NYU Department of Politics - Quant I
Fall 2013 - Prof. Patrick Egan

5 Lecture 5

5.1 A dip into multivariate analysis

- Although this first part of the course focuses on univariate analysis, we need to dip into multivariate analysis a bit in order to develop the theory we need governing how we draw inferences about a population from a sample. (You'll see why next week.)

- So let's take a brief excursion into the world of multivariate probability distributions. Until now, we've considered the probability distribution of one variable at a time. But sometimes we're interested in the probability of two or more variables. An example will help:

  - Let's imagine a hypothetical Congressional election where the Republican Party has a 73 percent chance of winning control of the House of Representatives, and an 18 percent chance of winning control of the Senate. We will consider these as two random variables, and call them Y_1 and Y_2, where $Y_i = 1$ if the G.O.P. wins control of the chamber and 0 if it does not.
  - In this context, let's call the event that the Republicans win control of either chamber a "success," and call Y_1 the Bernoulli experiment "observe whether the Republicans win a majority of seats in the House," and Y_2 the same experiment for the Senate.
  - We denote any particular realization of a pair of predictions as (y_1, y_2)–the "ordered pair"(y_1, y_2).
    - ∗ Just to be clear, the ordered pair $(y_1, y_2) = (y_2, y_1)$ if and only if $y_1 = y_2$.
  - We know $P(Y_1 = 1)$. It's .73. And we know $P(Y_2 = 1)$. It's .18. Do we necessarily know the probability that the Republicans win both chambers?
  - No. Define the event A as the intersection of the events $Y_1 = 1$ and $Y_2 = 1$.
    - ∗ i.e. $A = (Y_1 = 1 \cap Y_2 = 1)$.
    - ∗ What's $P(A)$? Is it $(.73)(.18) = .1314$?

    - ∗ Not necessarily. By the multiplicative law, it's $P(A) = P(Y_1 = 1 \cap Y_2 = 1) = P(Y_1 = 1) P(Y_2 = 1 | Y_1 = 1)$.
    - ∗ For $P(A) = (.73)(.18)$, it would have to be that control of the two chambers were *independent events*, allowing us to write
    - ∗ $P(A) = P(Y_1 = 1 \cap Y_2 = 1) = P(Y_1 = 1) P(Y_2 = 1)$.
    - ∗ And what's your sense about whether these are independent events?
    - ∗ If independent events:

		Y_1:GOP wins House		
		no $(Y_1 = 0)$	yes $(Y_1 = 1)$	totals
Y_2:	no $(Y_2 = 0)$	.22	.60	.82
GOP wins Senate	yes $(Y_2 = 1)$	.05	.13	.18
	totals	.27	.73	1

- – ∗ An example where not independent events, and we expect control of the two chambers to be dependent upon one another:

		Y_1:GOP wins House		
		no $(Y_1 = 0)$	yes $(Y_1 = 1)$	totals
Y_2:	no $(Y_2 = 0)$	.25	.57	.82
GOP wins Senate	yes $(Y_2 = 1)$	.02	.16	.18
	totals	.27	.73	1

- – ∗ Note how the probability assigned to the off-diagonals decreased, while that on the diagonals increased. But the probability of the two individual events remained the same.

  - What we're doing here is describing the **joint probability distribution** of the random variables Y_1 and Y_2.
  - As with univariate probability distributions, we describe joint probability distributions as the probabilities associated with all possible values of Y_1 and Y_2. Here we describe $P(Y_1 = y_1, Y_2 = y_2)$, or (to again use the shorthand) $p(y_1, y_2)$.

- More generically, if Y_1 and Y_2 are discrete RVs, the **joint probability function** for Y_1 and Y_2 is given by

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

- It is a function that returns the probability assigned to each of the possible ordered pairs, (y_1, y_2).
- In the example above, $p(y_1, y_2)$ doesn't reduce so nicely to a function. It makes more sense to represent this joint probability distribution with a table, as we have done.

- Similar rules from univariate world govern the assignment of joint probabilities. If Y_1 and Y_2 are discrete RVs with joint probability function $p(y_1, y_2)$, then (Axioms 1 and 2 again):

$$\begin{aligned} p(y_1, y_2) &\geq 0 \forall y_1, y_2 \\ \sum_{y_1, y_2} p(y_1, y_2) &= 1, \end{aligned}$$

where the sum is over all values of (y_1, y_2) assigned nonzero probabilities.

- Joint probability distributions can have **distribution functions**. (Notice difference between this and the **joint probability function** described above.) We write this function

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

- The joint distribution function is also at times called the **joint cumulative distribution function**, or the **joint CDF.**

- Just like in univariate world, joint distributions of random variables can be discrete or continuous.

