# Multivariate Probability distributions, chapter 6

Grethe Hystad
University of Arizona

April 13, 2011

# Multivariate Probability distributions

The National Highway Traffic Safety Administration is interested in the effect of seat belt use of saving lives. One study reported statistics on children under the age of 5 who were involved in motor vechilce accidents in which at least one fatality occurred. For 7060 such accidents between 1985 and 1989, the results were as shown:

| Seat Belt status | Survivors | Fatalities | Total |
|---:|---:|---:|---:|
| No belt | 1129 | 509 | 1638 |
| Adult belt | 432 | 73 | 505 |
| Children's car seat belt | 733 | 139 | 872 |
| Total | 2294 | 721 | 3015 |

# Multivariate Probability distributions

Define

$$X = \begin{cases} 0, & \text{if child survived;} \\ 1, & \text{if child did not survive} \end{cases}$$

which will keep track of number of fatalities per child. Children's car seats usually involve two belts. The regular adult belt in the vehicle, and the belt on the car seat itself. Define

$$Y = \begin{cases} 0, & \text{if no belt;} \\ 1, & \text{if adult belt used;} \\ 2, & \text{if child car seat belt used} \end{cases}$$

which is a random variable that keeps track of the number of belts.

# The Joint probability distribution

| | X | | |
|---|---|---|---|
| Y | 0 | 1 | |
| 0 | 0.38 | 0.17 | 0.55 |
| 1 | 0.14 | 0.02 | 0.16 |
| 2 | 0.24 | 0.05 | 0.29 |
| | 0.76 | 0.24 | 1.00 |

- Each entry in the table is $P(X = x, Y = y) = p(x, y)$.
- This is the joint probability function of $(x, y)$.
- For example $P(X = 0, Y = 2) = 0.24$, which is found as $\frac{733}{3015} \approx 0.24$ from the previous table.
- $P(X = 0, Y = 2)$ represent the probability that a child will both survive and use a children's carseat belt when involved in an accident in which at least one fatality occurred.

# The joint probability distribution

**Definition**

Let X and Y be discrete random variables. The joint probability distribution of X and Y is given by

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

defined for all real numbers x and y. A function $p(x, y)$ is a joint probability function iff

- $p(x, y) \geq 0$, $x, y \in \mathbb{R}$.
- $\sum_x \sum_y p(x, y) = 1$.

# The marginal probability function

- The probability that a child survived when involved in an accident in which at least one fatality occurred is

$$P(X=0) = P(Y=0, X=0) + P(Y=1, X=0) + \\ + P(Y=2, X=0) = 0.38 + 0.14 + 0.24 = 0.76.$$

- The probability that a child did not survive when involved in an accident in which at least one fatality occurred is

$$P(X=1) = P(Y=0, X=1) + P(Y=1, X=1) + \\ + P(Y=2, X=1) = 0.17 + 0.02 + 0.05 = 0.24.$$

- These are the Marginal probability functions of X.

## The marginal probability function

- Notice that $P(X = 0) + P(X = 1) = 0.76 + 0.24 = 1.00$.
- The table contains the marginal distributions in the row and column.

**Definition**

The marginal probability functions of X and Y, respectively, when X and Y are discrete random variables, are given by

$$p(x) = \sum_y p(x, y)$$

and

$$p(y) = \sum_x p(x, y)$$

# The conditional probability distribution

- For $p(y) > 0$, we define the conditional probability distribution of X given that $Y = y$ as

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

- Notice that

$$\sum_x P(X = x \mid Y = y) = 1.$$

# The conditional probability distribution

| | 0 | 1 | |
|---|---|---|---|
| 0 | 0.38 | 0.17 | 0.55 |
| 1 | 0.14 | 0.02 | 0.16 |
| 2 | 0.24 | 0.05 | 0.29 |
| | 0.76 | 0.24 | 1.00 |

$$P(X=0 \mid Y=0) = \frac{P(X=0, Y=0)}{P(Y=0)} = \frac{0.38}{0.55} = 0.69$$
$$P(X=1 \mid Y=0) = \frac{0.17}{0.55} = 0.31$$
$$P(X=0 \mid Y=1) = \frac{0.14}{0.16} = 0.875$$
$$P(X=1 \mid Y=1) = \frac{0.002}{0.16} = 0.125$$
$$P(X=0 \mid Y=2) = \frac{0.24}{0.29} = 0.828$$
$$P(X=1 \mid Y=2) = \frac{0.005}{0.29} = 0.172$$

