Chapter 2

An introduction to statistics

Statistics is the plural of the word *statistic*.

Definition 1 *A statistic is a function of one or more random variables.*

To know exactly what this means, one must first define *variable* and then a *random variable*. Put simply, too simply, a variable is something that varies. A variable can take on different numerical values (a realized value of a variable is some number), so varies. Each number represents a distinct *state* for that variable. For example, the variable G might represent gender state, where 1 corresponds to the state female and 0 corresponds to the state not-female. Or, if X is a variable such that $0 \leq x \leq 123$, X can take any numerical value between zero and 123, inclusive, where, for example, X might represent age of a human. Here, I have assumed all humans are not younger than zero and not older than 123 years (everyone would not agree on this lower bound).[1]

X , uppercase, is the name of the random variable (e.g. age, price, or amount of sexual activity), and x, lowercase, is a numerical value of X.[2] For example, if G is the variable gender then g is a specific gender.

[1] From Wikipedia: The longest unambiguously documented lifespan is that of Jeanne Calment of France (1875–1997), who died at age 122 years and 164 days. She met Vincent van Gogh at age 14.[1] This led to her being noticed by the media in 1985, at age 110. Subsequent investigation found that her life was documented in the records of her native city of Arles beyond reasonable question.[2] More evidence for the Calment case has been produced than for any other supercentenarian case, which makes her case a standard among the oldest people recordholders.[citation needed]. http://en.wikipedia.org/wiki/Oldest_people

[2] The notaional issue of how to distinguish between a random variable and a specific realization of that random variable can be confusing, and the literture is not consistent in how it notationally distinguishes between the two - I won't be either. I will try to use uppper case to denote the name of a random variable, and lower-case to denote a specific value of that random variable. However, I, and others, might use x to refer to both and hope the reader can determine which is meant by the context.

Expressing each separate state with a separate number is not restrictive, but one must take care when interpreting the numbers; their interpretation depends on whether the relationships between the states of the random variable have cardinal meaning, ordinal meaning or neither. Simply put, there are cardinal variables, ordinal variables and nominal (categorical) variables. Gender, for example is a nominal variable and one only has to specify a different number for each state. For example, 0 and 1, or 1 and 2, or 777 and −22.3. If one uses 1 and 2 to represent two gender states one would be foolish to say that Gender state 2 is better than state 1, or is twice in some sense. While one could, one should not use 777 to represent female and −22.3 to represent male.

In contrast, if variable P represents finishing place in the Tour de France bicycle race–first, second, third–the place states have ordinal meaning: 4 finished before 3, but one should not conclude that p was twice as fast as $2p$, only faster by some indeterminate amount. In contrast, age has cardinal meaning, so if A represents age, $2a$ is twice as old as a.[3]

Assume X is a **random** variable (I will define random variable in a second). In which case

$$y = f(x)$$

$$\beta = g(x)$$

and

$$m = 4 + 7x$$

are each statistics, all of the same random variable, X. The letters f and g are the names of particular functions.

Or more generally, imagine three random variables: X, Y and Z. In which case

$$\alpha = \alpha(x, y, z)$$

$$b = h(x, y, z)$$

$$\beta_1 = \beta_1(x, y, z)$$

and

$$\beta_2 = \beta_2(x, y, z)$$

are each statistics. So $c = m(b)$ is a rv.

Alternatively, one could let x refer to the rv and let x_i refer to value i of the rv. This approach will be pretty clear if there is only one rv being considered. But what if there are three rvs? Do I give them different names, like x, y and z? If so z_i refers to a realized value of z. But, what if instead of x, y and z, I had denoted the three random variables in the text x_1, x_2 and x_3 where the subscripts now refer to different rv's, not different observation on x. One must be viligant. One needs to be careful and figure out what is going on by the context.

[3] The differences between many statistical and economic models are often only the numerical properties (cardinal, ordinal, nominal) of the dependent and independent variables.

All statistics are random variables, but all random variables are not statistics, unless one defines $x = x$ as a function.

2.1 A random variable is

Definition 2 *X is a random variable if it is a variable and if it has a distribution. Said another way, X is a random variable if $\forall$ a and b one can determine the probability that $a \leq x \leq b$ if one knows the distribution of X*

X is distributed on some way. The above definition is not self-contained. It requires that we know what a distribution is, and we have yet to define that term, other than we have defined a *distribution* as something that allows us to calculate $\Pr(a \leq x \leq b)$. Note the definition requires that X has a distribution, but it does not require that we know what distribution.

The book, *Introduction to the Theory of Statistics* (Mood, Graybill and Boes) defines a continuous random variable as follows:

Definition 3 *The variable X is a one-dimensional, continuous random variable if there exists a function $f(x)$ such that $f(x) \geq 0 \; \forall \; x$ in the interval $-\infty \leq x \leq \infty$, and the probability that $(a \leq x \leq b)$ is*[4]

$$\Pr(a \leq x \leq b) = \int_a^b f(x)dx$$

The function $f(X)$ is called a *density function* (or a *probability density function*). The function, $f(X)$, describes the distribution of X.

Any function, $f(X)$, can serve as a density function as long as

$$f(x) \geq 0, \quad -\infty \leq x \leq \infty$$

and[5]

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Why do we care about density functions? Models typically assume outcomes (how much you drink, whether the interest rate will rise) are the result of some process with a random component: the model contains a random variable. Or said differently, the behavior of a variable in the model is described by some density function.

[4] Note the qualifiying adjective *continuous.*

[5] Note that $f(x) \leq 1$ is not a requirement (necessary condition). It is required for certain types of density functions, but not all of them. What types? What is required is that $\int_a^b f(x)dx \leq 1$, which follow from the restriction that $\int_{-\infty}^{+\infty} f(x)dx = 1$

A sample is the result of a random process–sampling–so a sample is a vector of random variables. Therefore, a function of a sample is a statistic.

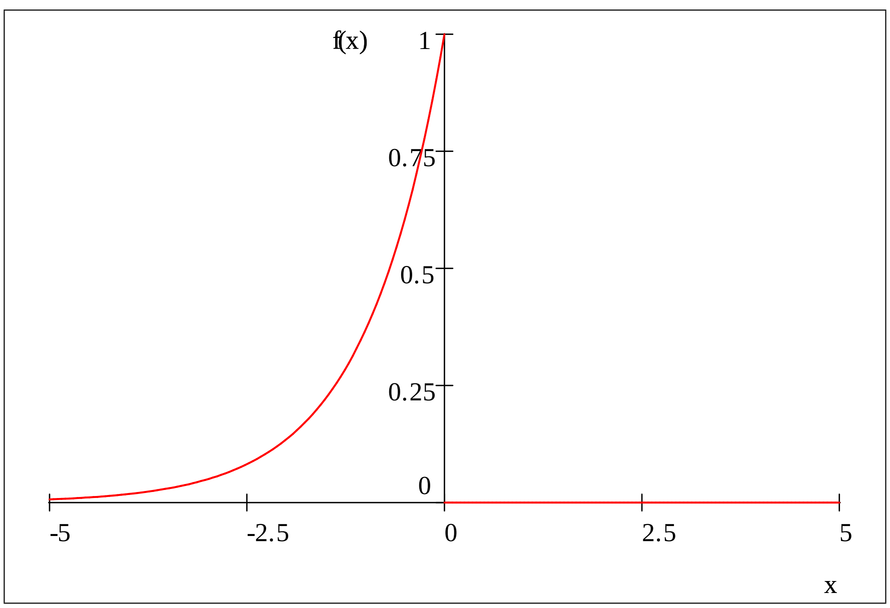
The realized value of a random variable is not a random variable: it is a fixed number; it does not vary, so not a variable. For example, consider the rv A, age at death. Assuming the determination of when one "kicks the bucket" has a random component, A is a random variable, but once the dice is thrown and Melvin "buys the ranch" a_{Melvin} is determined, fixed, and not a random variable. Up until that moment, a_{Melvin} was both a variable and random, but neither afterwards.[6] Consider how many hours you sleep each night, assuming its determination has a random component. Let S denote the number of hours you sleep in a night, so s_t is how many hours you sleep on night t. Before night t. s_t is a rv, but once the night is over, s_t is a fixed number.

