Political Science 100a/200a

Fall 2001

Probability, part II[1]

1 Random variables

- Recall idea of a variable and types of variables: ...
- Will now redescribe in probability theory terms as a *random variable*. Here is a technical/mathematical definition:

Def$^{\underline{n}}$: A *random variable* is a function that assigns a number to each point in a sample space S.

For social science purposes, a more intuitive definition is this: A random variable is a *process* or *mechanism* that assigns values of a variable to different cases.

- e.g.: Let $S = \{$a list of countries$\}$, with an arbitrary country in the sample space denoted $s \in S$.
  1. Then $X(s) =$ 1990 per capita GDP of country s is a random variable. For instance,

$$\begin{aligned} X(\text{Ghana}) &= \$902 \\ X(\text{France}) &= \$13,904 \\ \text{etc...} \end{aligned}$$

  2. Likewise, $Y(s) = 1$ is country s had a civil war in the 1990s, and $Y(s) = 0$ if not, is a random variable.
  3. $S = \{$list of survey respondents$\}$, $X(s) =$ Party ID of respondent s.
- Typically, we will drop the notation $X(s)$ and just write X and Y, leaving it implicit that this a random variable that associates a number to each point in a sample space.
- Why "random"?
  - Because we will be thinking about "experiments" where we draw a point or collection of points from the sample space at random or according to some probability distribution.

[1]Notes by James D. Fearon, Dept. of Political Science, Stanford University, October 2001.

- Because we may believe that the process that assigns a number like GDP (the value of the variable) to a case has a random or *stochastic* element.
- Some statisticians, like Friedman, actually define a random variable as "a chance procedure for generating a number," by which he means some *repeatable* process that has a well-understood source of randomness and for which "objective" probabilities can in principle be computed.
- For others (and most social scientists, I think) there is no necessary implication that the value of the variable (e.g., a country's per capita GDP) is actually produced by a stochastic process, although in fact this is often how we will think about it (treat as "as if" the result of a repeatable random process).
  * e.g.: an individual's propensity for suicide is assumed to be based on characteristics of the person (such as degree of social integration) plus "random noise"
  * e.g.: a particular measurement of the weight of an object is understood to be produced as equal to true weight + bias + noise.

- Observation: Things that are true of functions are also true, generally, of random variables.
  - Thus, if if $X(s)$ and $Y(s)$ are random variables (defined on the same sample space), then so are $X(s)+Y(s)$, $X(s)-Y(s)$, $X(s)Y(s)$, and $X(s)/Y(s)$ (provided $Y(s) \neq 0$ for all S).
  - In general: **Th$^{\underline{\mathbf{m}}}$:** If $\phi(z)$ is a function and $X(s)$ is a random variable, then $\phi(X(s))$ is a random variable as well.
  - How are such combinations and compositions of two random variables formed? Case by case. For example, we form $Z(s) = X(s) + Y(s)$ by adding the the values of $X(s) + Y(s)$ for each point s in the sample space.
- Important: Instead of working with a probability distribution or measure defined on the sample space S, it is often more convenient to work with the implied probability distribution on the different possible values of the random variable $X(s)$.
- Let

$$f(x) = P(\{s|X(s) = x\}).$$

  Thus, $f(x)$ is the probability that the random variable will take the value x. The set $\{s|X(s) = x\}$ is the set of all points in the sample space (possible outcomes of the experiment) such that $X(s) = x$. (Note: We need to do it like this because there might be many such $s \in S$.)
- So $f(x)$ is a probability distribution on possible outcomes of $X(s)$ which need not be the same as the probability distribution on the state space S.

– e.g.: Let $X(s)$ be the number of heads that appear in two tosses of a fair coin, with S taken as $\{hh, ht, th, tt\}$. Then

$$\begin{aligned} f(0) &= P(\{tt\}) = 1/4 \\ f(1) &= P(\{ht, th\}) = 1/2 \\ f(2) &= P(\{hh\}) = 1/4 \\ f(x) &= 0 \text{ for all other } x \neq 0, 1, 2 \end{aligned}$$

– e.g.: Let $X(s)$ be 1 for dyad years with a war and 0 for dyad years with no war, with S taken as a sample of dyad years. Then

$$\begin{aligned} f(0) &= P(\{s : \text{s is a peaceful dyadyear}\}) = .999 \\ f(1) &= P(\{s : \text{s is dyad year at war}\}) = .001 \end{aligned}$$

- From now on, we will often work directly with random variables and the associated probability distributions on them as defined here, bypassing the underlying sample space.

