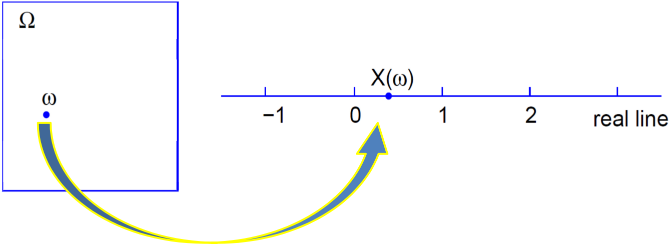
Statistics ST2334 Topic 4: Probability Distributions (part a)

2011/2012 Semester 2

Random Variables

- A random variable is any function that assigns a numerical value to each possible outcome.
- Mathematically, a random variable X is a real-valued function $X(\omega)$ over the sample space of a random experiment, i.e., $X : \Omega \to \mathbb{R}$



Random Variables

- Randomness comes from the fact that outcomes are random ($X(\omega)$ is a deterministic function of ω)
- Often, the random variable in random experiment are naturally defined. For example, when we roll a die and we want X to be the number rolled. However, the same outcome ω can have different numerical interpretations. For example, pick a student at random from the class, X could be weight of the student, while Y could be height of the student.

Some Notation

- We like to use upper case letters for random variables $(X(\omega), Y(\omega), ...)$
- We usually suppress the underlying event ω and simply write X, Y for $X(\omega), Y(\omega)$.
- We typically use lower case letters for values of random variables.
  - $X = x$ means that the random variable X takes on the value x
  - Technically, we are referring to the event ω where $X(\omega) = x$

Examples of Random Variables

1. Flip a coin 3 times. Here $\Omega = \{H, T\}$. Define the random variable $X \in \{0, 1, 2, 3\}$ to be the number of heads
2. Roll a 4-sided die twice.
   (a) Define the random variable X as the maximum of the two rolls ($X \in \{1, 2, 3, 4\}$)
   (b) Define the random variable Y to be the sum of the outcomes of the two rolls ($Y \in \{2, 3, ..., 8\}$)
   (c) Define the random variable Z to be 0 if the sum of the two rolls is odd and 1 if it is even
   (d) Flip coin until first heads shows up. Define the random variable $X \in \{1, 2....\}$ to be the number of flips until the first heads

Examples of Random Variables

3. Let $\Omega = \mathbb{R}$ and the exmeriment is pick a number randomly from $\mathbb{R}$. Define the two random variables

$$X(\omega) = \omega$$

$$Y = \begin{cases} +1 & \text{for } \omega > 0 \\ -1 & \text{otherwise} \end{cases}$$

4. Measure a person's height X and weight Y. Then $Z(\omega) = Y(\omega)/X(\omega)$ is also a random variable: a measure of how fit the person ω is.

Functions of Random Variables

If a, b and c are constants, then $a + bX$, $(X - c)^2$ and $X + Y$, are random variables defined by

$$\begin{aligned} (a + bX)(\omega) &= a + bX(\omega) \\ (X - c)^2(\omega) &= (X(\omega) - c)^2 \\ (X + Y)(\omega) &= X(\omega) + Y(\omega) \end{aligned}$$

In general, $g(X)$ (for any function g) is also a random variable and is defined by

$$g(X)(\omega) = g(X(\omega)).$$

Probability Distribution of a Random Variable

- The probability distribution describes the probability of a random variable X taking on certain values.
- This is governed by the probability function on the underlying events.
- For example, say we're interested in the probability of $X \in A$, where A is some subset of $\mathbb{R}$. Note that

$$\{X \in A\} \Leftrightarrow \{\omega : X(\omega) \in A\}$$

thus, when we say $P(X \in A)$, we really mean

$$P(\{X \in A\}) = P(\{\omega : X(\omega) \in A\}).$$

Classification of Random Variables

- Discrete: X can assume only one of a countable number of values. Such r.v. can be specified by a probability mass function (pmf). Examples 1, 2, 3(Y) are discrete r.v.s
- Continuous: X can assume one of a continuum of values and the probability of each value is 0. Such r.v. can be specified by a probability density function (pdf). Examples 3(X) and 4 are of continuous r.v.s.
- Mixed: X is neither discrete nor continuous. Such r.v. (as well as discrete and continuous r.v.s) can be specified by a cumulative distribution function (cdf)

