Statistics ST2334 Topic 4,5 addendum: Joint Distributions

2011/2012 Semester 2

Joint Distributions: Some Theory

Recall

- that a random variable X is a function that maps outcomes $\omega \in \Omega$ to $X(\omega) \in \mathbb{R}$.
- that a probability function P, maps events $A \subseteq \Omega$ to a number, and obeys the axioms of probability.
- that we can then apply P to X by specifying $P(X \in M) := P(\{\omega : X(\omega) \in M\})$ for $M \subseteq \mathbb{R}$.
- we then call this entire description of the values and the corresponding probability of X its probability distribution.

We can similarly define probability distributions for multiple random variables.

Joint Distributions: Some Theory

- Let X, Y be random variables, then

$$P(X \in M, Y \in N) := P(\{\omega : X(\omega) \in M\} \cap \{\upsilon : Y(\upsilon) \in N\})$$

where $\omega, \upsilon \in \Omega$.

- Note that there is nothing new going on on the RHS. It is applying the probability function P on a subset of Ω.
- This description of possible values X, Y can take and their corresponding probabilities is called the *joint* probability distribution of X and Y.
- We can of course extend this definition to any n number of random variables.

Joint Distributions: Discrete Case

- For discrete random variables, we extend our probability mass function to a function of multiple variables.
- E.g. For two discrete variables X_1 and X_2, their joint probability mass function is defined as

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

  where x_1, x_2 are possible values of X_1, X_2 respectively.
- Similarly, For n discrete random variables $X_1, \ldots, X_n$, their joint probability mass function is simply

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

Marginal Distributions: Discrete Case

- When given the joint distribution of say X_1, X_2, we call the distribution of X_1 (or X_2) alone the *marginal* distribution.
- Note that the marginal distributions are embedded in the joint distribution. We can recover the probability mass function f_1 of X_1 from the joint probability mass function f by

$$f_1(x_1) = P(X_1 = x_1) = \sum_{all \ x_2} P(X_1 = x_1, X_2 = x_2) = \sum_{all \ x_2} f(x_1, x_2)$$

- Similarly, if f is the joint probability mass function of $X_1, \ldots, X_n$, then the marginal probability mass functions are

$$f_i(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} f(x_1, \ldots, x_n)$$

Conditional Probability Distributions: Discrete Case

- Consistent with the definition of conditional probability of *events*, we can define the conditional probability distribution of X_1 given $X_2 = x_2$ as

$$f_1(x_1|x_2) = P(X_1 = x_1|X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$$
$$= \frac{f(x_1, x_2)}{f_2(x_2)}, \text{ provided } f_2(x_2) \neq 0$$

- If $f_1(x_1|x_2) = f_1(x_1)$ for all x_1 and x_2, then we say X_1 and X_2 are independent. This is equivalent to

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)$$

for all x_1 and x_2.

Example: Discrete Case

Let X_1 and X_2 have the joint probability distribution in the table below.

$f(x_1, x_2)$

		x_1		
		-1	0	1
x_2	-1	0.1	0.3	0.1
	1	0.15	0.25	0.1

- Find $P(X_1 + X_2 > 1)$.

$$P(X_1 + X_2 > 1) = P(X_1 = 1, X_2 = 1) = 0.1$$

- Find $P(X_1 X_2 = 0)$.

$$P(X_1 X_2 = 0) = P(X_1 = 0) = 0.3 + 0.25 = 0.55$$

Example: Discrete Case (continued)

- Find the marginal distributions of X_1 and X_2.

$f_1(x_1)$

x_1	-1	0	1
$f_1(x_1)$	0.25	0.55	0.1

$f_2(x_2)$

x_2	-1	1
$f_2(x_2)$	0.5	0.5

- Find the conditional distribution of X_2 given $X_1 = -1$

$$P(X_2 = x_2 | X_1 = -1) = f_2(x_2 | -1) = \frac{f(-1, x_2)}{f_1(-1)}$$

$f_2(x_2 | -1)$

x_2	-1	1
$f_2(x_2 \| -1)$	$0.1/0.25 = 0.4$	$0.15/0.25 = 0.6$

- Are X_1 and X_2 independent?

Joint Distributions: Continuous Case

- As we did before for single random variables, we introduce probability density functions to handle continuous random variables.
- Suppose $X_1, X_2, \ldots, X_n$ are n continuous random variables. Then the joint probability density function of these variables $f(x_1, x_2, \ldots, x_n)$ is such that

$$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, ..., a_n \leq X_n \leq b_n) \\ = \int_{a_1}^{b_1} \int_{a_2}^{b_2} ... \int_{a_n}^{b_n} f(x_1, x_2, ..., x_n) \; dx_1 dx_2 ... dx_n$$

- For the axions of probability to hold, we need
  - $f(x_1, x_2, \ldots, x_n) \geq 0$
  - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n) \; dx_1 dx_2 ... dx_n = 1$

Joint Distributions: Continuous Case

We define

- the joint cumulative distribution function

$$F(x_1, x_2, ..., x_n) := \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_n} f(u_1, u_2, ..., u_n)\ du_1 du_2 ... du_n$$

- the marginal probability density function for X_i

$$f_i(x_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(u_1, u_2, ..., u_n)\ du_1 ... du_{i-1} du_{i+1} ... du_n$$

- the marginal cdf for X_i

$$F_i(x_i) = \int_{-\infty}^{x_i} f_i(u_i) du_i$$

Conditional Distribution: Continuous Case

- Just as we did for the probability mass functions, we define the conditional probability density function

$$f_1(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

- And continuous random variables X_1, X_2 are independent if

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)$$

  for all values of x_1, x_2.
- This is equivalent to

$$F(x_1, x_2) = F_1(x_1)F_2(x_2)$$

  for all values of x_1, x_2.