# The conditional probability distributions

No car belt ($Y = 0$)

| Child survived ($X = 0$) | Child did not survive ($X = 1$) |
|---|---|
| 0.69 | 0.31 |

Adult car belt ($Y = 1$)

| Child survived ($X = 0$) | Child did not survive ($X = 1$) |
|---|---|
| 0.875 | 0.125 |

Children's car belt ($Y = 2$)

| Child survived ($X = 0$) | Child did not survive ($X = 1$) |
|---|---|
| 0.828 | 0.172 |

# The joint distribution function

**Definition**

The joint distribution function $F(x, y)$ for a bivariate random variable $(X, Y)$ is

$$F(x, y) = P(X \leq x, Y \leq y).$$

If X and Y are discrete,

$$F(x, y) = \sum_{s=-\infty}^{x} \sum_{t=-\infty}^{y} p(s, t),$$

where $p(s, t)$ is the joint probability function.

# The joint distribution function

A function, $F(x, y)$, is a joint distribution function iff

- $\lim_{x\to-\infty} \lim_{y\to-\infty} F(x, y) = 0$
- $\lim_{x\to\infty} \lim_{y\to\infty} F(x, y) = 1$
- $F(x, y)$ is nondecreasing; that is for each $a < b$ and $c < d$, $P(a < X \leq b, c < Y \leq d) =$ $F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0.$
- For fixed x or fixed y, $F(x, y)$ is right-hand continuous in the remaining variable. For example, for fixed y, $\lim_{x\to x_0^+} F(x, y) = F(x_0, y).$

# The joint probability density function

## Definition

Let X and Y be contnuous random variables. The joint probability density function $f(x, y)$ of X and Y is defined by

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) dx\, dy.$$

Notice that in order for $f(x, y)$ to be a probability density function, we must have

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

# The cumulative distribution function

## Definition

The cumulative distribution function, $F_{(X,Y)}(x, y)$, for a bivariate continuous random variable $(X, Y)$ is defined by

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y) dx\, dy.$$

# The marginal probability functions

**Definition**

The marginal probability functions of continuous random variables $X$ and $Y$, respectively, are given by

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

and

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

## Example

Verify that

$$f(x,y) = \begin{cases} \cos(x)\sin(x), & 0 \le x \le \frac{\pi}{2},\ 0 \le y < \frac{\pi}{2}; \\ 0, & \text{elsewhere.} \end{cases}$$

is a probability density function.

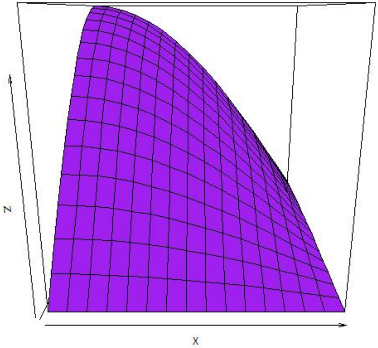


Figure: The function $f(x, y) = \cos(x)\sin(y)$

## Example

Let X and Y be continuous random variables with joint density function

$$f(x,y) = \begin{cases} e^{-(x+y)}, & 0 < x < \infty,\ 0 < y < \infty; \\ 0, & \text{elsewhere.} \end{cases}$$