2 Discrete random variables

Def$^{\underline{n}}$: A random variable X has a *discrete probability distribution* $f(x)$ if X can take only a finite number of values (or a countably infinite number of values).

- e.g.: Let X be the number of heads in n tosses of a fair coin (thus X is a random variable). Then for $x \in \{0, 1, 2, \ldots, 10\}$,

  $$f(x) = \binom{10}{x} \frac{1}{2^{10}},$$

  and for all other x, $f(x) = 0$.

  Plot with Stata ...

- Note that this is the same idea as for a histogram, which may take a more continuous variable, group it into categories, and display the probability associated with each of these categories.

- e.g.: Let X be the number of bills sent by Congress to the president in a given year. $f(x)$ would then refer to the probability that x bills are sent.

Another way to describe the probability distribution of a random variable ...

Def$^{\underline{n}}$: If a random variable X has a discrete probability distribution given by $f(x)$, then the *cumulative distribution function* (c.d.f) of X is

$$F(x) = P(X \leq x).$$

- In words, $F(x)$ is the probability that the random variable produces a value of less than or equal to x.
- e.g.: Graph the c.d.f. of a random variable distributed by a binomial distribution for $n = 2$ and $p = 1/2$. ...
- e.g.: for $n = 3$ ...

Def$^{\underline{n}}$: If an experiment or process produces two random variables X and Y, then the *joint probability distribution* of X and Y may be written as

$$f(x, y) = P(X = x \ \& \ Y = y).$$

- e.g.: You roll two dice, letting X be the random variable referring to the result on the first die, and Y the value that shows on the second. What is $f(x, y)$?
- recall defn of independence, example where it does not hold (e.g., income and life expectancy) ..
- Draw 3D picture of a discrete joint density ...

3 Continuous random variables

- Whenever you work with a sample space S that is finite, then you get a *discrete* distribution $f(x)$. The samples we observe in practice are always finite and thus have discrete distributions.
- For theoretical reasons, though, it is extremely useful to be able to represent the idea of a sample space that is (uncountably) infinite.
- e.g.: Take as the sample space all the points in the unit interval, so that $S = [0, 1]$, with typical element $x \in [0, 1]$.
- Suppose further that we consider a probability measure such that every point in this interval is equally likely. This is called a *uniform distribution* on $[0, 1]$, and is often denoted $U[0, 1]$.

- How characterize in terms of a probability distribution function like $f(x)$ for a discrete distribution? A paradox: If there are an infinite number of possible outcomes in the sample space, and each is equally likely, then what is the probability of any particular point, e.g., $s = .378$? ... if 0, won't work, if positive, won't work ...
- Instead, we define a probability measure that assigns positive probability numbers only to *sets of points in [0, 1]*, and in particular, only to what are called *measurable sets* of points in the sample space.
- e.g.: Intuitively, what would be the probability of drawing a point from the subinterval $[0, 1/2]$ if all points are equally likely? What about from $[1/4, 1/2]$?
- In general, we will characterize the uniform probability distribution on [0, 1] in terms of a *density function* $f(x)$. This is related to but NOT the same thing as $f(x) = P(X(s) = x)$ in the discrete case.

Def$^{\underline{n}}$: A *density function* $f(x)$ for a continuous random variable X has the property that the area under $f(x)$ for any interval $[a, b]$ is equal to the probability of drawing an $x \in [a, b]$.

- In the case of $U[0, 1]$, $f(x) = 1$ for all $x \in [0, 1]$.
- Show graphically ...
- Demonstrate that $P([a, b])$ equals area under $f(x)$ for $[a, b]$.
- What about the probability of drawing $x = .358$? Or what about the probability of drawing $x \in \{.358, .989\}$? Defined as having "measure zero" (i.e., zero probability). True for any collection of particular points, even if infinite.
- In calculus terms, a random variable X has a *continuous probability distribution* or *continuous density* if there exists a nonnegative function $f(x)$ defined on the real line such that for any interval $A \subset \mathbb{R}$,

$$P(X \in A) = \int_A f(x)dx.$$

  (For those who haven't had calculus, think of the S-like thing as another form of summation operator. This is like "the sum of $f(x)$timesdx, where dx is a very small number so that $f(x)dx$ is the area of a very thin rectangle, for all the points in the set A on the x-axis.)

- Unlike in the discrete case, keep in mind that with a continuous probability distribution $f(x)$ is NOT the probability of the specific point x. The probability of specific point in a continuous distribution is *zero*. Only intervals can have positive probability.

