Introduction to Probability and Inference

Alexandra Ding, HSSP Summer 2017

July 15, 2017

Please fill out the attendance sheet!
If you added this class in the last week: I will be sending out an email with last week's lecture notes.
Suggestions Box: Feedback and suggestions are important to the success of this class and my experience as a teacher, so please send comments to alexawding@gmail.com!

1 Lecture 1 Recap

- Randomness and Random Experiments

- Naive Definition of Probability

$$P(A) = \frac{\#\text{ favorable outcomes}}{\#\text{ unfavorable outcomes}}$$

- Non-Naive Probability: The Axioms of Probability
probability is a function that takes an event from a sample space and assigns it to a real number between 0 and 1. This function must satisfy the following rules:

  1. $P(A) \geq 0$
  2. $P(S) = 1$
  3. If $A_1, A_2...$ are disjoint (mutually exclusive) events, then

$$P(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$$

  Intuition: events that are more probable are "assigned" to a larger number. Probabilities of events in a sample space must sum up to 1, since the probability of being in the sample space must be 1 = certain. If two events are mutually exclusive, then the prob of either happening is the sum of the probs=

- Independence

- Conditional Probability and Bayes' Rule
Conditional Probability: probability of one event A occurring under the condition that we know the outcome of another event B. Compare Prior vs. Posterior Probability.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

  Bayes' Rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- PUZZLE: The Birthday Problem

2 Lecture 2

2.1 Warmup Puzzles

1. **More Disease Testing**
   Example taken from: Blitzstein & Huang: Introduction to Probability.

   Suppose you, an esteemed doctor, are testing for a disease that affects 1% of the population, and the test is advertised as "95% accurate", meaning that if we define T as "positive test", T^c as "negative test", D as "having disease" and D^c as "not having disease":

   $$P(T|D) = P(T^c|D^c) = 0.95$$

   Thomas Bayes, an esteemed statistician, walks into your office and gets tested. The test comes back positive. **What is the probability that he has the disease?**

2. **WARMUP PUZZLE: The Prosecutor's Fallacy**
   Example taken from: Blitzstein & Huang: Introduction to Probability. Also referenced on Wikipedia

   In 1998, Sally Clark was accused of killing her first child at 11 weeks of age and then her second child at 8 weeks of age. During her court trial, the expert witness testified that the probability of two children in the same family dying from SIDS (i.e. spontaneous natural causes) is about 1 in 73 million.

   The prosecutor says: "Since the probability of her children dying by chance is so low, that means that the probability that she committed a crime given that her children are dead is high". **What's wrong with this logic?**

2.2 Random Variables

- **Definition:** Let S be the sample space for an experiment. A function that maps the outcomes in S to the real line is called a random variable (RV)

- NOTATION for RVs: Use capital letters X, Y, Z

- **Distribution:** The distribution of a RV is the collection of all probabilities of the form $P(X \in C)$ for all sets of C of real numbers such that $X \in C$ is an event

- **Notation for Distributions:** For a RV X, we notate its distribution like this if the distribution has a name:

$$X \sim Blah(\text{ parameters })$$

- Probabilities of events concerning RVs are still probabilities, so they follow the axioms!

2.2.1 Continuous vs. Discrete

- **Discrete RV:** can only take on a finite number of different values, or at least a countably infinite sequence of different values

- **Continuous RV:** Uncountably infinite sequence (for example, an interval of the real numbers).

- **EXAMPLES:**

- **Support of a RV:** another name for the **sample space**. What is the sample space of all the above examples?

2.3 Probability Mass Function (PMF)

- For a discrete RV X, the PMF is defined as the function where, for x on the real number line:

$$f(x) = P(X = x)$$

  Sometimes $f(x)$ is written as $f_X(x)$ for clarity. PMF fully defines a distribution

- Revisiting Coin Flips and learning notation: Let Y ("Capital Y") be a Random Variable representing the outcomes of a coin toss

  A single coin flip is an example of a Bernoulli Distributed Event!

- **Bernoulli Distribution:** A RV takes on value 1 with probability **p**, and 0 with probability 1-p. In other words, the outcomes are Boolean. The parameter is **p** and takes on values from 0 to 1 (because it reflects probability).

- **Binomial Distribution:** A binomial RV is the *sum of independent Bernoulli RVs* with the same parameter. The Binomial has parameters n and p, where n is number of trials, p is probability of each of these trials taking on value 1.

