Probability & Statistics : A Brief Introduction

Ishwar Sethi
CSI5810, Fall2020

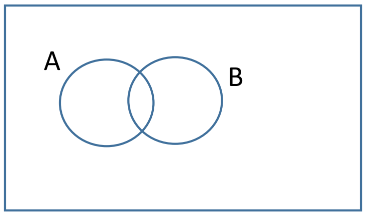
Probability and Statistics?

- Probability
  - Formally defined using a set of axioms
  - Seeks to determine the likelihood that a given event or observation, or measurement will or has happened
    - What is the probability of throwing a 7 using two dice?
- Statistics
  - Used to analyze the frequency of past events
  - Uses a given sample of data to assess a probabilistic model's validity or determine values of its parameters
    - After observing several throws of two dice, can we determine whether they are loaded

Probability

- Probability can be defined in terms of Kolmogorov axioms
  - The probability is a real-valued function defined on subsets A,B,… in sample space S

Intersection of A & B



$$\text{For every subset } A \text{ in } S,\ P(A) \geq 0$$

$$\text{If } A \cap B = 0, P(A \cup B) = P(A) + P(B)$$

$$P(S) = 1$$

  - This means the probability is a measure in which the measure of the entire sample space is 1

Two fair dice are tossed.
# of possible outcomes: 36
Let’s define event A as all throws where the sum of two faces is 4
Event B occurs whenever d2 greater or equal then d1

P(A) = ? 3/36
P(B) = ? 21/36

P(A ∧ B) = ? 2/36

P(B|A) = ? 2/3

P(A|B) = ? 2/21

Conditional probability formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Bayes' Theorem

- Bayes' theorem

Using $P(A \cap B) = P(B \cap A)$

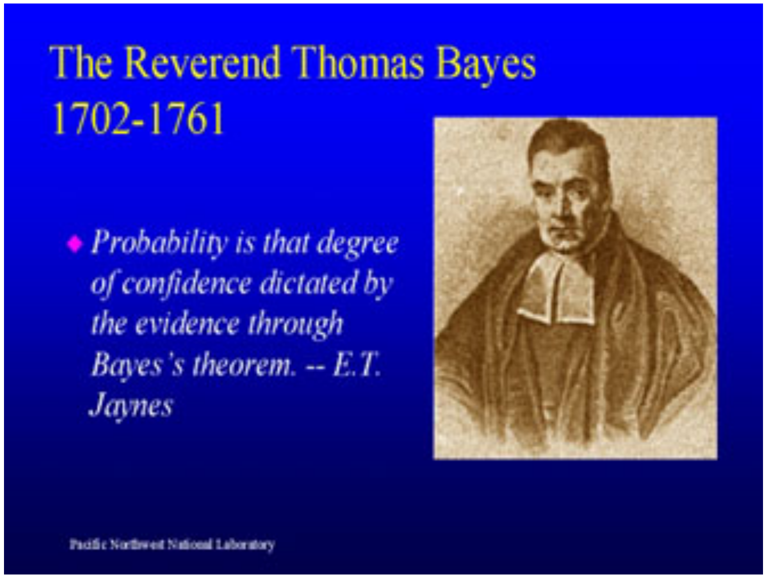
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- For disjoint A_i

$$P(B) = \sum_i P(B \mid A_i)P(A_i)$$

then

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{\sum_i P(B \mid A_i)P(A_i)}$$



Bayes' Theorem

- Suppose in the general population
  - P(disease) = 0.001
  - P(no disease) = 0.999
- Suppose there is a test to check for the disease
  - P(+, disease) = 0.98
  - P(-, disease) = 0.02
- But also
  - P(+, no disease) = 0.03
  - P(-, no disease) = 0.97
- You are tested for the disease and it comes back +. Should you be worried?

