Probability

Application to classification using Bayes' theorem
Application to discrete event simulation of queues using the Poisson distribution

David Vernon
Carnegie Mellon University Africa

vernon@cmu.edu
www.vernon.eu

Probability

Probability provides a mathematical foundation for many concepts

- Information
- Belief
- Uncertainty
- Confidence
- Randomness
- Variability
- Chance
- Risk

Probability

Probability Theory provides

- A framework for making inferences and testing hypotheses based on uncertain empirical data

- Building systems that operate in an uncertain world

  - Machine perception (speech recognition, computer vision)
  - Artificial intelligence

- Theoretical framework for understanding how the brain works

  - Many computational neuroscientists think the brain is a probabilistic computer build with unreliable components (i.e. neurons)

Probability

Probability Theory provides

- **A way of combining different sources of uncertain information to make rational decisions**

Probability

Three major interpretations of probability

Frequentist: probability as a relative frequency

- Probability of an event as the proportion of times such an event is expected to happen in the long run.
- The probability of an event E would be the limit of the relative frequency of occurrence of that event as the number of observations grows large

$$P(\mathrm{E}) = \lim_{n \to \infty} \frac{n_{\mathrm{E}}}{n}$$

Number of times the event is observed (n_{E})

Number of independent experiments (n)

Probability

Three major interpretations of probability

Bayesian or subjectivist: probability as uncertain knowledge

- "I will probably get an A in this class"

  By which we mean, "based on what I know about myself and about this class, I would not be very surprised if I get an A. However, I wouldn't bet my life on it, since there are a multitude of factors which are difficult to predict and that could make it impossible for me to get an A"

- This notion of probability is cognitive and does not need to be grounded in empirical frequencies

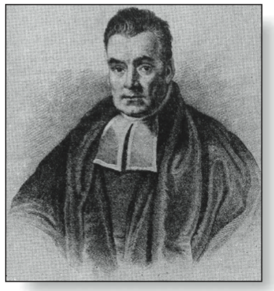
  "I will probably die poor" … not able to repeat that experiment many times and count the number of lives in which I die poor

Probability

Three major interpretations of probability

Bayesian or subjectivist:

- Useful in the field of machine intelligence
- Need to have knowledge systems capable of handling the uncertainty of the world
- Probabilists that are willing to represent internal knowledge using probability theory are called "Bayesian" (since Bayes was the first mathematician to do so)



THOMAS BAYES (1702–1761)

Probability

Three major interpretations of probability

Axiomatic or mathematical: probability as a mathematical model

- Rigorous definition
- Traceable to first principles
- Application of probability theory is not the main concern

Probability

- **Experiment**: a procedure that yields one of a given set of possible outcomes
  - E.g. rolling a die, tossing a coin, tossing a coin two times, observing the number of cars that arrive in a given period

- **Sample space** S: set of possible outcomes
  - E.g. Ω = {1,2,3,4,5,6}, Ω = {H, T}, Ω = {(H,H), (H, T), (T, H), (T, T)}

    Ω = {1 ,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

- **Event** E: subset of the sample space
  - Sets of outcomes
  - The set of all events is called the **event space**
  - E.g. Rolling an even number on a die: E = {2, 4, 6}

Probability

- **Experiment**: a procedure that yields one of a given set of possible outcomes
  - E.g. rolling a die, tossing a coin, tossing a coin two times, observing the number of cars that arrive in a given period

- **Sample space** S: set of possible outcomes
  - E.g. Ω = {1, 2, 3, 4, 5, 6}, Ω = {H, T}, Ω = {(H, H), (H, T), (T, H), (T, T)}
    Ω = {1 ,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

- **Event** E: subset of the sample space
  - Sets of outcomes
  - The set of all events is called the **event space**
  - E.g. Rolling an even number on a die: E = {2, 4, 6}

Probability

– **Experiment**: a procedure that yields one of a given set of possible outcomes
  - E.g. rolling a die, tossing a coin, tossing a coin two times, observing the number of cars that arrive in a given period

– **Sample space** S: set of possible outcomes
  - E.g. Ω = {1, 2, 3, 4, 5, 6}, Ω = {H, T}, Ω = {(H, H), (H, T), (T, H), (T, T)}

    Ω = {1 ,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

– **Event** E: subset of the sample space
  - Sets of outcomes
  - The set of all events is called the **event space**
  - E.g. Rolling an even number on a die: E = {2, 4, 6}

Probability

Definition of probability

For an event E, and a sample space S

The probability of E is $$p(E) = \frac{|E|}{|S|}$$

The probability of an event is between 0 and 1

– Example: an box contains four blue balls and five red balls; what is the probability that a ball chosen at random for the box is blue?

9 possible outcomes, four produce a blue ball, so probability is $4/9$

Probability

Probabilities of Complements and Unions of Events

For an event E, and a sample space S

The probability of the complement of E, $\overline{E} = S - E$, is given by

$$p(\overline{E}) = 1 - p(E)$$

- Example: A sequence of 10 bits is randomly generated. What is the probability that at least one of these bits is 0?

  Let E be the event with at least one of the 10 bits is 0

  Then $\overline{E}$ is the event that all the bits are 1.

  The sample space S is the set of all strings of length 10,

Probability

Probabilities of Complements and Unions of Events

For an event E, and a sample space S

The probability of the complement of E, $\overline{E} = S - E$, is given by

$$p(\overline{E}) = 1 - p(E)$$

– Example: A sequence of 10 bits is randomly generated. What is the probability that at least one of these bits is 0?

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{2^{10}}$$

$$= 1 - \frac{1}{1024} = \frac{1023}{1024}.$$

Probability

Probability measures

- You can think of probability as a function that assigns a number to a set: probability 'measures' a set (hence probability measures)

- If events $E_1, E_2, \ldots\ E_n$ are disjoint (i.e. no elements in common)

$$p(E_1 \cup E_2 \ldots \cup E_n) = p(E_1) + p(E_2) + \ldots\ p(E_n)$$

- Probability of rolling a die and getting a 1: $p\ (\{1\}) = 1/6$
  same for 2, 3, 4, 5, and 6.

$$p(\{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\}) = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1$$

Probability

Probabilities of Intersection of Events: Joint Probability

– For events E_1 and E_2 in a sample space S

$$p(E_1, E_2) = p(E_1 \cap E_2)$$

– The joint probability of two or more events is the probability of the intersection of those events

Probability

Probabilities of Intersection of Events: Joint Probability

- Consider the event $E_1 = \{2, 4, 6\}$ when rolling a die (rolling an even number)

- Consider the event $E_2 = \{4, 5, 6\}$ (rolling a number greater than 3)

- The joint probability (rolling an even number greater than 3) …

  - $p(E_1) = p(\{2\} \cup \{4\} \cup \{6\}) = 3/6$
  - $p(E_2) = p(\{4\} \cup \{5\} \cup \{6\}) = 3/6$
  - $p(E_1 \cap E_2) = p(E_1, E_2) = p(\{4\} \cup \{6\}) = 2/6$

  - Thus the joint probability of E_1 and E_2, $p(E_1, E_2)$, is $1/3$

Probability

Probabilities of Complements and Unions of Events

For events E_1 and E_2 in a sample space S

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

– Example: What is the probability that a positive integer selected at random from the set of positive integers less than or equal to 100 is divisible by either 2 or 5?

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

$$= \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5}.$$

Probability

Probabilities of outcomes of experiments where outcomes may **not** be equally likely

- Let S be a sample space of an experiment with a finite or countable number of outcomes

$p(s)$ is the probability of each outcome s

$$0 \leq p(s) \leq 1 \text{ for each } s \in S$$

$$\sum_{s \in S} p(s) = 1.$$

Probability

When there are n possible outcomes

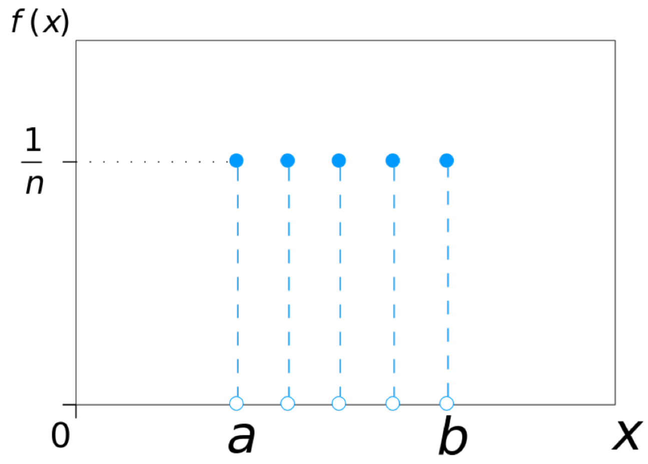
$$0 \leq p(x_i) \leq 1 \text{ for } i = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} p(x_i) = 1.$$

