Multivariate distributions

Introduction

- Generally interested in more than one random variable at a time

  1. n observations of a single characteristic from some population

     $X_1, \dots, X_n$ $\quad X_i \equiv$ weight of the ith Unit

  2. k different characteristics from a single individual

     One person $\quad Y_1 \; Y_2 \; Y_3 \cdots$

     $Y_1 \to$ age, $Y_2 \to$ hieght, $Y_3 \to$ gender

- *Notation:* $\underset{\sim}{X}$ = $\boldsymbol{X}$ = $\underbrace{(X_1, \dots, X_n)}$ is an n-dimensional random vector (r.v.)

  tilda

  In other words, $\boldsymbol{X}$ is a function from sample space $S \to \mathbb{R}^n$

  $$\underset{\sim}{X} : \Omega \longrightarrow \mathbb{R}^n$$
  $$S \longrightarrow \mathbb{R}^n$$
  $$\underset{\sim}{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) \in \mathbb{R}^n$$

- General plan:

  use the bivariate case (X_1, X_2) $(n = 2)$ to derive results and then extend to any n

  For notational simplicity, in the bivariate case, we'll denote a pair of random variables as (X, Y) instead of (X_1, X_2)

Multivariate distributions

Discrete random vectors & probability mass functions

Definition:

Definition: ~~If there exists a countable set of points $\{(x_i, y_i)\}$ such that~~

$$\cancel{\sum_{i=1}^{\infty} P[(X,Y) = (x_i, y_i)] = 1}$$

~~we say that the distribution of X and Y is **discrete** and the function~~

$X \in \{x_1, x_2, \dots, x_n\}$
$Y \in \{y_1, y_2, \dots, y_m\}$
$P(X,Y)$ is joint pmf
If $\sum_{j=1}^{m} \sum_{i=1}^{n} P(x_i, y_j) = 1$

$$f(x, y) = P(X = x, Y = y) \geq 0 \quad \text{for any } (x, y) \in \mathbb{R}^2$$

is called the **joint probability mass function** (joint pmf) for X and Y.

Example (artificial): (X, Y) with joint pmf given in tabular form as

		x		
		1	2	3
y	3	1/12	1/12	1/6
	2	1/12	1/6	1/12
	1	1/6	1/12	1/12

$f(x,y) = \mathbb{P}(X = x, Y = y)$
$f(1,2) = \mathbb{P}(X = 1, Y = 2) = 1/12$
$f(3,3) = 1/6$
$f(x,y) = 0$ for all other $x, y \in \mathbb{R}$

For (X, Y) jointly discrete with pmf $f(x, y)$,

$$P[(X,Y) \in A] = \sum_{(x,y) \in A} f(x, y), \quad \text{for } A \subset \mathbb{R}^2$$

$$\mathbb{P}(Y \geq X) = \sum_{\underbrace{(x,y): y \geq x}_{A}} f(x,y) = f(1,3) + f(2,3) + f(3,3) + f(1,2) + f(2,2) + f(1,1) = 3/4$$

Multivariate distributions

Another joint pmf example (i.e., discrete case again)

- Suppose there are n independent trials, where each trial has three possible outcomes: outcome a with prob. p_1, outcome b with prob. p_2 and outcome c with prob. $1 - p_1 - p_2$

- Let X =# of trials having outcome a & Y =# of trials having outcome b

X has the range from 0, 1, 2, ..., n
Y has the range from 0, 1, 2, ..., n
$0 \le X + Y \le n$

- Then the joint pmf of (X, Y) is

$$\mathbb{P}(X=x, Y=y) \quad f(x,y) = \binom{n}{x}\binom{n-x}{y} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

Out of n trials choose to x trials to "a"

$$= \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

(*)

Defined for $0 < p_1, p_2, p_1 + p_2 < 1$, and the support/range is given by all integer pairs (x, y) where $0 \le x, y, x + y \le n$

Otherwise

- If $n = 1$, then $f(1,0) = p_1$ $\quad f(0,1) = p_2$ $\quad f(0,0) = 1 - p_1 - p_2$

f(x,y) (y \ x)	0	1
0	$1-p_1-p_2$	p_1
1	p_2	0

- If $n = 2$, then $f(2,0) = p_1^2$, $f(1,1) = 2p_1p_2$, etc.

y \ x	0	1	2
0	$(1-p_1-p_2)^2$	$2p_1(1-p_1-p_2)$	p_1^2
1	$2p_2(1-p_1-p_2)$	$2p_1p_2$	0
2	p_2^2	0	0

- multinomial distribution with three outcomes: $(X, Y) \sim$Multinomial(n, p_1, p_2)

a binominal distribution is a multinomial distribution with two outcomes

Multivariate distributions

Continuous random vectors & probability density functions

Definition: If there exists a function $f(x,y) \geq 0$ for $(x,y) \in \mathbb{R}^2$ such that

$$\int\int_{\mathbb{R}^2} f(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

and

$$P((X,Y) \in A) = \int\int_A f(x,y)dxdy \quad \text{for } A \subset \mathbb{R}^2$$

then we say that the distribution of X and Y is **jointly (absolutely) continuous**. The function $f(x,y)$ is called the **joint probability density function** (joint pdf) for X and Y.

