MPO1: Quantitative Research Methods

Session 2: Random Variables and Probability Distributions

Thilo Klein

University of Cambridge
Judge Business School

Probability Rules

Gambling consult from last week

- Chevalier de Mere to Blaise Pascal : What is more likely?
  - Rolling at least one 6 in four throws of a single die
  - Rolling at least one double 6 in 24 throws of a pair of dice

CAMBRIDGE Judge Business School

Probability Rules

Review: Probability Rules

- $1 \geq P(A) \geq 0$; $P(S) = 1$
- Can combine events to make other events using logical operations: A and B, A or B, not A
- Probability of event A or B: Addition Rule
  $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
- If the events are A and B mutually exclusive:
  $P(A \bigcup B) = P(A) + P(B)$
- Probability of event AAnd$B$: Multiplication Rule
  $P(A \bigcap B) = P(A) \cdot P(B)$ if the events are independent
- If not: $P(A \bigcap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$
- For any event: $P(A) = 1 - P(\bar{A})$

CAMBRIDGE Judge Business School

Gambling consult

Solution to de Mere's problem

- Let E be getting *at least one Six* in 4 throws of a single die
- What is $P(E)$?
  - $\bar{E}$ is getting *no Sixes* in 4 throws
  - Let A_i be the event of getting *no Six* in the i^{th} throw
  - $P(A_i) = 5/6$, so $P(\bar{E}) = (5/6)^4 = 0.482$
- $P(E) = 1 - P(\bar{E}) = 0.518$

CAMBRIDGE Judge Business School

Gambling consult (cont'd)

Solution to de Mere's problem

- Let F be event of getting *at least one double Six* in 24 throws
- What is $P(F)$?
    - Let B_i be the event of *no double Six* in the i^{th} throw.
    - $P(B_i) =?$
    - $\bar{F} = B_1$ and B_2 and ... B_{24}
    - $P(\bar{F}) = (35/36)^{24} = 0.509$
- $P(F) = 1 - P(\bar{F}) = 0.491$
- So: P(*at least one Six* in 4 throws) = 0.518 > P(*at least one double Six* in 24 throws) = 0.491

CAMBRIDGE Judge Business School

Bayes' Theorem example

Question

- Example to illustrate conditional probability distributions, and hypothesis tests
  - A rare disease infects 1 person in a 1000
  - There is good but imperfect test
  - 99% of the time, the test identifies the disease
  - 2% of uninfected patients also return a positive test result
  - A patient has tested positive
  - What are the chances he actually has the disease?

CAMBRIDGE Judge Business School

Bayes' Theorem example

Data

- Event A: patient has the disease
- Event B: Patient tests positive
- $P(A) = 0.001$
- $P(B|A) = 0.99$ Note: conditional probability
- $P(B|\overline{A}) = 0.02$ cond. prob.; *False positive*
- Question is: $P(A|B)$?
- We should also note other type of error, other than false positive
- *False negative*, i.e., testing negative though ill: $P(\overline{B}|A) = ?$

Bayes' Theorem example

Sample space

	A: patient has disease	$\overline{A}$: patient does not have disease
B: patient tests positive	$A \bigcap B$	$\overline{A} \bigcap B$
$\overline{B}$: patient does not test positive	$A \bigcap \overline{B}$	$\overline{A} \bigcap \overline{B}$

CAMBRIDGE
Judge Business School

Bayes' Theorem example

Conditional probability

	A: patient has disease	$\overline{A}$: patient does not have disease
B: patient tests positive	$P(A \bigcap B)$ $= P(B\|A) \cdot P(A)$ $= 0.99 \cdot 0.001$ $= 0.00099$	$\overline{A} \bigcap B$
$\overline{B}$: patient does not test positive	$A \bigcap \overline{B}$	$\overline{A} \bigcap \overline{B}$

Recall: $P(A) = 0.001$; $P(B|A) = 0.99$; $P(B|\overline{A}) = 0.02$

Bayes' Theorem example

Conditional probability (cont'd)

	A: patient has disease	$\overline{A}$: patient does not have disease
B: patient tests positive	$P(A \bigcap B) = P(B\|A)P(A) = 0.99 \cdot 0.001 = 0.00099$	$P(\overline{A} \bigcap B) = P(B\|\overline{A})P(\overline{A}) = 0.02 \cdot 0.999 = 0.01998$
$\overline{B}$: patient does not test positive	$A \bigcap \overline{B}$	$\overline{A} \bigcap \overline{B}$