The function $p(s)$ or $p(x_i)$ from the set of all outcomes of sample space S is called a probability distribution

Probability

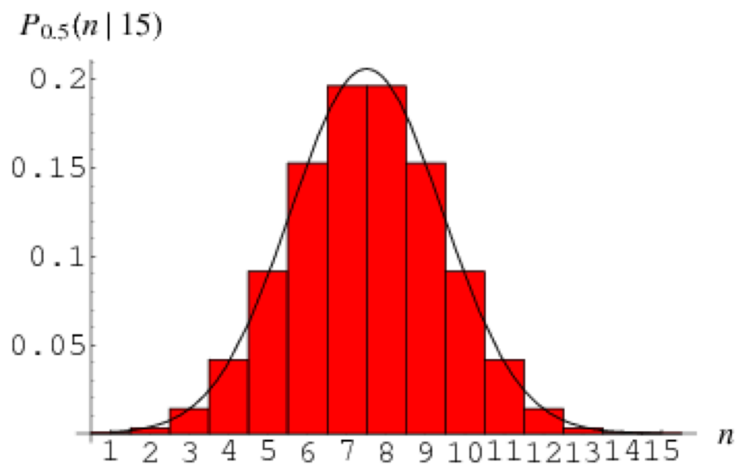
The discrete uniform distribution assigns the probability $1/n$ to each element of S



https://en.wikipedia.org/wiki/Discrete_uniform_distribution

Probability

The binomial distribution



http://mathworld.wolfram.com/NormalDistribution.html

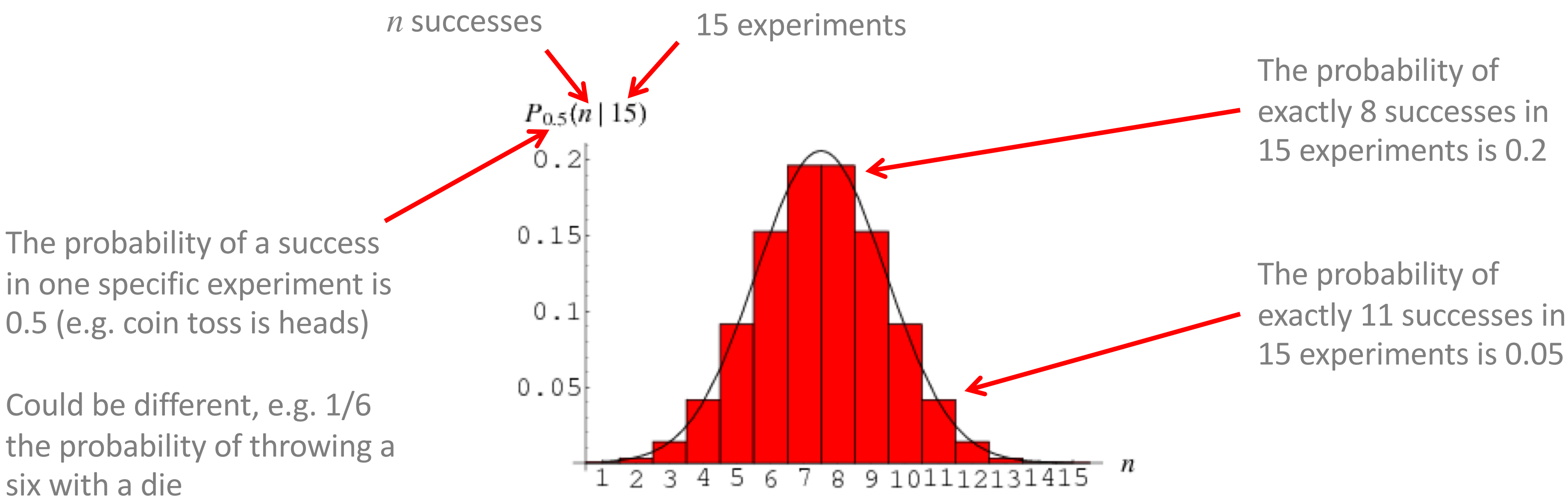
gives the probability of n successes in a sequence of N independent experiments
where the outcome of each experiment can have one of two values: either success or failure

hence binomial

The probability of a success in one specific experiment is p
The probability a failure is $q = 1 - p$
A single experiment is called a Bernoulli trial

Probability

The binomial distribution



http://mathworld.wolfram.com/NormalDistribution.html

gives the probability of n successes in a sequence of N independent experiments
where the outcome of each experiment can have one of two values: either success or failure

hence binomial

The probability of a success in one specific experiment is p
The probability a failure is $q = 1 - p$
A single experiment is called a Bernoulli trial

Probability

The parameters for this distribution are

N the number of trials (experiments)
p the probability success in one trial (experiment)

giving the probability of getting exactly n successes in N trials

Probability

For example, roll a die 10 times, what is the probability of throwing a four 0 times, 1 time, 2 times, 3 times, 4 times, ..., 10 times

N the number of trials (experiments) = 10

p the probability success in one trial (experiment) = $1/6$

giving the probability of getting exactly n (0, 1, 2, ... 15) successes in 15 trials

Probability

Confusingly, these parameters are often written another way:

n the number of trials (experiments)

p the probability success in one trial (experiment)

giving the probability of getting exactly k successes in n trials

$$\Pr(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

This is the probability mass function for the binomial distribution

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Sometimes referred to as the binomial coefficient

This is the number of combinations of k elements drawn from n items
Sometimes written ${}^{n}C_{k}$

E.g. from a pool of 5 people, how many committees can be formed with 3 people

Probability

For example, roll a die 10 times, what is the probability of throwing a four
0 times, 1 time, 2 times, 3 times, 4 times, ..., 10 times

N the number of trials (experiments) = 10

p the probability success in one trial (experiment) = 1/6

giving the probability of getting exactly n (0, 1, 2, ... 15) successes in 15 trials

We can use the probability mass function formula to work out

- The whole (discrete) probability distribution, or
- Just one case, e.g. the probability of throwing a four just twice ($k = 2$) in ten ($n = 10$) throws (trials)

Probability

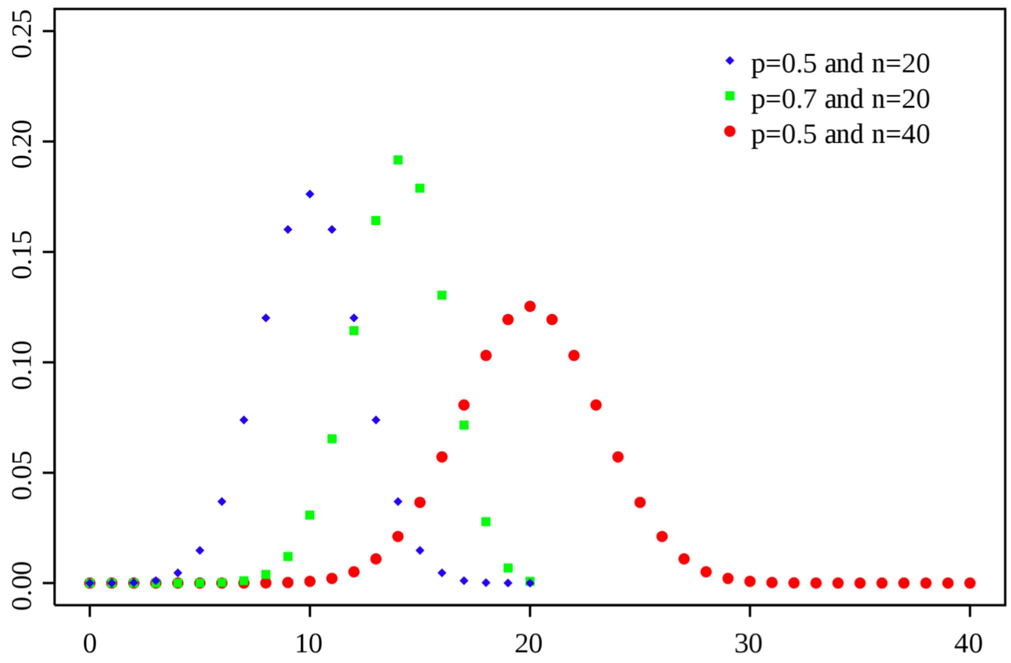
Given that 70% of people who purchase bicycles are men, and if 100 bike owners are randomly selected, what is the probability that exactly 7 are men?

