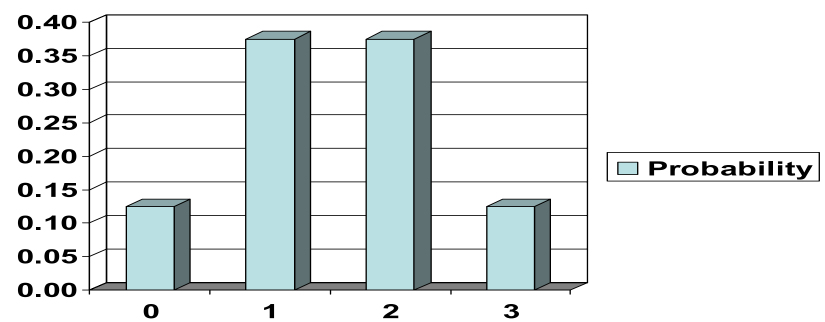
Lecture 3: Multivariate Distributions

Quantitative Methods for Regulation and Competition

Remember me?

- Experiment: tossing a “fair” coin three times
- Random variable X: count the number of heads
- Probability distribution: assign probabilities as

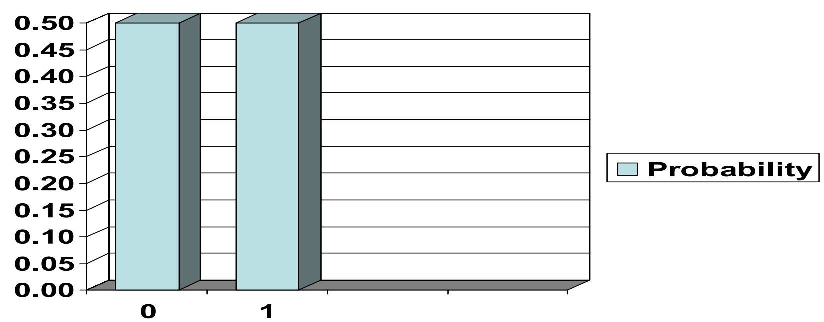
Number of heads	Probability
0	1/8
1	3/8
2	3/8
3	1/8



New Random Variable

- Experiment: same
- Random variable Y: number of heads in 3rd flip
- Assigned probabilities:

Number of heads	Probability
0	½
1	½



Bivariate distribution

- Are RV X and Y related?
- What about the probability that X=2 and Y=1? And X=3 and Y=0?
- Bivariate distribution of X and Y:

		X			
Y		0	1	2	3
	0				
	1				

Today's Lecture

- For discrete random variables:
  - Bivariate density and cumulative distributions
  - Marginal distributions and expectations
  - Conditional distributions and independence
  - Measures of relatedness: covariance and correlation
- Extension to continuous random variables

Bivariate discrete distribution

- Two random variables X and Y that can take values $\{x_1,\ldots,x_N\}$ and $\{y_1,\ldots,y_M\}$
- Bivariate (joint) probability distribution:
  - Attach a probability to each joint outcome $\{x_{i,}\ y_j\}$
  - Denoted as f(x,y) and plotted in 3D
- Usual properties:
  - All numbers are positive (or 0) and add up to 1
- Cumulative bivariate probability distribution:
  - $F(x,y) = P(X \leq x, Y \leq y)$
  - Example: $F(2,0) = P(X \leq 2, Y \leq 0) =$
  - Example: $F(2,0) = P(X \leq 3, Y \leq 1) =$

Marginal Distributions and Expectations

- If we want to go back to the univariate…
  - Example: P(X=2) if we only have joint distribution
  - Called "marginal distributions":

$$g(x) = P(X = x) = \sum_{\text{all y}} f(x, y) \qquad h(y) = P(Y = y) = \sum_{\text{all x}} f(x, y)$$