Bayes' Theorem

- Apply Bayes' theorem

$$P(disease,+) = \frac{P(+,disease)P(disease)}{P(+,disease)P(disease) + P(+,no\ disease)P(no\ disease)}$$

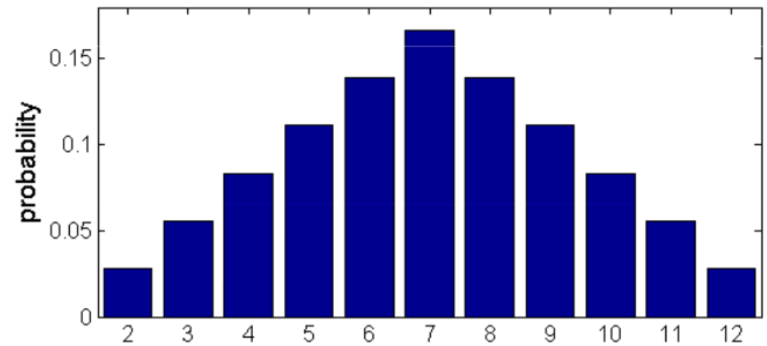
$$P(disease,+) = \frac{0.98 \cdot 0.001}{0.98 \cdot 0.001 + 0.03 \cdot 0.999} = 0.032$$

- 3.2% of people testing positive have the disease
- Your degree of belief about having the disease is 3.2%

Probability distributions

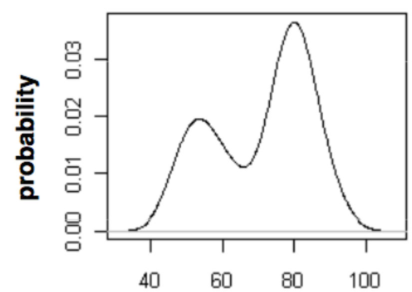
- Discrete: *probability mass function* (pmf)

  example:
  sum of two
  fair dice



- Continuous: *probability density function* (pdf)

  example:
  waiting time between
  eruptions of Old Faithful
  (minutes)



Probability Density Function

- Let the sample space S be the space of all possible outcomes of an experiment
- Let x be a possible outcome
  - Then P(x found in [x,x+dx]) = f(x)dx
  - f(x) is called the probability density function (pdf)
  - It may be called $f(x;\theta)$ since the pdf could depend on one or more parameters θ
    - Often, we will want to determine θ from a set of measurements
  - Of course x must be somewhere so

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Mean and Variance

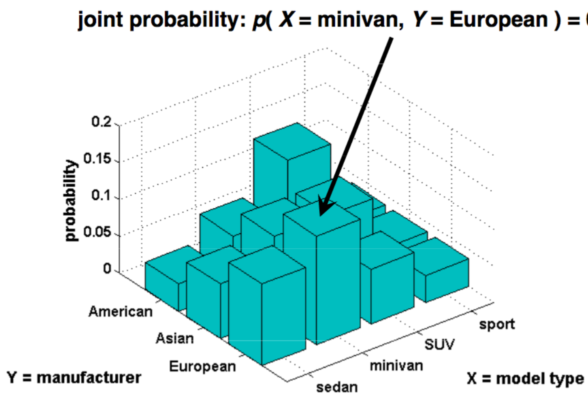
- Definitions of mean and variance are given in terms of expectation values

$$E[x] = \int x f(x) dx \equiv \mu$$

$$V[x] = E\left[(x - E[x])^2\right] = E\left[x^2\right] - \mu^2 \equiv \sigma^2$$

Multivariate probability distributions

- Scenario
  - Several random processes occur together
  - Want to know probabilities for each possible combination of outcomes
- Can describe as *joint probability* of several random variables
- Example: two processes whose outcomes are represented by random variables X and Y. Probability that process X has outcome x and process Y has outcome y is denoted as: $p(X = x, Y = y)$