- For two discrete random variables $Y_1$and Y_2, this is:

$$F(y_1, y_2) = \sum_{t_1 \leq y_1} \sum_{t_2 \leq y_2} p(t_1, t_2).$$

- Just to be clear, this equals the sum of the probabilities assigned to all the simple events in which (Y_1 takes on values $\leq y_1$ AND Y_2 takes on values $\leq y_2$).

- Two continuous random variables are said to be **jointly continuous** if their joint distribution function is continuous in both arguments. That is if there exists a nonnegative function $f(y_1, y_2)$, such that

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2) dt_2 dt_1 \quad \text{for } -\infty < y_1 < \infty, -\infty < y_2 < \infty,$$

then Y_1 and Y_2 are jointly continuous. And the function $f(y_1, y_2)$ is called the **joint probability density function** or the **joint density** or the **joint PDF**.

- As before, we can use joint CDFs to determine the probability that two random variables fall jointly into particular intervals, in this case $P(a < Y_1 \leq b, c < Y_2 \leq d)$, where

$$\begin{aligned} P(a &< Y_1 \leq b, c < Y_2 \leq d) = \int_c^d \int_a^b f(y_1, y_2) dy_1 dy_2 \\ &= F(b, d) - F(b, c) - F(a, d) + F(a, c). \text{ [Exercise.]} \end{aligned}$$

- Illustrate this on the board.

- Bivariate CDFs satisfy a set of properties that will look familiar:

$$F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0.$$

$$F(\infty, \infty) = 1.$$

$$y_1^* \geq y_1, y_2^* \geq y_2 \Rightarrow F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0, \text{ since}$$

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) = P(y_1 < Y_1 \leq y_1^*, y_2 < Y_2 \leq y_2^*) \geq 0.$$

- And furthermore, if Y_1 and Y_2 jointly continuous,

$$\begin{aligned} f(y_1, y_2) &\geq 0 \forall y_1, y_2, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 &= 1. \end{aligned}$$

5.2 Marginal probability distributions

- Note that all the bivariate events $(Y_1 = y_1, Y_2 = y_2)$, as represented by the ordered pairs (y_1, y_2), are mutually exclusive events.

- So the univariate event $(Y_1 = y_1)$ can be thought of as the **union** of bivariate events, with the union being taken over all possible values for y_2.

- So, consider the roll of two six-sided dice:

$$\begin{aligned} P(Y_1 &= 1) = p(1,1) + p(1,2) + ... + p(1,6) \\ &= 6 \times \frac{1}{36} = \frac{1}{6}. \\ &\text{Generically,} \\ P(Y_1 &= y_1) = p_1(y_1) = \sum_{all \quad y_2} p(y_1, y_2). \end{aligned}$$

- We call $p_1(y_1)$ the **marginal probability function** of the discrete RV Y_1. (What's the marginal probability function for Y_2?)

$$P(Y_2 = y_2) = p_2(y_2) = \sum_{all \quad y_1} p(y_1, y_2).$$

- In the continuous case, the **marginal density function** of the continuous RV Y_1 is:

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2.$$

  - (What's the marginal density function for Y_2?)

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1.$$

5.3 Conditional probability distributions

- Now we turn to the notion of **conditional distributions.** Recall that $P(A \cap B) = P(A)P(B|A)$ (multiplicative law).

- Well, the bivariate event (y_1, y_2) is of course just another way to describe the intersection of the two numerical events $Y_1 = y_1$ and $Y_2 = y_2$. So we may write

$$\begin{aligned} p(y_1, y_2) &= p_1(y_1)p(y_2|y_1) \\ &= p_2(y_2)p(y_1|y_2), \end{aligned}$$

where

$p_1(y_1), p_2(y_2)$ are (again) the marginal probability functions associated with y_1 and $y_2,$ and

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}, p_2(y_2) > 0.$$

- This defined as the **conditional discrete probability function** of Y_1 given Y_2.

- In the case of continuous RVs, we adjust the concept accordingly:

$$P(Y_1 \leq y_1|Y_2 = y_2) = F(y_1|y_2),$$

called the **conditional distribution function** of Y_1 given $Y_2 = y_2$.

- Similarly, we write the **conditional density function of** Y_1 given $Y_2 = y_2$ as

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}.$$

Note its similarity to the conditional probability function in the discrete case.

5.4 Independent random variables

- If the previous topic feels a bit rushed, that's because it was. You will be assigned exercises in the book to get you used to working with marginal and conditional probability functions. But we rushed to get to the material that is needed at hand to describe the way we make inferences from samples.

