Lecture: Probability Distributions

Probability Distributions

random variable - a numerical description of the outcome of an experiment.

There are two types of random variables –

(1) discrete random variables – can take on finite number or infinite sequence of values
(2) continous random variables – can take on any value in an interval or collection of intervals

ex) The time that it takes to get to work in the morning is a continuous random variable.
ex) The number of Bs. that you get in class this semester is a discrete random variable.

Discrete Random Variables

Probability Function (PF) $f(x)$ - is a function that returns the probability of x for discrete random variables – for continuous random variables it returns something else, but we will not discuss this now.

The probability density function describles the the probability distribution of a random variable. If you have the PF then you know the probability of observing any value of x.

Requirements for discrete PFs.

(1) $f(x) \geq 0$
(2) $\sum f(x) = 1$

Cumulative Distribution Function (CDF) $F(x)$ - is a function that returns the probability that a random variable X is less than or equal to a value x. The distribution function has the same interpretation for discrete and continuous random variables. The CDF is also sometimes called the distribution function (DF).

Requirements for CDFs

(1) $F(x) \geq 0$ everywhere the distribution is defined
(2) $F(x)$ non-decreasing everywhere the distribution is defined.
(3) $F(x) \to 1$ as $x \to \infty$

Example: Consider the probability distribution of the number of Bs you will get this semester

x	$f(x)$	$F(x)$
0	0.05	0.05
2	0.15	0.20
3	0.20	0.40
4	0.60	1.00

Expected Value and Variance

The expected value, or mean, of a random variable is a measure of central location.

$$E(x) = \mu = \sum x \cdot f(x)$$

In the formula above each value is weighted by probability that it occurs. The expectation is essentially a weighted average.

Some properties of expected values

- $E(a \cdot x) = a \cdot E(x) = a \cdot \mu$ where a is a constant
- $E(a \cdot x + b) = a \cdot E(x) + b = a \cdot \mu + b$ where a and b are constants
- $E(x + y) = E(x) + E(y)$

The variance is a measure of the variability of a random variable.

$$Var(x) = \sigma^2 = \sum (x - \mu)^2 \cdot f(x)$$

Some properties of variance

- $Var(x) = E(x^2) - \mu^2$
- $V(a \cdot x) = a^2 \cdot Var(x) = a^2 \cdot \sigma^2$ where a is a constant
- $E(a \cdot x + b) = a \cdot E(x) + b = a^2 \cdot Var(x) = a^2 \cdot \sigma^2$ where a and b are constants
- $Var(x + y) = Var(x) + Var(y) + 2 \cdot Cov(x, y) \Rightarrow Var(x + y) = Var(x) + Var(y)$ if x and y are independent.
- $Var(x - y) = Var(x) + Var(y) - 2 \cdot Cov(x, y) \Rightarrow Var(x - y) = Var(x) + Var(y)$ if x and y are independent

Lets compute the expected value and variance of the number of Bs you will get this semester.

$$E(x) = 0 * 0.05 + 1 * 0.15 + 2 * 0.20 + 3 * 0.60 = 2.35$$

$$Var(x) = (0-2.35)^2 * 0.05 + (1-2.35)^2 * 0.15 + (2-2.35)^2 * 0.20 + (3-2.35)^2 * 0.60$$
$$= 0.8275$$

<u>Some Common (and Useful) Discrete Probability Distributions</u>

Discrete Uniform Distribution

$f(x) = \frac{1}{n}$, where n is the number of values that x can assume

Binomial Distribution

Properties of a Binomial Experiment

(1) The experiment consist of n identical trials
(2) Two outcomes are possible on each trial – success or failure
(3) The probability of success, denoted by p, does not chance from trial to trial.
(4) The trials are independent.

Example: Flip a fair coin 4 times. What is the probability that I get two tails?

How many trials?
Are two outcomes possible?
Does the probability of a tail change?
Are the trials independent?