An interesting question is whether the world is inherently random–god rolls dice, as in quantum mechanics–or the world is deterministic and it just seems random from our perspective because we cannot observe or measure all of the things that determine things. While interesting, this distinction is not critical for studying statistics.

2.1.1 Making up density functions

The following three examples were provided by students. First,

$$f(x) = \begin{cases} 0 & if \quad x > 0 \\ e^x & if \quad x \leq 0 \end{cases}$$



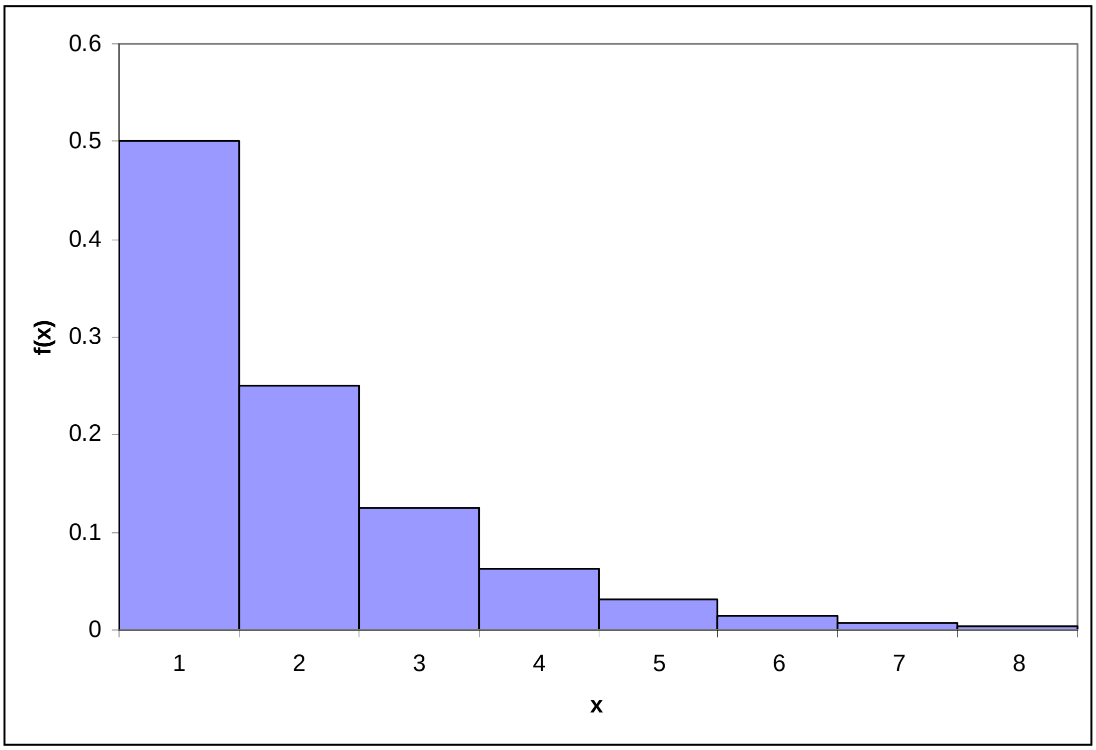
[6] "Kick the bucket" and "buy the ranch" are collequial expressions for dying. There are hundred of collequial expressions for dying.

It is obvious that $F(x) \geq 0 \forall x$ in the domain $(-\infty, +\infty)$. And $\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} e^x dx + \int_0^{+\infty} 0dx = e^x \mid_{-\infty}^{0} + 0 = e^0 - e^{-\infty} = 1 = 0 = 1$, so this is a density function.

Second example: *The Round-up*:

$$f(x) = \begin{cases} 0 & if \quad x < .5 \\ .5^{round(x)} & if \quad x \geq .5 \end{cases}$$

where $round(x)$ is defined as the integer closest to x, with .5 rounded up.



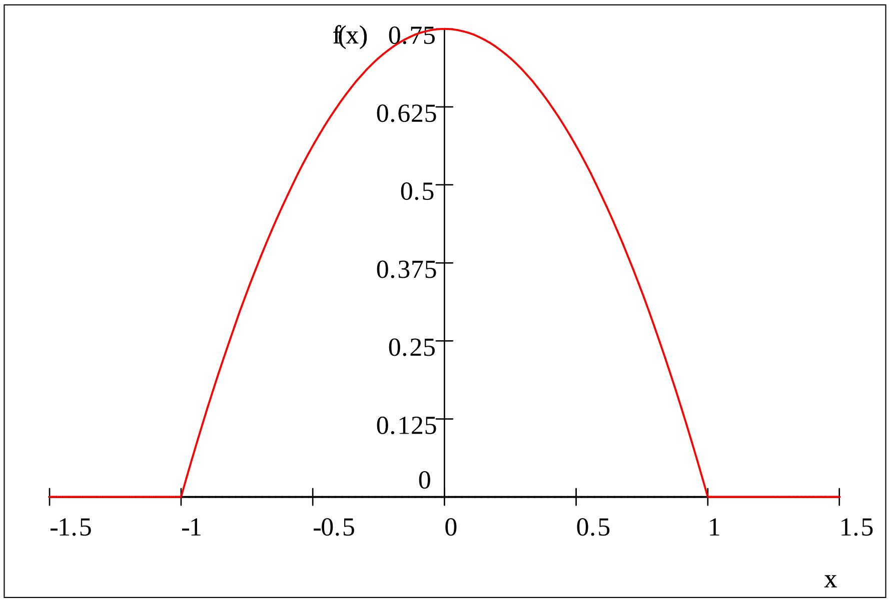
The Round-up density function

Prove this is a density function.

Third example: choose $a > 0$ and define $f(x)$ as follows:

$$f(x) = \begin{cases} 0 & x < -a \\ -\frac{3}{4a^3}x^2 + \frac{3}{4a} & -a \leq x \leq a \\ 0 & x > a. \end{cases}$$

Where the density has positive value, the graph is an upside-down parabola centered around the y-axis, which touches the x-axis at points $-a$ and a, and crosses the y-axis at $\frac{3}{4a}$. Here is what it looks like if $a = 1$.



Note the density outside of the $(-a, +a)$ range.

It is much easy to make up a density function $f(X)$ when X has a limited domain (examples 2 and 3), rather than an infinite domain: it is tough to find functions where the domain is infinite and the area under the function is one (example 1).

After one has chosen some population variable to study, how does /should one decide what to assume about its density function? For example, what if the rv is number of times living individuals have been married. What restrictions might one realistically impose on this distribution.

2.1.2 Using probability to define a random variable

A variable is a rv if there exists some **probability** that the variable lies in the interval a, b. It is sometimes easy to forget that statistics are all about determining or estimating probabilities

For example, in OLS regression analysis, something many of you have encountered, the probabilities are not always explicit. Many, unfortunately, fixate on the OLS parameter estimates, but the probabilities are there. For example, given the OLS parameter estimates, what is the probability that the true value of the parameter lies between a and b? We should be more interested in that question.

2.2 What do statisticians do?

Put simply, statisticians do statistics. Sam, the statistician, does statistics in the same sense that Tiger does golf, but Sam's probably does not do statistics as well as Tiger used to do golf.

Theoretical statisticians propose statistics to better understand random processes– most random processes are determined, in part, by other random processes. They then try to determine the properties of those statistics.

For example, for some random process with unknown parameters, statisticians develop statistics based on samples of random variables from that process, statistics that either help to describe the process or are estimates of the unknown parameters, or both. The goal is to find a statistic that describe the process and that has desirable properties for the question at hand. (To accomplish this goal, one has to obviously decide on what properties are, and are not, desirable–what is desirable also depends on the question at hand.)

A statistic is like a *significant other*, one wants one with desirable properties.

Applied statisticians use these desirable statistics, along with data, to estimate things about the world. Implicit in the approach is the idea that observed outcomes are the result of some random process. Remember that a statistic is a function of random variables, and the function has parameters, which the statistician wants to estimate–estimation means using data to estimate the parameters in statistics.

