STAT 311: LECTURE 13

Heavily based on lecture notes from Martina Morris

Random Variables

Logistics

- Homework due today
- Midterms to be passed back Friday
- Shiqing will be giving lectures on Monday and Wednesday
- Sam’s office hours will be on Friday from 1-3

Random Variables

From probabilities of specific events

To describing the full probability distribution

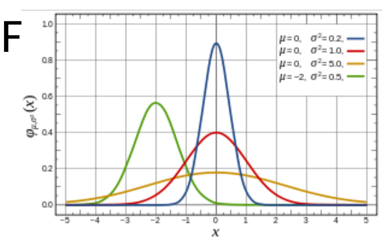
What is a random variable?

<u>*Formally*</u>:

- An event in a sample space

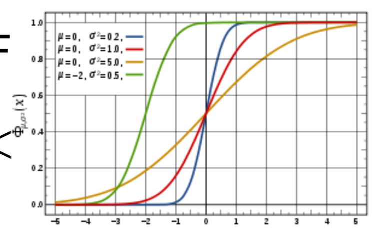
- That takes a value
  - Discrete or
  - Continuous

PDF



- With a certain probability
  - Can be described by the PDF: $\mathrm{P(X{=}k)}$
  - Or by the CDF: $\mathrm{P(X \leq k), P(X > k), P(j < X <}$

CDF



Examples

- Flip a coin: Heads or tails?
- How many cars will drive through an intersection in an hour?
- What is the height of a random individual?
- How much time until I get my next text message?

Random variable notation

- Random variables are denoted by capital letters
  - For example, X or Y

- The value the RV takes in a specific case is called a “realization” and denoted by a lowercase letter
  - For example: x, y *or* k

- $\mathrm{P}(\mathrm{X}=\mathrm{k})$
  - “The probability that the random variable X takes the value k”

- The sample space (set of all possible outcomes) is denoted by $\mathcal{X}$

Empirical vs Theoretical Distributions

- We have seen distributions of observed data. These are often called **Empirical Distributions**, because they are what we empirically observe
- Today we will begin to formally discuss distributions which are mathematical constructs used to model real world situations
- These distributions are typically a family of distributions which are specified by a mathematical equation and are governed by a set of **parameters**

Notation: statistics vs. parameters

Sample Statistics

- Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Variance

$$s_X^2 = \frac{1}{(n-1)}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

- Each observation has equal weight (1/n)

Theoretical parameters

- Expected value

$$\mu_x = \sum_{x_i \in \mathcal{X}} p_i x_i$$

- Variance

$$\sigma_x^2 = \sum_{x_i \in \mathcal{X}} p_i (x_i - \mu_x)^2$$

- Each possible outcome in the sample space receives its own weight (p_i)

Example

- If I roll a single fair dice

$$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$$

$$\begin{aligned}
\mu_x &= \sum_{x_i \in \mathcal{X}} x_i p_i \\
&= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) \\
&= 21/6 = 3.5
\end{aligned}$$

$$\begin{aligned}
\sigma_x^2 &= \sum_{x_i \in \mathcal{X}} (x_i - \mu_x)^2 p_i \\
&= (1 - 3.5)^2(1/6) + (2 - 3.5)^2(1/6) \ldots (6 - 3.5)^2(1/6) \quad = 2.92
\end{aligned}$$

What comes next

- Deriving expectations and variances for different distributions
  - Discrete (Bernoulli, Binomial, Poisson)
  - Continuous (Normal, Uniform, Exponential)

- For example: with coin tosses
  - Each toss is a random variable with outcomes {0, 1}
  - The sum of these outcomes over *n* tosses is *a linear combination of the random variables* for each toss, with outcome space {0, 1, 2, … , n}

- We start with the rules of expectations and variances for linear transformations and combinations of RVs

