Ch02_Basic Probability Theory

Prof. Cheolsoo Park

Basic Statistics

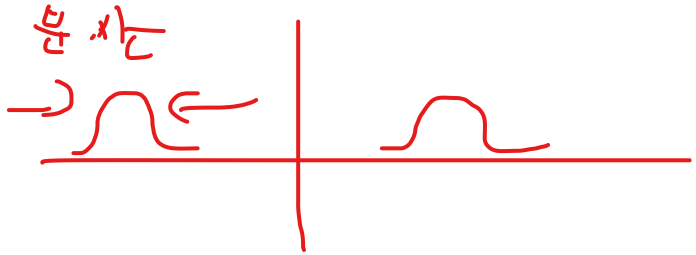
- Statistics is a study about an objective decision-making process through the data, converted to information

- In pattern recognition, various statistics method is utilized, and analysis based on prior knowledge is necessary

- Under the uncertainty, various probability theory is applied to classify unknown data into several categories

Basic Statistics

- Terminology of statistics
  - A population
    - Entire data for the analysis
  - Samples
    - Some part of data among all
  - Sample distribution
    - Distributions of sample data
- Statistical parameters
  - Mean (1st order)
    - A point of center for gravity
  - Variance (2nd order)
    - How much scattered numerical data is
- Standard deviation
  - Square root of variation



$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Eq. 1 numerical expression of mean

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Eq. 2 numerical expression of variance

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Eq. 3 numerical expression of STD

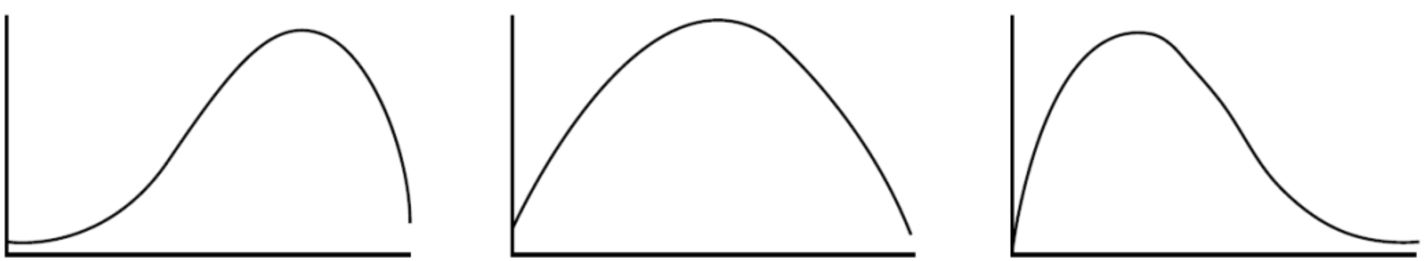
Basic Statistics

- Skewness (3rd oder)
  - A measure of the asymmetry of the data probability distribution
  - Negative skew : the left tail is longer; the mass of the distribution is concentrated on the right of the figure
  - Positive skew : the right tail is longer; the mass of the distribution is concentrated on the left of the figure
  - If the distribution is symmetry, skewness is 0

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, and the numerator m_3 is the sample third central moment.

Eq. 6 skewness



(a) Negative Skew (b) Skewness = 0 (c) Positive Skew

Basic Statistics

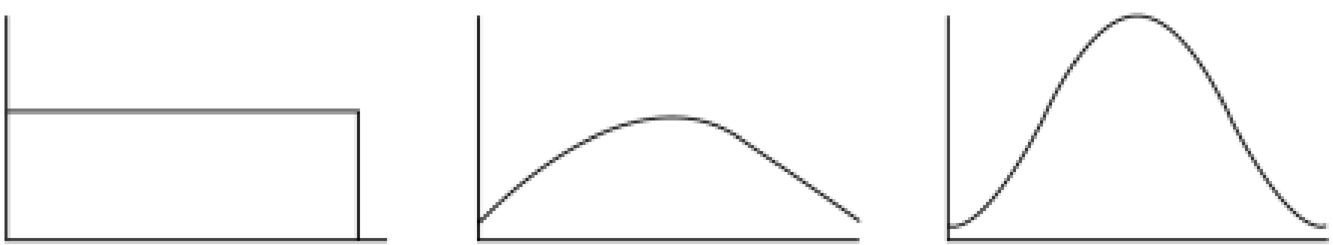
- Kurtosis (4th order)
  - Kurtosis is a descriptor of the shape of a probability distribution