Exercise:

Use the binomial probability mass function to compute this probability

Probability

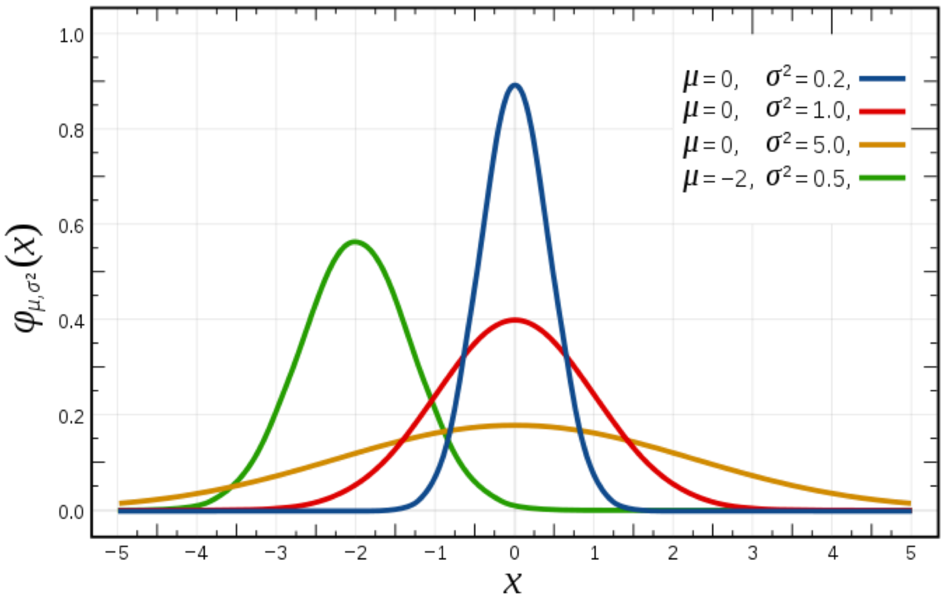
Different parameters, i.e. p and n, give different distributions



https://en.wikipedia.org/wiki/Binomial_distribution

Probability

The normal distribution is the limiting case of a discrete binomial distribution as the sample size N becomes large



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

This is the probability density function for the normal (Gaussian) distribution (its a continuous probability distribution)

https://en.wikipedia.org/wiki/Normal_distribution

Probability

The central limit theorem

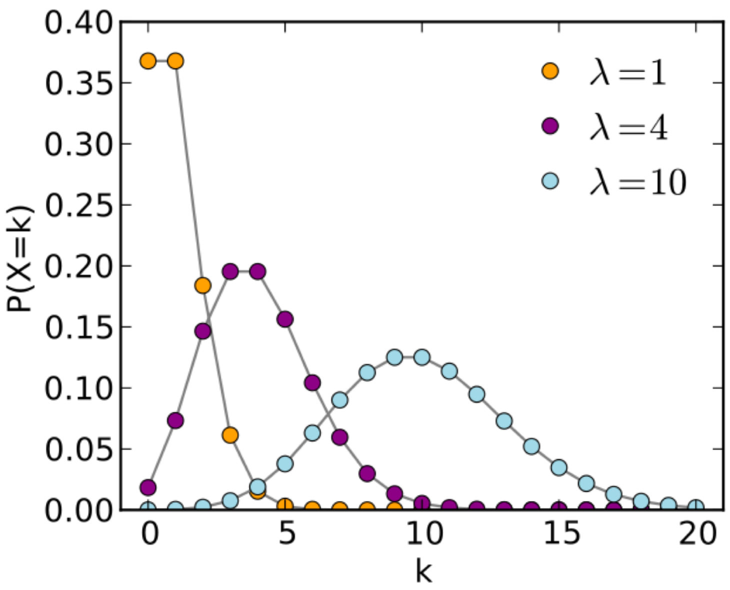
Observations (or processes) which are the result of a sum of a large number of random and independent influences have a distribution function closely approximated by that of a Gaussian random variable

This theorem applies to many natural observations: height, weight, voltage fluctuations, IQ...

https://en.wikipedia.org/wiki/Normal_distribution

Probability

If events occur at random in time then the number of events in a fixed interval of time has a Poisson distribution.



https://en.wikipedia.org/wiki/Poisson_distribution

Probability

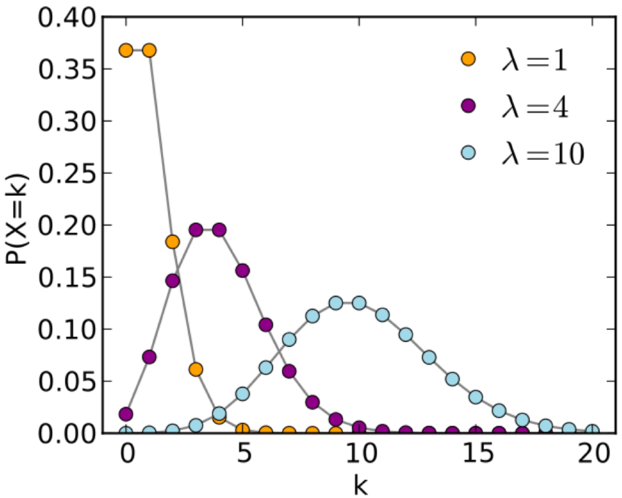
The Poisson distribution $P(k \text{ events in interval}) = e^{-\lambda} \frac{\lambda^k}{k!}$

λ is the average number of events per interval

e is the number 2.71828... (Euler's number) the base of the natural logarithms

k takes values 0, 1, 2, …

$k! = k \times (k-1) \times (k-2) \times \ldots \times 2 \times 1$ is the factorial of k.



https://en.wikipedia.org/wiki/Poisson_distribution

Probability

Conditions for a Poisson distribution

- An event can occur any number of times during a time period
- Events occur independently
- The rate of occurrence is constant
- The probability of an event occurring in proportional to the length of the time period

Probability

Definition of the probability of an event

$$p(E) = \sum_{s \in E} p(s)$$

The probability of an event E is the sum of the probabilities of the outcomes in E

Probability

Conditional Probability

- Let E and F be events with $p(F) > 0$

  The conditional probability of E given F, denoted $p(E \mid F)$, is defined as

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

← Joint probability $p(E, F)$

Probability

Conditional Probability

- What is the conditional probability that a family with two children has two boys, given that they have at least one boy?

  Assume that each of the possibilities BB, BG, GB, GG is equally likely.

  Let E be the event that the family with two children as two boys

  Let F be the event that a family with two children has at least one boy

  $E = \{BB\}$

  $F = \{BB, BG, GB\}$

  $E \cap F = \{BB\}$

  $p(F) = 3/4$ and $p(E \cap F) = 1/4$

  $$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Probability

Independence

- If $p(E \mid F) = p(E)$ it means F has no bearing on E
- We say E and F are independent events

Definition: the events E and F are independent if and only if

$$p(E \cap F) = p(E)\, p(F)$$

- Why?
  Because $p(E \mid F) = \dfrac{p(E \cap F)}{p(F)} = p(E)$

Probability

Independence

Let E be the event that the family with two children has two boys

Let F be the event that a family with two children has at least one boy

Are the two events independent?

$E = \{BB\}$ so $p(E) = 1/4$

$F = \{BB, BG, GB\}$ so $p(F) = 3/4$

Thus, $p(E)\ P(F) = 3/16$

$p(E \cap F) = 1/4$

Since $p(E \cap F) \neq p(E)\ P(F)$ the events are not independent

Probability

Random Variables

– Many problems are concerned with a numerical value associated with the outcome of an experiment

- E.g. the number of 1 bits in a randomly generated string of 10 bits
- E.g. the number of times a head comes up when you toss a coin 20 times
- E.g. some feature (e.g. weight) of a manufactured part

A **random variable** is a **function** from the sample space of an experiment to the real numbers

$$f: S \to \Re$$

Probability

Random Variables

- A **random variable** is a **function** from the sample space of an experiment to the real numbers
  - A random variable assigns a real number to each possible outcome
  - The input to a random variable is an elementary outcome, and the output is a number
  - We can think of random variable as numerical measurements of outcomes
  - A random variable is a function
    - It is not a variable
    - It is not random!

Probability

DEFINITION 5 A *random variable* is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

Remark: Note that a random variable is a function. It is not a variable, and it is not random!