- If dx is understood to be a very small number, e.g., .001, then it is correct to say that $f(x)dx$ is approximately the probability of a number in a very small interval around x. (Illustrate graphically.)
- If $f(x)$ is a probability density function, then
  1. $f(x) \geq 0$ for all $x \in \mathbb{R}$.
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$. (I.e., the area under the whole curve is $1 = P(S)$.)
- Thus, a *full* specification of $f(x)$ for $U[0,1]$ is

$$f(x) = \begin{cases} 1 & \text{for} \quad x \in [0,1] \\ 0 & \text{for} \quad x < 0 \text{ and } x > 1 \end{cases}$$

  (Note that the integral over $(-\infty, \infty)$ works out here...)
- Why all this fuss over continuous random variables? One major reason is that the important normal distribution is a continuous distribution: Illustrate graphically ...
- Just as we could define a *cumulative distribution function* or c.d.f. for a discrete probability distribution, we can do the same for a continuous distribution: $F(x) = Pr(X \leq x)$.
  - Thus, $F(x)$ is the probability that the random variable X realizes a value less than or equal to the number x. Illustrate graphically for normal curve ...
  - e.g.: What is the cdf of the uniform distribution on $[0,1]$? work through ... Draw $F(x) = x$ ...
  - Note how the cdf translates the density function into a different shaped curve that summarizes the same info a bit differently. e.g., if you have a density function like this (draw bell-shaped curve) ... Te area under the curve to the left of x is $F(x)$. Show how this gives an S-shaped cdf.
  - A general point about cdf's follows:
    Th$^{\underline{\text{m}}}$: For all cdf's, $F(x)$ is non-decreasing in x, that is,

$$x < x' \Leftrightarrow F(x) \leq F(x').$$

  - Cdf's do not have to be strictly increasing everywhere, just nondecreasing. Flat parts of cdf's correspond to intervals that have zero probability. Note that cdf's of discrete probability distributions will always have flat portions (in fact they will be all flat portions with "jumps").

4 Expectations and expected values

- Recall the idea of the *mean* or average as a measure of the central tendency of a variable X: $\mu = \frac{1}{n}\sum x_i$.
- This has an important generalization when we move to the idea of a random variable that has a probability distribution $f(x)$:

  The *expected value* or *expectation* of a random variable is like a mean of the variable where the values it takes *are weighted by the probability with which they occur.*
- Consider a random variable X.

  1. Suppose X has a discrete probability distribution function $f(x)$, and that the values X might assume are indexed $x_1, x_2, x_3, \ldots, x_n$. Then the expected value or expectation of X is defined as

     $$E(X) = \sum_{i=1}^{n} x_i f(x_i).$$

     We could also write $E(X(s)) = \sum_{s \in S} X(s)P(\{s\})$, in terms of the state space.
  2. Suppose X has a continuous distribution with density function $f(x)$. Then

     $$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

     (if this integral exists ...)
- The expectation $E(X)$ is thus *a weighted average* of the possible values of a random variable, where each value is weighted by its probability.
- e.g.: Let $X = 1$ if a fair coin turns heads, and $X = 0$ if it comes up tails. Then what is $E(X)$? Let $x_1 = 0$ and $x_2 = 1$. Then

  $$E(X) = \sum_{i=1}^{2} x_i f(x_i) = 0 * .5 + 1 * .5 = 1/2.$$
- e.g.: You are in a rush, and can only find parking in a tow zone. You will only be inside for 30 minutes. You think the probability of a getting a ticket in this time interval is .05, and the probability of being towed is .01. A ticket would cost \$60, and being towed would cost \$300 in time and fees. Let X be a random variable that represents your losses. What is $E(X)$?

  $$E(X) = \sum_{i=1}^{3} x_i f(x_i) = 0 * .94 + \$40 * .05 + \$300 * .01 = \$7,$$

where we let $x_1 = 0$ and $x_2 = \$40$, and $x_3 = \$300$.

How think about this quantity? If you like, as the long run average of what you would be losing if you could do this over and over again.

