9/27/2022

Brief introduction to multivariable calculus

Before we can talk about our last bit of probability distributions, we need to do a bit of a crash course in the basic of multivariable calculus. We'll stick to two variables in rectangular coordinates and no fancy integration techniques. This is just enough to get the flavor of some more advanced probability ideas that we'll need for some topics, especially regression next semester (we'll review these ideas at that time as well).

Partial derivatives

When you have a function of more than one variable, such as $f(x, y) = x^2 + 3xy - y^3$, there isn't only one derivative. There is a first derivative with respect to each variable. If we want to take the derivative of $f(x, y)$ with respect to x, the notation is $f_x(x, y)$ or $\frac{\partial f}{\partial x}$. We use this symbol ∂ (pronounced "del") to indicate that this is a partial derivative. If we take the derivative with respect to y, we write $f_y(x, y)$ or $\frac{\partial f}{\partial y}$.

The basic idea for partial derivatives from a process perspective is to pretend all other variables in the problem are constant (just as we would with an unknown parameter).

For our example function

$$f_x(x, y) = 2x + 3y$$

$$f_y(x, y) = 3x - 3y^2$$

We can take more derivatives the same way, and we are not required to stay in the same variable.

$$\frac{\partial^2 f}{\partial x^2} = f_{xx}(x, y) = 2$$
$$\frac{\partial^2 f}{\partial y^2} = f_{yy}(x, y) = -6y$$
$$\frac{\partial^2 f}{\partial y \partial x} = f_{xy}(x, y) = 3$$

Integration works similarly. We integrate with respect to one variable at a time, treating the other variable as a constant. One complication is that we usually have definite integrals rather than indefinite integrals, and the inner integral can take the outer variable in the limits, while the outer variable must be constant.

Suppose I want integrate our example function. It might look something like this.

$$\int_0^1 \int_0^x x^2 + 3xy - y^3 dydx$$

The limits of integration describe a region in the xy-plane. In this case the region between 0 and 1 in x, but under the line $y = x$ and the line $y = 0$.

The inside integral is done first. Here, we do the antiderivative in y, and again treat x as a constant.

$$\int_0^1 \left[x^2y+\frac{3}{2}xy^2-\frac{1}{4}y^4\right]_0^x dx=\int_0^1 x^3+\frac{3}{2}x^3-\frac{1}{4}x^4dx=\int_0^1 \frac{5}{2}x^3-\frac{1}{4}x^4dx$$

The last step is just like before to complete the integration.

$$\int_0^1 \frac{5}{2}x^3-\frac{1}{4}x^4dx=\frac{5}{8}x^4-\frac{1}{20}x^5\Big|_0^1=\frac{5}{8}-\frac{1}{20}=\frac{23}{40}$$

This is not a probability distribution, so we did not get 1. But if we multiplied by a scalar ($\frac{40}{23}$ in this case), then we could turn it into a probability distribution, as long as we were sure that the distribution was never negative on any part of the domain.

If the region in the plane is a rectangle, then both sets of limits are constant, but that isn't necessarily always the case.

Joint probability distributions

Just as with the single random variable case, we must deal with both the discrete case and the continuous case with probability density functions.

In the discrete case

$$P(X=x,Y=y)=p(x,y)$$

And which follows the usual rules that all values of $0\le p(x,y)\le 1$, and

$$\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)=1$$

Discrete probabilities can be expressed with piecewise function notation, but is often expressed in the form of a table (this becomes more difficult when there are more than two variables).

$p(x,y)$		y: 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

In the continuous case, the pdf is a function of two (or more) variables. In the two-variable case

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dydx=1$$

And $0\le f(x,y)\le 1$ for all allowable values of x and y.

To find the probability that (X, Y) is in some region, we integrate the function within those limits.

Note: If the function has more than two variables, it will need one integral for each variable, but we will stick with the 2D case here.

To break the probabilities down into their single variable cases, these are called marginal probabilities. In the discrete case

$p_X(x) = \sum_y p(x, y)$ for each value of x
$p_Y(y) = \sum_x p(x, y)$ for each value of y

In the continuous case

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Notice that the ideas are similar, we just replace the summations with integrals.

We'll do a complete integration example later, so let's look at the marginal distributions for our discrete case.

