MULTIVARIATE DISTRIBUTIONS

Introduction

Multivariate distributions

- So far, we have considered distributions of *single* random variables (often, one measurement or trial per person).
- But often we make multiple measurements per person
- Example 1: For each household we measure Y_1 = household income and Y_2 = number of children living there.

  (Y_1, Y_2)
- Example 2: For each patient we measure Y_1 = body mass index (BMI), Y_2 = systolic blood pressure, and Y_3 = age.

  (Y_1, Y_2, Y_3)
- We now consider probability distributions of *random vectors*, e.g., (Y_1, Y_2, Y_3).

Discrete bivariate probability distributions

- If Y_1 and Y_2 are both random variables then we can consider the bivariate random vector $\boldsymbol{Y} = (Y_1, Y_2)$.
- Definition 1: For two discrete random variables Y_1 and Y_2, the JOINT PROBABILITY MASS FUNCTION (joint pmf) is given by $P(Y_1 = y_1, Y_2 = y_2)$ for all possible combinations of (y_1, y_2).
- As with (univariate) pmfs, this could be given by a table or a formula.

Example 3

y_1

y_2 \ y_1	0	1	2	3
0	0.008	0.060	0.150	
1				0
2			0	0
3		0	0	0

$0 \le Y_1 + Y_2 \le 3$

- Three players are competing for prizes. For each trial, A has a 50% chance of winning, B has a 30% chance of winning, and C has a 20% chance of winning. Let Y_1 be the number of prizes won by A, and let Y_2 be the number of prizes won by B after three trials. Find the joint distribution of Y_1 and Y_2.

independent

$P(Y_1 = 0, Y_2 = 0) = P(\{CCC\}) = (0.2)(0.2)(0.2) = 0.008$ C must win 3

$P(Y_1 = 1, Y_2 = 0) = P(A \text{ win } 1 \text{ time}, B \text{ wins } 0 \text{ times})$

$= P(\{ACC\}) + P(\{CAC\}) + P(\{CCA\}) = 3(0.2)(0.2)(0.5)$

$= 0.060$

$.5 \; .2 \; .2 \qquad .2 \; .5 \; .2$

$P(Y_1 = 2, Y_2 = 0) = P(\{AAC\}) + P(\{ACA\}) + P(\{CAA\}) = 3(0.5)(0.5)(0.2)$

0.150

Discrete bivariate probability distributions

- Theorem 1: If Y_1 and Y_2 are discrete random variables with joint pmf $p(y_1, y_2)$, then
  1. $p(y_1, y_2) \geq 0$ for all y_1 and y_2
  2. $\sum_{y_1} \sum_{y_2} p(y_1, y_2) = 1$ (the sum is taken over all possible (y_1, y_2) pairs.
- Definition 2: For any* random variables Y_1 and Y_2 the JOINT CUMULATIVE DISTRIBUTION FUNCTION (joint cdf) is

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

prob.

AND

* discrete and/or continuous

Continuous bivariate probability distributions

- Definition 3: A random vector (Y_1, Y_2) is CONTINUOUS (or, Y_1 and Y_2 are JOINTLY CONTINUOUS) if their joint cdf $F(y_1, y_2)$ is continuous in both arguments.

- Definition 4: Suppose Y_1 and Y_2 are jointly continuous with joint cdf $F(y_1, y_2)$. If there is a nonnegative function $f(y_1, y_2)$ such that

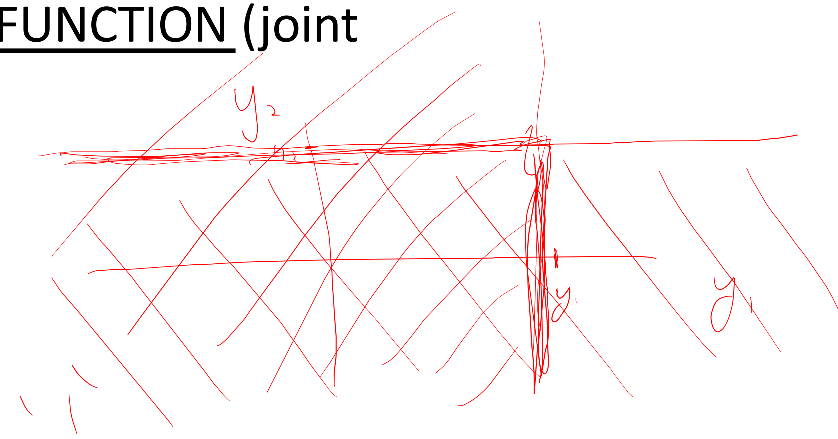
AND

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(u_1, u_2)\, du_2 du_1, -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

Then $f(y_1, y_2)$ is the JOINT PROBABILITY DENSITY FUNCTION (joint pdf) of (Y_1, Y_2).