Recall: $P(A) = 0.001$; $P(B|A) = 0.99$; $P(B|\overline{A}) = 0.02$

CAMBRIDGE Judge Business School

Review

Conditional Probability

- Conditional Probability of Event B given Event A has occurred:
  $P(B|A) = \frac{P(A \bigcap B)}{P(A)}$
- If A and B are mutually exclusive,
  $P(B|A) = 0 = P(A|B)$
- Events A and B are independent if: $\quad P(B|A) = P(B)$
- Of course, $P(A|A) = 1$
- Rearranging the expression for conditional probability, Probability of Event $A and B : P(A \bigcap B) = P(B|A)P(A)$
- Note: $\quad P(A|B)P(B) = P(B|A)P(A)$
- If A and B are independent, Multiplication Rule:
  $P(A \bigcap B) = P(A)P(B)$

CAMBRIDGE Judge Business School

Bayes' Theorem example

Marginal distribution

	A: patient has disease	$\overline{A}$: patient does not have disease	
B: patient tests positive	0.00099	0.01998	$P(B) = 0.02097$
$\overline{B}$: patient does not test positive	$A \bigcap \overline{B}$	$\overline{A} \bigcap \overline{B}$	$P(\overline{B}) = 0.97903$
	$P(A) = 0.001$	$P(\overline{A}) = 1 - P(A) = 0.999$	1

CAMBRIDGE Judge Business School

Bayes' Theorem example

Joint distribution

	A: patient has disease	$\overline{A}$: patient does not have disease	
B: patient tests positive	0.00099	0.01998	$P(B) = 0.02097$
$\overline{B}$: patient does not test positive	0.00001	0.97902	$P(\overline{B}) = 0.97903$
	$P(A) = 0.001$	$P(\overline{A}) = 0.999$	1

CAMBRIDGE Judge Business School

Bayes' Theorem

The Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$
$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$$

- Can compute $P(A|B)$ from $P(A)$, $P(B)$, and the inverse conditional probability $P(B|A)$
- $P(A|B) = P(A \bigcap B)/P(B) = 0.00099/0.02097 = 0.0472$
- Probability that a person who tests positive has the disease is ≈ 0.05 !

CAMBRIDGE Judge Business School

Bayes' Theorem example

Conclusions

- Only 5% of those who test positive have the disease !
- $P(\overline{A}|B) = P(\overline{A} \bigcap B))/P(B) = 0.01998/0.02097 = 95\%$ (though $P(B|A) = 99\%$)
  - Probability of *false positives*, $P(B|\overline{A}) = 0.02$ given
  - In a group of 1000, on average, only 1 will have the disease, but 21 will test positive
  - 20 false positives come from the much larger uninfected group
  - But with a positive test, the chance of having the disease goes up from 1 in 1000 to 1 in 21
- Probability of *false negatives*, $P(\overline{B}|A)$? (Probability of having the disease though test is negative?)
  - $= P(A \bigcap \overline{B})/P(A) = 0.00001/0.001 = 1\%$

CAMBRIDGE Judge Business School

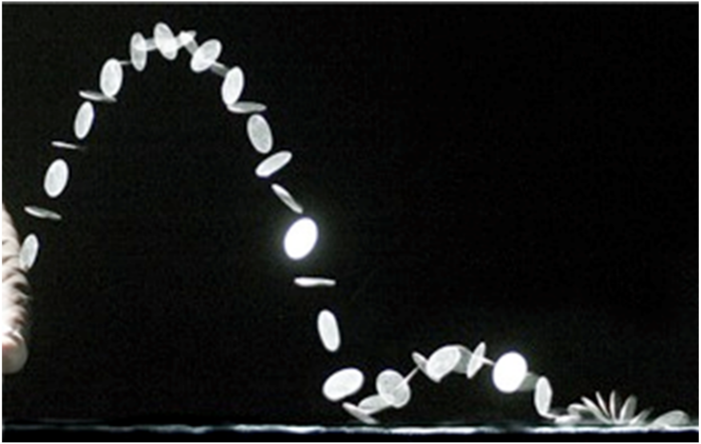
Random Variables

Random Variables: Review

- The outcome of an random "experiment" need not be a number
  - e.g., Coin toss : 'heads' or 'tails'
- To make progress we represent outcomes as numbers (but we pay attention to scale)
- We associate a unique real number with each elementary outcome of the experiment
  - Some set of real numbers represents the sample space of our random process
- The numerical value (outcome) will vary from trial to trial if the "experiment" is repeated
- The random variable (the experiment) is then characterized fully by its probability function