5 Properties of expectations

1. For a random variable $X = b$ where $b \in \mathbb{R}$ is a constant, $E(X) = b$. That is, the expectation of a constant is the constant. This follows directly from the definition of $E(X)$.
2. If $Y = aX + b$, where X and Y are random variables and a and b are constants, then
$$E(Y) = aE(X) + b.$$
   - Important: Thus, you can take the expectations operator $E(\cdot)$ "through" a linear combination of a random variable. This again follows directly from definition of $E(\cdot)$. Show ...
3. If X and Y are random variables, then $E(X + Y) = E(X) + E(Y)$. Prove ...
   - Important: It follows from the last two that *you can "take expectations through"* any linear combination of random variables. e.g.: $E(aX + bY + cZ) = aE(X) + bE(Y) + cE(Z)$.
   - e.g.: The basic regression model, $y_i = a + bx_i + \epsilon_i$, or written as random variables, $Y = a + bX + \epsilon$. Taking expectations we have
$$E(Y) = E(a + bX + \epsilon) = a + bE(X) + E(\epsilon) = a + bE(X)$$
since $E(\epsilon) = 0$.
4. Note that $E(X+Y) = E(X)+E(Y)$ *whether or not the random variables are independent of each other.* (Recall: Two random variables X and Y are independent of each other if $f(x, y) = f(x)f(y)$, where $f(x, y)$ is the joint density function.
5. But this is as far as it goes: In general, it is NOT true that if $\phi(z)$ is a function (e.g., $\phi(z) = z^2$), $E(\phi(X)) = \phi(E(X))$.

   That is: *you cannot take the expectations operator "through" a nonlinear function.*
   - e.g.: It is not generally true that $E(X^2) = (E(X))^2$.
6. Nor is it generally true that $E(XY) = E(X)E(Y)$.
   - This *will* be true if the random variables X and Y are independent. Prove ...
   - If $X_1, X_2, \ldots, X_n$ are n *independent* random variables, then
$$E(X_1X_2X_3 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n).$$

6 Variance

Just as we can think about means in terms of expectations, so we can think about variances:

Def$^{\underline{n}}$: Let $\mu = E(X)$, where X is a discrete random variable. Then the *variance of X* is defined as

$$\sigma^2(X) = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{n}(x_i - \mu)^2 f(x_i),$$

- e.g.: Consider one flip of a "coin" that comes up heads with probability .7, and the random variable $X = 1$ heads and $X = 0$ if tails. Note that $E(X) = .3 * 0 + .7 * 1 = .7$.

$$var(X) = E[(X-\mu)^2] = \sum_{i=1}^{n}(x_i-\mu)^2 f(x_i) = (0-.7)^2 * .3 + (1-.7)^2 * .7 = .21.$$

- e.g.: In general, for a Bernoulli trial with parameter p (a "coin" that comes up heads each time with probability p), $E(X) = p$ and

  $$var(X) = (0-p)^2(1-p)+(1-p)^2 p = (1-p)(p^2+(1-p)p) = (1-p)(p^2+p-p^2) = p(1-p).$$

  Discuss substantive meaning.

- e.g.: What are the mean and variance of a uniform distribution on the [0, 1] interval?

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 xdx = \frac{1}{2}x^2|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}.$$

$$var(X) = \int_0^1 (x-E(X))^2 f(x)dx = \int_0^1 (x-1/2))^2 dx = \frac{1}{3}(x-1/2)^3|_0^1 = \frac{1}{3}(1/8-(-1/8)) = 1/12.$$

Properties of variance

1. $Var(X) = 0$ if and only if $X(s) = c$ for all $s \in S$. That is, the variance of a constant is zero (since it doesn't vary).
2. For any $a, b \in \mathbb{R}$, $Var(aX + B) = a^2 Var(X)$.

   This is worth working through, because the proof illustrates several properties of expectations.

   Proof: By definition, $Var(aX+b) = E([aX+b-E(aX+b)]^2)$. But, from the properties of expectations. $E(aX+b) = aE(X)+b$. (You can take $E(\cdot)$ "through".) Let $\mu = E(X)$.

Thus,

$$\begin{aligned} Var(aX+b) &= E([aX+b-a\mu-b]^2) \\ &= E([a(X-\mu)]^2) \\ &= E(a^2(X-\mu)^2) \\ &= a^2E((X-\mu)^2) \\ &= a^2Var(X). \end{aligned}$$

3. $Var(X) = E(X^2) - (E(X))^2$.

   Proof:

$$\begin{aligned} Var(X) &= E((X-\mu)^2), \text{ where } \mu \equiv E(X) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2. \end{aligned}$$

   This is just a useful fact.

4. If $X_1, X_2, \ldots, X_n$ are independent random variables, then $Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$.

   In words, for *independent* random variables, the variance of their sum equals the sum of their variances.