- We begin to do this now by extending the notion of independent events to define the idea of an **independent random variable**. Recall that two events are independent if

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

- Now consider a event involving two random variables, the event:

$$(a < Y_1 \leq b) \cap (c < Y_2 \leq d).$$

- This is an event composed of the two events

$$a < Y_1 \leq b \text{ and } c < Y_2 \leq d.$$

- For consistency, we'd like it to be the case that if

$$Y_1, Y_2 \text{ independent} \Rightarrow P(a < Y_1 \leq b, c < Y_2 \leq d) = P(a < Y_1 \leq b)P(c < Y_2 \leq d).$$

  - That is, the joint probability of two independent RVs can be written as the product of their marginal probabilities.

- So we'll do just that: random variables Y_1 and Y_2 are defined to be **independent** iff

$$Y_1 \text{ and } Y_2 \text{ independent} \Leftrightarrow F(y_1, y_2) = F_1(y_1)F_2(y_2) \text{ for every pair } (y_1, y_2),$$

  where $F(y_1, y_2)$ is the joint CDF for $Y_1$and Y_2 and $F_1(y_1)$ is the CDF for Y_1 and $F_2(y_2)$ is the CDF for Y_2. If $Y_1$and Y_2 are not independent they are by definition **dependent**.

- If follows [proof omitted] that

$$\begin{aligned} Y_1, Y_2 \text{ independent} \quad &\Leftrightarrow \quad p(y_1, y_2) = p_1(y_1)p_2(y_2) \text{ [discrete RVs]} \\ &\Leftrightarrow \quad f(y_1, y_2) = f_1(y_1)f_2(y_2) \text{ [continuous RVs].} \end{aligned}$$

- One final result is that independence of Y_1, Y_2 implies that we can write the joint density of the two RVs as the product of functions only of y_1 and y_2:

$$Y_1, Y_2 \text{ independent} \Leftrightarrow f(y_1, y_2) = g(y_1)h(y_2),$$

where $g()$ and $h()$ are non-negative functions of $y_1$and y_2 alone. This means that if we want to prove two RVs are independent, we can do so by finding two functions that satisfy these properties.

5.5 The EV of a function of RVs

- Recall that in univariate world, we talk a lot about the function of a random variable Y, say $g(Y)$. We showed that the expected value of this function is

$$\begin{aligned} E[g(Y)] &= \sum_y g(y)p(y) \text{ [discrete RV } Y] \\ &= \int_{-\infty}^{\infty} g(y)f(y)dy \text{ [continuous RV } Y] \end{aligned}$$

- Well, we can also talk about functions of random variables (plural).
  - For example, one function of random variables $Y_1, Y_2, ..., Y_k$ about which we are particularly interested is their **mean**, which is the function

$$\overline{Y} = g(Y_1, Y_2, ..., Y_k) = \frac{1}{K}\sum_{i=1}^{k} Y_i$$

- And we often find ourselves interested in the expected value of such a function. Recall that we showed that the expected value of a function of one random variable, $g(Y)$, is:

$$E[g(Y)] = \sum_y g(y)p(y).$$

Well, analogously to univariate world, we define the expected value of a function of several random variables $g(Y_1, Y_2, ..., Y_k)$ as:

$$E[g(Y_1, Y_2, ..., Y_k)] = \sum_{y_k} ... \sum_{y_2} \sum_{y_1} g(y_1, y_2, ..., y_k) p(y_1, y_2, ..., y_k),$$

where $p(y_1, y_2, ..., y_k)$ is the joint probability function of the k random variables. (Note that we're just extending the notion of joint probability from two to k RVs here.) This, is of course, for the discrete case. In the continuous case, we write

$$E[g(Y_1, Y_2, ..., Y_k)] = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2, ..., y_k) f(y_1, y_2, ..., y_k) dy_1 dy_2 ... dy_k.$$

- Well, just as in the case of the expected value of one RV, analogous results hold for the functions of several random variables. Where $g(Y_1, Y_2)$ is a function of the RVs Y_1 and Y_2,

$$E[cg(Y_1, Y_2)] = cE[g(Y_1, Y_2)].$$

- And furthermore, where we have a total of k functions of these random variables $g_1(Y_1, Y_2), g_2(Y_1, Y_2)$, $...g_k(Y_1, Y_2)$, we can "distribute expectations" over the sum of these functions:

$$E[g_1(Y_1, Y_2) + g_2(Y_1, Y_2) + ... + g_k(Y_1, Y_2)] = E[g_1(Y_1, Y_2)] + E[g_2(Y_1, Y_2)] + ... + E[g_k(Y_1, Y_2)]$$

- A powerful result is that, if Y_1 and Y_2 are independent, and if $g(Y_1)$ and $h(Y_2)$ are functions only of $Y_1$and Y_2, then

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)].$$

- Proof is omitted here, but it is intuitive and found on page 260 of your text.
- (in the continuous case):