CAMBRIDGE Judge Business School

Diaconis on tossing coins



Discrete Random variables

Discrete: Countable number of elementary events

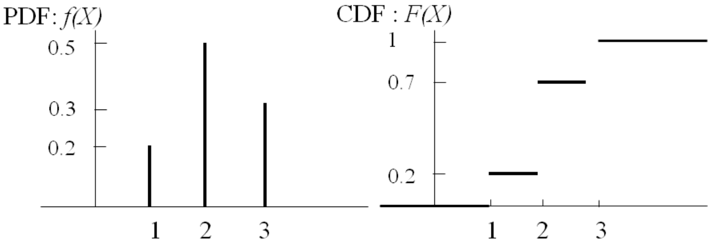
- Probability distribution function (Probability mass function): list of values of the discrete random variable with their chances of occurring
- $f(x) = Pr(X = x)$ Probability that random variable X takes value x
- Example: throwing a fair die, Sample space, $S = \{1, 2, 3, 4, 5, 6\}$
  $f(x_i) = 1/6$, $x_i = 1, 2, ..., 6$, $\sum_{i=1}^{6} f(x_i) = 1$

CAMBRIDGE Judge Business School

Cumulative Distribution Function

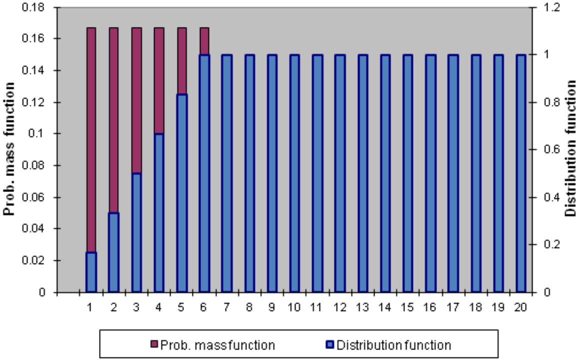
Probability that a discrete random variable X takes on a value less than or equal to x

- $F(x) = \sum_{X \leq x} f(X) = Pr(X \leq x)$
- $Pr(x_1 \leq X \leq x_2) = F(x_2) - F(x_1), x_2 \geq x_1$
- $Pr(2 \leq X \leq 5) = F(5) - F(2)$



Discrete uniform distribution

Fair dice: N=6, a=1, b=6



Quiz

Discrete random variable, probability and cumulative distribution functions

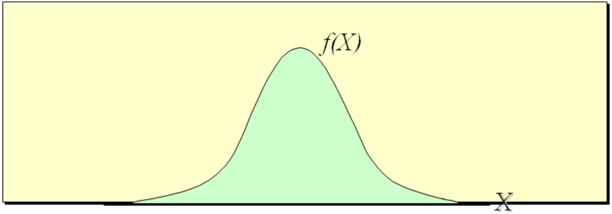
- Experiment: Throw a pair of fair dice
- X is a random variable defined as the *sum* of two die faces
- What does the probability distribution look like?
- What does the CDF look like?

CAMBRIDGE Judge Business School

Continuous Random variables

The variable (outcome) can take any real value in an *interval* on the real number line. This is the sample space

- Probability density function (probability density function) $f(X)$ is described graphically by a curve
- The area under the probability density function corresponds to probability: $\int_a^b f(X)dX = Pr(a \leq X \leq b)$
- $\int_{\text{Sample space}} f(X)dX = 1 \quad$ i.e., $\quad Pr(\text{sample space}) = 1$
- If sample space is the set of real numbers, $\int_{-\infty}^{\infty} f(X)dX = 1$