Example: Continuous Case

If two variables X_1 and X_2 have a joint density function

$$f(x_1, x_2) = \begin{cases} \frac{1}{18} e^{-x_1} x_2^2, & \text{for } x_1 > 0, -3 < x_2 < 3 \\ 0, & \text{otherwise} \end{cases}$$

find the probabilities that

(a) the first random variable will take on a value between 0 and 1 and the second random variable a value between 1 and 3;

$$\begin{aligned}
\int_0^1 \int_1^3 f(x_1, x_2)\, dx_1 dx_2 &= \int_1^3 \left\{ \int_0^1 \frac{1}{18} e^{-x_1} x_2^2 \, dx_1 \right\} dx_2 \\
&= \frac{1}{18} \int_1^3 \left[-e^{-x_1} \right]_0^1 x_2^2 \, dx_2 \\
&= \frac{1}{18} \cdot (1 - e^{-1}) \left[\frac{1}{3} x_2^3 \right]_1^3 \\
&= \ldots = 0.304
\end{aligned}$$

Example: Continuous Case continued

(b) find the joint cumulative distribution function

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(x_1, x_2)\ dx_1 dx_2$$

- we need to be careful about the limits: split into 3 regions

$$= \begin{cases} 0, & \text{if } x_1 \le 0, x_2 \le -3 \\ \int_0^{x_1} \int_{-3}^{x_2} \frac{1}{18} e^{-x_1} x_2^2 du_1 du_2, & \text{if } x_1 > 0, -3 < x_2 < 3 \\ \int_0^{x_1} \int_{-3}^{3} \frac{1}{18} e^{-x_1} x_2^2 du_1 du_2, & \text{if } x_1 > 0, x_2 \ge 3 \end{cases}$$

$$= \begin{cases} 0, & \text{if } x_1 \le 0, x_2 \le -3 \\ \frac{1}{54}(1 - e^{-x_1})(x_2^3 + 27), & \text{if } x_1 > 0, -3 < x_2 < 3 \\ 1 - e^{-x_1}, & \text{if } x_1 > 0, x_2 \ge 3 \end{cases}$$

Example: Continuous Case continued continued

(c) are X_1 and X_2 independent?

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, u_2) du_2 \\ &= \begin{cases} \int_{-3}^{3} \frac{1}{18} e^{-x_1} u_2^2 \, du_2 = e^{-x_1}, & \text{if } x_1 > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Similarly,

$$f_2(x_2) = \int_{-\infty}^{\infty} f(u_1, x_2) du_1 = \begin{cases} \frac{1}{18} x_2^2, & \text{if } -3 < x_2 < 3 \\ 0, & \text{otherwise} \end{cases}$$

Thus

$$f_1(x_1) f_2(x_2) = f(x_1, x_2)$$

They are independent.

Expectation of multiple random variables

- In the discrete case,

$$E[g(X_1, X_2, ..., X_n)] = \sum_{x_1} \sum_{x_2} ... \sum_{x_n} g(x_1, x_2, ..., x_n) f(x_1, x_2, ..., x_n)$$

- In the continuous case,

$$\begin{aligned} E[g(X_1, X_2, ...X_n)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(x_1, x_2, ..., x_n) \\ &\quad \times f(x_1, x_2, ..., x_n)\ dx_1 dx_2 ... dx_n \end{aligned}$$

Expectation of multiple random variables continued

- For the previous example, taking $g(x_1, x_2) = x_1 + x_2$, we have

$$E(X_1 + X_2) = \int_0^{\infty} \int_{-3}^{3} (x_1 + x_2)(\frac{1}{18} e^{-x_1} x_2^2) \; dx_1 dx_2$$

- You can work it out directly, but recall for linear functions we can use the following theorem:

  For any random variables $X_1, ..., X_n$ and constants $a_0, a_1, ..., a_n$, we have

$$E(a_0 + a_1 X_1 + ... + a_n X_n) = a_0 + a_1 E(X_1) + ... + a_n E(X_n)$$

Covariance

- Define the covariance of X_i and X_j

$$Cov(X_i, X_j) = E\{(X_i - EX_i)(X_j - EX_j)\}$$

- This can also be written as

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

- We have already seen that for discrete random variables: if X_i, X_j are independent, the above expression $= 0$ (Topic 4c slide 4).

Covariance continued

- Similarly, for continuous X_i, X_j, if they are independent, then

$$\begin{aligned}
E(X_iX_j) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_ix_jf(x_i,x_j)\ dx_idx_j \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_ix_jf_i(x_i)f_j(x_j)\ dx_idx_j \\
&= \int_{-\infty}^{\infty} x_if_i(x_i)\ dx_i \int_{-\infty}^{\infty} x_jf_j(x_j)\ dx_j \\
&= E(X_i)E(X_j)
\end{aligned}$$

Hence, X_i, X_j independent $\Rightarrow Cov(X_i, X_j) = 0$

- Note that the converse is NOT true. i.e. two random variables with covariance 0 are not necessarily independent.

Variance

- For any random variables $X_1, ..., X_n$, it's not hard to see

$$Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n) + 2\sum_{j>i} Cov(X_i, X_j)$$

- You can see now that our earlier identity follows: If $X_1, ..., X_n$ are independent, then

$$Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n)$$