Lets think about how we could compute this probability

The first thing that we need to do is think about how many ways I can get 2 tails in 4 flips

TTHH
THTH
THHT
HTTH
HTHT
HHTT

There are 6 ways. Each way there is to get 2 tails has a probability $P^2(1-P)^2$ of occurring. This the probability of getting exactly 2 flips in 4 trials is

$$f(x) = 6(0.5)^2(0.5)^2$$

The Binomial PF: $f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

The number of ways to get x successes out of n trials

Probability of each way actually happening

Example: Someone in the class play basketball? What was the free throw percentage? This will be your probability of success. Suppose that we have you shoot 10 free throws.

How many trials would we have here? 10
Are two outcomes possible on any trial? 2
Does the probability of success chance from trial to trial? Maybe
Are the trials independent? Maybe

Using the formula for the binomial PF we can figure out the probability of you making exactly 5 free throws assuming that the probability of success does not change trial to trial and that the trials are indpendent..

$$f(5) = \frac{10!}{5!(10-5)!} 0.80^5 (1-0.80)^5 = .026$$

For the binomial random variable

$E(x) = np$ - this one is pretty intuitive

$Var(x) = np(1-p)$ - this one you are just going to have you believe me on.

Example: Compute the expected number of made free throws and the variance of the number of made free throws

The Poisson Probability Distribution

The Poisson Distribution describes the probability of that x number events occur in an interval (time or space for example).

Example: The number of job offers that an unemployed worker receives in a week might have a Poisson Distribution.

The number of defective parts that come off the assembly line per hour might also have a Poisson Distribution

The number of Zs that we see per 60 miles of road might also have a Poisson Distribution.

Properties of the Poisson Distribution

(1) The probability of occurrence is the same for any two intervals of equal length.
(2) The occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval.

The Poisson PF: $f(x) = \dfrac{\mu^x e^{-\mu}}{x!}$

where $\mu = E(x) =$ the expected number of events occuring in an interval and $e = 2.71828$

Example: An unemployed worker receives an average of two job offers per week. What is the chance that he receives 4 offers in one week?

$$P(4) = f(4) = \frac{2^4 e^{-2}}{4!} \approx 0.09$$

Example: On average we see 4 Zs. per 100 miles of road. What is the probability of observing no Zs in a 100 mile stretch of highway?

$$P(0) = f(0) = \frac{4^0 e^{-4}}{0!} = 0.018$$

What about a 3Zs in 50 mile stretch of highway?

Because the probability is the same over any two intervals of equal length and the probability of the occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval you can always calculate the probability of x events-arrivals in by multiplying the mean be a factor of proportionality.

Lets calculate the probability of 3 Zs in 50 miles of highway. Because 50 miles is half 100 miles I multiply the original mean of 4 by 0.5.

$$\text{P(3 Zs in 50 miles)}=\frac{\mu^x e^{-\mu}}{x!}=\frac{2^3 e^{-2}}{3!}\approx 0.18$$

The Hypergeometric Distribution

The HD distribution is like the binomial distribution except that with the HD distribution the trials are not independent and the probability of success is not the same trial to trial.

Example: Suppose I have an urn with 4 red balls and 4 blue balls. Using the HD distribution I could calculate the probability of selecting exactly 2 blue balls in 3 trials assuming that I am sampling without replacement.

Why couldn't I just use the binomial distribution? The trials are not independent and the probability of success is not the same trial to trial. Because I am sampling without replacement the probability of selecting a blue ball chances conditional on the color of the balls already selected.

Lets think about this particular experiment.   How many ways can I select exactly 2 blue balls in 4 trials?

BBRR – (1/2)*(3/7)*(1/2)*(3/7) = ((1/2)^2)*((3/7)^2)
BRBR –
BRRB
RBBR
RBRB
RRBB

The all have the same probability.