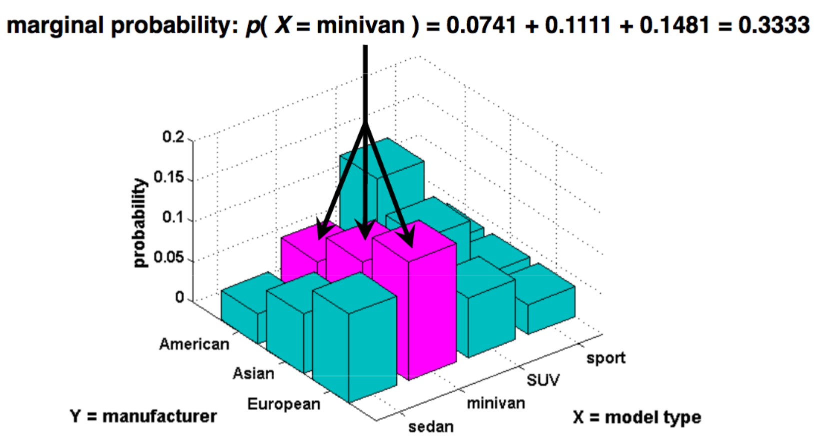


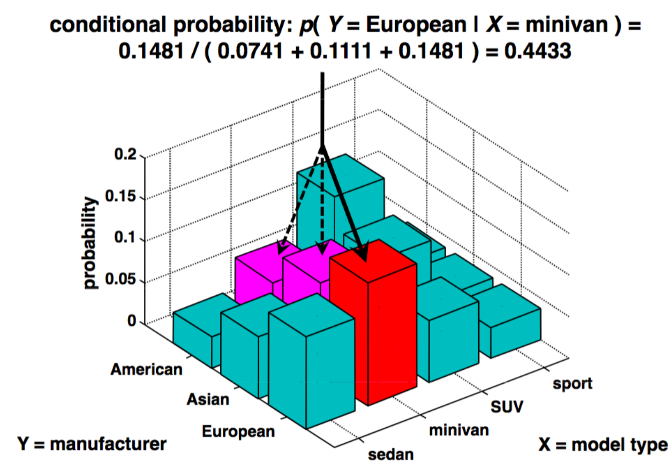
Multivariate probability distributions

- *Marginal* probability
  - Probability distribution of a single variable in a joint distribution
  - Example: two random variables X and Y:

    $p(X = x) = \sum_{b=\text{all values of } Y} p(X = x, Y = b)$
- *Conditional* probability
  - Probability distribution of one variable *given* that another variable takes a certain value
  - Example: two random variables X and Y:

    $p(X = x \mid Y = y) = p(X = x, Y = y) / p(Y = y)$





Covariance & Correlation

- Definitions of covariance and correlation coefficient

$$V_{xy} = \mathrm{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x \mu_y$$

$$\rho_{xy} = \frac{\mathrm{cov}[x, y]}{\sigma_x \sigma_y}$$

$$\text{if x, y independent then } f(x,y) = f_x(x) f_y(y)$$

$$\text{and then } E[x,y] = \iint xyf(x,y)dxdy = \mu_x \mu_y$$

$$\text{and so } \mathrm{cov}[x, y] = 0$$

Binomial Distribution

- Consider N independent experiments (Bernoulli trials)
- Let the outcome of each be pass or fail
- Let the probability of pass = p

$$\text{Probability of n successes} = p^n (1-p)^{N-n}$$

$$\text{But there are } \frac{N!}{n!(N-n)!} \text{ permuations for}$$

$$N \text{ distinguishable objects, grouping them n at a time}$$

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

An Example

100 examples from two classes are present in a training set. C1 is represented by 95% of the examples and class C2 by 5%. A random sample of 10 examples is selected. What is the probability that the selected sample of 10 examples has no example from class C2.

- First calculate the number of possible samples (ways in which 10 items out of 100 can be selected)
- This is given by 100!/10!90!
- The number of possible ways in which no C2 example is selected = (5!/0!5!)(95!/10!85!)
- Thus the probability that a randomly selected set of 10 examples has representation from class C2:
- = (5!/0!5!)(95!/10!85!)/(100!/10!90!) = 0.58375

Binomial Distribution

- For the mean and variance we obtain (using small tricks)

$$E[n] = \sum_{n=0}^{N} n f(n; N, p) = Np$$

$$V[n] = E[n^2] - (E[n])^2 = Np(1-p)$$

- And note with the binomial theorem that

$$\sum_{n=0}^{N} f(n; N, p) = \sum_{n=0}^{N} \binom{N}{n} p^n (1-p)^{N-n} = (p+1-p)^N = 1$$

Binomial Distribution

- It's baseball season! What is the probability of a 0.300 hitter getting 4 hits in one game?