Find $P(0 < X + Y < 3)$

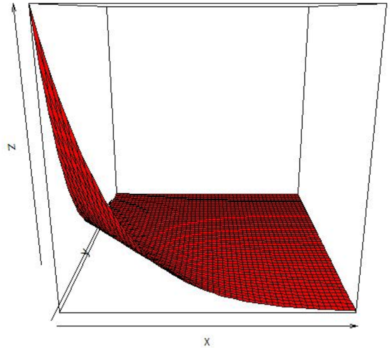


Figure: The function $f(x, y) = e^{-(x+y)}$

## Solution

$$
\begin{aligned}
P(0 < X + Y < 3) &= P(0 < Y < 3 - X) \\
&= \int_0^3 \int_0^{3-x} e^{-(x+y)} dy\, dx = 1 - 4e^{-3}.
\end{aligned}
$$

## Example

Let X and Y be continuous random variables with joint density function

$$f(x,y) = \begin{cases} x+y, & 0<x<1,\ 0<y<1; \\ 0, & \text{elsewhere.} \end{cases}$$

(A) Verify that $f(x,y)$ is a probability density function.
(B) Determine the marginal density function for X, where X is nonzero.
(C) Find $P(X < 2Y)$.

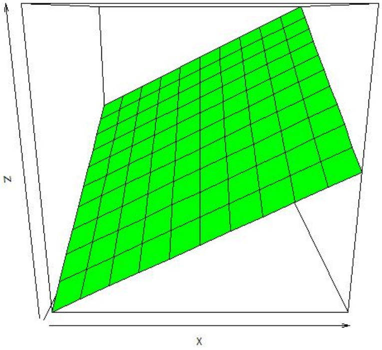


Figure: The function $f(x, y) = x + y$

## Solution

*(A)*

$$\int_0^1 \int_0^1 (x+y)dxdy = 1.$$

*(B)*

$$f_X(x) = \begin{cases} \int_0^1 (x+y)dy = \frac{1}{2} + x & 0 < x < 1; \\ 0, & \textit{otherwise} \end{cases}$$

*(C)*

$$P(X < 2Y) = P\left(\frac{x}{2} < Y\right)$$
$$= \int_0^1 \int_{x/2}^1 (x+y)dydx = \frac{19}{24}.$$

# Conditional probability distributions, chapter 6.2

## Theorem

*The conditional distribution of $X$ given $Y = y$ is*

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

*which is defined for all real values of $X$. Similarly, for all real values of $Y$, the conditional distribution of $Y$ given $X = x$ is*

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}.$$

# Conditional probability distributions

## Theorem

*Let $X$ and $Y$ be continuous random variables with joint probability density function $f(x,y)$ and marginal densities $f_X(x)$ and $f_Y(y)$, respectively. Then, the conditional probability density function of $X$ given $Y = y$ is defined by*

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f(x,y)}{f_Y(y)}, & f_Y(y) > 0; \\ 0, & \textit{elsewhere.} \end{cases}$$

*and the conditional density function of $Y$ given $X = x$ is defined by*

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f(x,y)}{f_X(x)}, & f_X(x) > 0; \\ 0, & \textit{elsewhere.} \end{cases}$$

# Independent random variables, chapter 6.3

**Theorem**

*Discrete random variables are said to be independent if*

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

*for all real number $x$ and $y$. Continuous random variables $X$ and $Y$ are said to be independent if*

$$f(x, y) = f_X(x)f_Y(y)$$

*for all real numbers $x$ and $y$.*

# Conditional probability distributions

## Example

Let X and Y be continuous, random variables with $X \leq Y$ and with joint probability density function

$$f(x,y) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq y,\, 0 \leq y \leq 2; \\ 0, & \text{elsewhere.} \end{cases}$$

(A) Find the conditional density of X, given $Y = y$.
(B) Find $P(X \leq 1/2 \mid Y = 1)$.