<table>
<tr><td rowspan="4">Y</td><td colspan="5">X</td><td>Marginal dist Y</td></tr>
<tr><td></td><td>0</td><td>1</td><td>2</td><td>3</td><td></td></tr>
<tr><td>0</td><td>1/8</td><td>¼</td><td>1/8</td><td>0</td><td></td></tr>
<tr><td>1</td><td>0</td><td>1/8</td><td>¼</td><td>1/8</td><td></td></tr>
<tr><td>Marginal dist X</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

- Expectation of X and Y given the joint distribution:

$$E(X) = \sum_{\text{all x}} x g(x) = \sum_{\text{all x}} \sum_{\text{all y}} x f(x, y) \qquad E(Y) = \sum_{\text{all y}} y h(y) = \sum_{\text{all y}} \sum_{\text{all x}} y f(x, y)$$

Conditional Distribution and Independence

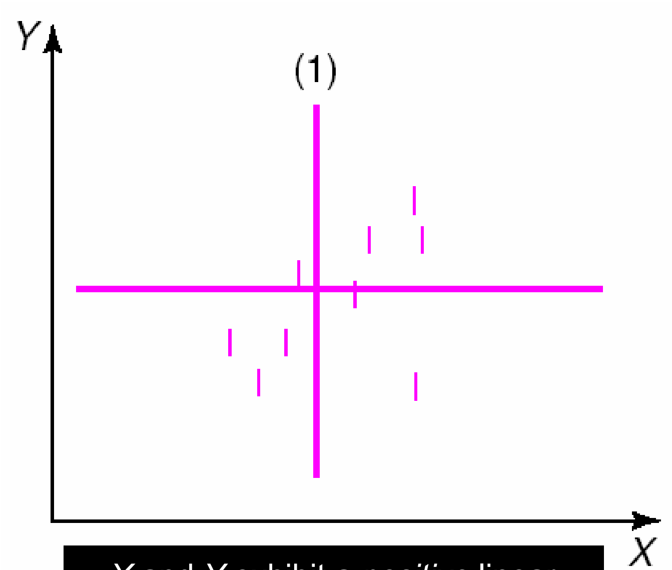
- Suppose that we know that Y=0, what is the probability of X=0? And X=1?
  - P(X=x|Y=0) or more generally P(Y=y|X=x)

$$f(x|y) = P(X = x|Y = y) = f(x, y) / h(y)$$

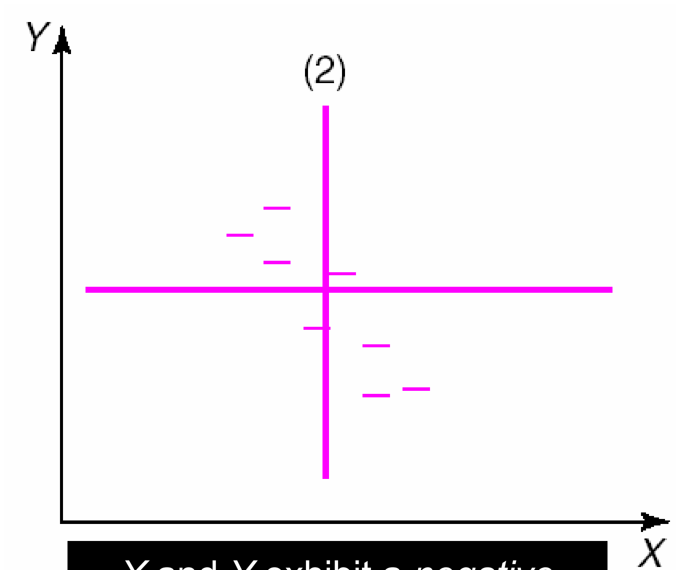
$$f(y|x) = P(Y = y|X = x) = f(x, y) / g(x)$$

  - Example: f(x|0)= P(X=x|Y=0)
- X and Y are independent whenever…
  - f(x,y) = g(x) h(y) or
  - f(x|y) = g(x)
- We can also compute conditional expectations