2.2.1 Urns

If one reads statistics books, one quickly gets the idea that statisticians have a "thing for" urns and drawing balls from urns. When entering the kitchen to make the kids' breakfast, the statistician takes the lid off the *breakfast urn* and draws a ping-pong ball. If the ball is red it is eggs, blue cereal, ... Maybe there is different urn for weekend breakfasts. Tonight will it be TV in bed or sex with the spouse, it all depends of the draw from the *bedroom urn*.

2.2.2 Doing statistics is more difficult than watching TV

A statistic is a function of random variables, so a random variable. Random variables have density functions, so a statistic has a density function. It is often damn difficult to determine that density function.

For example, imagine that X, Y and Z are each a random variable, such that X has the famous Guber distribution, Y has the famous Gomer distribution, and Z has the not-so-famous Snerd distribution.

Now define the statistic

$$\begin{aligned} s &= s(x, y, z) \\ &= \exp(\tan(x^2 y) + xy(\ln(.5z \exp(zy))) \end{aligned}$$

and figure out the density function for S, $f(s)$, NOT.

2.2.3 Econometricians are, one would hope, a subspecies of statisticians

Some quotes from Peter Kennedy's A Guide to Econometrics

> Econometrics is what econometricians do
>
> Econometrics is the study of the application of statistical methods to the analysis of economic phenomena
>
> What distinguishes an econometrician from a statistician is the former's preoccupation with problems caused by violations of statisticians standard assumptions, owing to the nature of economic relationships and the lack of controlled experimentation, these assumptions are rarely met.

Econometricians are often criticized, and often by other econometricians. They have a bad habit of ignoring the quality of their data. Again, some quotes from *A Guide to Econometrics.*

> Econometricians are often accused of using sledgehammers to crack open peanuts while turning a blind eye to data deficiencies and the many questionable assumptions required for the successful application of these many techniques.
>
> Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many milliseconds to bake the mixture at exactly 474 degrees of temperature. But when the statistical cook turns to raw materials, he finds that hearts of cactus fruit are unavailable, so he substitutes chunks of cantaloupe; where the recipe calls for vermicelli he used shredded wheat; and he substitutes green garment die for curry, ping-pong balls for turtle's eggs, and for Chalifougnac vintage 1883, a can of turpentine. (Valavanis)
>
> It is the preparation skill of the econometric chef that catches the professional eye, not the quality of the raw materials in the meal, or the effort that went into procuring them. (Griliches[7])
>
> The art of the econometrician consists in finding the set of assumptions which are both sufficiently specific and sufficiently realistic to allow him to take the best possible advantage of the data available to him (Malinvaud)[8]

[7] More about Zvi Griliches can be fournd at Wiki: http://en.wikipedia.org/wiki/Zvi_Griliches

[8] Edmond Malinvaud published in 1964, *Statistical Methods in Econometrics.* More details about Malinvaud can be found at Wiki- http://en.wikipedia.org/wiki/Edmond_Malinvaud

The applied econometrician: The applied econometrician, unlike the theoretical econometrician, needs to worry as much about her data as about the theory. The forecasts and predictions generated by the econometric model are only as good as the data that produced them. A well-known econometrician recently mentioned to me that he was hired by a group of wealthy gamblers to use his choice-modeling skills to predict the outcome of horse races. It might be important that he get it right.[9]

2.3 Other perspectives on statistics and statisticians

The following quote is from the front of *The Advanced Theory of Statistics*, Vol. 2, by M.G. Kendall and A. Stuart. They attributed it to the fictitious K.A.C. Manderville, *The Undoing of Lamia Gurdleneck*.)

> "You haven't told me yet," said Lady Nuttal, "what it is your fiancé does for a living."
>
> "He's an statistician." replied Lamia, with an annoying sense of being on the defensive.
>
> Lady Nuttal was obviously taken aback. It had not occurred to her that statisticians entered into normal social relationships. The species, she would have surmised, was perpetuated in some collateral manner, like mules.
>
> "But Aunt Sara, it's a very interesting profession," said Lamia warmly.
>
> "I don't doubt it," said her aunt, who obviously doubted it very much. "To express anything important in mere figures is so plainly impossible that there must be endless scope for well-paid advice on how to do it. But don't you think that life with an statistician would be rather, shall we say, humdrum?"
>
> Lamia was silent. She felt reluctant to discuss the surprising depth of emotional possibility which she had discovered below Edgar's numerical veneer.
>
> "It's not the figures themselves," she said finally, "it's what you do with them that matter."

Some additional quotes:

> To understand God's thoughts we must study statistics, for these are the measures of his purpose. (Florence Nightingale)

[9] It should not surprise that many statisticans and econometricans gamble; probability theory developed to improve one's odds in games of chance.

> Statistics are like a bikini. What they reveal is suggestive, but what they conceal is vital. (Aaron Levenstein)

> The first less you must learn is, when I call for statistics about the rate of infant mortality, what I want is proof that fewer babies died when I was Prime Minister than when anyone else was Prime Minister. That is a political statistics. (Winston Churchill)

> There are three kinds of lies, lies, damned lies, and statistics. (Benjamin Disraeli, but sometimes attributed to Mark Twain)

Too bad we can't e-mail Florence and ask here what the hell she meant. Maybe she meant that "casualty statistics," estimates of maimed and dead soldiers, are a "measure of [God's] purpose": part of God's big plan. Churchill suggests that statistics can be manipulated, Disraeli that they mislead.

2.3.1 How Takeshi Amemiya defines *statistics*

Takeshi Amemiya

Takeshi Amemiya is a professor of economics and classics at Stanford, an eminent econometrician, and the author of the well-known and well-used text, Introduction to statistics and econometrics [10]

Definition 4 *"Statistics is the science of assigning a probability to an event on the basis of experiments." (Amemiya, p. 2)*

Definition 5 *"Statistics is the science of observing data and making inferences about the characteristics of the random mechanism that has generated the data." (p. 3)*

The latter definition says "observing data" is required, but this is not strictly correct; one can develop statistics and investigate their properties without ever seeing data. The word "event" turns out to have a very specific meaning in statistics, as we will soon see.

[10] Hhome page http://economics.stanford.edu/faculty/amemiya

Implicit in his definitions, and in much of statistics, is the assumption that what we observe in the world is the result of draws from populations where different outcomes have different probabilities of occurring. Think in terms of drawing one or more balls from an urn, where the urn holds different colored balls in different proportions.

These two Amemiya's definitions are *what-is-it-for* definitions, "assigning a probability" and "making inferences." One might meld these definitions to my definition of a statistic to get

Definition 6 *Statistics is the study of statistics, functions of random variables. The field of statistics combines statistics with data to make inferences/predictions about the process that generated the data.*

Consider an urn that contains all dead smokers, pickled in brine, and one wants to determine the probability that a smoker will, at death, have lung cancer–one draws a sample of dead smokers from the urn.[11] One biopsies the lungs of each to determine whether the rv Cancer, C, takes a value of 0 or 1 for that individual. The result is a vector of random variables, $c_1, c_2, ..., c_s$, where c_3 indicates the cancer status of the third guy drawn. One plugs these observations, the sample, into a statistic to estimate the probability that a smoker will get lung cancer. In this case, the statistic of choice would likely be $\sum_{s=1}^{S} c_s/S$,where S is the sample size.

With his example in mind, consider Amemiya's definition of a random variable:

Definition 7 *"A random mechanism whose outcomes are real numbers is called a random variable." (p. 4).*[12]

Note how he defines random variable as a "mechanism," making no distinction between the process and the notation used to distinguish between different states of the process. He goes on to say, "The characteristics of a continuous random variable are captured by a *density function.*" (p. 4 - later in the book he provides a more technical definitions of a rv). Whether one has lung cancer is not a continuous rv, at least not given the way we define cancer, but we imagine there is a "random mechanism" for getting lung cancer such that smoking affects the probability of acquiring lung cancer.

On to his third definition of statistics:

Definition 8 *"Statistics is the science of estimating the probability distribution of a random variable on the basis of repeated observations drawn from the same random variable."*[13] *(p. 4)*

[11] Alternatively, imagine a cemetery where all and only smokers are buried. One digs up a bunch of the decomposing and takes from each a snip of lung tissue to see whether the smoker had lung cancer. Here Reference the study that dug up frozen guys from WWI to see what kind of flu they had.

[12] I would modify this to "whose outcomes can be expressed with real numbers." For example one would still have a random variable if the variable was hair color and one used letters of the alphabet, rather than numbers, to denote the different colors hair can take.