Probability of selecting 2 blue balls in 4 trials, sampling without replacement =6*((0.5)^2)*((3/7)^2)=.2755

Or we could use the HD PF to figure this out

$$\textbf{Hypergeometric PF: } f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \text{ for } 0 \le x \le r$$

where

n=the number of trials
N=the number of elements in the population
r=the number of elements in the population labled a success

Insights into this formula

Number of ways to select x success from a population of r successes

Number of ways to select n-x failures from a population of N-r failures

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \text{ for } 0 \le x \le r$$

Number of ways to select n objects from a population of N.

Example: Suppose that we have 6 La Follette Faculty Members.

Nichols
Haveman
Holden
Reschovsky
Engel
Wolf

What is the probability of forming a two person committee with at least one women if committee members are chosen randomly.

At least one means 1 or 2 female committee members thus we must calculate the probability of one and then of 2 female committee members

$$f(1) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{2}{1}\binom{4}{1}}{\binom{6}{2}} = \frac{8}{15}$$

$$f(2) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}} = \frac{1}{15}$$

Probability of at least one committee member $= f(1) + f(2) = 9/15 = 3/5$

Note we could also compute this probability as $1 - f(0)$

Continuous Probability Distributions

The Difference Between Continuous and Discrete PFs

Discrete PF tell us the probability of x while continuous PF do not.

First of all why can't they indicate the probability of that a random variable assumes a specific value?

The deal with continuous probability distributions is that the probability of any one point is zero.

This is because there are infinitely many units in a given interval. If there are infinitely many values then the probability of observing any one value is zero.

If continuous PF do not tell us the probability of x then what do they tell us?

PF for continuous random variables do indicate the probability that a random variable falls in a particular interval. This probability is given by the area under the PF in the interval. For this reason PF for continuous probability distributions are called probability density functions (PDFs).

Uniform Probability Distribution

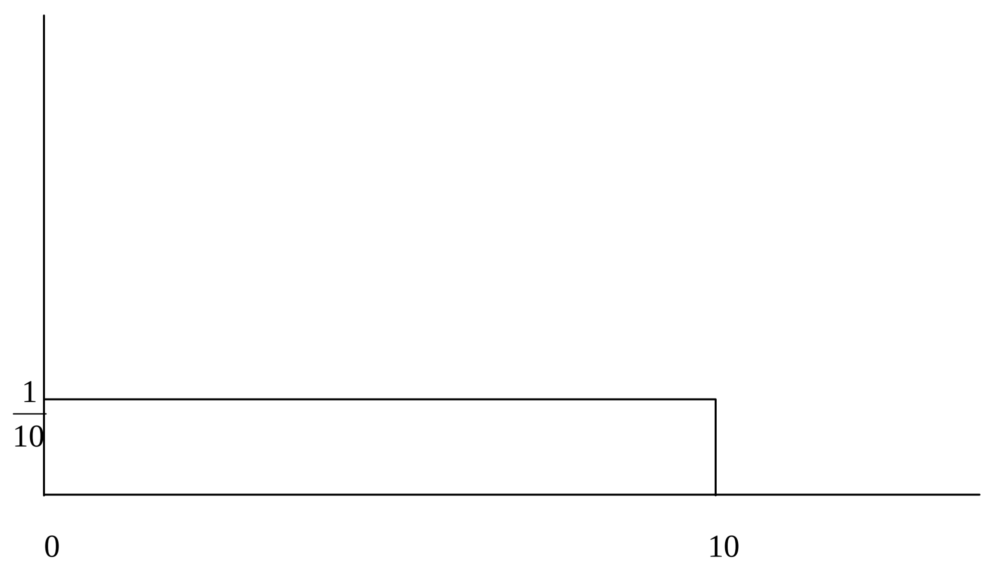
Continuous Uniform PDF: $f(x) = \dfrac{1}{b-a}$ for $a \le x \le b$

The distinguishing feature of the continuous uniform distribution is that the probability that a random variable falls in any two intervals of equal length is equal

Example: Suppose that the pdf associated with a continuous random variable is

$f(x) = 1/10$ for $0 \le \text{x} \le 10$

Is this rv uniform?
What are b and a?
What does this pdf look like graphically?