# Conditional probability distributions

## Solution

*(A) The marginal density function for Y is*

$$f_Y(y) = \begin{cases} \int_0^y \frac{1}{2}dx = \frac{1}{2}y, & 0 \le y \le 2; \\ 0, & \textit{elsewhere.} \end{cases}$$

*Then*

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{1/2}{y/2} = \frac{1}{y}, & 0 \le x \le y \le 2; \\ 0, & \textit{elsewhere.} \end{cases}$$

# Conditional probability distributions, solutions cont.

## Solution

*(B)*

$$f_{X|Y(x|y=1)} = \begin{cases} 1, & 0 \leq x \leq 1; \\ 0, & \textit{elsewhere.} \end{cases}$$

$$P(X < 1/2 \mid Y = 1) = \int_0^{1/2} (1) dx = \frac{1}{2}.$$

# Expected values of functions of random variables, chapter 6.4

## Definition

Suppose that the discrete random variables $(X, Y)$ have a joint probability function given by $p(x, y)$. If $g(X, Y)$ is any real-valued function of $(X, Y)$, then the expected value of $g(X, Y)$ is

$$E[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y).$$

If $(X, Y)$ are continuous random variables with probability density function $f(x, y)$, then

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

# Expected values of functions of random variables

## Theorem

*If $X$ and $Y$ are independent with means $\mu_X$ and $\mu_Y$, respectively, then*

$$E(XY) = E(X)E(Y).$$

*If $X$ and $Y$ are independent, $g$ is a function of $X$, and $h$ is a function of $Y$, then*

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

# The Covariance between two random variables

**Definition**

The Covariance between two random variables X and Y is given by

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

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

**Theorem**

*If $X$ has a mean $\mu_X$ and $Y$ has mean $\mu_Y$, then*

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

# The Covariance between two random variables

- If $cov(X, Y) > 0$, then X tends to be large when Y is large and X tends to be small when Y is small.
- If $cov(X, Y) < 0$, then X tends to be small when Y is large and large when X is small.
- The covariance measures the direction of the association between two random variables.

# The Correlation between two random variables

**Definition**

The correlation between two random variables X and Y is given by

$$\rho = \frac{cov(X, Y)}{\sqrt{V(X)V(Y)}}.$$

$\rho$ is called the correlation coefficient.

The correlation measures the strength of the association between two random variables.

# The Correlation between two random variables

- The correlation is a unitless quantity.
- The correlation takes on values beween $-1$ and $1$.
- If $\rho = \pm 1$, then Y is a linear function of X.
- If $\rho = +1$, the slope is positive and if $\rho = -1$, the slope is negative.
- The strength of the linear relationship becomes less as $\rho$ decreases from $+1$ to $-1$. It is smallest at $\rho = 0$.
- If X and Y are independent random variables, then

  $$cov(XY) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

  The converse is not necessarily true.

## Theorem

*Let $X_1, X_2, ..., X_n$ be random variables. Define*

$$X = \sum_{i=1}^{n} a_i X_i$$

*for constants $a_1, a_2, ..., a_n$. Then we have*

- $$E(X) = \sum_{i=1}^{n} a_i E(X_i).$$

- $$V(X) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum \sum_{i<j} a_i a_j cov(X_i, X_j).$$

# Problem 6.31

## Example

Bollworms are a pest of cotton that cause economic damage if uncontrolled. Lady beetles are natural pests of bollworms. If sufficient numbers of lady beetles are present in a particular cotton field, then chemical control may not be needed. Let X and Y denote the numbers of bollworms and lady beetles on a cotton plant, respectively. A researcher explored the relationship in these two vaiables in a cotton field and found that they had the following joint distribution, $p(x, y)$:

| y \ x | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0.03 | 0.04 | 0.11 | 0.10 |
| 1 | 0.04 | 0.05 | 0.09 | 0.06 |
| 2 | 0.09 | 0.07 | 0.06 | 0.02 |
| 3 | 0.12 | 0.10 | 0.01 | 0.01 |

# Problem 6.31 cont.

**Example**

(A) Find the mean number of bollworms on the plants in this cotton field.
(B) Find the mean number of lady beetles on the plants in this cotton field.
(C) Find the covariance between the numbers of lady beetles and bollworms on a cotton plant.
(D) Find the correlation between the numbers of lady beetles and bollworms on a cotton plant.