$$N = 4, p = 0.3$$

$$f(4;4,0.3) = \frac{4!}{4!0!} \cdot 0.3^4 \cdot 0.7^0 = 0.0081$$

$$E[n] = 4 \cdot 0.3 = 1.2$$

$$V[n] = 4 \cdot 0.3 \cdot 0.7 = 0.84$$

$$\text{Expect } 1.2 \pm 0.84 \text{ hits for a } 0.300 \text{ hitter}$$

Poisson Distribution

- Consider when

$$N \to \infty$$

$$p \to 0$$

$$E[n] = Np \to v$$

$$\text{The Poisson pdf is}$$

$$f(n;v) = \frac{v^n}{n!} e^{-v}$$

$$\text{and one finds}$$

$$E[n] = v$$

$$V[n] = \sigma^2 = v$$

Poisson Distribution

$$f(n;N,p)=\frac{N!}{n!(N-n)!}p^n(1-p)^{N-n}$$

$$\frac{N!}{(N-n)!}\approx N^n \text{ for N large}$$

$$(1-p)^{N-n}=1-p(N-n)+\frac{p^2(N-n)(N-n-1)}{2!}+\ldots$$

$$(1-p)^{N-n}\approx 1-Np+\frac{N^2p^2}{2!}+\ldots=e^{-Np}=e^{-v}$$

$$f(n;N,p)\rightarrow\frac{N^n}{n!}p^ne^{-v}=\frac{v^ne^{-v}}{n!} \text{ for large } N \text{ and small } p$$

Poisson Distribution

- Examples
  - Number of customers/hour arriving at a store
  - Cosmic rays observed in a time interval t
  - Number of entries in a histogram bin when data is accumulated over a fixed time interval
  - Infant mortality
  - QC/failure rate predictions

Gaussian Distribution

- Gaussian distribution
  - Important because of the central limit theorem
    - For n independent variables $x_1, x_2, \ldots, x_N$ that are distributed according to any pdf, then the sum $y = \sum x_i$ will have a pdf that approaches a Gaussian for large N
    - Examples are almost any measurement error (energy resolution, position resolution, …)

$$E[y] = \sum_i \mu_i$$

$$V[y] = \sum_i \sigma_i$$

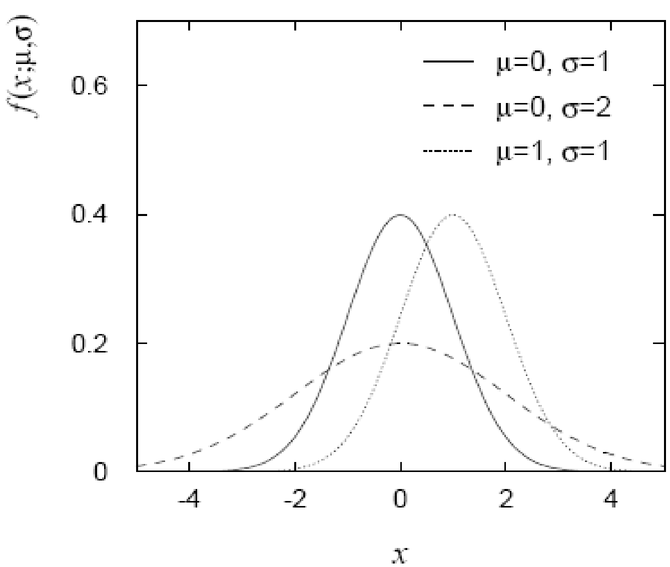
Gaussian Distribution

- The familiar Gaussian pdf is

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

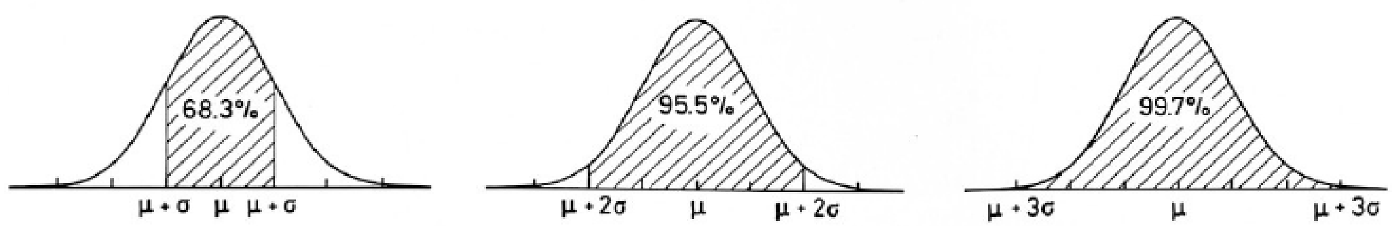
$$E[x]=\mu$$

$$V[x]=\sigma^2$$



Gaussian Distribution

- Some useful properties of the Gaussian distribution are
  - P(x in range $\mu \pm \sigma$) = 0.683
  - P(x in range $\mu \pm 2\sigma$) = 0.9555
  - P(x in range $\mu \pm 3\sigma$) = 0.9973
  - P(x outside range $\mu \pm 3\sigma$) = 0.0027
  - P(x outside range $\mu \pm 5\sigma$) = 5.7×10^{-7}