$$Kurtosis = \frac{\mu_4}{\sigma^4} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^2}$$

where μ4 is the fourth moment about the mean
and σ is the standard deviation.

Eq. 7 kurtosis

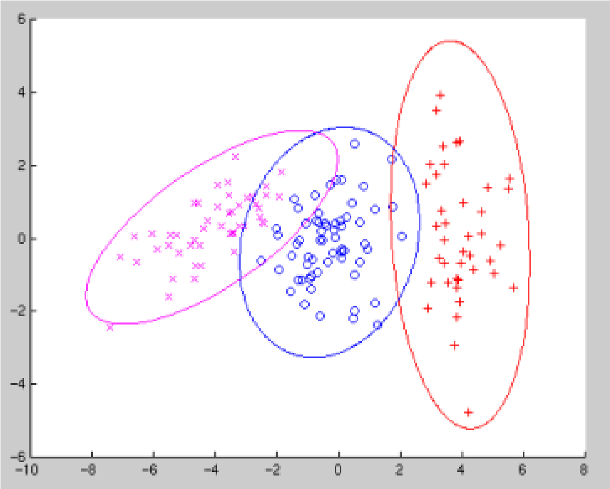
$$Kurtosis = \begin{cases} 3 \text{ less } (< 3): & \text{platykurtic} \\ 3(= 3) : & \text{normal} \\ 3 \text{ more } (> 3): & \text{leptokurtic} \end{cases}$$



(a) κ = 3 less (b) κ = 3 (c) κ = 3 more

Covariance and Correlation

- Covariance
  - How much related two random variables are with changing together
  - Zero covariance means no correlation between two



Example of 3 random variables

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$C(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \overline{y} - \sum_{i=1}^{n} \overline{x} y_i + \sum_{i=1}^{n} \overline{x}\,\overline{y}\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \overline{y} - \sum_{i=1}^{n} \overline{x} y_i + n\overline{x}\,\overline{y}\right)$$

