Modeling Uncertainty - 1

Decision Supports Systems 2017/18, Lecture 04

Marko Tkalčič

Alpen-Adria-Universität Klagenfurt

Table of Contents

Modeling Uncertainty

- uncertainty is a critical element of many decisions

Modeling Uncertainty

- uncertainty is a critical element of many decisions
- how do we model uncertainty?
  - probability

Modeling Uncertainty

- uncertainty is a critical element of many decisions
- how do we model uncertainty?
  - probability
- sometimes, the problem at hand is similar to some **prototypical situations**
  - we will look at some such standard models

Table of Contents

Basics of Probability

- a central principle in decision making is that we can **represent uncertainty through the appropriate use of probability**
- many uncertain events are quantitative (e.g. tomorrow's max temperature)
- if not quantitative, we can introduce a quantitative variable:
  - $X = 1$ if it rains
  - $X = 0$ if it does not rain

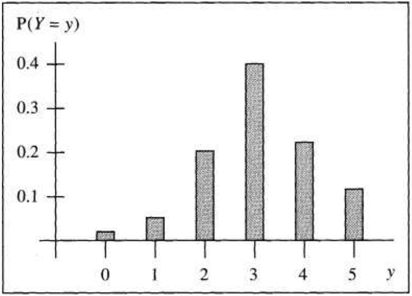
Basics of Probability

- a central principle in decision making is that we can **represent uncertainty through the appropriate use of probability**
- many uncertain events are quantitative (e.g. tomorrow's max temperature)
- if not quantitative, we can introduce a quantitative variable:
  - $X = 1$ if it rains
  - $X = 0$ if it does not rain

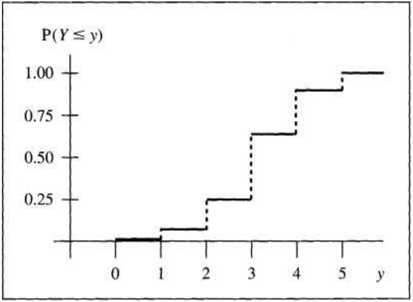
- the set of probabilities associated with all possible outcomes = **probability distribution**
- example:
  - we denote the number of raisins in an oatmeal cookie as Y
  - $P(Y = 0) = 0.02$
  - $P(Y = 1) = 0.05$
  - $P(Y = 2) = 0.20$
  - etc
- all probabilities ina probability distribution sum up to 1
- uncertain quantities (e.g. number of raisins Y) are called **random variables**

Discrete Probability Distributions

- the uncertain quantity can assume a **finite** or countable number of possible values
- example:
  - raisins in oatmeal cookie
  - precipitation (yes/no)
- we describe it with
  - probability mass function (PMF)
  - cumulative distribution function (CDF)



the probability that a discrete random variable Y is exactly equal to some value y



the probability that Y will take a value less than or equal to some value y

Expected value

- the **expected value** of a discrete random variable X is its probability-weighted average
- also average, mean, μ

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x)$$

- best guess

Variance and Standard Deviation

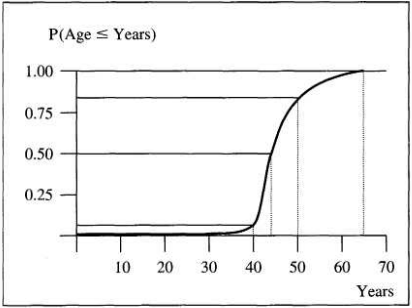
- variance = sum of squares of deviations from mean
- also $Var(X)$ or σ_X^2

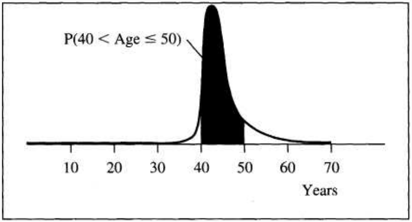
$$Var(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 \cdot P(X = x)$$

- standard deviation = square root of variance
- also σ_X
- best guess of how far the outcome might lie from $E(X)$
- a large variance or standard deviation indicates that the outcome is highly variable and hard to predict

Continuous Probability Distributions

- the uncertain quantity (represented by the random variable X) can take a value within a range (as opposed to discrete distributions)
- example
  - tomorrow max temperature
- we typically speak about interval probabilities $P(a \leq Y \leq b)$
- we describe it with
  - probability density function (PDF)
  - cumulative distribution function (CDF)





Statistical Moments

- expected value

$$E(X) = \int_{x-}^{x+} x f(x) dx$$

- variance

$$Var(X) = \sigma_X^2 = \int_{x-}^{x+} [x - E(X)]^2 f(x) dx$$

Table of Contents

Theoretical Probability Models

- sometimes, the problem at hand is similar to some **prototypical situations**
  - theoretical probability models

Theoretical Probability Models

- sometimes, the problem at hand is similar to some **prototypical situations**
  - theoretical probability models

- when we look at the nature of the uncertain event ...
- ... we can find a similar existing protoypical distribution