What is probability of that the rv is less than 5? We can use the cdf

F(5)=(1/10)5=0.5

What is the probability that the rv is between 3 and 6? Here again we can use the cdf

The probability that the rv is between 3 and 6 equals the probability the rv is less than 6 minus the probability that the rv is less than 3.

$P(3<x<6)=F(6)-F(3)=3/10$.

What is the probability the rv is greater than 8?

$P(x>8)=1-F(8)=2/10$

What is the probability that the rv equals 4?

0 because it is a continuous rv

Lesson from this example is that for continuous rvs having the cdf is critical to computing probabilities.

The Normal Probability Distribution

This will be the most important distribution in this class. You need to get very comfortable with dealing with the tables that describe probabilities associated with each distribution.

The Normal pdf: $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ where

$\mu = \text{mean}$

$\sigma^2 = \text{variance}$

π=3.14159=ratio of the circumfrance to diameter

$e = 2.71829$

Important things about at the normal distribution

(1) There are infinitely many variations of the normal distribution differentiated by μ and σ^2.
(2) The highest point of a normal is at the mean which is also the median.
(3) The normal distribution is symmetric. This implies that

$$F(x) = 1 - F(-x)$$

The Standard Normal Distribution

One very important variant of the normal distribution is the standard normal. The standard normal distribution has a mean of 0 and a variance of 1. This is an important distribution because tables describing the probabilities associated with the standard normal distribution are commonly available, but these tables are not available for other variants of the normal distributions.

How to use the Normal Table

Some normal tables are like the one in the book which gives us the value of F(X). Others give F(X)-F(0). Suppose that we want to know the probability that a standard normal rv (z) is less than or equal to 0.5.

Example: This probability is simply F(0.5) so we can just look down the rows of the table to where z=0.5 and to the column that equals 0.00 because 0.5+0.00=0.5.

Example: If we wanted the probability that a standard normal rv (z) was smaller than or equal to 0.55 then we would look down to the row that 0.50 and over to the column that reads 0.05 because 0.5+0.05=0.55.

Example: Suppose that we wanted the probability that a standard normal rv was smaller than 1.655. We would look at the normal table at 1.6 (row) and 0.05 (column). We would also look at the normal table at the 0.06 column. We would take the midpoint between the 0.05 and 0.06 column at 1.6 which is 0.9510.

Example: Suppose that we wanted the probability that a standard normal rv was between 1.78 and 0.47. Simply take F(1.78)-F(0.47).

Example: Suppose that we wanted the probability that a standard normal rv was between 1.78 and –0.47. Take F(1.78) minus 1-F(0.47) because 1-F(0.47)=F(-0.47) (symmetry).

Converting any Normally Distributed RV Into a Standard Normal RV

First a note about Expectations and Variance

It is useful to be able to take expectations (expected values) and variances of transformations of random variables

Consider the following linear transformations of the RV x

$y = ax + b$ where a and b are constants

If $E(x) = \mu$ and $Var(x) = \sigma^2$ then $E(y) = a \cdot \mu + b$ and $Var(y) = a^2\sigma^2$

Rule: If a RV z is a linear transformation of a normal RV x then z is also normal.

What this means is that I can take a normal RV with mean μ and variance σ^2 and make it into a standard normal by centering it on its mean and dividing by the standard deviation.

Let $z = \dfrac{x-\mu}{\sigma}$

If x is normal we know that z must also be normal because z is a linear transformation of x. The expected value of z is

$$E(z) = \frac{1}{\sigma}\left(E(x) - \mu\right) = 0$$

The variance of z is

$$Var(z) = Var\left(\frac{x-\mu}{\sigma}\right) = Var\left(\frac{x}{\sigma} - \frac{\mu}{\sigma}\right) = Var\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma^2}Var(x) = 1$$

Thus z is a standard normal RV.