$F(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2)$

$Y_1 \le y_1 \cap Y_2 \le y_2$



$P(Y_1 \le -\infty, Y_2 \le -\infty) = 0$

$P(Y_1 \le y_1, Y_2 \le -\infty)$

0

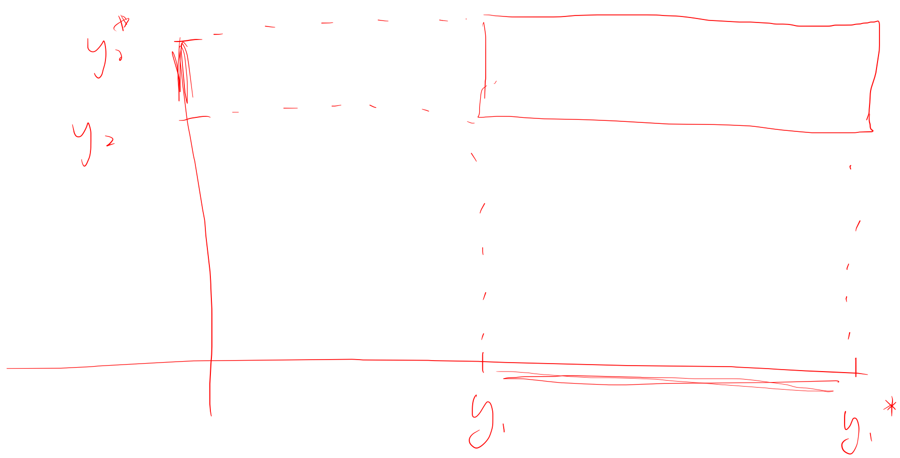
Properties of joint cdf

- If Y_1 and Y_2 are random variables with joint cdf $F(y_1, y_2)$, then
    1. $F(-\infty, -\infty) = F(y_1, -\infty) = F(-\infty, y_2) = 0$ for all y_1 and y_2.
    2. $F(\infty, \infty) = 1$ $\quad P(Y_1 \le \infty, Y_2 \le \infty) = 1$
    3. If $y_1^* \ge y_1$ and $y_2^* \ge y_2$ then

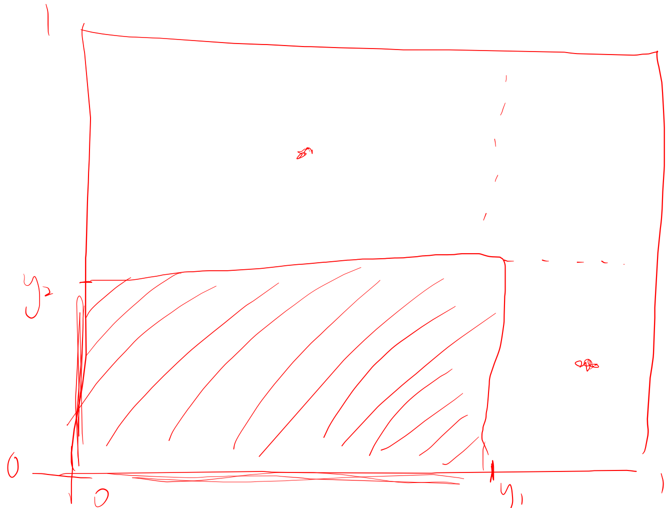
$$P(y_1 < Y_1 \le y_1^*, y_2 < Y_2 \le y_2^*) = F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2)$$

check this!

AND



Example 4



- Y_1 and Y_2 have joint pdf

$$f(y_1, y_2) = y_1 + y_2 \text{ for } 0 \le y_1 \le 1,\ 0 \le y_2 \le 1$$

Find the joint cdf.