Theoretical Probability Models

- sometimes, the problem at hand is similar to some **prototypical situations**
  - theoretical probability models

- when we look at the nature of the uncertain event ...
- ... we can find a similar existing protoypical distribution

- we make a subjective assessment on which model could fit the nature of our random variable

Theoretical Probability Models

- sometimes, the problem at hand is similar to some **prototypical situations**
    - theoretical probability models

- when we look at the nature of the uncertain event ...
- ... we can find a similar existing protoypical distribution

- we make a subjective assessment on which model could fit the nature of our random variable

- it is important to have a good overview of the main distributions and their natural applications

Distributions

- there is A LOT of distributions
- we will cover the following distributions
  - discrete
    - binomial distribution
    - Poisson distribution
  - continuous
    - exponential distribution
    - normal distribution
    - beta distributions
- for each distribution:
  - typical example
  - parameters
  - shape
  - no equations in this course

Binomial Distribution

- discrete distribution
- example
  - you are in a race for mayor of your hometown, and
  - you wanted to find out how you were doing with the voters.
  - you might take a sample and count the number of individuals who indicated a preference for you
  - in this situation, each voter interviewed can be either for you or not.

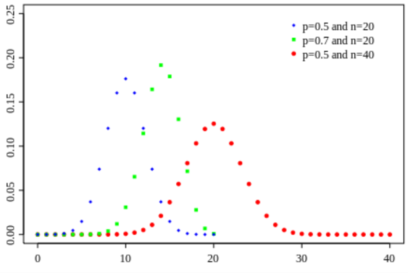
Binomial Distribution

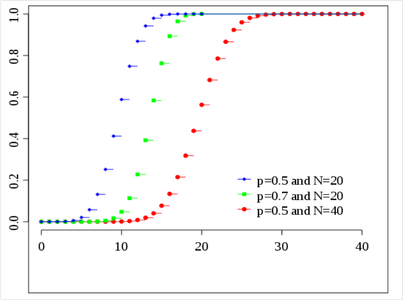
- discrete distribution
- example
  - you are in a race for mayor of your hometown, and
  - you wanted to find out how you were doing with the voters.
  - you might take a sample and count the number of individuals who indicated a preference for you
  - in this situation, each voter interviewed can be either for you or not.

- Binomial distribution is applicable in cases where:
  - outcomes are *dichotomous*: sequential, each can be only true/false
  - *constant probability* p at each event (trial) the probability of true is always the same
  - the outcome of each trial is *independent* of the previous ones

Binomial Distribution

- parameters of the binomial distribution $B(n, p)$ are
  - $n \in N_0$ — number of trials
  - $p \in [0, 1]$ — success probability in each trial
- when to use?
  - The binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N.
  - e.g. when having N voters, what is the chance that n will vote for you if the probability of voting is p?





Poisson distribution

- discrete distribution
- representing occurrences of a particular event over time or space
- example
  - you are interested in the number of customers who arrive at a bank in one hour.
  - this is an uncertain quantity; there could be none, one, two, three, etc.

Poisson distribution

- discrete distribution
- representing occurrences of a particular event over time or space
- example
  - you are interested in the number of customers who arrive at a bank in one hour.
  - this is an uncertain quantity; there could be none, one, two, three, etc.

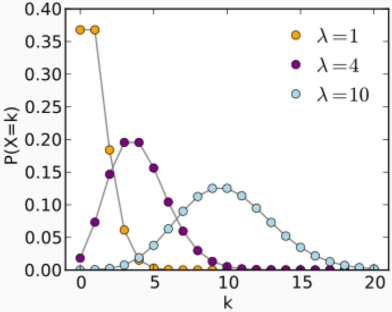
- The Poisson distribution is an appropriate model if the following assumptions are true.
  - k is the number of times an event occurs in an interval and k can take values 0, 1, 2, ...
  - The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
  - The rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals.
  - Two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.
  - The probability of an event in a small sub-interval is proportional to the length of the sub-interval.

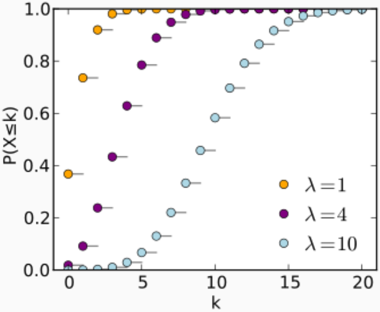
Poisson distribution

- parameters of the distribution $P(\lambda)$
  - $\lambda > 0$ real

Poisson distribution

- parameters of the distribution $P(\lambda)$
  - $\lambda > 0$ real
- when to use?
  - for modelling the number of times an event occurs in an interval of time or space.
  - e.g. what is the chance that k people will walk into the bank between 10am and 11am?