Example: Toss a coin

- $\Omega = \{H, T\}$.
- If the coin is fair, then $P(H) = 0.5, P(T) = 0.5$.
- Define a r.v. $X(H) = 1, X(T) = 0$. Then we have

$$\{X = 1\} = \{H\}, \{X = 0\} = \{T\}, \text{ and so,}$$

$$P(X = 1) = P(\{H\}) = 0.5, \quad P(X = 0) = P(\{T\}) = 0.5.$$

- The probability distribution of X is given by the probability mass function f, where $f(0) = f(1) = 0.5$.

Probability Mass Function (pmf)

The probability mass function (pmf) of a discrete random variable X is given by

$$f(x_i)[\text{or} \;\; p(x_i) \text{ or } \;\; p_i] = P(\{X = x_i\})$$

Properties of any probability mass function

1. $f(x_i) \geq 0$ for every x_i
2. $\sum_{\text{all } x_i} f(x_i) = 1$
3. $P(X \in E) = \sum_{x_i \in E} f(x_i)$

Bernoulli Trials

- A Bernoulli trial is an experiment with TWO outcomes. One of which we consider a "success". Examples include
  - A Coin Toss: Say we want heads, then H="heads" is success, and T="tails" is failure
  - Rolling a Die: Say we only care about rolling a 6. Then, the outcome space is binarized to "success"$=\{6\}$ and "failure" $=\{1,2,3,4,5\}$
  - Polls: Choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.
- A Bernoulli Process consists of repeatedly performing independent and identical Bernoulli trials.

Bernoulli Random Variable

- Mathematically, a Bernoulli trial is modeled by a random variable
$$X = \begin{cases} 1, & \text{success} \\ 0, & \text{failure} \end{cases}$$
- Let p denote the probability of success. The Bernoulli r.v. has pmf given by
$$f(x) = \begin{cases} p, & \text{when } x = 1, \\ 1 - p, & \text{when } x = 0 \end{cases}$$
- We can also express the pmf by a table

x	0	1
$P(X = x)$	$1 - p$	p

Binomial Random Variables

A Binomial random variable counts the number of successes in n trials in a Bernoulli Process. That is, suppose we have n trials where

- the probability of success for each trial is the same p,
- the trials are independent,
- and n is fixed.

then $X = \{\text{number of successes in the } n \text{ trials}\}$ is a Binomial random variable.

- We say X has a binomial distribution.

Examples of Binomial Random Variables

- Roll a standard die ten times. Let X be the number of times 6 turned up
- A student randomly guesses at 5 multiple-choice questions. Let X be the number of questions the student guessed correctly.
- Randomly pick a family with 4 kids. Let X be the number of girls amongst the kids.
- Urn has 4 black balls and 3 white balls, draw 5 balls with replacement. Let X be number of black balls.

Binomial Distribution

- Let X be a binomial random variable with parameters n (sample-size) and p (probability of success at one trial).
- We write $X \sim B(n, p)$.
- The probability of getting exactly x successes is given by the Binomial probability mass function:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, ..., n$$

x	0	1	...	k	...	n
$P(X = x)$	$(1-p)^n$	$np(1-p)^{n-1}$	...	$\binom{n}{x} p^k (1-p)^{n-k}$	...	p^n

Example: Guessing on Exam

Pat Statsdud failed to study for the next stat exam. Pat's exam strategy is to rely on luck for the next quiz. The quiz consists of 10 multiple-choice questions (n=10). Each question has five possible answers, only one of which is correct (p=0.2). Pat plans to guess the answer to every question.