Always draw a picture!

$$F(y_1,y_2) = P(Y_1 \le y_1, Y_2 \le y_2) = \int_0^{y_1}\left(\int_0^{y_2} f(u_1,u_2)\,du_2\right)du_1$$

$$= \int_0^{y_1}\left(\int_0^{y_2}(u_1+u_2)\,du_2\right)du_1 = \int_0^{y_1}\left(u_1u_2 + \frac{1}{2}u_2^2\Big|_{u_2=0}^{y_2}\right)du_1$$

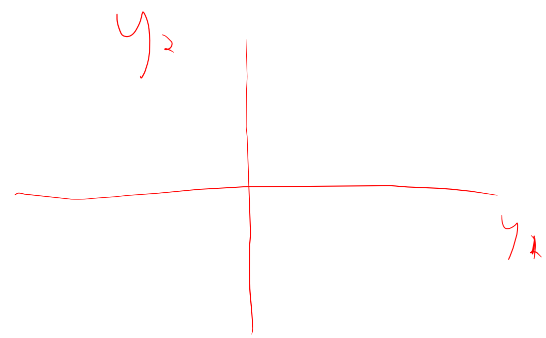
$$= \int_0^{y_1}\left(u_1y_2 + \frac{1}{2}y_2^2 - 0 - 0\right)du_1 = \int_0^{y_1}\left(u_1y_2 + \frac{1}{2}y_2^2\right)du_1 = \frac{1}{2}y_2u_1^2 + \frac{1}{2}y_2^2u_1\Big|_0^{y_1}$$

$$= \frac{1}{2}y_2y_1^2 + \frac{1}{2}y_2^2y_1 - 0 - 0$$

$$\boxed{= \frac{1}{2}y_1y_2(y_1+y_2) \quad 0 \le y_1 \le 1,\ 0 \le y_2 \le 1}$$

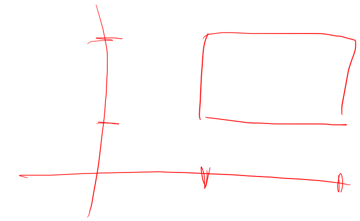
Properties of joint pdf

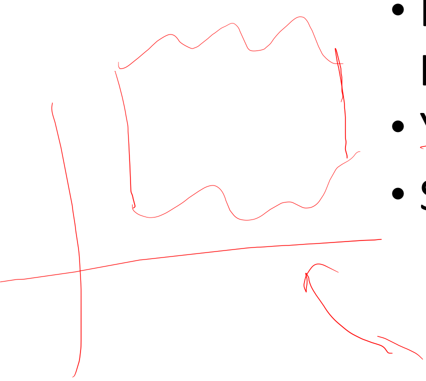
- If Y_1 and Y_2 are random variables with joint pdf $f(y_1, y_2)$, then
  1. $f(y_1, y_2) \geq 0$ for all y_1 and y_2.
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 dy_2 = 1$

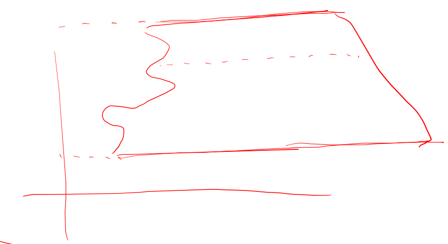


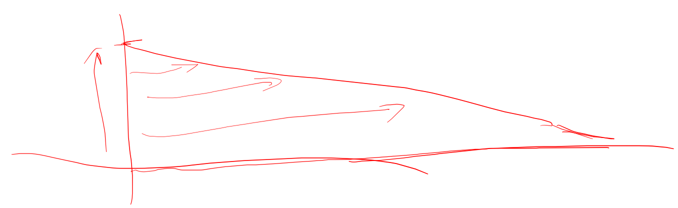
Finding probabilities with joint pdfs

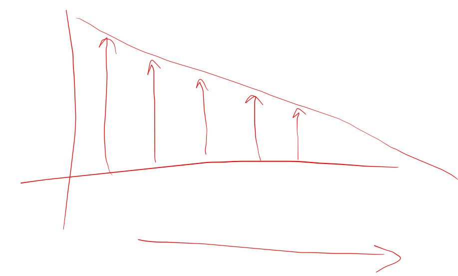
- Probabilities will involve integrating over a *region* in the (y_1, y_2) plane.
- You should always make a sketch.
- Some types of regions:
  - Rectangular
  - Vertically simple
  - Horizontally simple
  - Complex
- Some regions (e.g., triangles) could be treated either as VS or HS.