[13] The term density function is typically only used to describe the distribution of the rv

2.4 Many statistics of interest are call *estimators*

An *estimator* is a type of statistic. Specifically, it, with data, generates an *estimate* of a parameter, or parameter range, in the data-generating process.

We assume members of our population-of-interest are generated by a process, a process with a random component - a *data-generating process*. And assume members of the population can be characterized in terms of some small number of random variables. A member is simply a realization of those random variables.

For example, assume that the rv of interest is y and y_i is a realization of y.[14] One might assume that $y_i = ax^{\varepsilon_i}$ where $\varepsilon\tilde{\ } N(0, \sigma)$, a is a constant, and x is a variable, but not a random variable. Or assume the rv of interest is W, glasses of wine a day, assuming $f(w) = \frac{\mu^w e^{-\mu}}{w!}$ where the constant $\mu > 0$.

If the population of interest is all humans now alive, we might be interested in how this population varies in terms of age, gender, height and weight. We might be willing to assume these four variables are random variables - their realized values are generated by a random process. Each of you can be described as realized value of that random process: you have some age, gender, height and weight.

Or, one's interest in this population might be in finding someone to date. You are writing out your application for the online dating service and have gotten to the question about what kind of person you want to date. You are writing out your application for the online dating service and have gotten to the question about what kind of person you want to date. You want a female between 25 and 30, over six feet tall and less than 150 pounds, but are not sure you should be this restrictive, maybe they only occur rarely in the population. So, you ask yourself, "what is the probability someone has signed up who is female, between 25 and 30 years old, over 6 ft. tall, and weighs less than 150 pounds." To answer, you need to learn about/estimate the joint density function for humans that have signed up for this dating service, and then use it to determine the probability that your dream date exists–keep in mind that you it is unlikely that you are her dream date.

To say that members of the population are generated by a process with a random component is equivalent to saying that each member is a draw from some density function. That density function has some functional form and we want to estimate its form and its parameters. Said another say, populations have properties and we want to come up with estimates of those properties. Said another way, what we observe is the outcome of a process that is driven by parameters, and we want to estimate those population parameters.

---

if the rv varys continuously over some range (it is a *continous rv*). If a rv can take only a countable number of values (it is a *discrete rv*), each with some probability, we don't call its distribution a *density function*. Rather we call it a *discrete distribution*. The term *probability distribution* refers to either a density function, a discrete distribution, or some combination of the two.

[14] Note that I have broken my "rule" about uppercase for the name of the rv.

2.4.1 Consider something some people do: smoke

Define the random variable c as the number of cigarettes smoked per day by an individual, where c_i denotes the number of cigarettes smoked by individual i: c is a random variable and c_i is a realized value of this random variable.[15]

We might want to learn about the distribution of this random variable in our population of interest: determine its density function. The data generating process is draws from that density function.

Note the term *population of interest.* For example, the density function for cigarettes smoked by residents of Italy is very different from the density function for cigarettes smoked by residents of the U.S. And the density function for cigarettes smoked by foreign, male graduates students in Boulder Colorado is different from the density function for all Boulder residents.

What properties, if any, must these density function have? Can the number of cigarettes smoked take any value or must it be an integer? Can it be a negative number? Can it be zero? Can it be 1000 a day? Someone want to check the world record for number of cigarettes smoked in 24 hours? Go to http://www.jimmouth.com/tv04_body.html to see some idiot smoke 159 cigarettes at once

More specifically, we might want to estimate what the distribution of c would be when cigarettes cost a dollar each, and compare this to what the distribution would be if cigarettes cost 10 cents each: this is a problem in demand estimation. Make sure you understand why this is a problem in **demand estimation.**

Econometricians want to estimate the properties of populations (humans, smokers, interest rates, prices). We do this by taking a sample(s) from the population - we sample random variables that describe the population.

We then propose statistics of the sampled values of those random variables , statistics that will hopefully be good estimators of population properties. That is, we want to estimate population parameters. **The statistics that we will use to estimate population parameters are called estimators**. We want our estimators to do a good job of estimating the population parameter.

An estimator is a function of random variables. If one plugs particular values of the random variables into the function one gets an *estimate.* Note the difference between *estimators* and *estimates* - estimators are functions, estimates are realized values of an estimator - estimates are outcomes/numbers.

2.4.2 The estimated mean, a popular estimate

Consider some rv H. If the population is small we can sometimes observe the whole population. If so, we can calculate (not estimate) the population mean.

But, most of the time we do not observe the whole population, so are limited to estimating the mean of the rv H. The function that we use to estimate the mean is an estimator. The inputs into this function are rv's: plugging in a vector

[15] Not that for many i, $c_i = 0$, particularly residents of Boulder. The only people in Boulder who smoke appear to be foreign graduate students.

of realized values of the rv's, out comes an estimate of mean H - remember that the mean of H is not a rv, but our estimate of it is a rv.

Different realizations of the random variables will generate different estimates of the population mean of H – the estimated mean will vary from sample to sample, have *sampling variation*.

For example, assume the goal is estimating the mean weight in the U.S. population, ω.[16]

Sample four weights in the population, denote the weight of the first person observed, w_1, second person w_2, third w_3 and fourth w_4.[17] Every time we sample, we get four different observations: a different sample. In the U.S. population there is a very large number of different possible samples (different sets of 4 people). Let $\boldsymbol{w}^s \equiv (w_1^s, w_2^s, w_3^s, w_4^s)$ denote sample s. $\boldsymbol{w}^s$ is a vector of four random variables, so any function of $\boldsymbol{w}^s$ is a statistic.[18]

Consider the following three statistics

$$\widetilde{w} = f(w_1, w_2, w_3, w_4) = w_1 + 3w_2 + (w_3 w_4)^2$$

$$\widehat{w} = g(w_1, w_4) = \frac{w_1 + w_4}{2}$$

$$\overline{w} = h(w_1, w_2, w_3, w_4) = (.25 \ln w_1 + .25 e^{w_2})^{w_3} + w_4$$

Where did these three statistics come from? I made up three functions of the four random variables.

I now declare each of these an estimator of ω; anything can be an estimator of anything, so what I declare is not untrue, each is an estimator of ω. That said, they may be lousy estimators of ω.

Every time we plug in values from a different sample, we will get new estimates. For example $\overline{w}^s = h(w_1^s, w_2^s, w_3^s, w_4^s) = (.25 \ln w_1^s + .25 e^{w_2^s})^{w_3^s} + w_4^s$ is the estimate of $\overline{w}$ for sample s.

If God said that $\widehat{w} = g(w_1, w_4) = \frac{w_1 + w_4}{2}$ was the best estimator of ω, the wise, applied statistician/econometrician would always use this estimator to estimate ω, no matter their sample.[19]

God is either unavailable or unwilling; so, we need to decide which of all feasible estimators is the preferred estimator (which has the most desirable properties). To determine which is the preferred estimator, from those available, we ask the theorists what properties we would like an estimator to have (not all theorists agree), and which estimator has the most of those properties.

[16] Note that we have assumed that there is a mean weight. Not all random processes have finite means. The mean weight in the U.S. population continuously increases. Maybe it will eventually reach infinity.

According to "WolframAlpha" the mean is 180 and the median is 173.

[17] Note that here the subscript refers to different observations on w, not to four different rv's. But, that said, one could think of them as four different rv's. For example in every sample there will be a first observation, w_1, and this will vary from sample to sample.

[18] Note my use of **bold** to denote a vector.

[19] This would be an interesting God. If she wanted to be helpful, why didn't she just tell us ω?

Since we can never know the true population mean of H, ω, we cannot judge an estimate of mean H by how close it is to the true value. (If we knew the true value, we would not need to do estimation. We judge estimators, not estimates, this point is lost on many souls–those souls should be dammed to Purgatory, maybe the third level of Purgatory.

Words that come to mind when we think about the properties of an estimator include *simple*, *linear*, *unbiased* (vs. biased), *efficient*, *asymtotically unbiased*, *consistent*, and easy to estimate.

2.4.3 So, how do *estimators* relate to the familiar *ordinary least squares* (OLS)?

You have to wonder.