5. Discuss why $var(X+Y) \neq var(X)+var(Y)$ when the two variables are not independent. What would happen if, say, they were positively correlated?

7 Examples of the application of these properties

Question: Suppose we toss a fair coin n times. What is the expected number of heads, and what is the variance (or standard deviation) of the number of heads that might appear?

- We are treating X = number of heads in n tosses as a random variable, and asking what is $E(X)$ and $V(X)$ (or $sd(X) = \sqrt{V(X)}$).
- Formally,

$$E(X) \equiv \sum_{i=1}^{n} x_i f(x_i) = \sum_{i=1}^{n} x_i \binom{n}{x_i} (.5)^{x_i} (1-.5)^{n-x_i} \text{ and}$$

$$V(X) \equiv \sum_{i=1}^{n} (x_i - E(X))^2 f(x_i) = \sum_{i=1}^{n} (x_i - E(X))^2 \binom{n}{x_i} (.5)^{x_i} (1-.5)^{n-x_i}.$$

- Yuck.
- But there is a *much* easier way: Think of X, the random variable representing the total number of heads in n tosses *as the sum of n random variables, one for each toss.*
- Let X_i be a random variable for the ith toss, so that $X_i = 1$ with probability .5 and $X_i = 0$ with probability .5 (1 represents a head, 0 a tail). Observe that $X = X_1+X_2+\ldots+X_n$, and note that $E(X_i) = 1/2$.
- From the result above,

  $$E(X) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n) = 1/2 + 1/2 + \ldots + 1/2 = n/2.$$

  So the expected number of heads in n tosses of a fair coin is, naturally, enough, half the number of tosses.
- And since the n tosses are independent, we have that

  $$V(X) = V(X_1) + V(X_2) + \ldots + V(X_n).$$

  But what is $V(X_i)$? Use the definition:

  $$V(X_i) = \sum_{j=1}^{2}(x_j - E(X_i))^2 f(x_j) = (0 - .5)^2 * .5 + (1 - .5)^2 * .5 = 1/4.$$
- So the variance of the number of heads in n tosses is just $n/4$, and the standard deviation is $\sqrt{n}/2$.
- So, if it happens (as we will show soon) that in a large number of tosses – say 100 – the number of heads that appears has a distribution that is approximately normal, then about 68% of the time the number of heads that appears will be between 45 and 55 (since $\sqrt{100}/2 = 5$. In 1000 tosses, 68% of the time the number of heads should fall within about 16 of 500, since $16 \approx \sqrt{1000/4}$.
- The above analysis was for a *fair* coin, but we can easily generalize it to the case of a sequence of *Bernoulli trials*, which are like flips of a coin that lands heads with probability $p \in (0, 1)$. Let X be the total number of "successes," which occur on each trial with probability p. The results are:

  $$\begin{aligned} E(X) &= np. \\ V(X) &= np(1 - p). \end{aligned}$$

  I *strongly* urge that you work through this on your own to convince yourself that this is true (just follow the same steps as above). This comes in very handy for hypothesis testing when the data you have is in the form of proportions.

8 Some important distributions

A quick pass through some important probability distributions, plus how to generate variables with these distributions in Stata ...

8.1 The uniform distribution

- As we saw, a random variable X is distributed uniformly on $[0,1]$ (i.e., $X \sim U[0,1]$) when its probability density function is

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \in [0,1] \\ 0 & \text{if } x > 1 \end{cases}$$

- Intuitively, this is like drawing a number from [0, 1] interval where all points in the interval are equally likely.
- Illustrate with Stata: **set obs 10000**, **gen x = uniform()**, **sum x**. What will **graph x ,bin(10)** look like?

8.2 Tossing coins with Stata

- We can use the random uniform distribution generator in Stata to simulate coin tossing: **set obs 10000**, **gen x = 1 if uniform() > .5**, **replace x = 0 if x == .**
- Checking the binomial distribution: Treat these 10,000 flips as 1000 "experiments" in which we toss the coin ten times and record the number of heads each time.
  - **gen caseno = _n**, **gen trial = int((caseno-1)/10)**, **egen heads = sum(x) ,by(trial)**, **graph heads ,bin(11)**, **tab heads**

8.3 The Poisson distribution

- Imagine a sequence of random events occuring in time, such as incoming calls at a telephone exchange, wars or international disputes occurring around the globe, cabinet dissolutions in a parliamentary government, or annual number killed in the Prussian army by horsekicks.
- Consider the random variable defined as the number of events occuring in some fixed interval of time, which we can arbitrarily set equal to 1. Now imagine dividing this

interval into equal segments of length $1/n$, where n is large. If it is substantively appropriate to suppose that the probability of an event occurring in any given subinterval is *constant*, and that as n grows large the probability of two or more events occuring in the same subinterval approaches zero sufficiently quickly, then this random variable X will have a Poisson distribution:

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

where $f(x)$ is the probability that x events occur, and λ ("lambda") is a parameter.