K. H. Rosen, *Discrete Mathematics and Its Applications*, 2012, p. 408.

Probability

Random Variables

– For example, toss a coin three times

Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome

What are the outcomes? $HHH, HHT, HTH, THH, TTH, THT, HTT, TTT$

$X(HHH) = 3$

$X(HHT) = 2$

$X(HTH) = 2$

$X(THH) = 2$

$X(TTH) = 1$

$X(THT) = 1$

$X(HTT) = 1$

$X(TTT) = 0$

Probability

Random Variables

The distribution of a random variable X on a sample space S is the set of pairs

$(r, p(X = r))$

For all $r \in X(S)$, where $p(X = r)$ is the probability that X takes the value r

For the previous example,

$p(X = 3) = 1/8$

$p(X = 2) = 3/8$

$p(X = 1) = 3/8$

$p(X = 0) = 1/8$

Hence, the **distribution** of $X(t)$ is the set of pairs

$(3, 1/8), (2, 3/8), (1, 3/8), (0, 1/8)$

Probability

Bayes' Theorem / Rule

- Shows how to revise probability of events in the light of new data

- For example, we can determine the probability that a particular incoming email is spam using the occurrence of words in the message

- To do this we need to know
  - The percentage of incoming emails that are spam
  - The percentage of **spam** messages in which these words occur
  - The percentage of messages that are **not spam** in which each of these words occur

Probability

Bayes' Theorem / Rule

Let E be an event from a sample space S

$F_1, F_2, \ldots F_n$, are mutually exclusive events such that

$F_1 \cup F_2, \ldots \cup F_n = S$

$p(E) \neq 0$ and $p(F_i) \neq 0$

Likelihoods

Prior probability

Posterior probability

$$p(F_j \mid E) = \frac{p(E \mid F_j)p(F_j)}{\sum_{i=1}^{n} p(E \mid F_i)p(F_i)}$$

Probability

Sometimes we write $p(F_j \mid E)$

as $p(H_j \mid D)$

to be read as "The probability that hypothesis H_j is true given data D"

which is computed from the prior probabilities that each hypothesis is true

$$p(H_\mathrm{j})$$

and the conditional probabilities (likelihoods) of that data occurring in the case of each hypothesis

$$p(D \mid H_j)$$

Probability

A friend of yours believes she has a 50% chance of being pregnant.

She decides to take a pregnancy test and the test is positive.

You read in the test instructions that out of 100 non-pregnant women, 20% give false positives (the result of the test is positive when it should be negative).

Moreover, out of 100 pregnant women 10% give false negatives (the result is negative when it should be positive).

Help your friend upgrade her beliefs, i.e. calculate the probability that she is pregnant, given that the test is positive.

Probability

See also:

K. H. Rosen, *Discrete Mathematics and Its Applications*, 2012.

J. R. Movellan, *Introduction to Probability Theory and Statistics*, 2008.

Classification

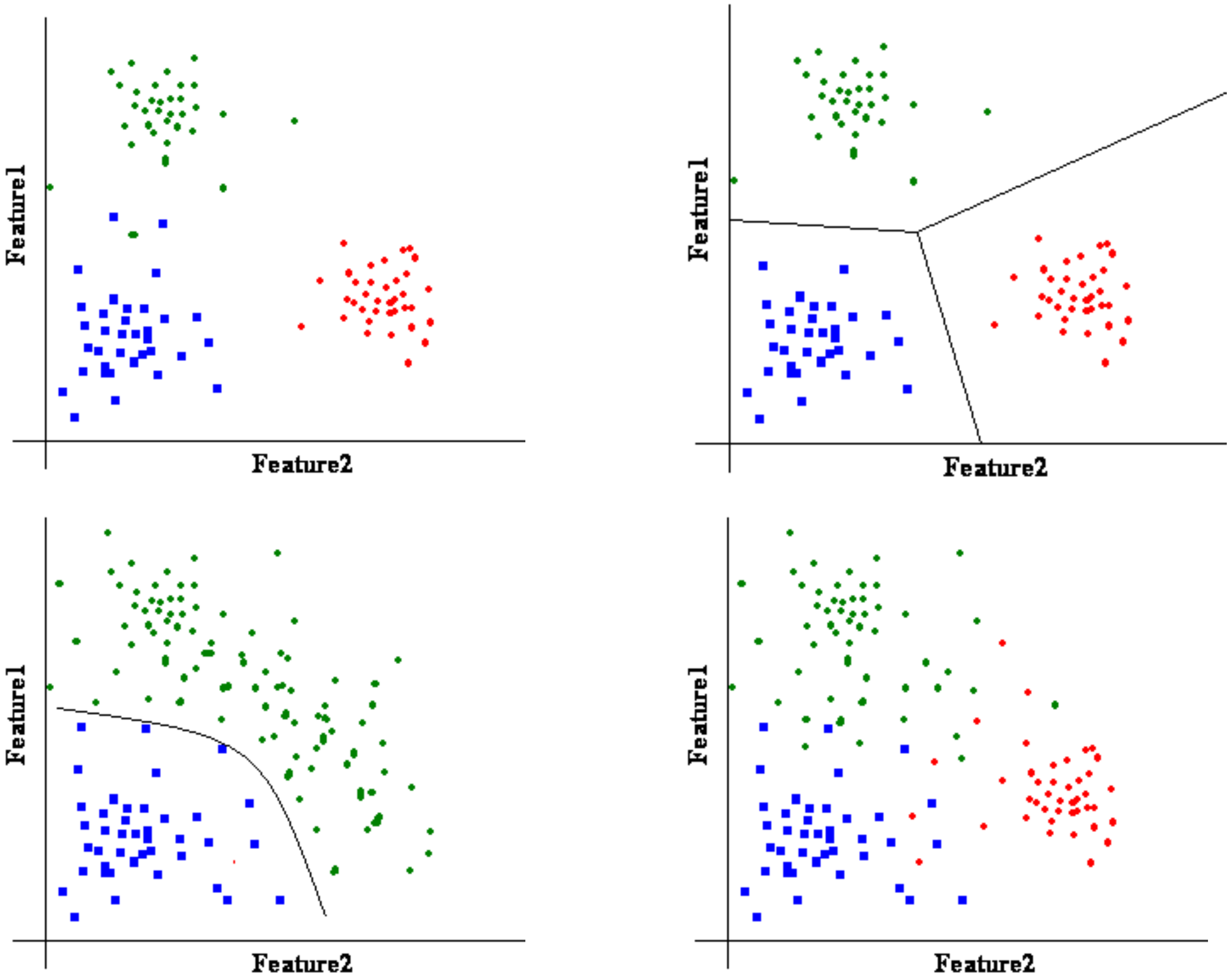
Maximum Likelihood Classifier / Naive Bayes Classifier

Classification

- Object recognition
  - R classes: $w_1, w_2, \ldots w_R$
- Classifier
  - n features: input pattern / feature vector: $x_1, x_2, \ldots x_\mathrm{n}$
- Feature space
  - Choosing the features
  - Clusters in feature space
- Separability
  - Linear separability
  - Hyper-surfaces
  - Inseparable classes
- Classifiers
  - Nearest Neighbour Classifier / Minimum Distance Classifier
  - Linear Classifier
  - (Naive) Maximum Likelihood Classifier / Naive Bayes Classifier
  - ...

Credit: Kenneth Dawson-Howe, A Practical Introduction to Computer Vision with OpenCV, © Wiley & Sons Inc. 2014

Classification

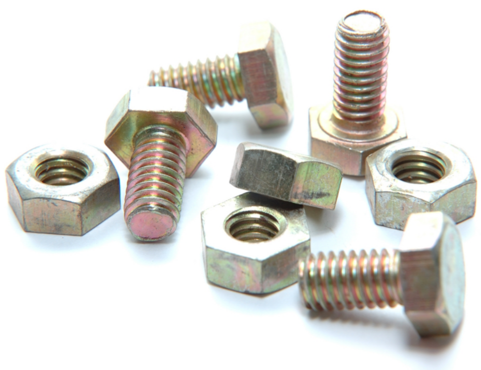


Credit: Kenneth Dawson-Howe, A Practical Introduction to Computer Vision with OpenCV, © Wiley & Sons Inc. 2014