Consider a one-parameter version of the classic linear-regression model

$$c_i = c(p_i, \varepsilon_i) = 25 + \beta p_i + \varepsilon_i \tag{2.1}$$

where c_i is the number of cigarettes consumed by individual i, and p_i is the price of cigarettes for individual i (assumed a variable but not a random variable). ε is assumed a random variable (rv) and ε_i is a random draw from ε. Assume that the density function of ε is normal with mean 0 and variance σ^2. β is a parameter, not a variable, it has some fixed value in the population of interest. An estimation problem only exists because we do not know β of σ

First note that ε is a rv, so the c is a rv, and $c(p, \varepsilon) = 25 + \beta p + \varepsilon$ is a statistic (a function of a rv).

Equation 2.1 describes the process by which cigarette consumption is determined. Note that a statistical model/process has been assumed: the process/model has a random component and we have assumed the density function for this rv belongs to the family of normal distributions.[20]

We have assumed most of the estimation problem away. All that is unknown about the population is the values of parameters β and σ^2.[21] These we want to estimate.

We want an estimator for β. We will use that estimator, along with realized values of the rv's that are the variables in the estimator, to get an estimate of β.

In OLS, we make our estimator of β a function of a sample drawn from the assumed population. In this case, one observation in the sample is the *ith* pair drawn, (c_i, p_i) - we don't observe the ε_i. A sample consists of N drawn pairs: $(c_1, p_1), (c_2, p_2), ..., (c_N, p_N)$. And, the OLS **estimator**/statistic of β is the b

[20] Note that this is a pretty stupid (unrealistic) model because most people smoke no cigarettes, in addition no one smokes a negative number of cigarettes, so consumption cannot be normally distributed.

[21] Econometricans like to assume away most of the estimation problem. We impose a lot of assumptions on our models, often independently of anything the data might suggest.

that minimizes

$$\sum_{i=1}^{N}(c_i - bp_i)^2$$

Denote this esimator b_{OLS}. Every sample taken will generate a different b_{OLS} estimate of β. Note that b_{OLS} is a rv - β is not a rv; it is a constant. Let b_{OLS}^s denote the OLS estimate generated by sample s.

As applied econometricians, we often mistakenly concentrate on the obtained estimate rather than keeping in mind that our b_{OLS}, b_{OLS}^1, is just one draw from a distribution of b_{OLS}.[22] That is, b_{OLS} is a random variable with some density function $f(b_{OLS})$.

Much of the work underlying the classical linear regression model has to do with deriving that density function. Once we have it, we can answer question such as "Given β, what is the probably that an estimate, b_{OLS}, will be between $(\beta - \alpha)$ and $(\beta + \alpha)$?" Or, of more relevance, "What is the probability that β is between $(b_{OLS} - \alpha)$ and $(b_{OLS} + \alpha)$?"

So, put simply, the OLS esimator is a special type of statistic, an estimator. And, OLS estimates are rv's with some distribution. We like OLS estimates–when we assume the classical linear-regression model–because we can show that the OLS estimator has nice properties: it is, if one buys the assumptions, "BLUE" (a Best Linear Unbiased Estimator).

Note that what has nice properties is the estimator, **not** any particular estimate generated by the estimator. Our actual OLS estimate of β often sucks.

In all of my years as an applied econometrician I not published a paper that report the results of a linear regression. There is much more to statistics and econometrics than linear regressions. Remember urns. Urns might sound far afield from what econometrician do, but they're not. For example, drawing a sample is akin to drawing balls from urns. Consider a sample of c, p pairs. One could view the world as consisting of a number of urns, each corresponding to a different price of cigarettes, for example, urn sixteen might includes cigarette consumption by everyone who faces a price of $3.00 a pack for cigarettes.

[22] b_{OLS}^1 is the esimate from the first sample. I assume that most of the time we only collect one sample. When we do simulations, we will collect many samples.

Chapter 3

Quick on set theory

A *set* is a collection of things. I will use letters, often uppercase, to denote a set. Three examples are

$S \equiv \{$all squirrels who live on the University of Colorado campus$\}$

$A \equiv \{$all male students at C.U. Boulder who are math majors with BMWs$\}$

and

$$T \equiv \left\{ \begin{array}{c} \text{tennis sets: each element is a sequence of at least six games} \\ \text{between two individuals with the last two won by the same person} \end{array} \right\}$$

$\{\ \}$ are called "braces." Some symbols are reserved for particular sets: $\emptyset$ denotes the empty set (a set with no members), my set of friends, and Ω denotes the universal set (the set that includes everything, or at least everything being considered).

3.1 As statisticians we care about set theory because?

Put simply (more later), we care about set theory because it is a foundation of probability theory. Consider the outcome of an experiment. Then consider all of the possible outcomes - the set of outcomes. A particular outcome is an element in the set of possible outcomes.

We are often interested in the likelihood (probability) that an outcome will have some property (e.g. the patient dies, interest rates go up, 10 cigarettes are consumed). For example, a treatment might characterized in terms of two random variables: patient dies within a year of treatment (or does not), and patient loses at least ten pounds (or not). We might be interested in the likelihood that someone who gets the treatment lives and loses weight. All the outcomes with the properties "patient dies" and "patient loses at least ten pounds" are a subset of all the possible outcomes of the experiment. We want to know the likelihood that the outcome will belong to that subset.

3.2 Some notation about the relationship between two sets

- $X \cup Y$ is the *union* of the sets X and Y; that is, the set that includes all the elements of X and Y. $\cup$ can be intrepreted as "or", as in, the element x is in either X **or** Y.
- $X \cap Y$ is the *intersection* of the X and Y; that is, those elements that belong to both X and Y. Alternative notation is XY and "X and Y". $\cap$ can be intrepreted as "and", as in, the element x is in X **and** Y.
- $X \backslash Y$ is the elements that belong to X but not to Y. An alternative notation is $X - Y$, called a *set difference*.
- $\overline{X} \equiv \Omega \backslash X \equiv \Omega - X$. $\overline{X}$ is called the compliment of X. Sometimes you will see the notation X^c to denote the compliment of X. So $X \cup \overline{X} = \Omega$ and $X \cap \overline{X} = \emptyset$.

Some additional notation and concepts:

- $x \in A$ means x is an element in the set A. So, for example, $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- $M \subset K$ means the set M is a subset of the set K.[1]
- Sets X and Y are said to be equivalent if $X \subset Y$ and $Y \subset X$.

3.2.1 Some examples:

Unions:

$\{1, 2, 3\} \cup \{a, b, c\} = \{1, 2, 3, a, b, c\}$

$\{1, 2, 3\} \cup \{3, 5\} \cup \{7\} = \{1, 2, 3, 5, 7\}$

$\{\sqrt{2}, \pi, 3.9, r\} \cup \{a, b, c\} = \{\pi, r, a, b, c, 3.9, \sqrt{2}\}$

Intersections:

$\{1, 2, 3\} \cap \{2, 4, 6\} = \{2\}$

$\{a, b, c, d\} \cap \{d, e, f\} = \{d\}$

$\{1, 2, 3\} \cap \{a, b, c\} = \emptyset$

$\{1, 2, 3\} \cap \{\} = \emptyset$

[1] Sometimes we distinguish between subsets and strict subsets, using $\subseteq$ to denote subset and $\subset$ to denote a strict subset. With this more precise notation $A \subset B$ mean that A is a subset of B but there are elements in B that are not in A, so A is a strict subset of B. Whereas, $A \subseteq B$ allows for the possibility that $A \equiv B$.
Note that $(A \subseteq B \text{ and } B \subseteq A) \Longleftrightarrow (A \equiv B)$.
In the notes, we are using $\subset$ to mean $\subset$ or $\subseteq$.

Note that $\cap$ and $\cup$ are algebraic commands but that they apply to sets rather than to variables. Think of *algebra* for sets, *set algebra*. If two sets have no elements in common their intersection is the *empty set*, denoted by empty brackets $\{\}$ or the symbol $\emptyset$.