  Intuition/Story: Binomial Distribution reflects the number of successes in n independent trials where the probability of success on each trial is p

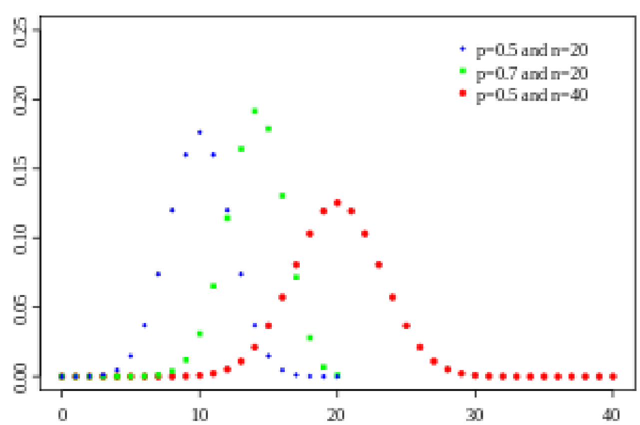
$$X \sim Bin(n, p)$$

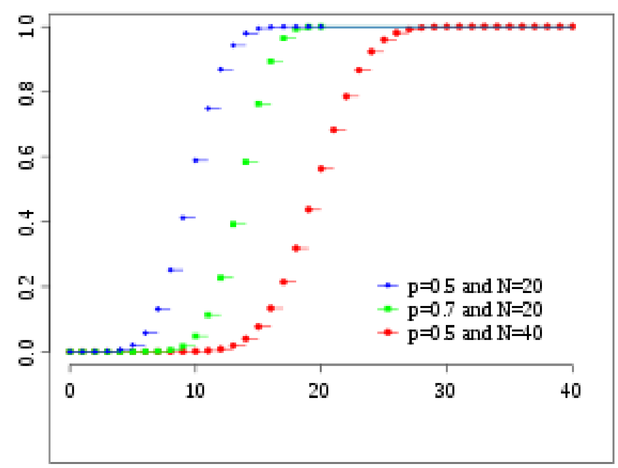
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- **Cumulative Distribution Function (CDF)**: Reflects the probability of the RV being less than x. Always non-decreasing, and right-continuous

$$F_X(x) = P(X \leq x)$$

- PMF and CDF are two ways to describe the distribution of a discrete random variable. *Diagrams from Wikipedia*





2.4 Continuous Random Variables

- Continuous: Possible values of the RV takes on a subset of the real number line (ex: Blood Pressure)
- Suppose you have a RV X which is distributed according to a continuous distribution. Then what is $P(X = x)$ or $f_X(x)$ for some x in the support?

- **Probability Density Function:** a function that gives the "probability density"/ likelihood (not the probabilty!) of observing some value of a random variable. We usually have to evaluate **integrals** of PDFs, to find the probability that the RV takes on a value in a certain interval.
- PDFs must integrate to 1 over the support. *Intuition: the "Area under the whole curve" must be 1, just like discrete probabilities must sum to 1*
- The **Normal Distribution** is a very special continuous distribution (see below)
- **VISUAL EXAMPLE:** Let Z be a RV describing the blood pressure of a randomly selected person.

2.5 Independence of RVs

- Independence of Events: A and B are independent ($A \perp B$) if

$$P(A \cap B) = P(A)P(B)$$

- **Notation for AND/Intersection:** $A \cap B$ is sometimes written as A, B
- **Random Variables X and Y are independent if**

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for Discrete RV X and Y}$$

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \text{ for both Discrete and Continuous RV X and Y}$$

- **Independent and Identically Distributed (IID)**: RVs are independent and each distributed the same way
- In other words, Binomial is a sum of IID Bernoullis!