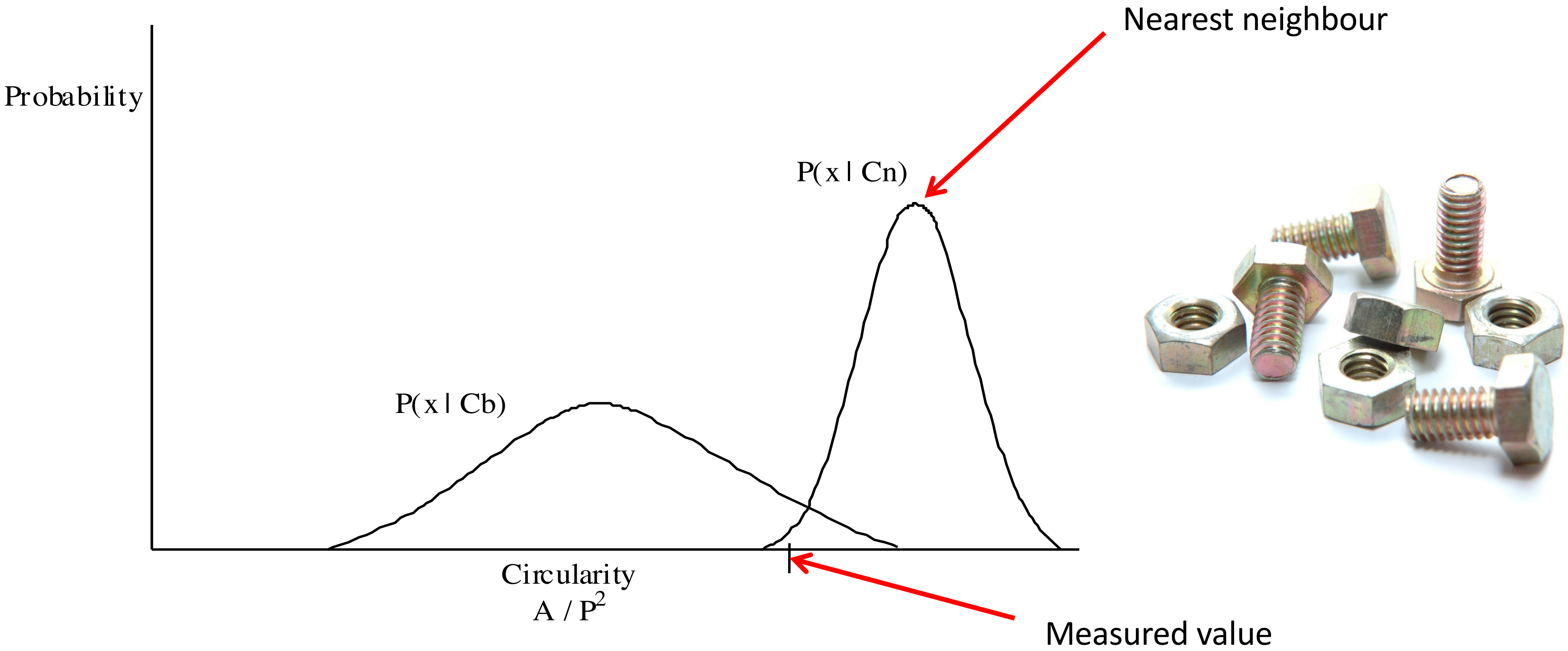
Classification

Example: Maximum Likelihood Classifier

- Let's design a system that can classify two parts: nuts and bolts
- Two classes $C_b\, C_n$
- Let's decide to use a feature 'circularity' x to distinguish nuts from bolts (nuts are more circular than bolts)

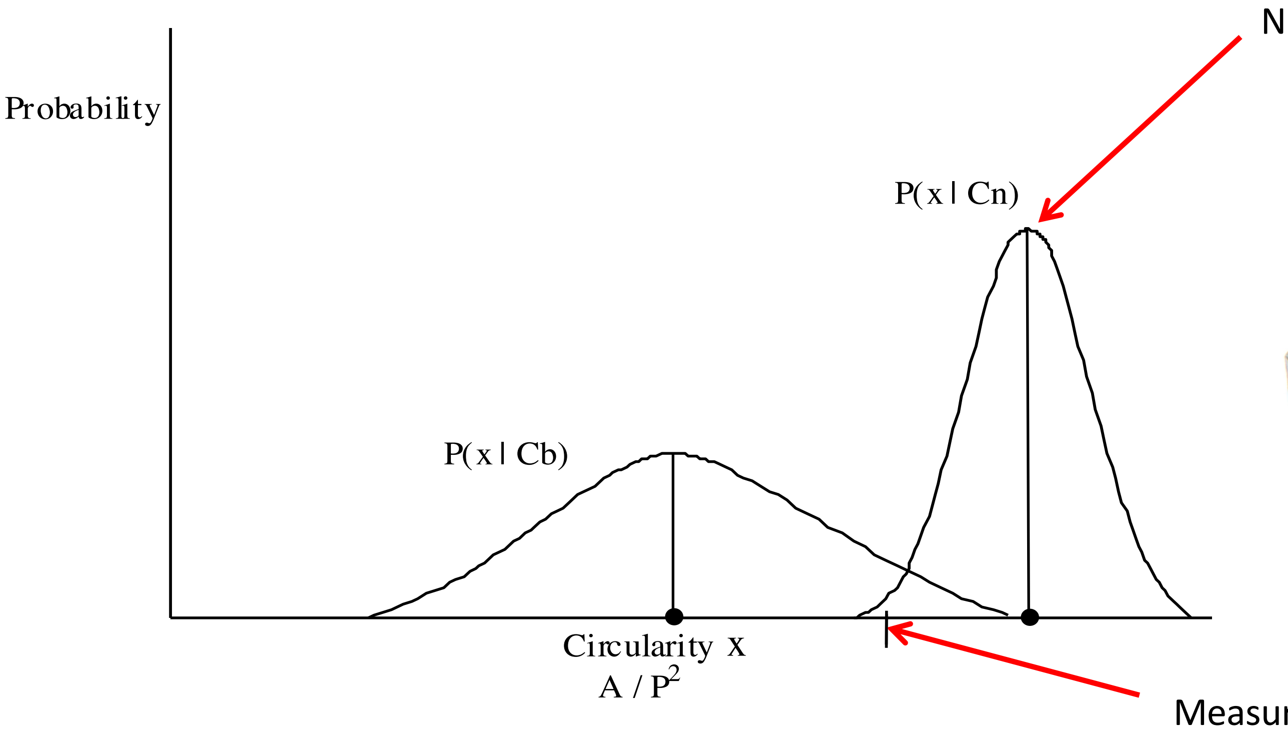


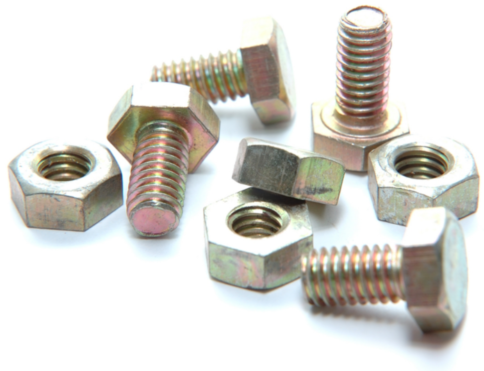
Classification



We can do better than nearest neighbour if we use knowledge about the probability of an object having some feature value $P(x|C_i)$ and the probability of that object being there at all $P(C_i)$

Classification





We can do better than nearest neighbour if we use knowledge about
the probability of an object having some feature value $P(x|C_i)$
and the probability of that object being there at all $P(C_i)$

Classification

Example: Maximum Likelihood Classifier

- Neither of these probabilities are what we are interested in!
- We want the probability that an object belongs to a particular class, *given that a particular value of x has occurred* $P(C_i|x)$
- Thus, we classify the object as a bolt if

$$P(C_b|x) > P(C_n|x)$$

- We use Bayes' Theorem to convert the probabilities we know (or can measure) to the ones we need

Classification

Example: Maximum Likelihood Classifier

- The *posterior* probability, $P(C_i|x)$, that the object belongs to a particular class i is given by Bayes' Theorem:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}$$

where

$$P(x) = \sum_{i=1}^{2} P(x|C_i)P(C_i)$$

Classification

Example: Maximum Likelihood Classifier

- The first thing that is required is the probabilities for each of these two classes, *i.e.*, a measure of the probabilities that an object from a particular class will have a given feature value

- Since it is not likely that we will know these *a priori*, we will have to estimate them

Classification

Example: Maximum Likelihood Classifier

- Let S be the space of circularity values that we can measure with our computer vision system

- Let $X_n(x)$ be the random variable that equals the number of times a given circularity value appears when x is the outcome (i.e. when the circularity of a nut is measured)

- Let $X_b(x)$ be the random variable that equals the number of times a given circularity value appears when x is the outcome (i.e. when the circularity of a bolt is measured)

- We need the probability distribution of random variables X_n and X_b

Classification

Example: Maximum Likelihood Classifier

- We have used discrete random variables here
  - The distribution is called a **Probability Mass Function**
- However, since the circularity value is going to vary continuously (i.e. it won't have a finite set of values), we should really use a continuous random variable
  - The distribution is call a **Probability Density Function (PDF)**

Classification

Example: Maximum Likelihood Classifier

- The PDF for nuts can be estimated in a relatively simple manner
  - measuring the value of x for a large number of nuts
  - plotting the histogram of these values
  - smoothing the histogram
  - normalising the values so that the total area under the histogram equals 1
    - The normalisation step is necessary because certainty has a probability value of 1 and the sum of all the probabilities (for all the possible circularity measures) must necessarily be equal to a certainty of having that object, *i.e.*, a probability value of 1

Classification

Example: Maximum Likelihood Classifier

- The PDF for the bolts can be estimated in a similar manner.
- Next problem: the probability of each class occurring
  - We may know, for instance, that the class of nuts is, in general, likely to occur twice as often as the class of bolts
  - In this case we say that the prior (or *a priori*) probability of the two classes are :

$P(C_n) = 0.666$ and $P(C_b) = 0.333$

  - In fact, in this case, it is more likely that they will have the same *a priori* probabilities (0.5) since we usually have a nut for each bolt

Classification

Example: Maximum Likelihood Classifier

- The PDFs tell us the probability that the circularity x will occur, given that the object belongs to the class of nuts C_n in the first instance and to the class of bolts C_b in the second instance