Eq. 4 covariance of bivariate

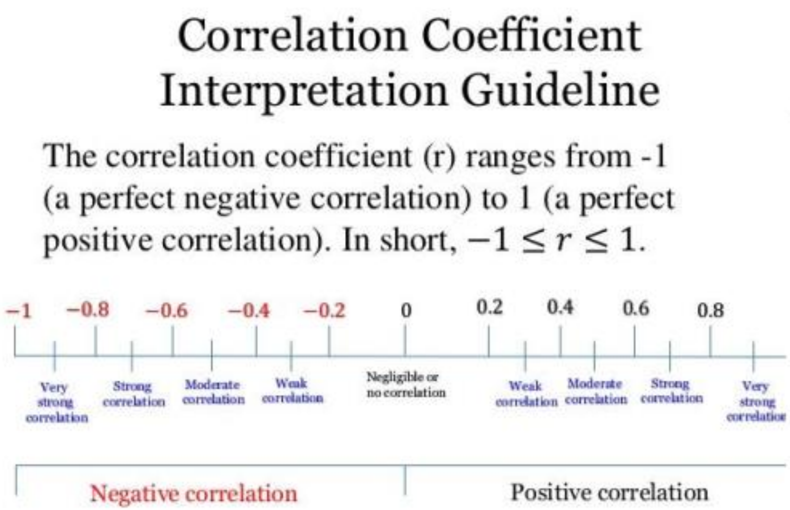
Covariance and Correlation

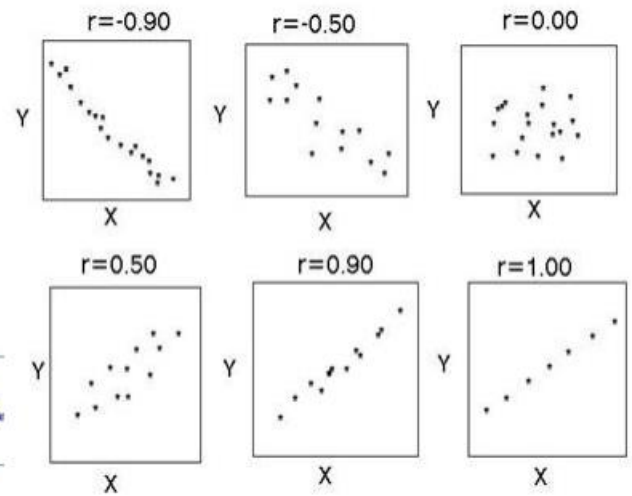
- Correlation
  - Normalised covariance by the standard deviation of two variables

  ➔ Variated from -1 to 1

$$0 \le \left|\rho_{xy}\right| \le 1$$

$$\rho = \frac{C(x,y)}{\sigma_x \sigma_y} = \frac{C(x,y)}{\sqrt{V_{(x)}}\sqrt{V_{(y)}}}$$ ※ σ_x, σ_y **is STD of variable** ***x, y***





linearity for 'r'

Basic Statistics

x_1: hours spending in library per a week
x_2: Test score of machine learning exam
x_3: days of class absence

x_1	x_2	x_3
35	11	0.5
5	12	0
0	5	10
20	9	4.5
15	8	5

	x_1	x_2	x_3
평균	15	9	4
분산	187.5	7.5	16.37

① x_1 , x_2 correlation coefficient

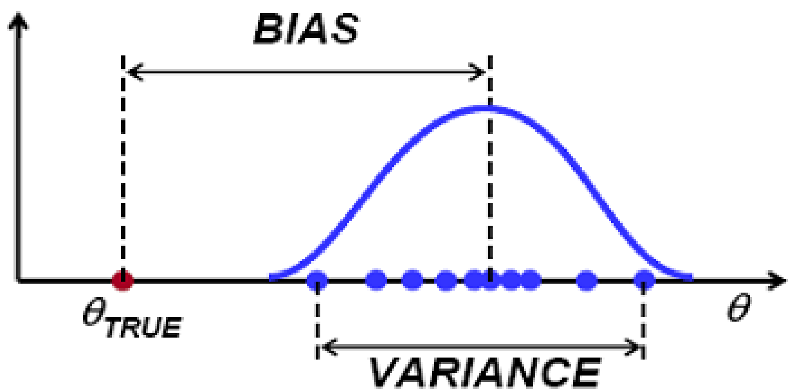
$$\rho_{12} = \frac{(35-15)(11-9)+(5-15)(12-9)+(0-15)(5-9)}{3\sqrt{187.5}\sqrt{7.5}} = 0.4667$$