Combinations of union and intersections:

$\{1, 2, 3, c\} \cap (\{2, 4, 6\} \cup \{a, b, c\}) = \{2, c\}$

$(\{1, 2, 3, c\} \cap \{2, 4, 6\}) \cup (\{1, 2, 3, c\} \cap \{a, b, c\}) = \{2, c\}$

3.3 Set diagrams (Euler and Venn diagrams)

Euler and Venn diagrams, often confused, are a useful to to pictorially represent the relationships between sets (unions, intersections, disjoint, etc). This section presents mostly Eulder diagrams, which I, until recently, mistakingly referred to them as Venn diagrams. They are named for the famous Swiss mathematician and physicist, Leonard Euler, but he probably should not get all the credit. I prefer Euler diagrams to Venn diagrams.

Leonard Euler

Consider first Euler diagrams that are not Venn diagrams. The universal set, Ω, is often represented with a rectangle in two-dimensional space: the rectangle represents all possibile outcomes, but sometimes Ω is simply implicit. The dimensions of the rectangle need not have cardinal or ordinal meaning, but can. The objective of the Euler diagram is to provide a visual representation of the **relationships** between **two** or more sets in terms of the set properties $\cup$ and $\cap$, $\setminus$, and . The sets to be considered (e.g. X,Y, and Z) are each represented an area in the rectangle. Each set is typically represented with a closed curve (often a circle); the interior of the curve represents the elements contained in that set, and the exterior represent elements not in that set.

For example, one might represent set X , a strict subset of Ω, with[2]

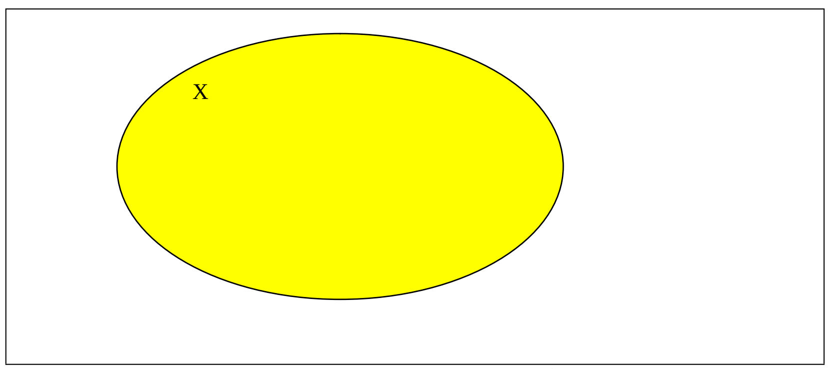


Figure 2_1: Euler diagram of set X

Two sets are represented, X and its compliment. The magnitude of the area used to represent X need not matter. For example, in the diagram, set X should be interpreted as a strict subset of Ω because it does not include all of Ω , but area (size) as a proportion of the rectangle need not reflect the proportion of the elements in Ω that are in X, but it could, depending on the intent of the diagram's creator. Consider now an Euler of two non-intersecting sets

[2]Figures 2_1, 2_2, 2_4, and 2_6 through 2_9 were drawn in WORD using "Shapes." Figure 2_3 and 2_5 were draw in WORD using "SmartArt." Neither WORD tool is ideal. *Shapes*, for example, does not allow the intersection of sets X and Y to have a different color (or shade) than the colors of X and Y. SmartArt has basically one Venn shape, "BasicVenn," with which one cannot shade different intersections differently.

Online "demonstration" applets to create Venn and Euler diagrams include http://apollo.cs.uvic.ca/euler/DrawEuler/index.html
and http://www.cs.kent.ac.uk/people/staff/pjr/EulerVennCircles/EulerVennApplet.html

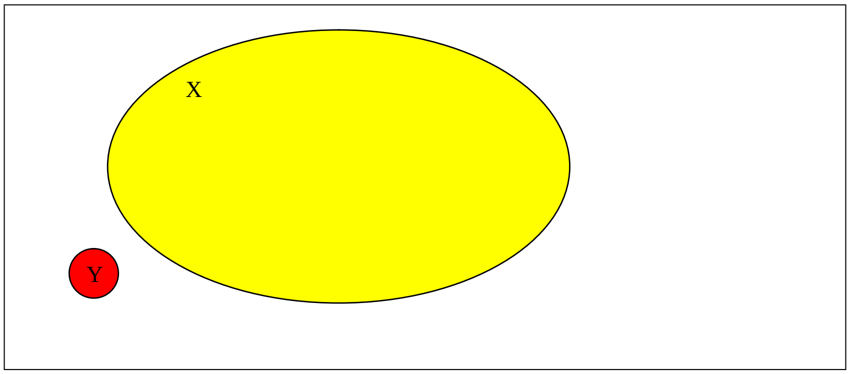


Figure 2_2: Euler of non-intersecting X and Y

The critical information conveyed is that X and Y do not intersect: they are disjoint. That said, If I were viewing Figure 2_2, but were not its creator, my tendency would be to assume set Y has fewer elements that set X.

In comparision, Figure 2_3 conveys that X and Y have common elements, but that each has elements that are not common.

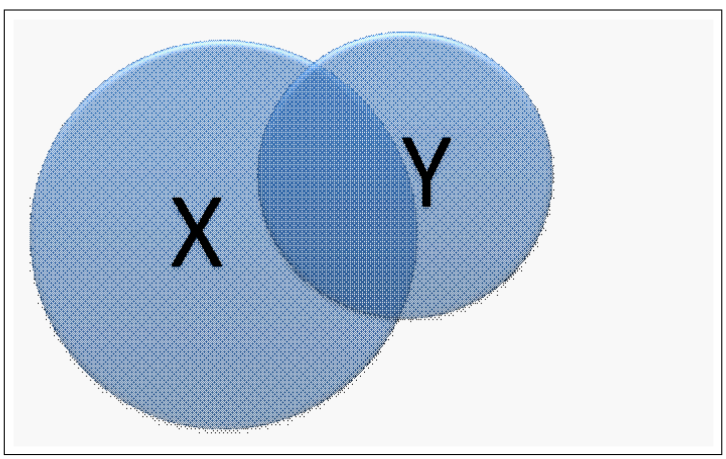


Figure 2_3: Euler of intersecting X and Y

While the shapes used to represent sets (circles, squares, rectangles, or more exotic shapes such as stars and donuts) typically have no significance, if the creator gives a set an exotic shape, the viewer will likely try to attach meaning to the shape. "Why does the shape used to represent all males in the U.S. look like the "Starship Enterprise?"". For example, Figure 2_4 is likely to confuse, unless Y represents Custer and his soldiers and Y represents the Indian war-party right before the *Battle of Little Bighorn* commenced.

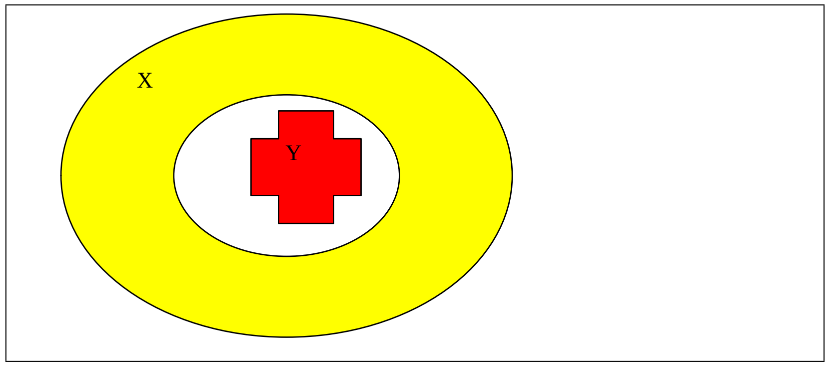


Figure 2_4: Weird Euler

A Venn diagram, attributed to John Venn circa 1880, is different from but conveys the same information as an Euler diagram. A Venn diagram for M sets shows, by definition, all the possible regional relationships between the M sets, whereas a Euler diagram for M sets shows only the non-empty regions.[3] The same information is conveyed by both diagrams because the Venn diagram shades dark all of the regions that are empty. Put simply all Venn diagrams for M sets look the same, except for the dark shading, but there are many possible Euler diagrams for M sets.[4] Note that Figures 2_1 through 2_4 are not Venn diagrams because only the non-empty regions are drawn.