Rules of expectations and variances

For linear transformations and combinations of random variables

1. Variance-Mean relationship

$$\sigma_x^2 = \mathrm{var}(x) = \sum_i p_i (x_i - \mu_x)^2 \quad = \sum_{x_i \in \mathcal{X}} (x_i^2 - 2x_i\mu_x + \mu_x^2) p_i$$

$$= \sum_{x_i \in \mathcal{X}} (x_i^2 p_i - 2x_i\mu_x p_i + \mu_x^2 p_i)$$

$$= \sum_{x_i \in \mathcal{X}} x_i^2 p_i - \sum_{x_i \in \mathcal{X}} 2x_i\mu_x p_i + \sum_{x_i \in \mathcal{X}} \mu_x^2 p_i)$$

$$= \sum_{x_i \in \mathcal{X}} x_i^2 p_i - \mu_x \sum_{x_i \in \mathcal{X}} 2x_i p_i + \mu_x^2 \sum_{x_i \in \mathcal{X}} p_i)$$

$$= E(X^2) - 2\left(E(X)\right)^2 + \left(E(X)\right)^2$$

$$= E(X^2) - \left(E(X)\right)^2$$

Note the theoretical or population variance formula here, not the sample variance.

Transformations and combinations of RVs

- Examples:
  - *Transformation*: converting degrees F to degrees C
  - *Combination*: adding your midterm and final exam scores

- Linear transformations and combinations have simple expressions for their expected values and variances
  - Linear transformations : $\mathrm{Y} = \mathrm{a} + \mathrm{X}, \quad \mathrm{Y}{=}\mathrm{bX}, \quad \mathrm{Y} = \mathrm{a} + \mathrm{bX},$
  - Linear combinations: $\mathrm{Z} = \mathrm{X} + \mathrm{Y}, \quad \mathrm{Z} = \mathrm{aX} + \mathrm{bY}$

2. Linear Transformations of RVs

- If a and b are constants, and X is a random variable, then:

$$E(a) = a \qquad Var(a) = 0$$

$$E(bX) = b\,E(X) \qquad Var(bX) = b^2\,Var(X)$$

$$E(a + bX) = a + b\,E(X) \qquad Var(a + bX) = b^2\,Var(X)$$

- These are easily proven if you start with the definitions on slide 8 and work through the algebra *(try it…)*.

3. Linear combinations of independent RVs

UH covers this:

Let a, b and c be constants, and X and Y be independent random variables

$$E(X+Y)=E(X)+E(Y)$$

$$E(X-Y)=E(X)-E(Y)$$

$$E(a+bX+cY)=a+bE(X)+cE(Y)$$

$$Var(X+Y)=Var(X)+Var(Y)$$

$$Var(X-Y)=Var(X)+Var(Y)$$

$$Var(a+bX+cY)=b^2Var(X)+c^2Var(Y)$$

Again, these are easily proven

4. Linear combinations of dependent RVs

- Not covered in UH, but straightforward

Let a and b be constants, and X and Y be random variables

$$E(X+Y)=E(X)+E(Y)$$

No difference in the mean

$$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$$
$$Var(X-Y)=Var(X)+Var(Y)-2Cov(X,Y)$$
$$Var(aX+bY)=a^2Var(X)+b^2Var(Y)+2abCov(X,Y)$$

But the variance changes

Why the difference for correlated RVs?

- Suppose 10 individuals are deciding whether or not to show up to a party. Each has a 50/50 chance of going

- If the individuals all decide independently, we tend to end up with very few extreme events (ie 0 show up or all 10 show up) and usually will have around 5 individuals

- If the individuals all text each other and decide to either all show up or all not show up (attendance for each individual is now dependent), we will always have either 0 or 10 individuals. The average is still 5, but the outcomes are more extreme

Summary

- There are simple rules for expectations and variances
- When we transform or combine random variables
  - As long as the transformation/combination is linear
- And these are the foundation for what comes next
  - Derive expected values and variances for some common distributions

19

Discrete random variables

Exploring the derivation of discrete probability distributions and their properties:

Bernoulli, Binomial and Poisson

The goal

- To define a theoretical **Probability Density Function** for a random variable

“Probability that the RV $X=k$, given p”

“The RV X is distributed as f with parameter p”

Where p is one or more *parameters* that determine the outcome of the random variable.

- And use the PDF to derive expected values, variances, and probabilities

With discrete distributions, $f(x)$ is technically called a “*probability mass function*” or *PMF*, but PDF is also used, and we will use it here.