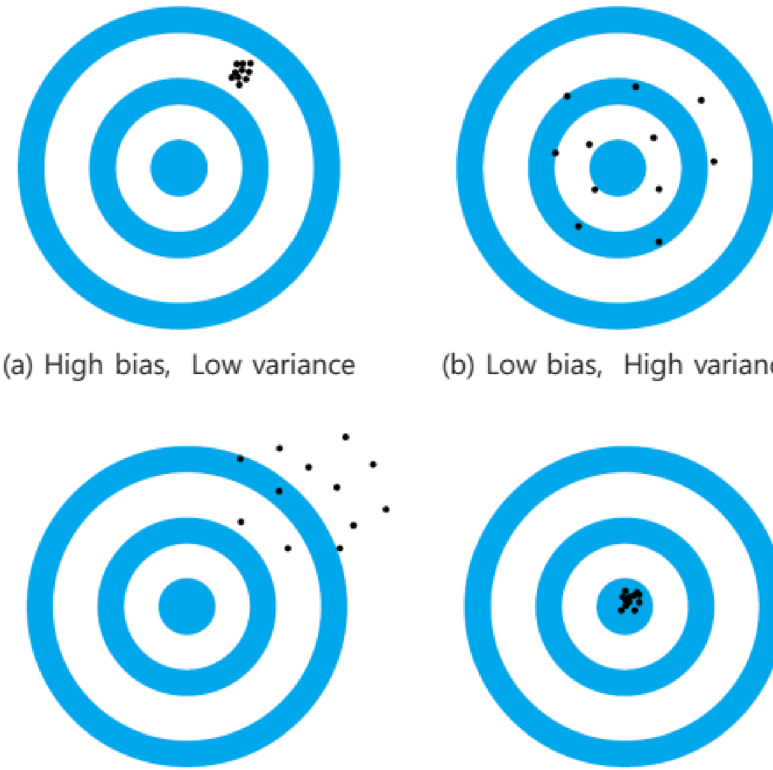
② x_2 , x_3 correlation coefficient

$$\rho_{23} = \frac{(11-9)(0.5-4)+(12-9)(0-4)+(5-9)(10-4)+(8-9)(5-4)}{4\sqrt{7.5}\sqrt{16.37}} = -0.9926$$

Bias and Variance

- Bias and Variance





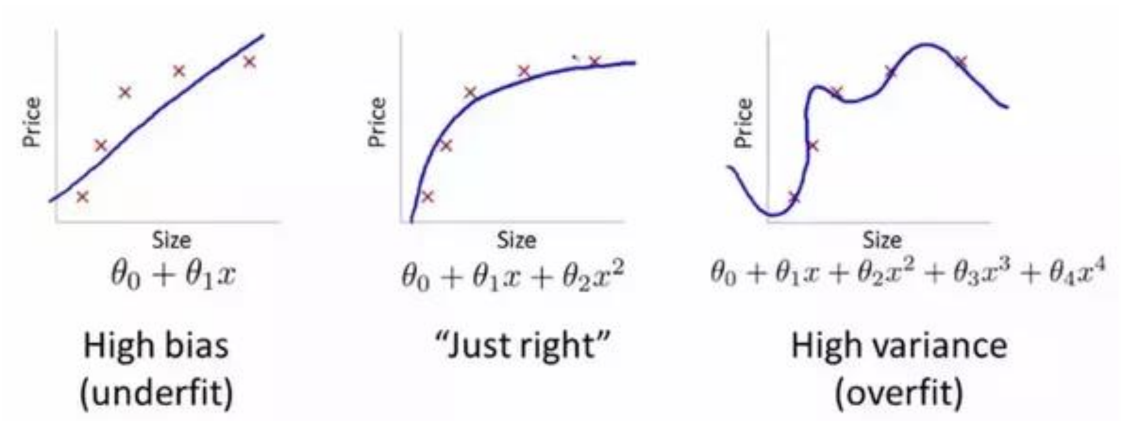
(a) High bias, Low variance (b) Low bias, High variance