$E[g(Y_1)h(Y_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1)h(y_2)f(y_1, y_2)dy_2dy_1$ [definition of the expected value of a function of rando

$$\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1)h(y_2)f_1(y_1)f_2(y_2)dy_2dy_1 \text{ (since } Y_1\text{, } Y_2 \text{ independent)} \\
&= \int_{-\infty}^{\infty} g(y_1)f_1(y_1)\left[\int_{-\infty}^{\infty} h(y_2)f_2(y_2)dy_2\right]dy_1 \text{ (pulling functions of } y_1 \text{ out of second integral)} \\
&= \int_{-\infty}^{\infty} g(y_1)f_1(y_1)E[h(y_2)]dy_1 \text{ (definition of expected value)} \\
&= E[h(y_2)]\int_{-\infty}^{\infty} g(y_1)f_1(y_1)dy_1
\end{aligned}$$

($E[h(y_2)]$ is a constant with regard to y_1 and can be pulled out of integral)

$$= E[g(Y_1)]E[h(Y_2)] \text{ (definition of expected value again.)}$$

5.6 Covariance of Two Random Variables

- Let's take a breath and reconsider the notion of independence in the context of our definition of a random variable. You'll recall that a RV is a function for which the domain is a sample space. It maps every sample point/simple event to a real number.

- Draw diagram here of RV Y_1 with sample space S_1.

- We we say that two random variables Y_1 and Y_2 are **independent**, we are saying that:

  - their joint probability function is equal to the product of their individual probability functions [discrete world].

  - Or we say that their joint PDF is equal to the product of their individual PDFs [continuous world].

- But (now draw diagram of RV Y_2 with sample space S_2) in the context of the definition of a random variable, we are saying that the realization of Y_2 is unrelated to the realization of Y_1. These are two separate processes.

- What happens, however, if the two realizations *are* related? That is, given that you know the value of Y_1, you are able to make a better than random guess about Y_2. Well, we have a way to describe how much the two processes are related, and it is the property of **covariance**.

- We define covariance as

$$COV(Y_1, Y_2) \equiv E[(Y_1 - \mu_1)(Y_2 - \mu_2)],$$

where μ_1, μ_2 are the means of RVs Y_1 and Y_2.

- To get a feel for what we mean by covariance, consider three hypothetical distributions of the observed values of random variables Y_1 and Y_2.

  ftbpF4.1494in5.7218in0incovariance.tif

- In panel A, we say that Y_1 and Y_2 have a _______________ relationship (positive).

- In panel B, we say that Y_1 and Y_2 have a _______________ relationship (negative).

- In panel C, we say that they have **no** relationship.

- Now consider two quantities: $(y_1 - \mu_1)$ and $(y_2 - \mu_2)$,(called *deviations*, or deviations from the mean) and their product, $(y_1 - \mu_1)(y_2 - \mu_2)$.

- Let's talk about the **sign** of this product.

  - In panel A: When the first multiplicand is positive, so is the second. When the first is negative, so is the second. The sign of this product is therefore always positive.
  - Now consider panel B. For similar reasons, the product here is always negative.
  - And panel C? We don't know: not clear.

- Now let's talk about the **magnitude** of this product.

  - In panel A, big deviations on y_1 are paired with big deviations on y_2. This makes the product bigger than if, say, big deviations on y_1 were paired with small deviations on y_2 and vice-versa.
  - What about panel B? (Same.)
  - And C? Well exactly what makes the product smaller: big deviations are not necessarily paired with big deviations.

- Now, you can see the logic of using $E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$ as our definition of covariance. It is, literally:
  - how much the size of the deviations of Y_1 and Y_2 from their means tend to vary with one another,
  - signed in the direction of the relationship between the two variables.
- Now draw a panel D that looks like C but with bigger axes.
- *Ceteris paribus*, what can we say about the covariance of these two variables versus those in C? It's larger–simply because we've changed the scale, not because they covary to a greater degree. This is problematic for comparing variables on different scales.
- To handle this challenge, we often standardize the value of a covariance by the product of the two variables' standard deviations. We call this standardized value a **correlation coefficient**, defined as

$$\rho_{(Y_1,Y_2)} \equiv \frac{COV(Y_1, Y_2)}{\sigma_1 \sigma_2}.$$

- A proof that it is always the case that $-1 \leq \rho \leq 1$ will be on your homework.
- Note that covariance and correlation are good at detecting/measuring the strength of a *linear* relationship. Not at measuring the strength of other relationships. (Draw curvilinear relationship on board.) Thus (we'll see later), while independence of Y_1, Y_2 implies $COV(Y_1, Y_2)$ = 0, the converse is not true.

6 Lecture 6

6.1 Some additional helpful results regarding the math of expectations

- Go over handout: Some Additional Helpful Results Regarding the Math of Expectations. Have students do proofs on board (without benefit of handout). [This can easily take 45 min-hour.]