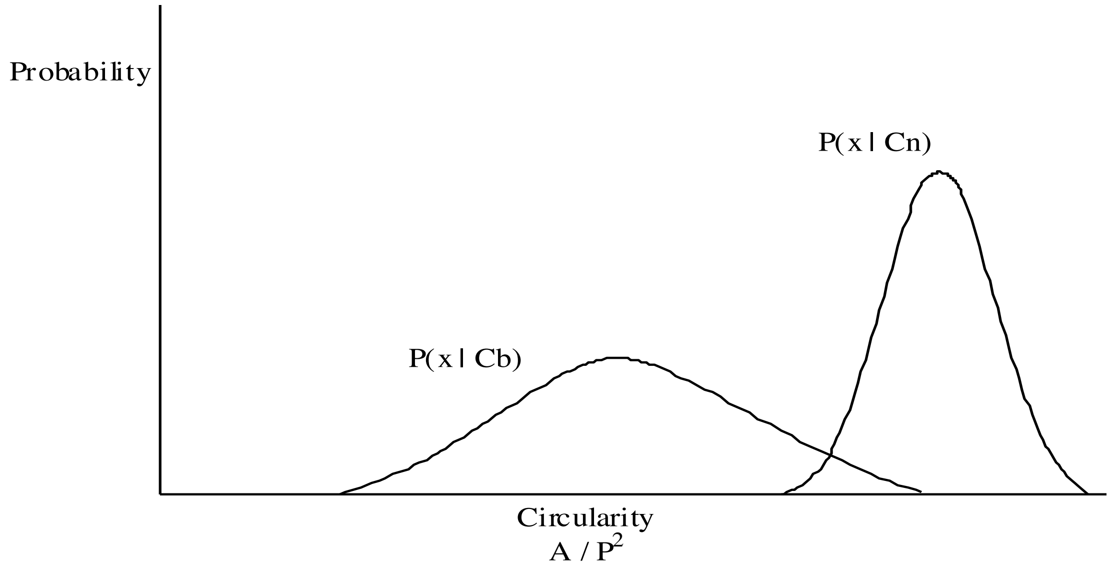
- As we know, this is termed the conditional probability of an object having a certain feature value, given that we know that it belongs to a particular class

Classification

Example: Maximum Likelihood Classifier

- Thus, the conditional probability, $p(x\,|\,C_b)$ enumerates the probability that a circularity x will occur, given that the object is a bolt.

- The two conditional probabilities $p(x\,|\,C_b)$ and $p(x\,|\,C_n)$ are shown below



Classification

Example: Maximum Likelihood Classifier

- This is not what are interested in …
- We want the probability that an object belongs to a particular class, given that a particular value of x has occurred (*i.e.* been measured), allowing us to establish its identity

Classification

Example: Maximum Likelihood Classifier

- This is called the posterior (or *a posteriori*) probability, $p(C_i | x)$ that the object belongs to a particular class C_i, given that a particular value of x has occurred
- It is given by Bayes' Theorem

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}$$

where

$$P(x) = \sum_{i=1}^{2} P(x|C_i)P(C_i)$$

Classification

Example: Maximum Likelihood Classifier

– $p(x)$ is a normalisation factor which is used to ensure that the sum of the *a posteriori* probabilities sum to one, for the same reasons as mentioned earlier

Classification

Example: Maximum Likelihood Classifier

– In effect, Bayes' theorem allows us to use

  - the *a priori* probability of objects occurring in the first place

  - the conditional probability of an object having a particular feature value given that it belongs to a particular class and …

  - The actual measurement of a feature value (to be used as the parameter in the conditional probability) to estimate the probability that the measured object belongs to a given class

  - Once we can estimate the probability that, for a given measurement, the object is a nut and the probability that it is a bolt, we can make a decision as to its identity, choosing the class with the higher probability

Classification

Example: Maximum Likelihood Classifier

- This is why it is called the maximum likelihood classifier
- Thus, we classify the object as a bolt if :

$$P(C_b|x) > P(C_n|x)$$

Classification

Example: Maximum Likelihood Classifier

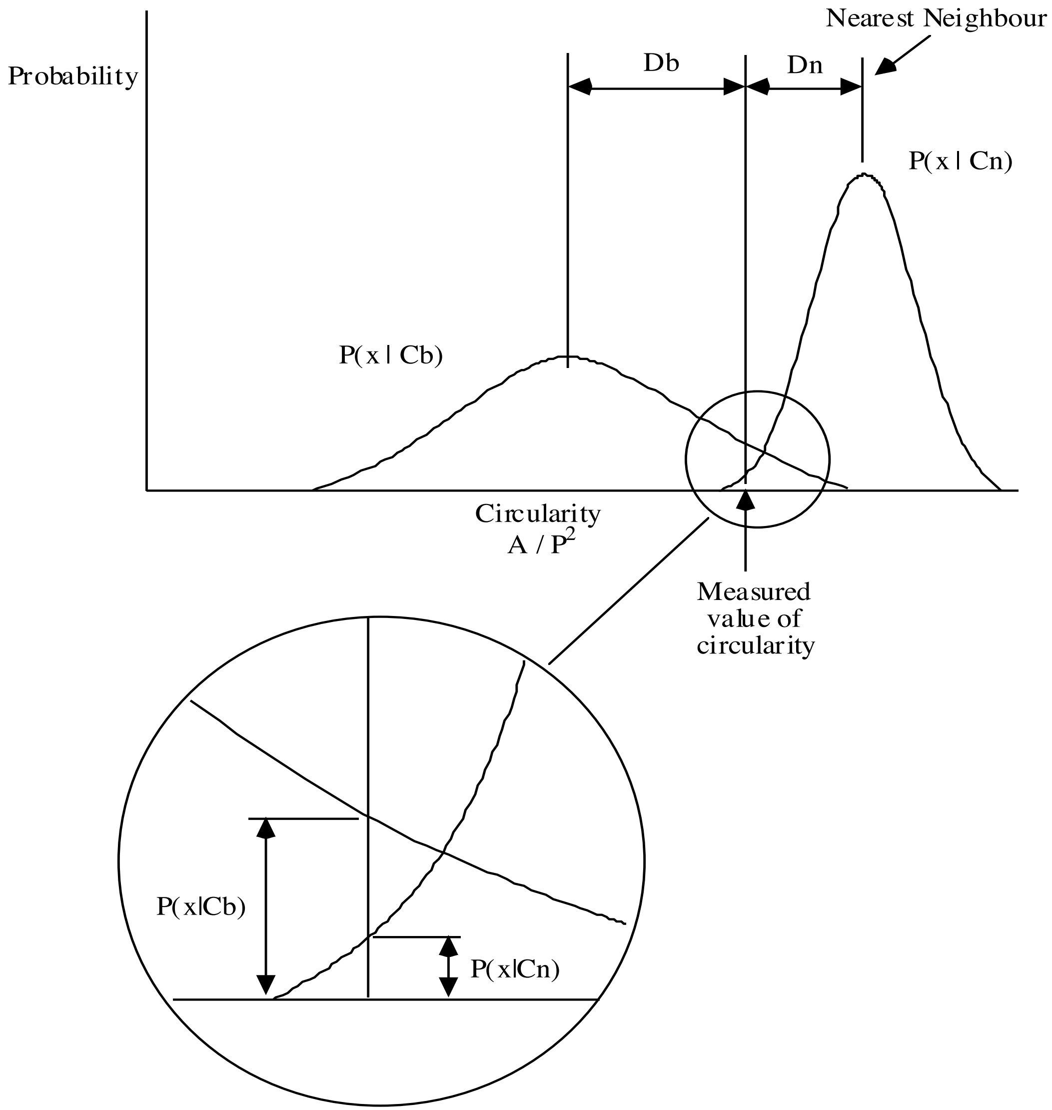
- Using Bayes' Theorem again, and noting that the normalising factor $p(x)$ is the same for both expressions, we can rewrite this test as

$$P(x|C_b)P(C_b) > P(x|C_n)P(C_n)$$

- If we assume that the chances of an unknown object being either a nut or a bolt are equally likely (*i.e.* $P(C_b) = P(C_n)$), then we classify the unknown object as a bolt if :

$$P(x|C_b) > P(x|C_n)$$

Classification



Classification

Example: Maximum Likelihood Classifier

- For the example shown $p(x \mid C_b)$ is indeed greater than $p(x \mid C_n)$ for the measured value of circularity and we classify the object as a bolt

- If, on the other hand, we were to use the nearest neighbour classification technique, we would choose the class whose mean value "is closer to" the measured value

