STAT 400 : DISCUSSION SECTION BD3 AND BD4 HANDOUT 6

SUBHADEEP PAUL

1. Theory and Discussions

1.1. **Bivariate distributions:**

1.1.1. *Important terms.*

- Set Up : We have a 2 dimensional real euclidean space and x and y represents two random variables. We will consider probabilities for points and events in 2D space. Later we will extend some of the notions to n dimensional space.
- In discrete case , Joint $f(x, y) = P(X = x, Y = y)$., Conditional $f(x|y) = P(X = x|Y = y)$ Marginal $f(x) = P(X = x)$. These are true probabilities. Probability of a set is defined as sum of the probabilities over all points ((x, y) pairs) in the set as before. Probability of the sample space is 1 and probabilities of exhaustive disjoint (mutually exclusive) events sum to 1 as before.
- In continuous case pdf is not a true probability. Probability of a set is defined as volume under a curve instead. Volume of the entire sample space is 1.

1.1.2. *Relationship between the terms.*

- Discrete case: Easy to derive from what we know about probabilities, i.e conditional=joint/marginal and marginal= sum of joint over all y, i.e $\sum f(x, y)$
- Continuous case, extends similarly, $f(x|y) = \frac{f(x,y)}{f(y)}$ is the conditional density, $f(x) = \int f(x, y)dy$.$f(y) = \int f(x, y)dx$ are the marginal densities.

1.1.3. *Expectation of different functions.*

- $E(g(x, y)) = \int\int g(x, y)f(x, y)dydx$. $E(xy) = \int\int xyf(x, y)dydx$, $E(x) = \int\int xf(x, y)dydx = \int xf(x)dx$, $E(y) = \int\int yf(x, y)dydx = \int yd(y)dy$.
- Conditional mean of x given y, $E(X|Y) = \int xf(x|y)dx$, $E(Y|X) = \int yf(y|x)dy$

1.1.4. *Independence.*

- The notion of independence is extended as, two RVs X and Y being independent iff their joint PMF/ PDF is the product of their marginal PMF/PDF at ALL points i.e $f(x, y) = f(x)f(y)$, or equivalently the conditional pmf/pdf of X given Y is same as the marginal pmf/pdf of X. i.e, $f(x|y) = f(x)$ for all x, y .

1.1.5. *Covariance/Correlation.*

- $\sigma_x^2 = E(x^2) - (E(x))^2$, $\sigma_y^2 = E(y^2) - (E(y))^2$, $\sigma_{xy} = E(xy) - E(x)E(y)$, $r_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$
- $Cov(aX + bY, cX + dY) = acVar(X) + (ad + bc)Cov(X, Y) + bdVar(Y)$

1.2. **We Prepare to start Statistics ! Sample and its distribution, CLT.**

- What is a population and what is a Sample? Statistics is the science of infering from data. What method I need to use depends on my data and the questions I am trying to answer. e.g suppose, my question is what is the mean of heights of students in this class. So the class (of 50 students) is my population and the height is my random variable which follows a certain distribution (say normal). I am interested in finding mean $\mu$of this random variable. Depening upon my time and resources I can take two approaches. 1) Record every one's height and take the average 2) select 10 students randomly, record their heights and take the average.
- A random sample is always i.i.d (independently and identically distributed). We compute different functions of the sample, which are called "statistic" to "estimate" population characteristics. e.g one such function of the sample is the sample mean $\bar{x}$. We will talk more about estimation later.
- $E(\bar{X}) = \mu$, $Var(\bar{X}) = \sigma^2/n$. (these is true irrespective of the distribution of X).
- $\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$ follows a standard normal distrbution under two conditions : (a) n is large X has any distibution, (b) n is any value, X has normal distribution.

2. Problems

Problem 1. (A question from final exam of stat 400, fall 2012.) Suppose two random variables X and Y have the following joint density function:

$$f(x,y) = \begin{cases} Cxy & 0 < x < 1, x + 4y \leq 1, \\ 0 & otherwise \end{cases}$$

(1) What must the value of C be so that this is a valid probability density function?
(2) Find the marginal density functions, $f(X)$ and $f(Y)$.
(3) Find $E(X)$ and $E(Y)$
(4) Find $\sigma^2(x)$ and $\sigma^2(y)$
(5) Find correlation coefficient between x and y .

Note. Here the marginals and the expectations for X and Y would be different since the domains are different. (otherwise they would have been same)

Problem 2. (Another problem from last fall's finals) Let X and Y have joint density $f(x, y) = Csin(6xy)$ for $0 < x < 1$ and $0 < y < 2$ and 0 otherwise, for appropriate constant C (which cannot be computed explicitly). In terms of C compute each of the following.

(1) Find the marginal density functions, $f(X)$ and $f(Y)$.
(2) Find $E(X)$ and $E(Y)$

Problem 3. (Midterm 2, last fall) Suppose $\sigma_x^2 = 20$, $\sigma_y^2 = 20$, $\sigma_{xy} = 20$, Find $cov(2x + 3y, 2x - 4y)$

Use direct formula to compute the covariance

Problem 4. (Midterm 2, last fall) Suppose you roll one fair six-sided die and then flip as many coins as the number showing on the die. (For example, if the die shows 4, then you flip four coins.) Let X be the number showing on the die, and Y be the number of heads obtained.

(1) What is the joint pmf of X and Y, i.e find an expression for $P(X = j, Y = k)$ in terms of j and k
(2) Find $P(X = Y)$