- Illustrate with Stata ... this is a discrete distribution potentially relevant for data on variables that are *counts* of some event.
- It is possible to show that if X has a Poisson distribution,

$$\mu(X) = var(X) = \lambda.$$

8.4 The normal distribution

- You have already seen that the probability density function for a normally distributed random variable X is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

  where $\mu = E(X)$ and $\sigma^2 = V(X)$.
- The notation $X \sim N(\mu, \sigma^2)$ reads "X has a normal distribution with mean μ and variance σ^2."
- When $\mu = 0$ and $\sigma = 1$, it is called a *standard normal distribution.*
- graph with Stata, review implications ... (area ...)
- How to draw a normally distributed random variable in Stata:
  - e.g.: **set obs 10000**, **gen x = invnorm(uniform())**, **graph x ,bin(50) normal**. Explain ...
- Some remarkable features of the normal distribution:
  - If $X \sim N(\mu, \sigma^2)$, then $Y = a+bX \sim N(a+b\mu, b^2\sigma^2)$. In words, a linear combination of a normal random variable is also a normal random variable.
    * This is related to the fact that the normal distribution depends on arbitrary mean and variance terms, so that multiplying by a constant and adding constant affects these but not the underlying "shape" of the distribution.

    * This is also the basis for the important property that FPP emphasize in many exercises: If you have a normally distributed variable X with mean μ and variance σ^2, you can calculate the probability that a draw from the distribution will have a value less than x by converting x to standard units ($z = (x - \mu)/\sigma$) and using a standard normal table.

  - If $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, then $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.
    So, sums of normal random variables are also normally distributed. Show with Stata. **gen x = invnorm(uniform())**, **gen y = invnorm(uniform())**, **gen z = x + y**, **graph z ,bin(50) normal**. Show also **graph y x, s(.)** and **graph z x ,s(.)**. Interpret ...

8.5 How to generate normally distributed variables with correlation ρ.

- Using above results, **corr z x**. Now **gen w = x + .3*y**, **corr w x**, **graph w x ,s(.)**. What happens? Why?
- Suppose you want to create two standard normal variables that have correlation ρ. First create two uncorrelated $N(0, 1)$ variables (like x and y above). Then

$$\textbf{gen z} = \rho\textbf{x} + \sqrt{1-\rho^2}\textbf{y}.$$

- Illustrate with Stata ...
- Where does this come from?
  - First, observe that

$$\begin{aligned} var(Z) &= var(\rho X + \sqrt{1-\rho^2}Y) \\ &= \rho^2 var(X) + (1-\rho^2)var(Y) \\ &= \rho^2 + 1 - \rho^2 \\ &= 1. \end{aligned}$$

    So Z will have variance of one also (in fact, regardless of whether X and Y are normal).

– Second, note that

$$
\begin{aligned}
\rho = \frac{cov(X,Z)}{\sqrt{var(X)var(Z)}} &= \frac{cov(X,\rho X+\sqrt{1-\rho^2}Y)}{\sqrt{var(X)var(\rho X+\sqrt{1-\rho^2}Y)}} \\
&= \frac{\frac{1}{n}\sum x_i(\rho x_i+\sqrt{1-\rho^2}y_i)}{\sqrt{var(X)(\rho^2 var(X)+(1-\rho^2)var(Y))}} \\
&= \frac{\frac{1}{n}\sum\left(\rho x_i^2+\sqrt{1-\rho^2}x_i y_i\right)}{\sqrt{\rho^2+(1-\rho^2)}} \\
&= \frac{\rho\frac{1}{n}\sum x_i^2+\sqrt{1-\rho^2}\frac{1}{n}\sum x_i y_i}{\sqrt{\rho^2+(1-\rho^2)}} \\
&= \frac{\rho var(X)+\sqrt{1-\rho^2}cov(X,Y)}{\sqrt{\rho^2+(1-\rho^2)}} \\
&= \frac{\rho}{\sqrt{\rho^2+(1-\rho^2)}} \\
&= \rho.
\end{aligned}
$$