Measures of Association: Some Examples

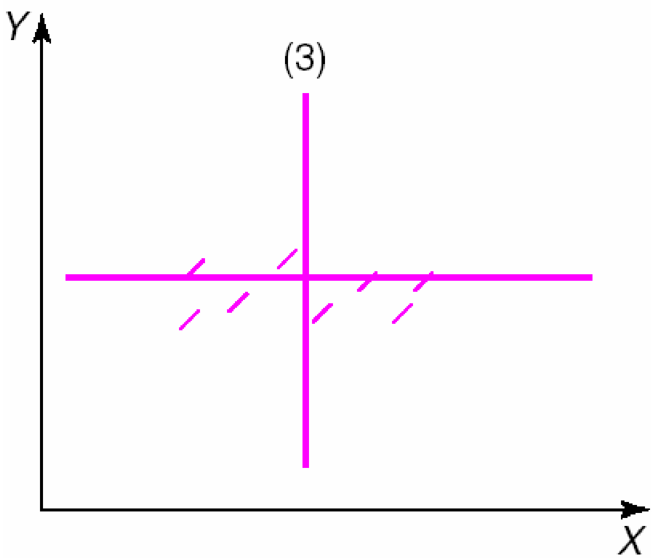


X and *Y* exhibit a *positive* linear relationship—when *X* increases in magnitude *Y* typically also increases in magnitude.

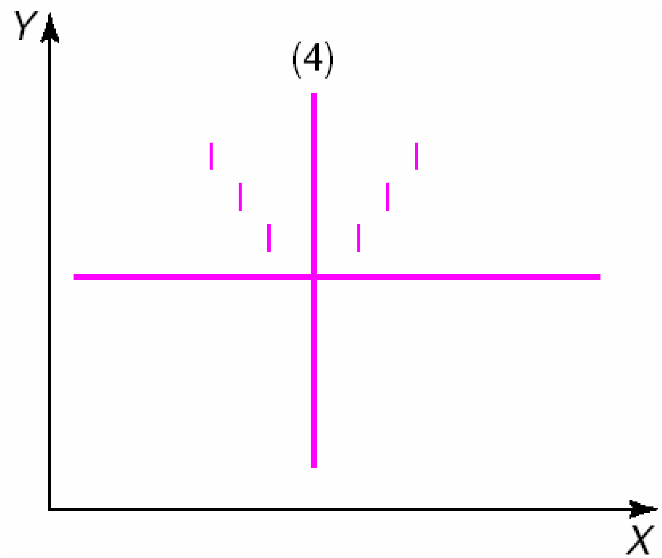


X and *Y* exhibit a *negative* linear relationship—when X gets larger Y tends to get smaller

Measures of Association: Some Examples



There appears to be little relation between the two random variables.



There is a nonlinear relationship between the two variables.

Measures of Association: Covariance and Correlation

- Covariance: $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$
- If discrete then: $Cov(X,Y) = \sum_{\text{all x}} \sum_{\text{all y}} [(x - E(X))(y - E(Y)) f(x,y)]$
- Example and properties
- Good for direction but depends on the units

- Correlation:

$$\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

- Unit free and good measure of magnitude
- Properties: $-1 \leq \rho \leq 1$

Extension to continuous (1)

- f(x,y) joint probability density function:

$$P(a \le X \le b \text{ and } c \le Y \le d) = \int_a^b \int_c^d f(x,y)\,dy\,dx$$

- F(x,y) cumulative distribution function:

$$F(x,y) = P(X \le x \text{ and } Y \le y) = P(-\infty \le X \le x \text{ and } -\infty \le Y \le y)$$

$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,dt\,ds$$

- Marginal density functions:

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

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

Extension to continuous (2)

- Expectations: $$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dydx$$
$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx$$

- Covariance:

$$\text{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (X - E(X))(Y - E(Y))f(x,y)dydx$$