With these caveats and warnings in place, consider a Euler diagram to represent $X \cup Y$, $(X \cup Y)\backslash Z$, and $(X \cup Y) \cap Z$. The intent is to represent the relationship between the sets. $X \cup Y$ is the bluish and greenish areas, including their overlap and their parts that are orangeish; $(X \cup Y)\backslash Z$ is the blue and green excluding the orangeish, and $(X \cup Y) \cap Z$ is the area that is simultaneously orangeish and either bluish or greenish.

[3] Regions are sometimes called "zones." There are 2^M regions. For example if $M = 2$, there are 4 regions: $X\backslash Y$, $Y\backslash X$, $X \cap Y$, and $\overline{X \cup Y}$

[4] Venn diagrams for two or three sets can be drawn with intersecting circles, but if there are more than three sets, circles must be abandoned for more exotic shapes–e.g., for four sets one cannot represent all the possible regions with four intersecting circles.

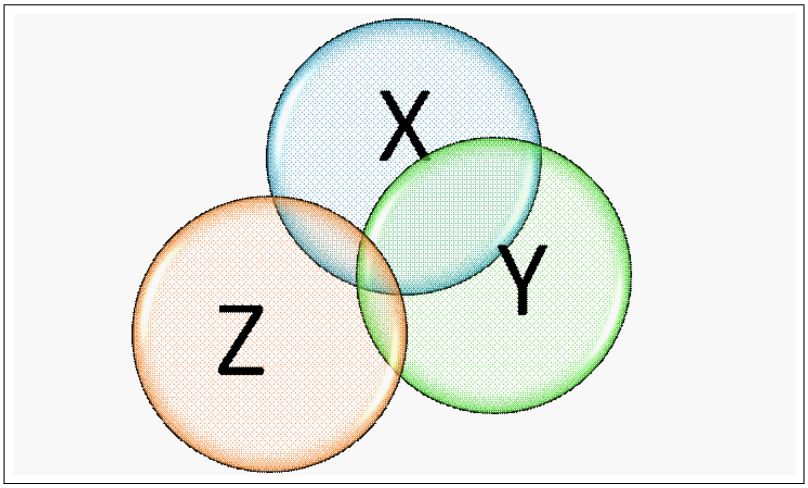


Figure 2_5: Euler of three overlapping sets

Figure 2_5, being an Euler diagram, indictes that all of the subsets have members. If, for example, $X \cap Z = \phi$, one would have to redraw the diagram so that X and Y had no intersection. One could convert Figure 2_5 into a Venn diagram where X and Y had no intersection by simplying shading dark that intersection.

Euler diagrams prove useful for understanding the relationships between sets, for proving that certain relationships eliminate the possibility of other relationships, for intuiting equivalent ways to represent the same relationship, and for disproving conjectures about set relationships (by presenting a counter-example with an Euler diagram). For example, that $\overline{A \cup B} = \overline{A} \cap \overline{B}$ (De Morgan's Law) is, to me, visually intuitive in terms of an Euler diagram.

3.3.1 An example: guys named Edward

Now consider the following sets: all homo sapiens who have ever lived, all male homo sapiens who ever lived, all homo sapiens currently alive, and all homo sapiens ever named Edward. Assume that females are never named Edward and that everyone is either a male or female. Consider an Euler diagram, Figure 2_6 that indicates that most humans with the name Edward are dead.[5] Not that the adjective "most" indicates that region sizes will have significance.

[5] Given the recent popularity of the Twilight series (http://en.wikipedia.org/wiki/Twilight_(series), with Edward, its hearthrob vampire, the number of living males named Edward should see a dramatic rise.

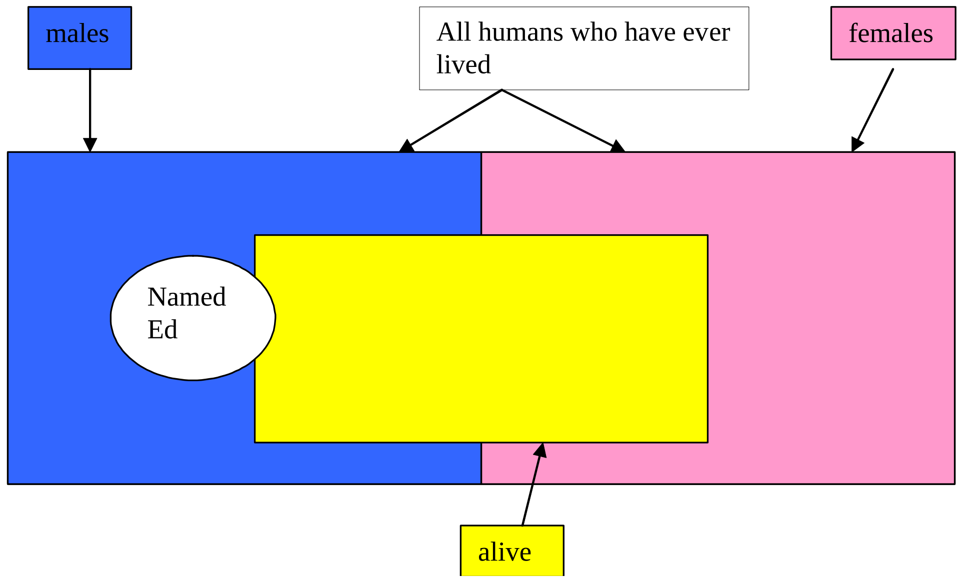


Figure 2_6: Euler of guys named Ed

I have assumed everyone was or is either a male or female–not quite true–and that half the births are female, so made the two gender regions of equal size. Those alive are a strict subset of those who have lived. I drew the "Ed set" so that most of them are in the dead-male category.

3.3.2 An example: cows in the valley

Now consider the following bad joke about deduction. Assume three statisticians: a statistics professor at CU, an undergraduate statistics major at CU, and a statistics professor at MIT. They are out hiking, come to a ridge and look down into a valley. They see a bunch of cows, all of which look black. The professor at CU concludes on the basis of this observation that "All cows are black." The undergraduate say "wrong", then goes on to say that the observation proves only that "All cows in this valley are black." The MIT prof then asserts that what has been proven is that "All cows in this valley are black on at least one side."

The Euler diagram, Figure 2_7, demonstrates that only the professor at MIT is necessarily correct.

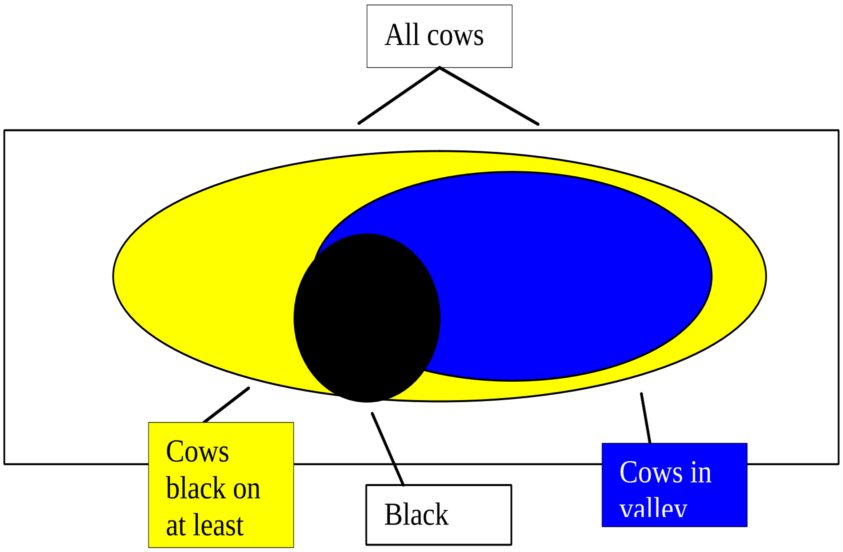


Figure 2_7: Euler of cows in the valley

The white area is all cows that are not black on at least one side.[6]

If all cows in the valley are black, one needs the set of "cows in the valley" to be a subset of "black cows." That is,

