Math 3070, Applied Statistics

Section 1

October 14, 2019

Lecture Outline, 10/14

Section 5.2

- Double Integration
- Expected Value
- Correlation and Covariance

Double Integration

Due to course prerequisites, the following adjustment will be made.

- Brief introduction to double integral will be presented.
- Lectures will contain double integrals.
- Homework problems that cannot be reduced to single integrals will be graded on completion.
- Quiz and exam questions will not require double integration. **Questions may reference multiple random variables and joint PDFs.** In these question, random variables will usually be assumed to be independent.

Double Integration, Quiz and Exam Questions

X, Y are continuous and independent, $f(x,y) = f_X(x)f_Y(y)$.
Events:

$$P(a < X < b, c < Y < d) = P(a < X < b)P(c < Y < d)$$
$$= \int_a^b f_X(x)dx \int_a^b f_Y(y)dy$$

Conditional and marginal PDFs, and conditioned events:

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

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

$$P(a < X < b | c < Y < d) = \frac{P(a < X < b)P(c < Y < d)}{P(c < Y < d)}$$
$$= P(a < X < b)$$

Double Integration, Procedure

Using double integrals

1. Identify the limits of integration from the region of integration. May have to write limit of one variable in terms of the other. PDFs may impose more restrictions. Drawing a picture may help.
2. Integrate the inner integral. Consider the other variable to be a constant.
3. Integrate the outer integral.

✓✓ If any variables are left over, the integral was likely set up incorrectly in step 1. Probabilities are less than 1.

Double Integration, Example

X and Y are lifetimes of components and have the following joint PDF. Determine K.

$$f(x,y) = \begin{cases} K(x^2+y^2), & 0 < x, y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Want: $1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy$
Step 1:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^1 \int_0^1 K(x^2+y^2)dxdy$$

Double Integration, Example

Step 2:

$$\int_0^1 \int_0^1 K(x^2+y^2)dxdy = K\int_0^1 \left(\frac{x^3}{3} + xy^2\bigg|_{x=0}^{1}\right)dy$$
$$= K\int_0^1 \frac{1}{3} + y^2 dy$$

Step 3:

$$K\int_0^1 \frac{1}{3} + y^2 dy = K\left(\frac{y}{3} + \frac{y^3}{3}\bigg|_{x=0}^{1}\right)$$
$$= K\left(\frac{1}{3} + \frac{1}{3}\right)$$

$$1 = K\frac{2}{3} \rightarrow K = \frac{3}{2}$$

Double Integration, Example with Sum

Consider X and Y to be independent exponential random variables with parameter $\lambda = 1$. Compute the probability that $X + Y < 3$. Independence implies

$$f(x,y) = \begin{cases} e^{-x}e^{-y}, & x,y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Step 1: $X < 3 - Y$ or $Y < 3 - X$ can both be used as a restriction.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{3-y} f(x,y)dxdy \quad \text{or} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{3-x} f(x,y)dydx$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{3-y} f(x,y)dxdy = \int_{0}^{3}\int_{0}^{3-y} e^{-x}e^{-y}dxdy$$

Double Integration, Example with Sum

Step 2:

$$\int_0^3 \int_0^{3-y} e^{-x} e^{-y}\, dxdy = \int_0^3 e^{-y} \left(-e^{-x} \Big|_{x=0}^{3-y} \right) dy$$

$$= \int_0^3 e^{-y} (1 - e^{-(3-y)}) dy$$

$$= \int_0^3 e^{-y} - e^{-3} dy$$

Step 3:

$$\int_0^3 e^{-y} - e^{-y-3} dy = -e^{-y} - ye^{-3} \Big|_{y=0}^{3}$$

$$= 1 - 4e^{-3} \approx 0.800851727$$

Expected Value

The expected value of function of discrete random variables is

$$E[h(X, Y)] = \sum_x \sum_y h(x, y) f(x, y)$$

For continuous random variables the expected value is

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy$$

The average value of a function of random variables h. For example, h could be a length times a length, a random rate times a random time, the sample mean and so on...

Expected Value, Example 1

The joint PMF of X and Y are given below. Determine the expected value of XY.