  - P(x in range $\mu \pm 0.6745\sigma$) = 0.5

Multivariate Normal Distribution

Let

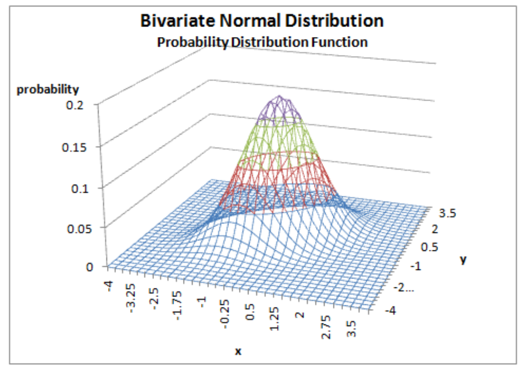
$$f(x_1, x_2) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x_1, x_2)}$$

where

$$Q(x_1, x_2) = \frac{\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}}{1-\rho^2}$$

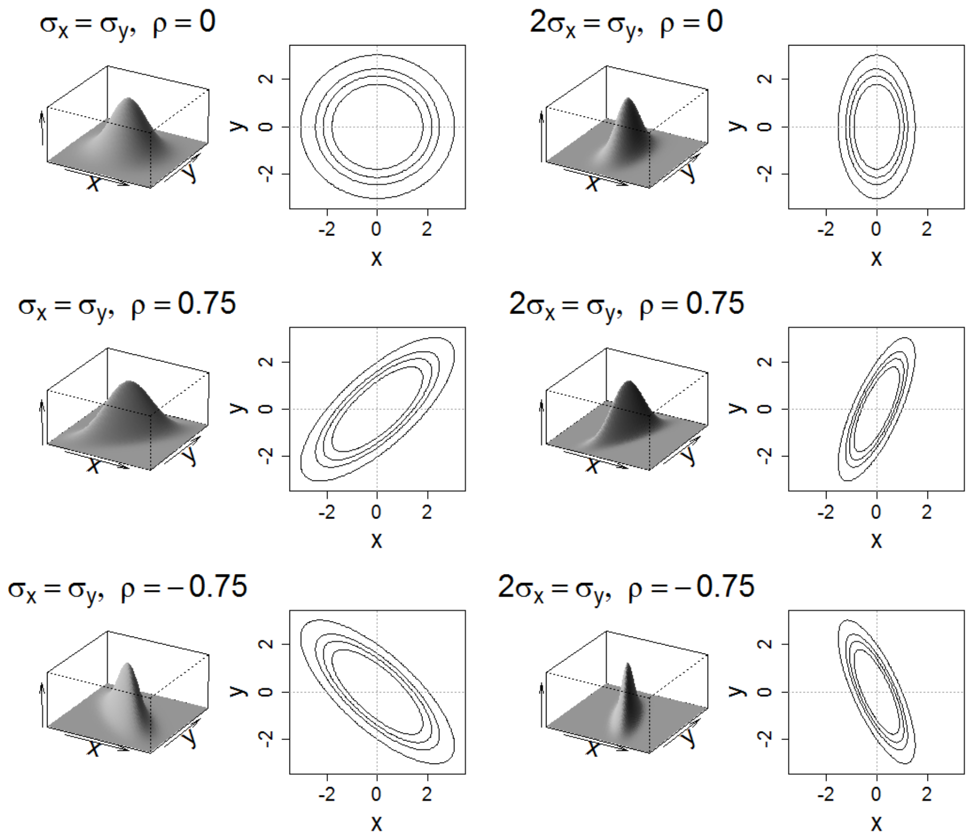
This distribution is called the **bivariate Normal distribution.**

The parameters are μ_1, μ_2, σ_1, σ_2 and ρ.



$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

$$= \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{|2\pi\boldsymbol{\Sigma}|}},$$


σx = σy, ρ = 0
2σx = σy, ρ = 0
σx = σy, ρ = 0.75
2σx = σy, ρ = 0.75
σx = σy, ρ = −0.75
2σx = σy, ρ = −0.75
x
y
-2
0
2

For more on probability and statistics, download the following slides.

http://users.encs.concordia.ca/~doedel/courses/comp-233/slides.pdf