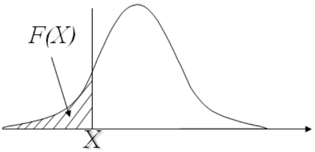


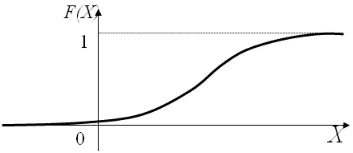
CAMBRIDGE Judge Business School

Cumulative distribution functions

Continuous random variables and Cumulative distribution functions

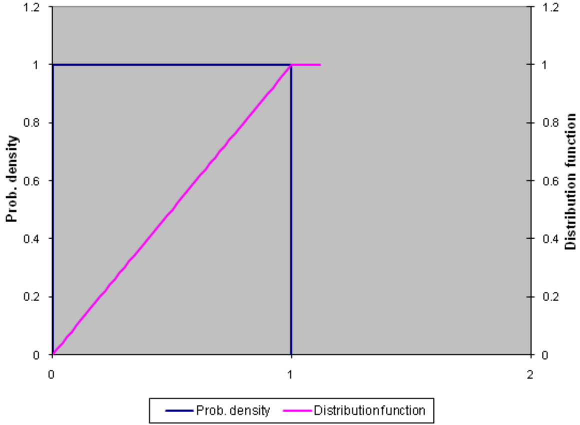
- The cumulative distribution function $F()$
- $F(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(V)dV$
- $F(b) - F(a) = Pr(a \leq X \leq b) = \int_{a}^{b} f(X)dX$
- Probability density f is the derivative of F





Uniform distribution

Continuous uniform (rectangular) distribution. $U[0,1]$

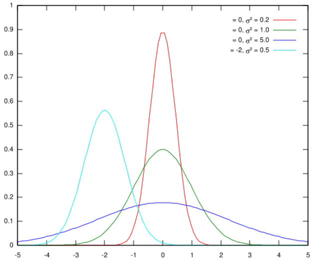


Normal distribution

$X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

μ=mean, σ= st. dev., $\pi = 3.14...$, $e = 2.71...$

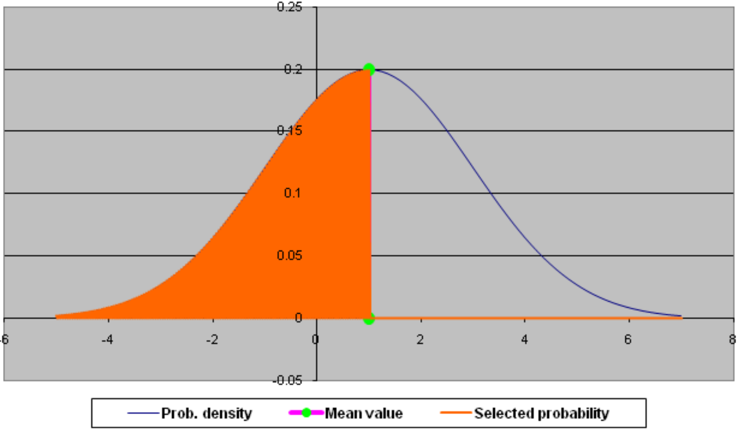


The normal distribution is, in fact, a family of distributions

CAMBRIDGE Judge Business School

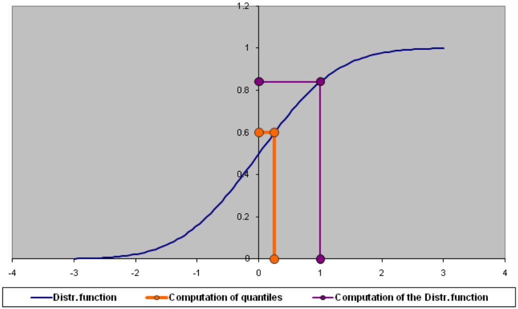
Normal distribution PDF

Normal distribution PDF. $\mu = 1, \sigma = 2)$



Standard Normal distribution

Standard Normal, N(0,1): CDF (e.g., $F(1) = 0.84$) and Quantiles (e.g., $F^{-1}(.6) = 0.25$)



Why is the Normal Distribution so common? Central Limit Theorem. Example

Moments of a Random variable

Expectation

- Characterizing the random variable in the *population* (not sample)
- The first moment of a discrete random variable X: *mean* / *expected value* / *expectation*

$$E(X) = \mu = x_1 p_1 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

- p_i = probability that $X = x_i$ in the population

CAMBRIDGE Judge Business School

Moments of a Random variable (cont'd)

Expectation of functions of a Random variable, E(f(X))

- Is a function of a random variable a random variable?
- Expected value of functions of X

$$E(X^2) = x_1^2 p_1 + ... + x_n^2 p_n = \sum_{i=1}^{n} x_i^2 p_i$$

$$E[g(X)] = g(x_1)p_1 + ... + g(x_n)p_n = \sum_{i=1}^{n} g(x_i)p_i$$

Second Central Moment

The Second Central Moment of a Random variable

- Central moments: moments about the mean
- Second moment: Variance
- For a discrete random variable:

$$Var(X) = \sigma^2 = E\left((X-\mu)^2\right)$$

$$Var(X) = (x_1-\mu)^2 p_1 + \cdots + (x_n-\mu)^2 p_n = \sum_{i=1}^{n} (x_i-\mu)^2 p_i$$