$$X_1 ... X_n \sim Bern(p)$$

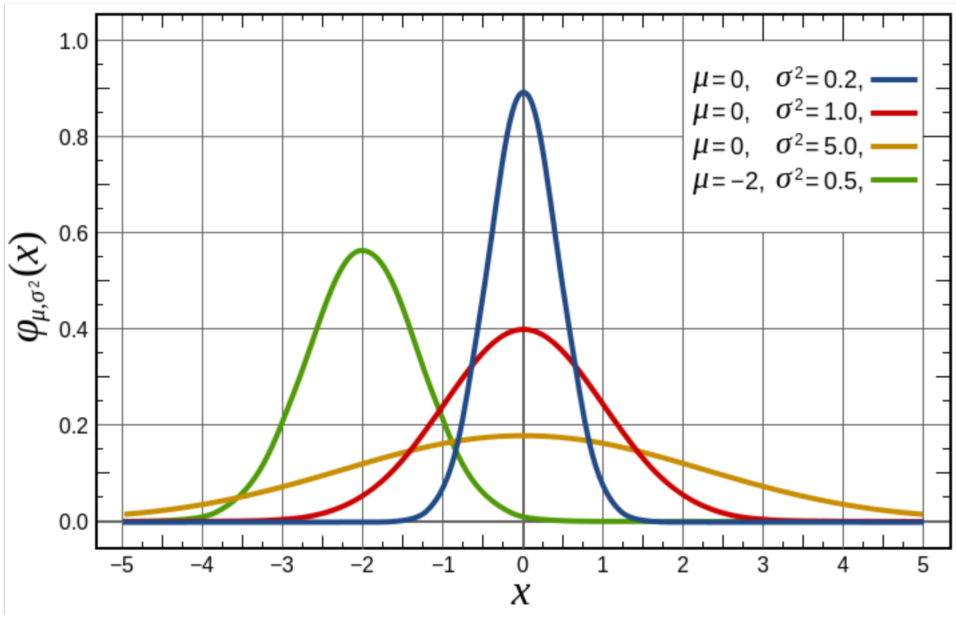
$$Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$$

2.6 Normal Distribution: A very special continuous distribution

- Also known a THE BELL CURVE
- Let X be a RV, and let

$$X \sim Normal(\mu, \sigma^2)$$

  This means that X is distributed normal with parameters mean $= \mu$ and variance σ^2



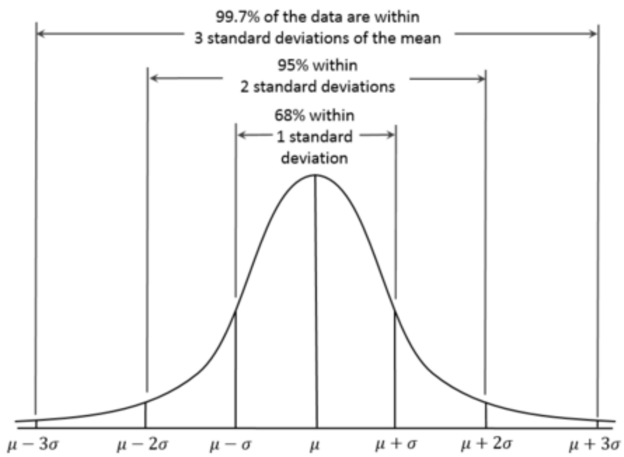
- **Normal Distribution Density:** is Symmetrical, peaks in the middle at the mean. The Variance reflects how "spread out" the distribution is.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

- **Z Score:** How many standard deviations (σ) away from the mean μ is your value?

$$Z = \frac{x - \mu}{\sigma}$$

- **68, 95, 99.7 Rule**: Reflects the probabilty that an observation of a RV falls within 1, 2 and 3 standard deviations from the mean. In a diagram:



Suppose that blood glucose in a patient population is distributed Normally with mean 15 and variance 4. In other words:

$$X \sim N(\mu = 15, \sigma^2 = 4)$$

What is the probability that blood glucose is between 13 and 17?

What is the probability that blood glucose is between 11 and 15?

- Normal Distribution describes many natural phenomena due to the **Central Limit Theorem**
- **Central Limit Theorem (Casual Definition):** given a sufficiently large random sample from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population.
- **Central Limit Theorem (Less Casual Definition):** the sum of IID random variables (with finite variance) tends toward a normal distribution, even if the RVs themseleves are not normally distributed.
- Normal approximations of distributions (ex: Binomial)

2.7 Executive Summary

- Random variables map outcomes of an experiment to the real number line

- For discrete RVs, the Probability Mass Function describes the probability of observing certain outcomes.
- Bernoulli Distributed Rvs, which are discrete, take on values 1 or 0 with some probability p. p is the only parameter.
- Binomial Distributed RVs, which are also discrete, are the sum of independent Bernoulli RVs. They represent the number of 1s in n independent trials here the probability of getting a 1 in each trial is p. Thus, n and p are the parameters of a Binomial.
- Continuous RVs are described by a Probability Density Function, rather than a Probabilty Mass Function. We usually evaluate the probability of observing a value in some interval, rather than the probability of observing a given value.
- The Normal Distribution is a continuous distribution with parameters: mean and variance.
- The Normal Distribution has special properties, including the Empirical Rule and usefulness in describing natrual phenomena.