Example

- Coin tosses:
  - Each individual toss is an RV with 2 outcomes
  - Let X be the random variable for each toss: $\mathcal{X} = \{H,T\}$

- Pass Stat 311:
  - Each individual student is an RV with 2 outcomes
  - Let X be the random variable for each toss: $\mathcal{X} = \{\text{Pass, No Pass}\}$

The Bernoulli distribution

- Let X be a random variable with two outcomes:
  $\mathcal{X} = \{0,1\}$ (we have to decide what is 1 and what is 0)

$X \sim$ Bernoulli (p)

$$f(x;p) = p^x(1-p)^{(1-x)}$$

$$f(1;p) = p^1(1-p)^{(1-1)} = p$$

$$f(0;p) = p^0(1-p)^{(1-0)} = 1-p$$

Derivation of $E(Y)$ and $Var(Y)$

X ~ Bernoulli (p)

$$\begin{aligned}\mu_x &= \sum_{x_i \in \mathcal{X}} x_i p_i \\ &= 1(p) + 0(1-p) = p\end{aligned}$$

$$\begin{aligned}\sigma_x^2 &= \sum_{x_i \in \mathcal{X}} (x_i - \mu_x)^2 p_i \\ &= (1-p)^2(p) + (0-p)^2(1-p) \\ &= (1-2p+p^2)(p) + p^2(1-p) \\ &= (p-2p^2+p^3) + p^2 - p^3 \\ &= p - p^2 = p(1-p)\end{aligned}$$

Getting more complicated

- If there are 50 students in a class, how many total students will show up to class.
  - Assume each student shows up with probability .8
  - Assume the attendance of each student is independent of other students
- Each student's attendance is a Bernoulli trial
- Want to know the sum of the Bernoulli trials
- How can we do this?

Repeated Bernoulli trials: *the Binomial*

- Define a general probability distribution for the sum of n **independent** Bernoulli trials

- Let $X = \{0, 1, 2, \ldots, n\}$ the <u>count</u> of the number of 1's.

$X \sim \text{Binomial}(n; p)$

Example: 3 Coin tosses with H=1

	TTT	HTT THT TTH	HHT HTH THH	HHH
Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

The Binomial distribution

- The Binomial describes the probability distribution of <u>*counts* of successful trials</u>

  - It is a linear combination (sum) of Bernoulli RVs

  - So both the number of trials, n, and the probability of each trial, p, outcome influence the result

  - And the outcome space is now {0, 1, 2, … , *n*}

Binomial probabilities

There are 3 elements to the calculation:

1. Define the probability of each individual outcome in the sample space.
2. Identify the number of outcomes in the set of interest (i.e., that satisfy the condition $X=k$).
3. Multiply the probability of the outcome by the number in the set

What is the probability of each outcome?

- Start with a single trial:
  - What is the probability of each outcome Y?

$$p^y(1-p)^{1-y}$$

Our friend the Bernoulli

- What about
  - What is the probability of each outcome

$$p^k(1-p)^{2-k}$$

- And in general: $\ldots \quad p^k(1-p)^{n-k}$

How many outcomes satisfy $X = k$

How many outcomes satisfy

- This is a counting problem

This is a counting problem

- How many ways to get k successes in n trials when order does not matter?

How many ways to get successes in trials when order does not matter?

- Out of our N trials, we need to choose k trials to be the successes

Out of our N trials, we need to choose k trials to be the successes

- Use the combination rule: $C_N^k = \binom{N}{k}$

Use the combination rule:

The "binomial coefficient"

The number of outcomes that satisfy the condition:

$$
\begin{array}{ccccc}
 & & \text{HTT} & \text{HHT} & \\
 & & \text{THT} & \text{HTH} & \\
 & \underline{\text{TTT}} & \underline{\text{TTH}} & \underline{\text{THH}} & \underline{\text{HHH}} \\
\binom{n}{k} = & \binom{3}{0}, & \binom{3}{1}, & \binom{3}{2}, & \binom{3}{3} \\
= & 1, & 3, & 3, & 1
\end{array}
$$

- is referred to as the binomial coefficient in this context

The Binomial distribution

Putting these all together

Let Y be a random variable with two outcomes, Y={0,1}

Let X be the number of successes in n trials, and p be the probability of success on each trial. Then:

$$X \sim Bin(n; p) \quad P(X = k) = \binom{n}{k} \underbrace{p^k (1 - p)^{n-k}}$$