Standard deviation $= \sqrt{\text{Variance}} = \sigma$

3rd and 4th Central Moments

3rd and 4th Central Moments of a Random variable

$$Skewness = \frac{E[(X-\mu)^3]}{\sigma^3}$$

$$Kurtosis = \frac{E[(X-\mu)^4]}{\sigma^4}$$

CAMBRIDGE Judge Business School

Linear transformed Random variable

Quiz:

$$X_{new} = aX + b$$

- Mean of X_{new} = ?
- Median of X_{new} = ?
- Variance of X_{new}= ?

Adding independent random variables

Quiz:

- Mean and Variance of a sum of independent random variables
- Many useful *statistics* are linear combinations of data, i.e., of random variables
- If X and Y are *independent*:
  - $E(X+Y) = ?$
  - $Var(X+Y) = ?$

CAMBRIDGE Judge Business School

Moments of a Random variable

Quiz:

- Experiment: Throw a pair of fair dice, independently.
- Outcome $= X =$ sum of the faces
- Mean?
- Variance?
- Skewness?

CAMBRIDGE Judge Business School

Joint distribution of Two Discrete Random variables

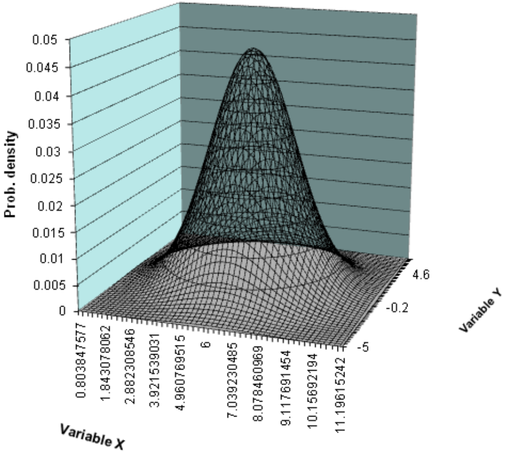
Example: X = the sum of two independent die faces

green \ red	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

CAMBRIDGE Judge Business School

Joint distribution: Bivariate Normal distribution

Independent r.v.s



Joint distribution of two continuous random variables

Danny Quah, 2000

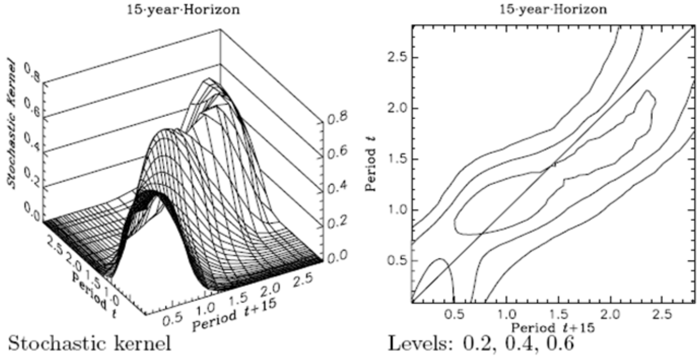


Fig. 3: Distribution dynamics across countries (Relative output per worker) The right panel contains contour plots of the 15-year stochastic kernel in the left panel.

Joint distributions and Covariance

The *covariance* between random variables X and Y:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \mu_{XY}$$

- Measure of linear association between X and Y;
- Units: Units of $X\times$ Units of Y
    - Positive linear relation between X and Y: $Cov(X, Y) > 0$
      Negative: $(Cov(X, Y) < 0)$
    - If X and Y independently distributed: $Cov(X, Y) = 0$
    - But not vice versa!! (Why?)
- The Covariance of a r.v. with itself is its variance:

$$Cov(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2] = \sigma_X^2$$

CAMBRIDGE Judge Business School

Covariance of functions of r.v.s

Covariance between linear functions of r.v.s:

- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
- $Cov(X,Y)$:

$$\begin{aligned}
\sigma_{XY} &= E\left((X-\mu_X)(Y-\mu_Y)\right) \\
&= E\left(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y\right) \\
&= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y
\end{aligned}$$

- Likewise, can show: $Cov(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}$