x axis: k . . . number of occurrences

Exponential distribution

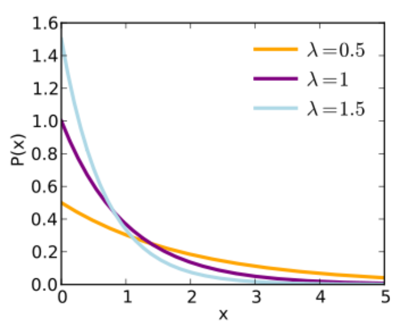
- continuous distribution
- also known as negative exponential distribution
- related to the Poisson distribution
  - describes the time between events in a Poisson process
- example:
  - Poisson: number of arrivals in a time window
  - exponential: time between arrivals

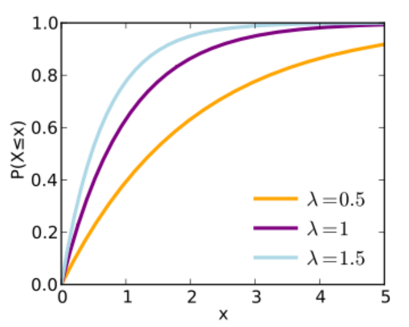
Exponential distribution

- continuous distribution
- also known as negative exponential distribution
- related to the Poisson distribution
  - describes the time between events in a Poisson process
- example:
  - Poisson: number of arrivals in a time window
  - exponential: time between arrivals
- has the same requirements as the Poisson distribution

Exponential distribution

- parameters :
  - $\lambda > 0$ real



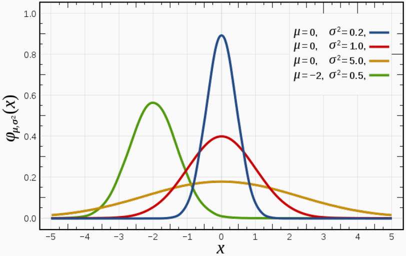


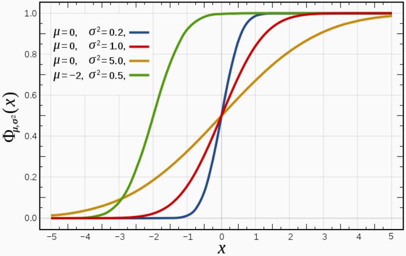
Normal Distribution

- continuous
- when uncertainty is due to several sources of uncertainty
- example:
  - measurement errors are due to environmental conditions, equipment malfunctions, human error, etc.
- central limit theorem:
  - averages of samples of observations of random variables independently drawn from independent distributions converge in distribution to the normal when the number of observations is sufficiently large.

Normal Distribution

- parameters of the $\mathcal{N}(\mu, \sigma^2)$
  - μ - mean
  - $\sigma^2 > 0$ - variance
- rules of thumb:
  - $P \approx 0.68$ that a normal random variable is within one standard deviation of the mean
  - $P \approx 0.95$ that it is within two standard deviations of the mean.



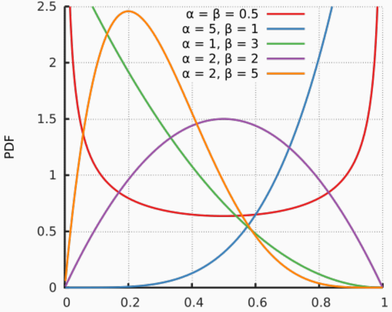


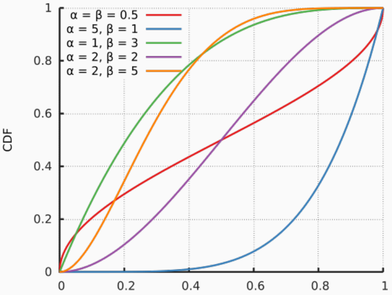
Beta Distribution

- a family of continuous probability distributions defined on the interval $[0, 1]$
- applicable when a random variable is in a limited interval
- example:
  - the proportion of people who will vote for candidate A

Beta Distribution

- parameters of $Beta(\alpha, \beta)$
  - $\alpha > 0$ shape
  - $\beta > 0$ shape





Quiz

Which distributions are good choices for the following examples of uncertainty

1. Educational and intelligence tests [normal]
2. In many market research studies, a fundamental issue is whether a potential customer prefers one product to another [binomial]
3. How many defects are acceptable in a finished product? In some products, the occurrence of defects, such as bubbles in glass or blemishes in cloth happens from time to time. [Poisson, exponential]
4. How to provide adequate service (e.g. how many cashiers should be open) when the arrival of customers is uncertain. [Poisson, exponential]

References

Part of the material has been taken from the following sources. The usage of the referenced copyrighted work is in line with fair use since it is for nonprofit educational purposes.

- Robert Clemen, Making Hard Decisions, 2nd Edition, 1996, Brooks Cole Publishing
- https://en.wikipedia.org/wiki/Binomial_distribution
- https://en.wikipedia.org/wiki/Poisson_distribution
- https://en.wikipedia.org/wiki/Exponential_distribution
- https://en.wikipedia.org/wiki/Normal_distribution
- https://en.wikipedia.org/wiki/Beta_distribution