Our $p_X(x, y)$ becomes

x	100	250
$p_X(x)$	0.5	0.5

Our $p_Y(x, y)$ becomes

y	0	100	200
$p_Y(y)$	0.25	0.25	0.5

If the probabilities are independent, then the product of the marginal probabilities is the same as the original probabilities.

$$p(x, y) = p_X(x)p_Y(y)$$
$$f(x, y) = f_X(x)f_Y(x)$$

Our discrete case is not independent. $P(X = 100) = 0.5, P(Y = 0) = 0.25$, but $P(X = 100, Y = 0) = 0.20 \neq (0.5)(0.25)$.

We can talk about conditional probabilities in the joint case.

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

Thus, the conditional probabilities are the two-variable distribution divided by the marginal distribution. We can generalize this to the $x|y$ case, and the discrete case.

The expected values are found similarly to the one variable case.

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dydx$$

$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx$$

In the discrete case

$$E(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p(x_i, y_j)$$

$$E(Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} y_j p(x_i, y_j)$$

The variances are calculated similarly as before, by multiplying our pdf by $(x-\mu_X)^2$ or $(y-\mu_Y)^2$ respectively.

Covariance

One new idea we have with two variables is the covariance, which measures how two the variables are related to each other. This is calculated as

$$Cov(X,Y) = E[(x-\mu_X)(y-\mu_Y)]$$

As with variance, we have an alternative formulation that produces an equivalent result.

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y$$

Covariance leads us to an idea that will be important when we tackle regression next semester: correlation coefficient.

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

Correlation is a value that can only take on values in the interval [-1,1]. Values closer to 1 or -1 show a stronger relationship, while values closer to 0 show a weaker linear relationship.

Linear transformations of X and Y do not change the correlation value which is useful since it is scale invariant.

Let’s look at a complete continuous example.

Consider the joint pdf $f(x,y) = Kxy, 0 \leq x \leq 1, 0 \leq y \leq 1$.

1. Find the value of K that makes this a valid probability distribution.

$$\int_0^1 \int_0^1 Kxydydx = \int_0^1 K\left(\frac{1}{2}xy^2\right]_0^1 = \int_0^1 \frac{K}{2}xdx = \frac{K}{4}x^2\Big|_0^1 = \frac{1}{4}K = 1$$

   So, we set $K = 4$.

2. Find the marginal pdf for x and y respectively.

$$f_X(x) = \int_0^1 4xydy = 2xy^2|_0^1 = 2x$$

$$f_Y(y) = \int_0^1 4xydx = 2x^2y|_0^1 = 2y$$

3. Are the variables independent?
   Yes, since $f_X(x)f_Y(y) = (2x)(2y) = 4xy = f(x,y)$

4. What is the expected value of x and y respectively?

$$E(X) = \int_0^1 \int_0^1 x(4xy)dydx = \int_0^1 \int_0^1 4x^2ydydx = \int_0^1 2x^2dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3}$$

The math for $E(Y)$ is exactly the same.

5. Let's calculate the variance.

$$E(X^2) = \int_0^1 \int_0^1 x^2(4xy)dydx = \int_0^1 \int_0^1 4x^3ydydx = \int_0^1 2x^3dx = \frac{1}{2}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

The math for $V(Y)$ is exactly the same.

6. Let's find the covariance.

$$E(XY) = \int_0^1 \int_0^1 xy(4xy)dydx = \int_0^1 \int_0^1 4x^2y^2dydx = \int_0^1 \frac{4}{3}x^2dx = \frac{4}{9}$$

$$Cov(X) = E(XY) - \mu_X\mu_y = \frac{4}{9} - \frac{2}{3}\left(\frac{2}{3}\right) = 0$$

7. The correlation is therefore also 0.

Review for exam #1

Format:

The exam will include a “take-home” portion. You will be given a dataset to analyze in R, which you will do at home. You’ll be directed to perform certain types of analyses, build graphs, etc. You will bring these materials back with you on test day. Part of the exam will be to answer questions about these analyses. Ideally, you will have the answers already prepared. You may be asked to provide a code chunk used to perform a computation. The second part of the exam will be analyses to complete in class, explanations, etc. Questions that require only the use of a basic calculator, or no calculations at all (interpreting, for example).

Topics to focus on:

- Sampling methods
- Variable types
- Experiments and ethics
- Calculating descriptive statistics
- Making graphs
- Interpreting graphs
- Basic probability
- Bayes’ rule
- Binomial and normal distributions
- Continuous distributions (calculus)
- Joint probability distributions

Next class is Exam #1

References:

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. http://betsymccall.net/prof/courses/mathnotes/handouts/statistics/2470MLEs.pdf