CAMBRIDGE Judge Business School

Correlation coefficient

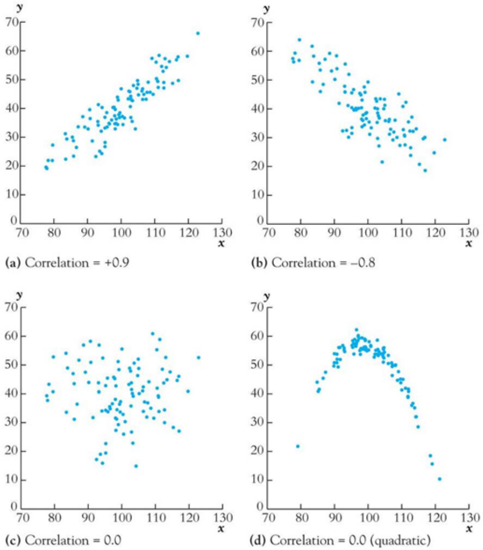
Correlation coefficient: standardized Covariance

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \rho_{XY}$$

- $-1 \leq \rho_{XY} \leq 1$
- Perfect positive linear dependence: $\rho_{XY} = 1$
  - $Y = \beta_0 + \beta_1 X$ for some constants β_0 and $\beta_1 > 0$
- Perfect negative linear dependence: $\rho_{XY} = -1$
  - $Y = \beta_0 + \beta_1 X$ for some constants β_0 and $\beta_1 < 0$
- No linear dependence: $\rho_{XY} = 0$

CAMBRIDGE Judge Business School

The correlation coefficient measures linear association



(a) Correlation = +0.9

(b) Correlation = –0.8

(c) Correlation = 0.0

(d) Correlation = 0.0 (quadratic)

CAMBRIDGE
Judge Business School

Joint distributions, review

Example

	$X = -100$	$X = 100$	
$Y = -50$	0.00099	0.01998	
$Y = 50$	0.00001	0.97902	

CAMBRIDGE Judge Business School

Marginal distributions

Example

	$X = -100$	$X = 100$	Marginal distribution of Y
Y=-50	0.00099	0.01998	0.00099+0.01998=**0.02097**
Y=50	0.00001	0.97902	0.00001+0.97902=**0.97903**
	0.001	0.999	

- **Marginal** distribution of X:
  $P(X = x) = \sum_j P(X = x, Y = y_j)$
- **Marginal** distribution of Y:
  $P(Y = y) = \sum_i P(X = x_i, Y = y)$

CAMBRIDGE Judge Business School

Conditional Distribution, Mean and Variance

Example

- Conditional distribution of Y, conditional on $X = -100$

	$X = -100$
Y=-50	$P(Y = -50 \mid X = -100) = 0.00099/0.001 = \mathbf{0.99}$
Y=50	$P(Y = 50 \mid X = -100) = 0.00001/0.001 = \mathbf{0.01}$

- Conditional Mean of Y, conditional on $X = -100$:

$$-50 * 0.99 + 50 * 0.01 = \mathbf{-49}$$

- Conditional Variance of Y, conditional on $X = -100$:

$$(-50 - (-49))^2 * 0.99 + (50 - (-49))^2 * 0.01 = \mathbf{99}$$

CAMBRIDGE Judge Business School

Conditional Distribution, Conditional Mean

Conditional distribution of Y given X:

- Distribution of Y, given the value of X: $P(Y = y|X = x)$

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

$$P(A|B) = \frac{P(A \text{ And } B)}{P(B)}$$

Conditional Mean=Mean of Conditional distribution

- Basic concept in regression

$$E(Y|X = x) = \sum_j y_j P(Y = y_j|X = x)$$

Conditional Variance

Conditional Variance=Variance of Conditional distribution

$$\sigma_Y^2(x) = Variance(Y|X = x)$$

- denote $E(Y|X = x) = \mu_Y(x)$

$$\sigma_Y^2(x) = \sum_j (y_j - \mu_Y(x))^2 \times P(Y = y_j|X = x)$$

CAMBRIDGE Judge Business School

Independence

If X and Y are independent:

- knowledge of X provides no information on Y, and vice versa
- $P(Y = y|X = x) = P(Y = y); P(A|B) = P(A)$
- $P(X = x|Y = y) = P(X = x); P(B|A) = P(B)$
- $P(X = x, Y = y) = P(X = x)P(Y = y)$; $P(A\&B) = P(A)P(B)$

CAMBRIDGE Judge Business School