## Problem 6.31

### Solution

*(A) The marginal distribution of X:*

| $x$ | $p(x)$ |
|---|---|
| 0 | 0.28 |
| 1 | 0.26 |
| 2 | 0.27 |
| 3 | 0.19 |

$$\begin{aligned} p(0) &= p(0,0) + p(0,1) + p(0,2) + p(0,3) \\ &= 0.03 + 0.04 + 0.09 + 0.12 = 0.28. \end{aligned}$$

*The mean number of bollworms on the plants is:*

$\mu_X = \sum_{x=0}^{3} xp(x) = 0(0.28) + 1(0.26) + 2(0.27) + 3(0.19) = 1.37.$

# Problem 6.31 continue

## Solution

*(A) The marginal distribution of Y:*

| $y$ | $p(y)$ |
|---|---|
| 0 | 0.28 |
| 1 | 0.24 |
| 2 | 0.24 |
| 3 | 0.24 |

$$p(0) = p(0,0) + p(1,0) + p(2,0) + p(3,0)$$
$$= 0.03 + 0.04 + 0.11 + 0.10 = 0.28.$$

*The mean number of lady beetles on the plants is:*
$\mu_Y = \sum_{y=0}^{3} yp(y) = 0(0.28) + 1(0.24) + 2(0.24) + 3(0.24) = 1.44.$

# Problem 6.31 continue

## Solution

*(C)*

$$
\begin{aligned}
E(XY) &= \sum_y \sum_x xyp(x,y) \\
&= (0\times 0)p(0,0) + (1\times 0)p(1,0) + (2\times 0)p(2,0) + \\
&\quad (3\times 0)p(3,0) + \cdots + (0\times 3)p(0,3) + (1\times 3)p(1,3) + \\
&\quad (2\times 3)p(2,3) + (3\times 3)p(3,3) = 1.36
\end{aligned}
$$

*The covariance between the numbers of lady beetles and bollworms on a cotton plant is* $Cov(X,Y) = E(XY) - \mu_X\mu_Y = 1.36 - (1.37)(1.44) = -0.61.$

## Problem 6.31 continue

**Solution**

*(D)*

$V(X) = E(X^2) - \mu_X^2 = 1(0.26) + 4(0.27) + 9(0.19) - (1.37)^2 = 1.17.$

$V(Y) = E(Y^2) - \mu_Y^2 = 1(0.24) + 4(0.24) + 9(0.24) - (1.44)^2 = 1.29.$

*The correlation between the numbers of lady beetles and bollworms on a cotton plant is*

$$\rho = \frac{-0.61}{\sqrt{(1.17)(1.29)}} = -0.50.$$

# Conditional Expectations, chapter 6.5

## Definition

Let X and Y be any two random variables. If X and Y are jointly continous then the conditional expectation of X given that $Y = y$ is defined by

$$E(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx.$$

If X and Y are jointly discrete then the conditional expectation of X given that $Y = y$ is defined by

$$E(X \mid Y = y) = \sum_{x} x p(x \mid y).$$

# Conditional Expectations

**Theorem**

*Let $X$ and $Y$ denote random variables. Then*

$$E(X) = E[E(X \mid Y)]$$

*and*

$$V(X) = E[V(X \mid Y)] + V[E(X \mid Y)].$$

# Problem 6.50

## Example

Suppose that X is the random variable denoting the number of bacteria per cubic centimeter in water samples and that for a given location, X, has a Poisson distribution with mean $\lambda$. But $\lambda$ varies from location to location and has a gamma distribution with parameters $\alpha$ and $\beta$. Find the expressions for $E(X)$ and $V(X)$.

# Problem 6.50, solution

## Solution

*For a given $\lambda$, $X$ has a Poisson distribution. Hence $E(X \mid \lambda) = \lambda$ and $V(X \mid \lambda) = \lambda$. $\lambda$ has a Gamma distribution with parameters $\alpha$ and $\beta$. Then*

$$E(X) = E[E(X \mid \lambda)] = E(\lambda) = \alpha \cdot \beta$$

*and*

$$\begin{aligned} V(X) &= E[V(X \mid \lambda)] + V[E(X \mid \lambda)] \\ &= E(\lambda) + V(\lambda) = \alpha \cdot \beta + \alpha \cdot \beta^2 = \alpha \cdot \beta(1 + \beta). \end{aligned}$$