(c) High bias, High variance (d) Low bias, Low variance

Ex) designing optimal size of mobile phone corresponding to users' preference

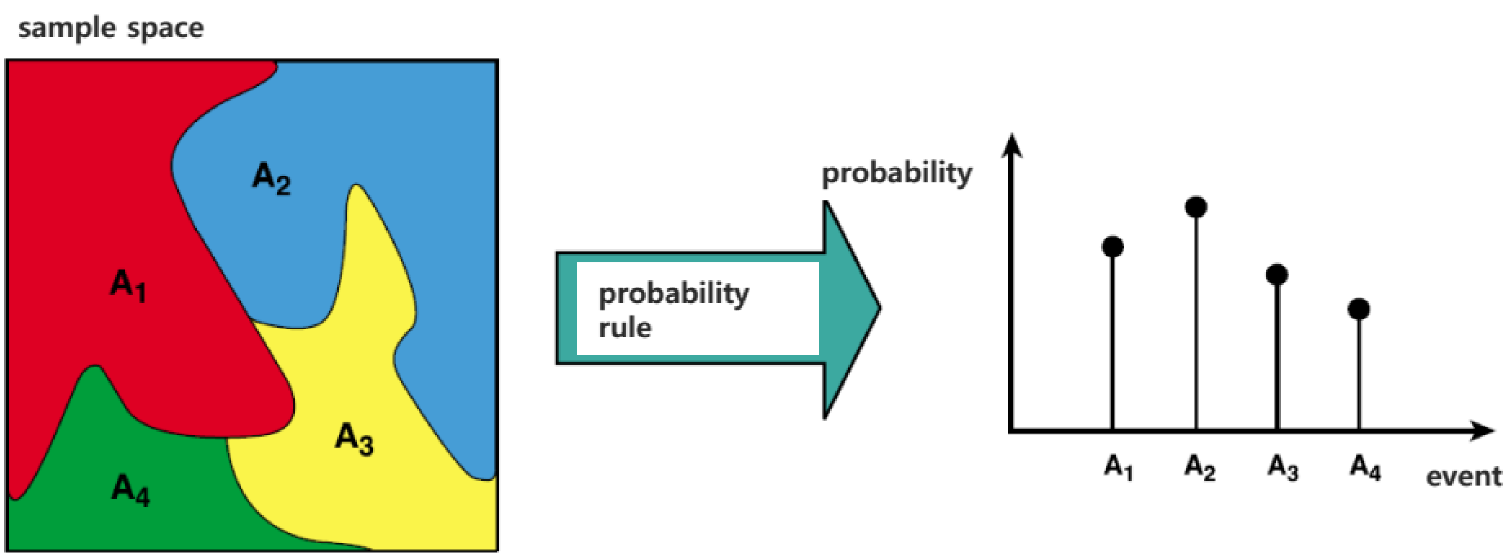
Bias and Variance

- Error(X) = noise(X)+bias(X)+variance(X)
  - noise : irreducible error due to its intrinsic property of data
  - bias and variance : reducible error depending on the model
- Bias and Variance Tradeoff
  - The bias is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).
  - The variance is an error from sensitivity to small fluctuations in the training set. High variance can cause overfitting: modeling the random noise in the training data, rather than the intended outputs.



Basic Probability

- Probability
  - Probability is the measure of the likelihood that an event will occur through statistic phenomenon
  - Probability is quantified as a number between 0 and 1
  - Probability rule
    - Probability rule is assigned probability of each events in random trial
    - Sample space of random trial, 'S' is all sets



Basic Probability

- Numerical Probability
  - When the possibility is equal that each event of sample, ‘S’, will occur, it is called numerical probability of event A about n(A)/n(S)
  - Think about the case of ‘dice’

- Statistic Probability
  - Not only the same probability of natural phenomenon or society phenomenon is rare, but also probability of each event has uncertainty
  - In this case, experiment is performed several times and event probability can be assumed through relative frequency
  - If event occur n of r, relative frequency can be expressed n/r

Basic Probability

- Event and exclusive events
  - Interesting event can be occurred by individual or multiplicative (complex) factors
    - ➔ this complex results are called event
  - If event A and B can not be occurred simultaneously, they are called exclusive events

Basic Probability

- Sample space and probability space
  - Sample space of an experiment or random trials is the set of all possible outcomes or results of the experiment
  - Elements of sample space is called ‘sample point’
  - Partial set of sample space is named as ‘event’, and only one sample point event is named as ‘fundamental event’
  - Probability space is set of <u>all result that can be occured</u>

Basic Probability

- Theorem of probability

theorem $0 \le P[A_i]$

theorem $P[S] = 1$

theorem **If** $A_i \cap A_j = \varnothing$, $P[A_i \cup A_j] = P[A_i] + P[A_j]$

- Characteristics of probability

! 1: $P[A^c] = 1 - P[A]$

! 2: $P[A] \le 1$

! 3: $P[\varnothing] = 0$

! 4: Given $\{A_1, A_2, \dots A_N\}$. if $\{A_i \cap A_j = \varnothing \ \forall i, j\}$, then $P[\bigcup_{k=1}^{N} A_k] = \sum_{k=1}^{N} P[A_k]$