Number of outcome combinations that have k successes

Probability of each n-trial outcome with k successes

Derivation of $E(X)$ and $Var(X)$

$X \sim Bin(n;\, p)$ the sum of n independent Bernoulli trials: $X = \sum_{i=1}^{n} Y_i$

$$
\begin{aligned}
E(X) &= \mu_X \\
&= E\left(\sum_{i=1}^{n} Y\right) \\
&= \sum_{i=1}^{n} E(Y)^{*} \\
&= n\mu_Y \\
&= np
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= \sigma_X^2 \\
&= Var\left(\sum_{i=1}^{n} Y\right) \\
&= n\sigma_Y^2 \quad * \\
&= np(1-p)
\end{aligned}
$$

* Using E(Z+W) = E(Z) + E(W)

* Using Var(Z+W) = Var(Z) + Var(W) for independent RVs Z and W

Binomial Summary

- For ***any*** repeated Bernoulli trial, the count of successes, X, has a Binomial distribution:

$$X \sim \text{Bin}\,(n; p)$$

$$f(x; n, p) = \binom{n}{x} p^x (1 - p)^{n - x}$$

$$\mu_X = np$$

$$\sigma_X^2 = np(1 - p)$$

We can calculate the mean, variance, and the probability of any value of X from just two values: n and p

Other models

- Suppose I know that on average 10 cars pass through a specific intersection near my house each hour
- The number of cars that pass through for a given hour is random
- It is discrete (we're only considering whole cars)
- No set number of trials, or maximum value that this rv can take
- How might we describe this process?

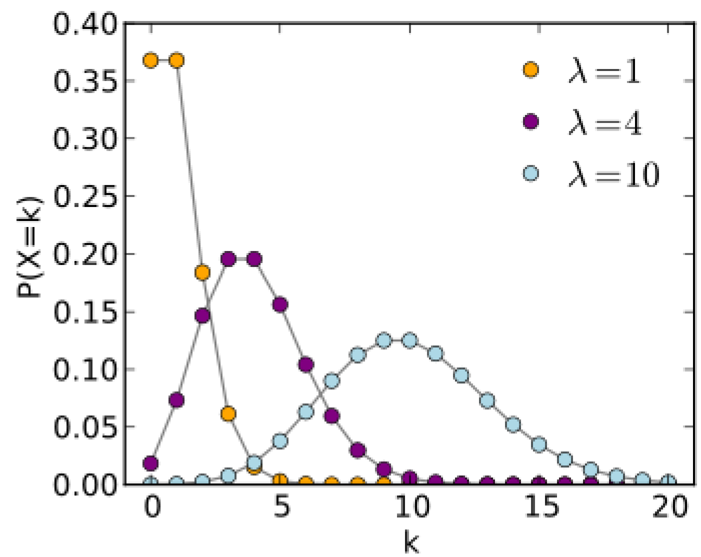
Poisson Distribution

- **Poisson** – Used for counts of events as rates
  - *n* (the number of trials) is large, *not fixed*,
  - λ approximates a rate of events (e.g., per time unit, or per capita)

for x ≥ 0

$$f(x;\lambda) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$\mu_X = \sigma_X^2 = \lambda$$



Poisson Distribution

- The numerator is always positive, so there is a positive probability for all $x \geq 0$, although it gets very small as x becomes large
- We assume that there is a constant rate
- We assume that all ``arrivals" are independent of each other

Example

- Suppose I know that on average 10 cars pass through a specific intersection near my house each day
  - What is λ?
  - What is the probability that 8 cars pass through the intersection in an hour?

$$f(8; \lambda = 10) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{10^8 e^{-10}}{8!} = .1125$$

  - What is the probability that 15 cars pass through the intersection in an hour?

$$f(15; \lambda = 10) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{10^{15} e^{-10}}{15!} = .0.0347$$

- Why might a Poisson assumption be wrong for this model?

Summary:

- A discrete random variable takes specific values
  - Typically integer counts
  - Each with a certain probability

- If you can represent the underlying stochastic process as a mathematical function
  - You can calculate almost anything you want for the RV
  - $\mathrm{E(X)}, \quad \mathrm{Var(X)}, \quad \mathrm{P(X{=}k)}, \quad \mathrm{P(X{\leq}k)}$ or $\quad \mathrm{P(i < X < j)}$

- The distribution is defined by the stochastic process
  - The details of the process matter, and are reflected in the formal definition of the distribution