(a) What is the probability that Pat gets two answers correct?

$$P(X=2) = \binom{10}{2}(0.2)^2(0.8)^8 \approx 0.302$$

(b) What is the probability that Pat fails the quiz? (suppose it is considered a failed quiz if a grade on the quiz is less than 50% , i.e. 5 questions out of 10)

$$\begin{aligned} P(\text{fail quiz}) &= P(X \leq 4) \\ &= P(X=0) + P(X=1) + P(X=2) \\ &\quad + P(X=3) + P(X=4) \\ &\approx 0.967 \end{aligned}$$

Example: Three pointers

Ray Allen hits 60% of three-point shots in catch-and-shoot situations. In a typical game, he takes 5 of such shots. Assuming each attempt is independent of the rest, and that he takes exactly 5 such shots, what are the probabilities that he

(a) hits four of the five shots?
Let X be the number of catch-and-shoot three-pointers Ray Allen makes out of five $\Rightarrow X \sim B(5, 0.6)$

$$P(X = 4) = \binom{5}{4}(0.60)^4(1 - 0.60)^{5-4} \approx 0.259$$

(b) hits at least four of the five shots?

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) \\ &\approx 0.259 + \binom{5}{5}(0.60)^5(1 - 0.60)^{5-5} \\ &\approx 0.337 \end{aligned}$$

Hypergeometric Distribution

- Suppose we draw n samples from a lot containing N units, of which a of them are defective. Let X be the number out of n which are defective.
- If we draw the samples with replacement, then $X \sim B(n, p)$.
- However, more often than not, in the above scenario, we are drawing without replacement.
- In this "sampling without replacement" case, X has the Hypergeometric Distribution. The probability mass function is given by

$$P(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

Hypergeometric Distribution

- The formula can be inferred by an equally likely outcomes argument.
- Note that the draws are not independent and the probability of success changes from trial to trial.
- Example: I just removed 2 dead batteries from a flashlight and inadvertently mixed them with a box of 5 good batteries. If I randomly pick 2 batteries, what is the probability that they are both good?

More Chicken Nuggets

- Recall that Alvin dislikes round chicken nuggets. Suppose Alvin buys a box of 20 chicken nuggets to share with his friends. He notes that 4 of them are round, but decides not to be rude, taking his share of 8 nuggets at random. What is the probability that
  - he gets 0 round nuggets?
  - he gets 1 round nugget?

Poisson Random Variables and Experiments

- Poisson Distribution is a discrete probability distribution that expresses the probability of the number of events occurring in a fixed period of time (or fixed region).
- Denoted by $Poi(\lambda)$, where parameter $\lambda > 0$ is the expected number of occurrences during the given period/region.
- For $X \sim Poi(\lambda)$, the Poisson probability mass function is given by

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

where k is the number of occurrences of an event

x	0	1	...	k	...
$P(X = x)$	$e^{-\lambda}$	$\frac{\lambda e^{-\lambda}}{1!}$	...	$\frac{\lambda^k e^{-\lambda}}{k!}$	...

Poisson Approximation of Binomial Distribution

- If $X \sim B(n, p)$, and n is very large and p is very small, then the distribution may be approximated by the Poisson distribution $X \sim Poi(np)$.
- More precisely, let

$$X_n \sim B(n, \lambda/n), \quad Y \sim Poi(\lambda)$$

  Then for all $0 \leq k \leq n$

$$\lim_{n \to \infty} P(X_n = k) = P(Y = k)$$

- The approximation is good when either $n \geq 20$ and $p \leq 0.05$, or if $n \geq 100$ and $np \leq 10$.

Examples of events that may be modeled by Poisson Distribution

1. The number of cars that pass through a certain point on a road (sufficiently distant from traffic lights) during a given period of time.
2. The number of spelling mistakes one makes while typing a single page.
3. The number of phone calls at a call center per minute.
4. The number of times a web server is accessed per minute.
5. The number of road kill (animals killed) found per unit length of road.
6. The number of mutations in a given stretch of DNA after a certain amount of radiation exposure.

Examples of events that may be modeled by Poisson Distribution

7. The number of unstable atomic nuclei that decayed within a given period of time in a piece of radioactive substance.
8. The number of pine trees per unit area of mixed forest.
9. The number of stars in a given volume of space.
10. The distribution of visual receptor cells in the retina of the human eye.
11. The number of light bulbs that burn out in a certain amount of time.
12. The number of viruses that can infect a cell in cell culture.
13. The number of inventions of an inventor over their career.

Example: Infections

The number of infections [X] in a hospital each week has been shown to follow a poisson distribution with mean 3.0 infections per week.

- $P(X = 0) =$
- $P(X < 4) =$
- $P(X > 4) =$