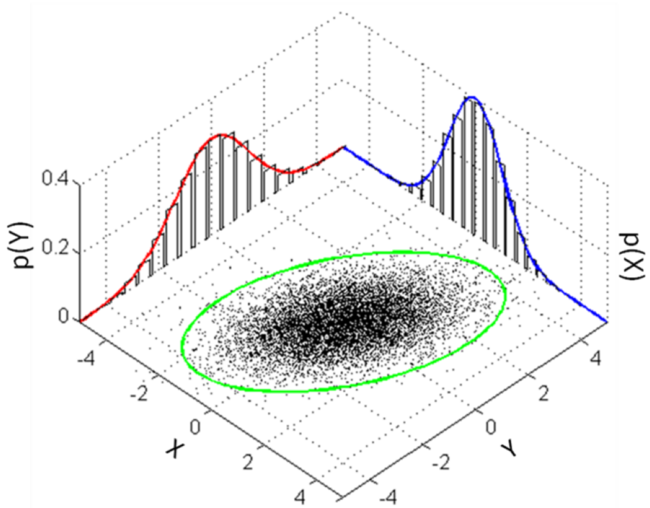
  - In this case, the distance D_n from the measured value to the mean of the PDF for nuts is less than D_b, the distance from the measured value to the mean of the PDF for bolts; we would erroneously classify the object as a nut

Classification

Example: Maximum Likelihood Classifier

- This was a simple example with just one feature and a 1-D PDF

- However, the argument generalizes directly to n-dimensions, where we have n features in which case the conditional probability density functions are also n-dimensional



Classification

Example: Maximum Likelihood Classifier

- If we assume that the features are independent then we can use the theory we've just outlined, multiplying together the conditional probabilities for each class
  - This is known a **Naïve Bayes Classifier**
  - It may be naïve, but it works surprisingly well
- If we don't assume independence, then we need a more complex theory

Queueing

Discrete Event Simulation using the Poisson Distribution

Three Questions

1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

    a) Two cars will arrive (in an interval of one minute)
    b) Three or more cars will arrive

2. In any given period of time, how many cars actually arrive?

3. Which question – 1 or 2 – is the right question to ask if we want to simulate the behaviour of traffic queues?

Three Questions

1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   a) Two cars will arrive (in an interval of one minute)
   b) Three or more cars will arrive

2. In any given period of time, how many cars actually arrive?

3. Which question – 1 or 2 – is the right question to ask if we want to simulate the behaviour of traffic queues?

Three Questions

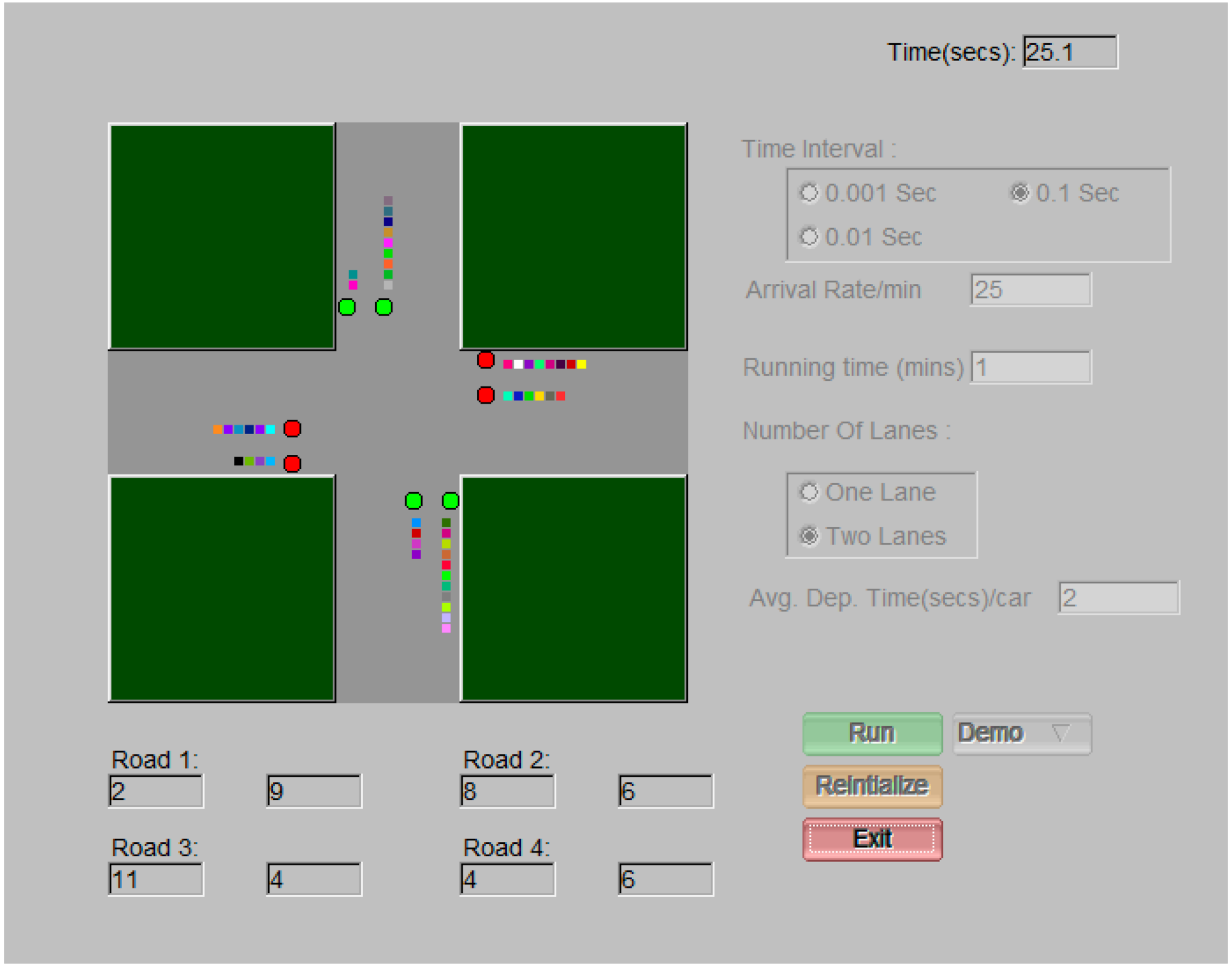
1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   a) Two cars will arrive (in an interval of one minute)
   b) Three or more cars will arrive

2. In any given period of time, how many cars actually arrive?

3. Which question – 1 or 2 – is the right question to ask if we want to simulate the behaviour of traffic queues?

Discrete Event Simulation of Traffic Queues



Three Questions

1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   a) Two cars will arrive (in an interval of one minute)
   b) Three or more cars will arrive

2. In any given period of time, how many cars actually arrive?

3. Which question – 1 or 2 – is the right question to ask if we want to simulate the behaviour of traffic queues?

Probability

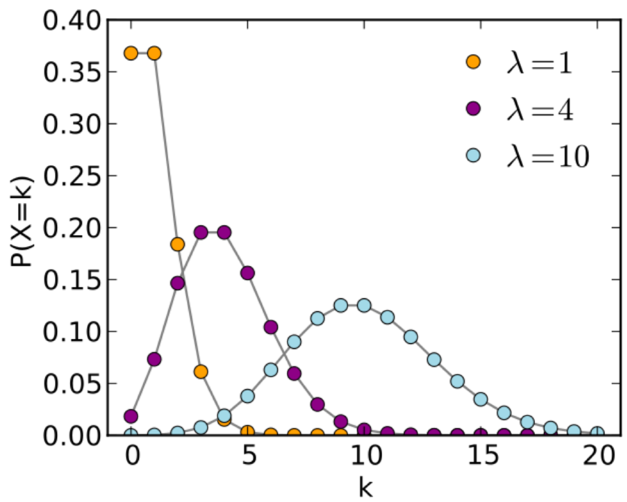
The Poisson distribution $$P(k \text{ events in interval}) = e^{-\lambda}\frac{\lambda^k}{k!}$$

λ is the average number of events per interval

e is the number 2.71828... (Euler's number) the base of the natural logarithms

k takes values 0, 1, 2, …

$k! = k \times (k-1) \times (k-2) \times \ldots \times 2 \times 1$ is the factorial of k.



https://en.wikipedia.org/wiki/Poisson_distribution

Three Questions

1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   a) Two cars will arrive (in an interval of one minute)

$$P(X = 2) = \frac{1.6^2 e^{-1.6}}{2!} \approx 0.258$$

https://brilliant.org/wiki/poisson-distribution/

Three Questions

1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

    b) Three or more cars will arrive

https://brilliant.org/wiki/poisson-distribution/

Three Questions

1. Given that the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   b) Three or more cars will arrive

      The problem is to find $P(X \geq 3)$
      but there is no upper limit on k so we can't compute it directly