$f(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	.20	.10	.20
$x=1$	.05	.15	.30

$$\begin{aligned}
E[XY] &= \sum_x \sum_y xyf(x,y) \\
&= (0)(0)0.20 + (0)(1)0.10 + (0)(2)0.20 \\
&+ (1)(0)0.05 + (1)(1)0.15 + (1)(2)0.30 \\
&= 0.75
\end{aligned}$$

Expected Value, Example 2

The joint PDF of X and Y are given below. Determine the expected value of X.

$$f(x,y) = \begin{cases} \frac{3}{2}(x^2+y^2), & 0 \leq x, y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy \\ &= \int_0^1\int_0^1 x\frac{3}{2}(x^2+y^2)dxdy = \frac{3}{2}\int_0^1\int_0^1 x^3+xy^2dxdy \\ &= \frac{3}{2}\int_0^1\left(\frac{x^4}{4}+\frac{x^2}{2}y^2\Big|_{x=0}^{1}\right)dy = \frac{3}{2}\int_0^1 \frac{1}{4}+\frac{y^2}{2}dy \\ &= \frac{3}{2}\left(\frac{y}{4}+\frac{y^3}{6}\Big|_{y=0}^{1}\right) = \frac{5}{8} \end{aligned}$$

Expected Value and Independence

Assume X and Y are independent.

$$E[h(X) + g(Y) + c] = E[h(X)] + E[g(Y)] + c$$

Continous Random Variables (discrete case replaces $\int$ with $\sum$):

$$E[X + Y] = \int \int (h(x) + g(y) + c) f_X(x) f_Y(y) dxdy$$

$$= \int \int h(x) f_X(x) f_Y(y) dxdy + \int \int g(y) f_X(x) f_Y(y) dxdy$$

$$+ c \int \int f_X(x) f_Y(y) dxdy$$

$$= \int f_Y(y) \int h(x) f_X(x) dxdy + \int g(y) f_Y(y) \int f_X(x) dxdy + c$$

$$= \int f_Y(y) E[h(X)] dy + \int g(y) f_Y(y) \int f_X(x) dxdy + c$$

$$= E[h(X)] \int f_Y(y) dy + \int g(y) f_Y(y) dy + c = E[h(X)] + E[g(Y)] + c$$

Expected Value and Independence

Assume X and Y are independent.

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)]$$

Continous Random Variables (discrete case replaces $\int$ with $\sum$):

$$\begin{aligned}
E[h(X)g(Y)] &= \int \int h(x)g(y)f_X(x)f_Y(y)dxdy \\
&= \int g(y)f_Y(y) \int h(x)f_X(x)dxdy \\
&= \int g(y)f_Y(y)E[h(X)]dy \\
&= E[h(X)] \int g(y)f_Y(y)dy \\
&= E[h(X)]E[g(Y)]
\end{aligned}$$

Variance and Independence

Assume X and Y are independent.

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$$

$$\begin{aligned}
E[(aX + bY + c)^2] &= E[a^2X^2 + bY^2 + c^2 \\
&+ 2abXY + 2caX + 2bcY] \\
&= a^2E[X^2] + bE[Y^2] + c^2 \\
&+ 2abE[XY] + 2caE[X] + 2bcE[Y] \\
E[(aX + bY + c)]^2 &= (aE[X] + bE[Y] + c)^2 \\
&= a^2E[X]^2 + b^2E[Y]^2 + c^2 \\
&+ 2abE[X]E[Y] + 2caE[X] + 2bcE[Y] \\
Var(aX + bY + c) &= E[(aX + bY + c)^2] - E[(aX + bY + c)]^2 \\
&= a^2(E[X^2] - E[X]^2) + b^2(E[Y^2] - E[Y]^2) \\
&+ E[XY] - E[X]E[Y] \\
&= a^2Var(X) + b^2Var(Y)
\end{aligned}$$

Example, Expected Value and Variance of a Binomial

Recall that a Binomial random variable can also be defined to be the sum of n independent Bernoulli random variables each with parameter p. Use this definition to compute the mean and variance.

$X_i \sim Bern(p), \;\; 1 \dots n.$
$Y = \sum_{i=1}^{n} X_i.$
$E[X_i] = p, \;\; Var(X_i) = p(1-p)$

$$E[Y] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$

$$Var(Y) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Example

X is a the midterm 1 score of a randomly selected student and Y is their midterm 2 score. Assume that X and Y are independent, $X \sim N(85, 10)$ and $Y \sim N(75, 14)$. Compute the expected value and variance of the student's average score.

$$\text{student's midterm average} = Z = \frac{X+Y}{2}$$

$$E\left[\frac{X+Y}{2}\right] = \frac{E[X]+E[Y]}{2} = \frac{85+75}{2} = 80$$

$$Var\left(\frac{X+Y}{2}\right) = \frac{Var(X)}{4} + \frac{Var(Y)}{4} = \frac{10+14}{4} = 6$$

Summary

- Discrete:

$$E[h(X, Y)] = \sum_x \sum_y h(x, y) f(x, y)$$

- Continuous:

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dxdy$$

- Assuming Independence,

$$E[g(X) + h(Y) + c] = E[g(X)] + E[h(Y)] + c$$

$$E[XY] = E[X]E[Y]$$

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$$

Covariance

Given random variables X and Y with means μ_X and μ_Y respectively, their **covariance** is

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

Expected change in X times change in Y. Simliar to the chapter 1 version.