! 5: $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$

! 6: $P[\bigcup_{k=1}^{N} A_k] = \sum_{k=1}^{N} P[A_k] - \sum_{j<k}^{N} P[A_k \cap A_j] + \dots + (-1)^{N+1} P[A_1 \cap A_2 \cap \dots \cap A_N]$

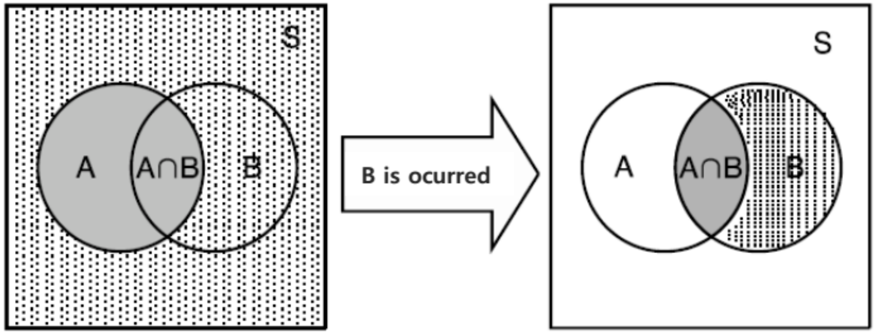
! 7: If $A_1 \subset A_2$, $P[A_1] \le P[A_2]$

Basic Probability

- Conditional Probability
  - Conditional probability is a measure of the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred
  - If the event of interest is A and the event B is known or assumed to have occurred, “the conditional probability of A given B”, or “the probability of A under the condition B,” is usually written as P(A|B)

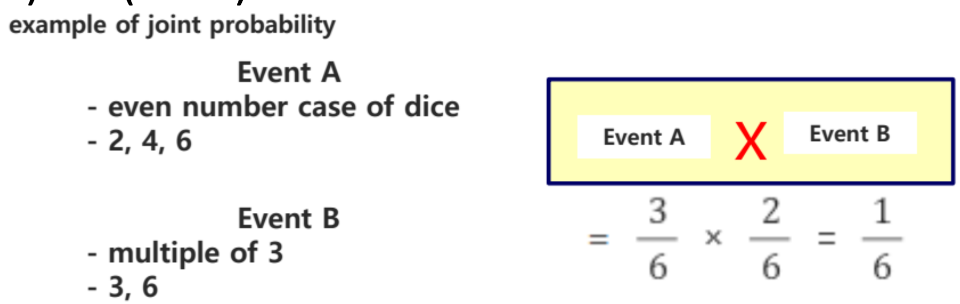
$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} \qquad ※\ P[B] > 0$$

Probability of A for Given B



Basic Probability

- Joint Probability
  - Event A and B are occurred simultaneously, called 'product rule'
  - If the event A and B are independent each other, $P(A|B)=P(A)$ and $P(A) \times P(B) = P(A \cap B)$



Ex) P(A) = lunch for today, P(B) = lunch for tomorrow

$P(A \cap B) = P(A) \times P(B)$ or $P(A) \times P(B|A)$ depending on your behavior

Ex) Restaurant event R:

P(R1) and P(R2) are the chance for restaurant 1 and 2

P(Tg) and P(Tb) are good and bad outcomes in a toilet

Using P(Tg|R1), P(Tb|R1), P(Tg|R2) and P(Tb|R2),

calculate P(Tg,R1), P(Tb,R1), P(Tg, R2) and P(Tb, R2)

* Possible for P(Tg|R1) = P(Tg|R2) = P(Tg), when they are independent

Basic Probability

- Chain Rule
  - The chain permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities
  - The chain rule produces this product of conditional probabilities:

$$P(A_1, A_2, A_3, A_4, \cdots, A_n)$$
$$= P(A_1 \mid A_2, A_3, A_4, \cdots, A_n) \times P(A_2 \mid A_3, A_4, \cdots, A_n) \times \cdots \times P(A_{n-1} \mid A_n) \times P(A_n)$$

Ex) Urn 1 has 1 black ball and 2 white balls. Urn 2 has 1 black ball and 3 white balls. We will choose the urn 1 and 2 with the same probability.

