.94$$

$$\sigma_{\hat{\lambda}} = \sqrt{\lambda / t} \overset{\text{plug in } \hat{\lambda}}{=} \sqrt{15.94 / 100} = 0.40$$

Geometric Distribution

- Going back to a series of independent Bernoulli trials, each with the same probability of success *p*

- Suppose we are interested in the random variable X which represents how many trials we have to perform until we have our first success
  - Example: I will roll a die until it comes up as a 6. How many rolls will there be?

- Then X follows the **geometric distribution** with parameter *p*

X ~ Geometric(p)

- The probability mass function of X is:

$$P(X = x) = \begin{cases} p(1-p)^{x-1} & \text{if x is a positive integer} \\ 0 & \text{otherwise} \end{cases}$$

- The mean and variance of X are:

$$\mu_X = \frac{1}{p}$$

$$\sigma_X^2 = \frac{1-p}{p^2}$$

Geometric Distribution – Illustration

- http://keisan.casio.com/exec/system/1245037158
- Shape of the distribution (PMF) is determined by the probability of success parameter p
  - Note that the range of x is all positive integers, even though the demo only shows a finite range of x-values
  - As you decrease p, the distribution becomes more skewed to the right

Example – Carnival Game

- We are going to play the 'duck pond' game at a carnival
- To play we get to pick one duck at random and get a prize if the duck has a star underneath
- Suppose that 5% of the ducks have a star and that there are enough ducks to assume independence of plays

1. On average, how many times do we have to play to win a prize?

   $X \sim Geom(0.05)$

   $E(X) = 20$

2. What is the probability that we'll win with only 2 plays?

   $P(X \leq 2) = 0.0975$

Distinguishing Discrete Distributions

- We've studied four different discrete distributions:
  - **Bernoulli** – one trial with two possible outcomes
  - **binomial** – a series of n independent Bernoulli trials
  - **Poisson** – events occurring in a fixed interval of time or space
  - **geometric** – a series of independent Bernoulli trials until a success
- Be able to determine which distribution(s) would apply to the following discrete RVs:
  - Number of hits to a website per hour
  - Whether it will snow tomorrow or not
  - The number of snow days 2014
  - The number of days until the next snowfall
  - The number of wins after playing 25 games of 'duck pond'

Next

- Note that we are only covering pgs 233-234 of section 4.4 (the subsection 'The Geometric Distribution')
  - skim over the hypergeometric, negative binomial, and multinomial distributions

- Normal and other common continuous distributions

- Central Limit Theorem (Normal approximation)