Example: Male height in the US is normally distributed with a mean 69 inches and variance 36 inches. What is the probability of randomly selecting a male from the population with a height between 5’8’’ and 5’4’’?

Example: What is the probability that a randomly selecting a male from the population with a height between 6’1’’ and 5’10’’?

Example: The natural log of income has a normal distribution with mean 10.35 and variance 2.5. What is the probability that a person has income below $10,000?

This is a tougher one. First lets find out what the natural log of 10,000. Its 9.21. Next we need the standard deviation of log income. Its 1.58.

Now we can form a z score.

$$z = \frac{x-\mu}{\sigma} = \frac{9.21-10.35}{1.58} = -0.72$$

What is $P(z = -0.72)$?

This is just $1 - P(z < 0.72) = 1 - F(0.72) = 1 - 0.7642 = 0.2458$

Cumulative Distribution Function F(Z)

F(Z)= Probability (Area)
from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Exponential Distribution

The exponential distribution is most commonly used to model waiting times and there is a relationship between the Poisson distribution and the exponential distribution.

This relationship is that such that if the mean number of events occurring in an interval of time has a Poisson distribution with mean μ, the mean waiting time between events is exponentially distributed with mean $\frac{1}{\mu}$.

Another way to describe the relationship between the means is

$$\frac{1}{\mu_{Exponential}} = \mu_{Poisson} \text{ or } \frac{1}{\mu_{Exponential}} = \mu_{Poisson}$$

Example: If an unemployed worker receives an average of 7 job offers per week and the number of offers per week is a Poisson random variable, then the average waiting time between job offers is $\frac{1}{7}$ weeks (or 1 day). Then the waiting time until a job offer is exponentially distributed with mean 1day.

JUST REMEMBER THAT μ IS THE EXPONENTIAL MEAN (THE MEAN WAITING TIME)

Exponential pdf: $f(t) = \frac{1}{\mu} e^{\frac{-t}{\mu}}$ where μ is the mean waiting time between events and $\frac{1}{\mu}$ is the mean rate at which events occur.

Exponential cdf: $F(x) = 1 - e^{\frac{-t}{\mu}}$

Survivor function: $S(t) = e^{\frac{-t}{\mu}}$ **=** probability of surviving at least t years.

Example: The number of cars that pass by a particular intersection in a minute is distributed as a Poisson random variable with mean 0.5 cars per minute. What is the probability of waiting 5 minutes without a car having passed?

Here we can use the relationship between the means

$$\frac{1}{\mu_{Poisson}} = \mu_{Exponential} \Rightarrow \mu_{Exponential} = \frac{1}{0.5} = 2$$

Thus the average waiting times between cars is exponentially distributed with a mean waiting time of 2 minutes.

The probability of waiting al least 5 minutes for a car to pass is

$$P(t \geq 5) = 1 - F(5) = S(5) = e^{\frac{-1}{\mu_{Exponential}}t} = e^{\frac{-1}{2}5} = e^{-0.5(5)} \approx 0.082$$

Example: What is the probability of waiting less than 30 seconds for a car to pass?

$$P(\leq 30\sec) = P(\leq 0.5\min) = F(0.5) = 1 - e^{-0.5(0.5)} \approx 0.221$$

Example: What is the probability of waiting less than 5 minutes, but more that 30 seconds?

$$P(30\sec \leq t < 5\min) = P(0.5\min \leq t \leq 5\min) = F(5) - F(0.5)$$
$$= e^{-0.5(0.5)} - e^{-0.5(5)} \approx 0.221 - 0.082$$

Example: The length of time that a particular welfare recipient spends on welfare is exponentially distributed with mean 24 months. The state in which she resides limits the number of months in which a person can receive welfare consecutively to 20 months. What is the probability that she will leave welfare before her 20 months of eligibility elapses.

$$P(\leq 20 months) = F(20) = 1 - e^{-\frac{1}{24}(20)} = 1 - e^{-\frac{5}{6}}$$