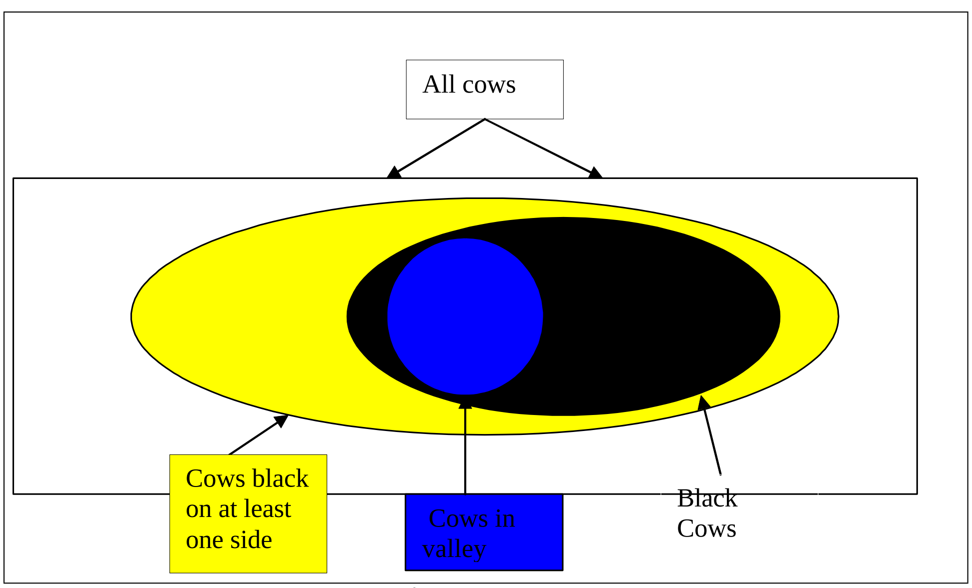


Figure 2_8: Euler if cows in the valley all black

[6]Note here that the universal set is not everything; it is all cows.

In Figure 2_8, all cows in the valley are black but not all cows are black. Note that this graph implies something not implied by what the three saw. It requires the added assumption that if one is black, one is black on all sides; an assumption made by the CU professor and student. If, more restrictively, all cows are black, the Euler diagram, Figure 2_9, is

Figure 2_9: Euler if all cows black

A stupid thing to conclude if you only see a couple of cows who happen to be black on the side you are viewing.

3.4 Theorems from the properties of sets

1. Communitve law: $A \cup B = B \cup A$ and $A \cap B = B \cap A$
2. Associate law: $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4. $\overline{(\overline{A})} = A$: the compliment of the compliment of A is A. That is $not(notA) = A$
5. $A \cap \Omega = A$, $A \cup \Omega = \Omega$, $A \cap \emptyset = \emptyset$, and $A \cup \emptyset = A$
6. $A \cap \overline{A} = \emptyset$, $A \cup \overline{A} = \Omega$, $A \cap A = A$ and $A \cup A = A$
7. $\overline{A \cup B} = \overline{A} \cap \overline{B}$ and $\overline{A \cap B} = \overline{A} \cup \overline{B}$ called *De Morgan's laws*[7]

[7] From Wikipedia (http://en.wikipedia.org/wiki/De_Morgan's_laws) "The law is named after Augustus De Morgan (1806–1871) who introduced a formal version of the laws to classical propositional logic. De Morgan's formulation was influenced by algebraization of logic undertaken by George Boole, which later cemented De Morgan's claim to the find. Although a similar observation was made by Aristotle and was known to Greek and Medieval logicians

8. $A/B = A \cap \overline{B}$
9. $A = AB \cup A\overline{B}$ and $AB \cap A\overline{B} = \emptyset$
10. If $A \subset B$, then $A \cap B = A$, and $A \cup B = B$

The above are algebraic relationships in terms of sets. You should be able to convince someone that each of these theorems follow from the assumed properties of sets. Euler diagrams might help you to be convincing, so would formal proofs, if your audience understands such things. Consider, for example theorem 10, use an Euler diagram to convince that this is true. It is pretty obvious. These ten theorems, along with the basic properties of sets are an *algebra*, an algebra of sets.

Venn diagrams are also convincing with respect to De Morgan's laws. Or, one can formally prove each of De Morgan's laws.

Proof. of $\overline{A \cap B} = \overline{A} \cup \overline{B}$. ■

Note that one has to show both that $\overline{A \cap B} \Longrightarrow \overline{A} \cup \overline{B}$ and that $\overline{A} \cup \overline{B} \Longrightarrow \overline{A \cap B}$

left to right

- $\forall\ x \in \overline{A \cap B} \Longrightarrow x \notin A \cap B$
- $\Longrightarrow (x \notin A)$ or $(x \notin B)$ (either it does not belong to A or it does not belong to B: it cannot belong to both)
- $\Longrightarrow (x \in \overline{A})$ or $(x \in \overline{B})$ $(x \notin A \Longleftrightarrow x \in \overline{A})$
- $\Longrightarrow x \in \overline{A} \cup \overline{B}$ ("$\cup$" means "or")
- Therefore $\overline{A \cap B} \Longrightarrow \overline{A} \cup \overline{B}$

right to left

- $x \in \overline{A} \cup \overline{B} \Longrightarrow (x \in \overline{A})$ or $(x \in \overline{B})$ (x is in either $\overline{A}$ or $\overline{B}$)
- $\Longrightarrow (x \notin A)$ or $(x \notin B)$
- $\Longrightarrow x \notin A \cap B$ (if x is not in A or not in B it can't be simaltaneously in both of them, the intersection)
- $\Longrightarrow x \in \overline{A \cap B}$
- Therefore $\overline{A} \cup \overline{B} \Longrightarrow \overline{A \cap B}$
- So, $\overline{A \cap B} \Longleftrightarrow \overline{A} \cup \overline{B}$ q.e.d.

(in the 14th century William of Ockham wrote down the words that would result by reading the laws out). De Morgan is given credit for stating the laws formally and incorporating them in to the language of logic. De Morgan's Laws can be proved easily, and may even seem trivial. Nonetheless, these laws are helpful in making valid inferences in proofs and deductive arguments."

Proof. of $\overline{A \cup B} = \overline{A} \cap \overline{B}$ ■

left to right

- $\forall\ x \in \overline{A \cup B} \Longrightarrow x \notin A \cup B$
- $\Longrightarrow x \notin A$ and $x \notin B$ (if x is in neither A or B it is not in A **and** it is not in B)
- $\Longrightarrow x \in \overline{A}$ and $x \in \overline{B}$
- $x \in \overline{A} \cap \overline{B}$ (if x belongs to both $\overline{A}$ and $\overline{B}$, it must be in both, the intersection)
- Therefore $\overline{A \cup B} \Longrightarrow \overline{A} \cap \overline{B}$

right to left

- $\forall\ x \in \overline{A} \cap \overline{B} \Longrightarrow x \notin A$ and $x \notin B$ (it belongs to neither A nor B)
- $\Longrightarrow x \in \overline{A}$ and $x \in \overline{B}$
- $\Longrightarrow x \in \overline{A \cup B}$ (x belongs to neither A nor B)
- Therefore $\overline{A} \cap \overline{B} \Longrightarrow \overline{A \cup B}$
- So, $\overline{A \cup B} \Longleftrightarrow \overline{A} \cap \overline{B}$ q.e.d.

3.5 Repetition: set theory is the foundation of probability theory

Again, and put simply, set theory and the "algebra" of sets is a foundation of probability. In probability theory, we define the *sample space* as the set of all possible outcomes of an experiment or sampling. Uses Ω to denote this set. Each element in Ω , ω, is a specific outcome/sample. Note that Ω is typically defined as the universal set, here in the context of all possible experiments.

One can view sampling as the outcome of a data-generating process. Each time the process is run, out pops a sample/outcome. The goal of probability theory is to determine, or estimate, the probability of different events. Let A represent some *event* (all possible outcomes that share the property of interest): a set of outcomes.

For example, conside r rolling a die. One possible event it that the number is 3, another event is the number is odd. Note that only one sample can produce the event 3, but 3 samples can produce the event odd. For a different experiments one event might be the outcome the individual is dead, another, the individual is fat. (Later, we will define events more precisely.)

Consider some experiment and denote the set of all possible events, $\mathcal{A}$. (My experience is that identifying all possible events is often a difficult task.) We are concerned with the probability that event A will occur where A is a subset of $\mathcal{A}$. A topic for the immediate future will be defining and studying the sets Ω and $\mathcal{A}$, and the relationship between them.