As for variance, there is a shortcut formula for covariance:

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

Proof of the following similar those from expected value section.

1. $Cov(X, X) = V(X)$
2. $Cov(Y, X) = Cov(X, Y)$
3. $Cov(cX, Y) = cCov(X, Y)$
4. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$

Correlation

The **correlation** between two random variables X and Y is

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

Correlation is a measure of how strongly two random variables follow a linear relation. It has several properties:

1. $-1 \le \rho \le 1$
2. If X and Y are independent, then $Corr(X, Y) = 0$.
3. Changing the scale of X and/or Y does not affect the correlation, i.e. for any $c \neq 0$,

$$Corr(cX, Y) = Corr(X, Y) = Corr(X, cY)$$

Covariance and Correlation, and Independence

Assume X and Y are indpendent.

$$
\begin{aligned}
Cov(X, Y) &= E[XY] - E[X]E[Y] \\
&= E[X]E[Y] - E[X]E[Y] = 0
\end{aligned}
$$

- Covariance and Correlation will not appear much for this class. But, correlation and covariance can be incorporated into models.
- Specifically, the bivariate Normal distribution is frequently used to model correlated variables.
- Measures correlation and covariance based on these these ideas appear in data methods such as PCA and Factor analysis.

Example

In the bank example, the joint pdf of two random variables X and Y was given by

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the correlation of X and Y.

The marginal pdfs of X and Y are, for $0 \le x \le 1, 0 \le y \le 1$,

$$f_X(x) = \int_0^1 \frac{6}{5}(x+y^2)\ dy = \frac{6}{5}\left(xy + \frac{y^3}{3}\right)\Big|_{y=0}^{1} = \frac{6}{5}\left(x + \frac{1}{3}\right)$$

$$f_Y(y) = \int_0^1 \frac{6}{5}(x+y^2)\ dx = \frac{6}{5}\left(\frac{x^2}{2} + xy^2\right)\Big|_{x=0}^{1} = \frac{6}{5}\left(\frac{1}{2} + y^2\right)$$

Example (continued)

From the marginal pdfs $f_X(x) = \frac{6}{5}(x + \frac{1}{3})$ and $f_Y(y) = \frac{6}{5}(\frac{1}{2} + y^2)$,

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx = \frac{6}{5}\int_0^1 \left(x^2 + \frac{x}{3}\right) dx = \frac{6}{5}\left(\frac{1}{3} + \frac{1}{6}\right) = \frac{3}{5}$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \frac{6}{5}\int_0^1 \left(\frac{y}{2} + y^3\right) dy = \frac{6}{5}\left(\frac{1}{4} + \frac{1}{4}\right) = \frac{3}{5}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\ dx = \frac{6}{5}\int_0^1 \left(x^3 + \frac{x^2}{3}\right) dx = \frac{6}{5}\left(\frac{1}{4} + \frac{1}{9}\right) = \frac{13}{30}$$

$$E(Y^2) = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \frac{6}{5}\int_0^1 \left(\frac{y^2}{2} + y^4\right) dy = \frac{6}{5}\left(\frac{1}{6} + \frac{1}{5}\right) = \frac{11}{25}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{13}{30} - \left(\frac{3}{5}\right)^2 = \frac{11}{150}$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{11}{25} - \left(\frac{3}{5}\right)^2 = \frac{2}{25}$$

Example (continued)

Finally, from $f(x,y) = \frac{6}{5}(x+y^2)$ we calculate

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\ dx\ dy = \frac{6}{5}\int_0^1\int_0^1 (x^2y + xy^3)\ dx\ dy$$
$$= \frac{6}{5}\int_0^1 \left(\frac{x^3y}{3} + \frac{x^2y^3}{2}\right)\Big|_{x=0}^1\ dy$$
$$= \frac{6}{5}\int_0^1 \left(\frac{y}{3} + \frac{y^3}{2}\right)\ dy = \frac{6}{5}\left(\frac{1}{6} + \frac{1}{8}\right) = \frac{7}{20}$$
$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{7}{20} - \frac{3}{5}\cdot\frac{3}{5} = \frac{-1}{100}$$

Therefore, from $V(X) = 11/150$ and $V(Y) = 2/25$,

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{-1/100}{\sqrt{11/150}\sqrt{2/25}} \approx -.131$$