What's the probability of choosing the first urn and a white ball from it.

Basic Probability

- Chain Rule
  - The chain permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities
  - The chain rule produces this product of conditional probabilities:

$$P(A_1, A_2, A_3, A_4, \cdots, A_n)$$
$$= P(A_1 \mid A_2, A_3, A_4, \cdots, A_n) \times P(A_2 \mid A_3, A_4, \cdots, A_n) \times \cdots \times P(A_{n-1} \mid A_n) \times P(A_n)$$

Ex) Urn 1 has 1 black ball and 2 white balls. Urn 2 has 1 black ball and 3 white balls. We will choose the urn 1 and 2 with the same chance.

What's the probability of choosing the first urn and a white ball from it.

=> P(B|1) = 1/3, P(W|1) = 2/3, P(B|2) = 1/4, P(W|2) = ¾

P(1, W) = P(1) x P(W|1) = ½ x 2/3

Basic Probability

- Marginal Probability
  - The marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset
  - It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables

		Mon	Tue	Wed	Thurs	
	Y \ X	x_1	x_2	x_3	x_4	$p_y(Y)\downarrow$
Football	y_1	$4/32$	$2/32$	$1/32$	$1/32$	$8/32$
Basketball	y_2	$2/32$	$4/32$	$1/32$	$1/32$	$8/32$
Baseball	y_3	$2/32$	$2/32$	$2/32$	$2/32$	$8/32$
Swim	y_4	$8/32$	0	0	0	$8/32$
	$p_x(X) \rightarrow$	$16/32$	$8/32$	$4/32$	$4/32$	$32/32$

Joint and marginal distributions of a pair of discrete, random variables X,Y having nonzero mutual information I(X; Y). The values of the joint distribution are in the 4×4 square, and the values of the marginal distributions are along the right and bottom margins.

Bayes' Theorem

- Bayes' theorem

$$P[B_j \mid A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A \mid B_j] \cdot P[B_j]}{\sum_{k=1}^{N} P[A \mid B_k] \cdot P[B_k]}$$

$$P[\omega_j \mid \mathrm{x}] = \frac{P[\mathrm{x} \mid \omega_j] \cdot P[\omega_j]}{\sum_{k=1}^{N} P[\mathrm{x} \mid \omega_k] \cdot P[\omega_k]} = \frac{P[\mathrm{x} \mid \omega_j] \cdot P[\omega_j]}{P[\mathrm{x}]}$$

※ ω_j : j -th class

x : Feature Vector

- $P[\omega_j]$: **the probabilities (prior probability)**
- $P[\omega_j \mid \mathrm{x}]$: **a conditional probability, is the probability of observing event A given that B is true.**
- $P[\mathrm{x} \mid \omega_j]$: **the probability of observing event B given that A is true.**
- $P[\mathrm{x}]$: **prior probability of X. P(B) =∑P(B|A)P(A)**


21
INSPIRED BY A TRUE STORY.
© 2007 Sony Pictures Digital Inc. All rights reserved.
THIS FILM IS NOT YET RATED
COLUMBIA PICTURES

Bayes' Theorem

- Example of Bayes' theorem

P(S) : probability of single, P(M) : probability of married

P(C1) : probability of area 1, P(C2) : probability of area2

Goal : P(S|C1) and P(S|C2)

P(S) and P(M) : from the population statistics

P(C1|S), P(C2|S), P(C1|M) and P(C2|M) : survey from people

=> $P(S|C1) = \frac{P(S)P(C1|S)}{P(C1)} = \frac{P(S)P(C1|S)}{\sum_{X=S,M} P(C,X)}$

$$= \frac{P(S)P(C1|S)}{P(S)P(C1|S) + P(M)P(C1|M)}$$