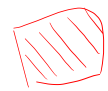
$f(y_1, y_2) = y_1 + y_2, \quad 0 \le y_1 \le 1, \quad 0 \le y_2 \le 1$

Example 4 revisited

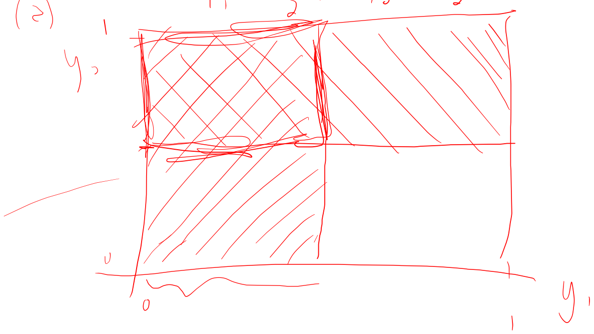
- For the previous joint pdf, calculate
  - (a) $P\left(Y_1 < \frac{1}{2}, Y_2 > \frac{1}{2}\right)$
  - (b) $P(Y_1 + Y_2 < 1)$.

$$P\left(Y_1 < \tfrac{1}{2}, Y_2 > \tfrac{1}{2}\right) = \int_{1/2}^{1} \int_{0}^{1/2} (u_1 + u_2)\, du_1\, du_2$$

$P\left(Y_2 > \frac{1}{2}\right)$



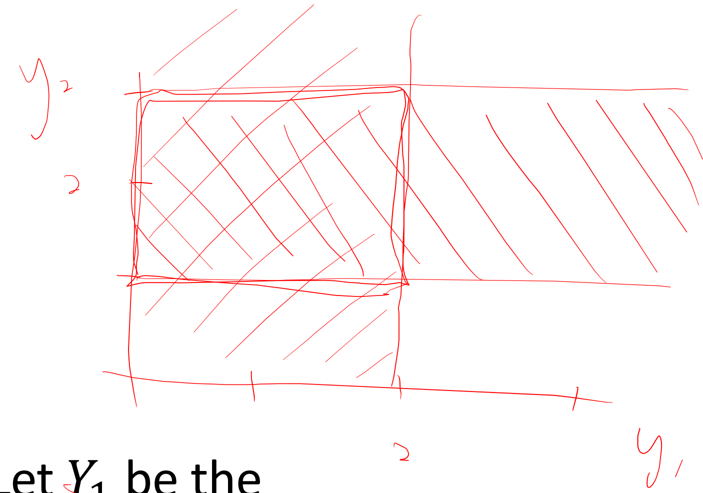
(a) $Y_1 < \frac{1}{2} \cap Y_2 > \frac{1}{2}$



$P\left(Y_1 < \frac{1}{2}\right)$

(b) $P(Y_1 + Y_2 \le 1)$

$$\int_0^1 \int_0^{1-y_2} (y_1 + y_2)\, dy_1\, dy_2$$



Example 5

- A hospital employee has to complete Task 1 and Task 2. Let Y_1 be the amount of time to complete Task 1 and Y_2 be the amount of time to complete Task 2, and suppose the joint pdf of Y_1 and Y_2 is

$\boxed{f(y_1, y_2) = e^{-(y_1+y_2)}}$ for $y_1 > 0$ and $y_2 > 0$.

a. Find the probability that s/he will take less than 2 hours on Task 1 and between 1 and 3 hours on Task 2.

b. Find the probability that Task 2 will take longer than Task 1.

(a) $P(Y_1 < 2, 1 \le Y_2 < 3) = \int_0^2 \int_1^3 e^{-(y_1+y_2)} \, dy_2 \, dy_1$

$$P(Y_2 > Y_1)$$

$$= \int_0^\infty \left(\int_{y_1}^\infty e^{-(y_1+y_2)} \, dy_2 \right) dy_1$$