      Instead, we compute $P(X \leq 2)$
      and compute $P(X \geq 3)$ as 1 - $P(X \leq 2)$

https://brilliant.org/wiki/poisson-distribution/

Three Questions

1. Given the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   b) Three or more cars will arrive

$$P(X=0) = \frac{1.6^0 e^{-1.6}}{0!} \approx 0.202$$

$$P(X=1) = \frac{1.6^1 e^{-1.6}}{1!} \approx 0.323$$

$$P(X=2) = \frac{1.6^2 e^{-1.6}}{2!} \approx 0.258$$

$$\Rightarrow P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) \approx 0.783$$

Remember
$p(E_1 \cup E_2 \ldots \cup E_n) = p(E_1) + p(E_2) + \ldots p(E_n)$

$$\Rightarrow P(X \geq 3) = 1 - P(X \leq 2) \approx 0.217$$

Remember
$p(E) = 1 - p(E)$

Three Questions

2. In any given period of time, how many cars actually arrive?

Three Questions

2. In any given period of time, how many cars actually arrive?

   To answer this question, we need to sample the Poisson probability distribution

   – i.e. generate the value of a Poisson-distributed random variable

   – i.e. we need to find the value of X, a random variable, i.e. the value of x (or k), given the Poisson parameter λ

Three Questions

General idea

- The time between arrivals in a Poisson process has an exponential distribution

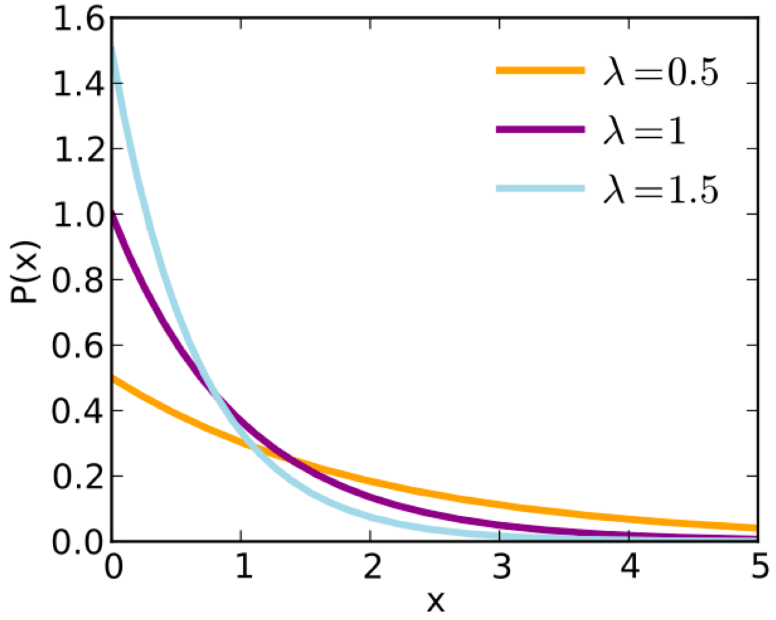
- Determine how many exponential events (i.e. random samples of the exponential distribution) just fit into a unit time interval

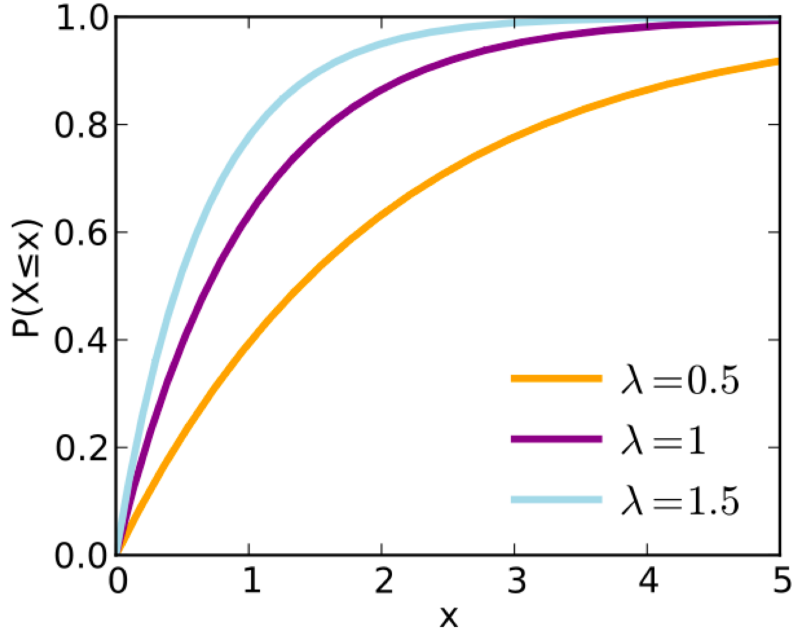
  - iteratively calculate the times between arrivals (sample the exponential distribution)

  - adding these times up until the time sum exceeds the interval

- That number will have a Poisson distribution: its value will be the required (Poisson) random sample, i.e. the number of cars that arrived

Three Questions





https://en.wikipedia.org/wiki/Exponential_distribution

On the Simulation of Random Events
or
"How to calculate the number of cars arriving at an intersection in any given period of time"

David Vernon
Carnegie Mellon University Africa

July 29, 2018

Abstract

This short note describes a software function that will create input data for a discrete event simulation of a system where input varies with time. In particular, it shows how to sample a Poisson probability distribution to compute the number of events that occur in a given period of time. This assumes that events happen randomly. Such situations arise very often and in this note we focus on the number of cars that arrive in a traffic queue in every finite period of time. However, the same theory and sampling technique applies to other applications such as packet traffic on a communications network.

```c
#include <stdlib.h>
#include <stdio.h>
#include <time.h>
#include <math.h>

int samplePoisson(double lambda) {

   /* Generate a random sample from a Poisson distribution with a given mean, lambda */
   /* Use the function rand to generate a random number                              */

   static bool first_call = true;
   int count;
   double product;
   double zero_probability;

   /* Seed the random-number generator with current time so */
   /* that the numbers will be different every time we run  */

   if (first_call) {
      srand( (unsigned)time( NULL ));
      first_call = false;
   }

   count = 0;
   product = (double) rand() / (double) RAND_MAX;

   zero_probability = exp(-lambda);

   while (product > zero_probability) {
      count++;
      product = product * ((double) rand() / (double) RAND_MAX);
   }
   return(count);
}
```

```
/* example usage of samplePoisson */

int      arrival_rate_input;  // the arrival rate in cars per minute

double   arrival_rate;        // the arrival rate in cars per millisecond

long int increment;           // the period of each simulation interval in milliseconds

int      count;               // the number of cars that arrive in any given
                              // simulation interval (i.e time increment)

double   lambda;              // the mean number of cars that arrive in any one
                              // similation interval (i.e. time increment)

/*  arrival_rate_input is in cars per second so convert to cars per millisecond   */

arrival_rate = ((float) arrival_rate_input) / (60 * 1000);


/* the Poisson distribution mean, lambda, is the arrival rate of cars during   */
/* the simulation interval, i.e. arrival rate per millisecond multiplied by    */
/* the number of milliseconds in each simulation interval                      */

lambda = arrival_rate * increment;


/* Compute the number of cars that have arrived in the current simulation interval */

count = samplePoisson(lambda);
```

Caveat

- The computational complexity of this approach is proportional to the the value of the Poisson parameter λ
  - Inefficient for large value of λ
- For more efficient approaches see

  A. C. Atkinson, "The Computer Generation of Poisson Random Variables", Journal of the Royal Statistical Society Series C (Applied Statistics), Vol. 28, No. 1, pp. 29–35, 1979.

Applications

Examples of phenomena that obey a Poisson distribution

- the number of mutations on a given strand of DNA per time unit
- the number of bankruptcies that are filed in a month
- the number of arrivals at a car wash in one hour
- the number of network failures per day
- the number of file server virus infection at a data center during a 24-hour period
- the number of Airbus 330 aircraft engine shutdowns per 100,000 flight hours
- the number of asthma patient arrivals in a given hour at a walk-in clinic
- the number of hungry persons entering McDonald's restaurant per day
- the number of work-related accidents over a given production time
- the number of birth, deaths, marriages, divorces, suicides, and homicides over a given period of time
- the number of customers who call to complain about a service problem per month
- the number of visitors to a web site per minute
- the number of calls to consumer hot line in a 5-minute period
- the number of telephone calls per minute in a small business
- the number of arrivals at a turnpike tollbooth per minute between 3 A.M. and 4 A.M. in January on the Kansas Turnpike.

Source: J. Letkowski, *Applications of the Poisson probability distribution*. Retrieved July 28, 2018 from http://www.aabri.com/SA12Manuscripts/SA12083.pdf.

Three Questions

1. Given the average rate at with cars arrive at a traffic intersection is 1.6 cars/minute, what is the probability that

   a) Two cars will arrive (in an interval of one minute)
   b) Three or more cars will arrive

2. In any given period of time, how many cars actually arrive?

3. Which question – 1 or 2 – is the right question to ask if we want to simulate the behaviour of traffic queues?