Note: As in the univariate case, any function $f(x,y)$ with

1. $f(x,y) \geq 0$ for all $(x,y) \in \mathbb{R}^2$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$ (continuous case)

   or $\sum_{(x,y)\in\mathbb{R}^2} f(x,y) = 1$ (discrete case) $\quad \sum_x \sum_y \mathbb{P}(X=x, Y=y)$

specifies the joint pdf or pmf of some bivariate random vector (X,Y)

Calculus we need for the Course

$$\int x^m \, dx = \frac{1}{m+1} x^{m+1}$$

$$\int e^x \, dx = e^x$$

$$\ast \quad \int U \, dV = UV - \int V \, dU$$

$$\int e^{-x} x^{m-1} \, dx = \Gamma(m)$$

$$\left(f(g(x))\right)' = g'(x) f'(g(x))$$

$$\left[e^{f(x)}\right]' = f'(x) e^{f(x)}$$

$$\left(\log f(x)\right)' = \frac{f'(x)}{f(x)}$$

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'g - g'f}{g^2}$$

$$\left[x^m\right]' = m x^{m-1}$$

$$\left[\left(f(x)\right)^m\right]' = m f'(x) \left[f(x)\right]^{m-1}$$

$$\int_{-a}^{a} f(x)\,dx \underset{f \text{ is}}{\overset{\text{even}}{=\!=}} 2\int_0^a f(x)\,dx$$

$$\int_{-a}^{a} f(x)\,dx \underset{f \text{ is}}{\overset{\text{odd}}{=\!=}} 0$$

We Say f is even If $f(x) = f(-x)$ e.g. $f(x) = x^2$

We Say f is odd If $f(-x) = -f(x)$ e.g $f(x) = x^3$, $f(x) = \sin x$

$$(\sin u)' = u' \cos u$$

$$(\cos u)' = -u' \sin u$$

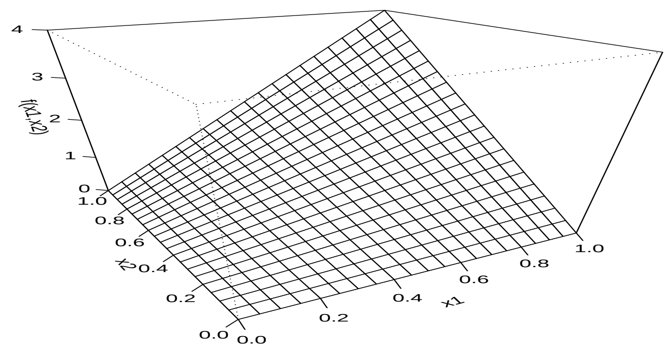
$$B(a,b) = \frac{\Gamma(b)\Gamma(a)}{\Gamma(a+b)}$$

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(b)\Gamma(a)}{\Gamma(a+b)}$$

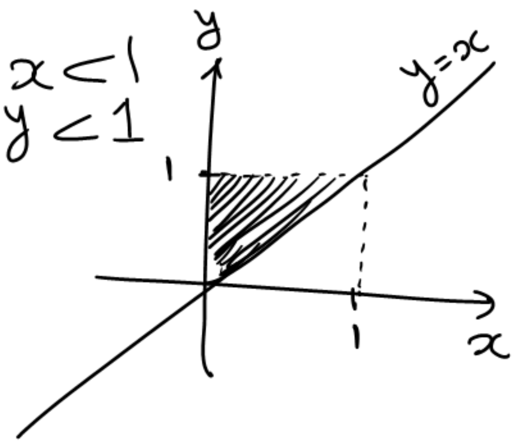
Multivariate distributions

Joint pdf example (i.e., continuous case)

Suppose (X, Y) have joint pdf $f(x, y) = 4xy$ for $0 < x, y < 1$



$0 < x < 1$
$0 < y < 1$



Find $P(X < Y)$

$$\mathbb{P}(X<Y) = \iint_{A} \underbrace{f(x,y)}\, \underset{\smile}{d}x\,dy$$

$A = \{(x,y) : y - x > 0\}$

$$= \int_0^1 \left[\int_0^y \underbrace{4xy\, dx} \right] dy$$

$$= \int_0^1 \left[y \left(2x^2\right)\Big|_0^y \right] dy = \int_0^1 y(2y^2) = 2\int_0^1 y^3\, dy$$

$$= 2/4\, y^2 \Big|_0^1 = 1/2$$

$x + y < 1 \Rightarrow \boxed{x < 1 - y}$

Find $P(\underbrace{X + Y < 1})$

$$\int_0^1 \int_0^{1-y} 4xy\, dx\, dy$$