PROBABILITY

Damian Klimke

Outline

- What is probablility? And Why?
- Random Variables
- Probability Distributions
  - Discrete Variables & Mass Functions
  - Continuous Variables & Density Functions
- Marginal Probability (sigma)
- Conditional Probability
- Chain Rule of CP
- Independence and Conditional Independence
- Expectation, Variance and Covariance

- Common Probability Distributions
  - Bernoulli Dis.
  - Mulitnoulli Dis.
  - Gaussian Dis.
  - Laplace
  - Dirac Dis. and Empirical Dis.
  - Mixtures of Dis.
- Useful Properties of Common Functions
- Bayes´ Rules
- Technical Details of Continuous Variables
- Information Theory
- Structured Probabilistic Models
- Monte Carlo
- Markov Chains
- Hashing Algorithms

What is probablility? And Why?

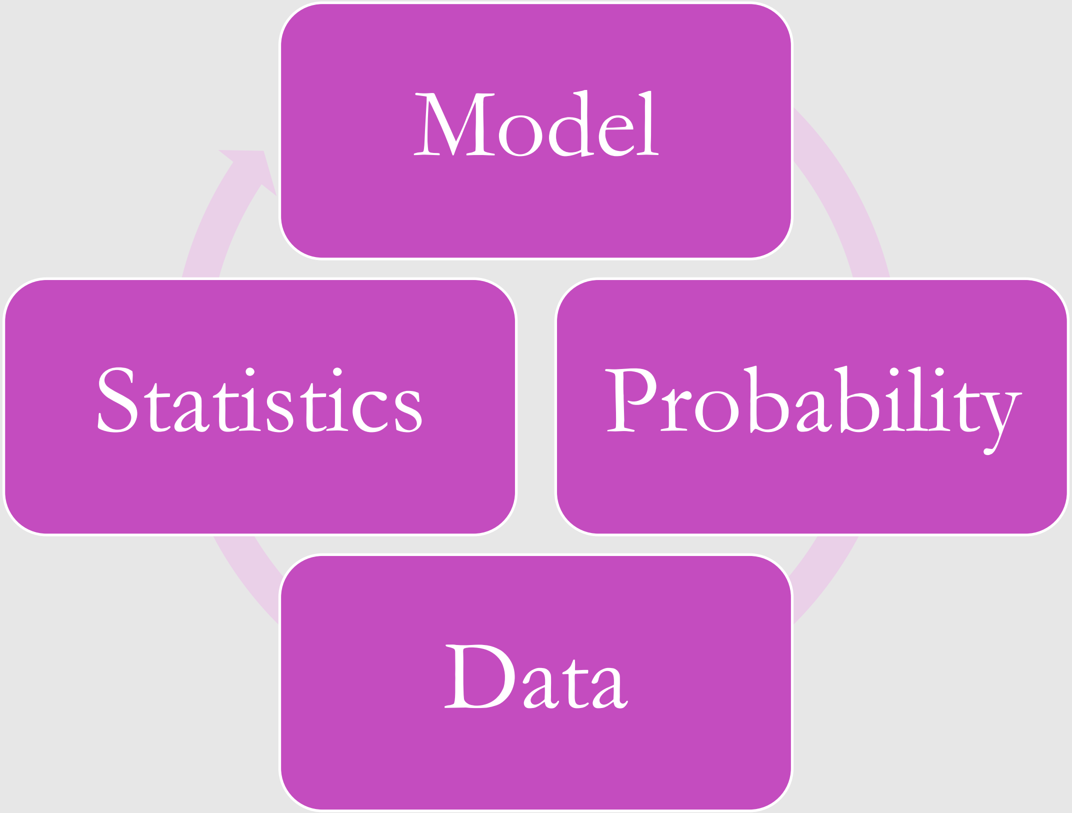
- Life is probability

- Why do we study probability theory?
  - An effective model of uncertainty
  - Decision Making under uncertainty
- Examples:
  - Measurement sensors
  - Waiting time at a Bank's teller.
  - Value of a stock at a given day.
  - Outcome of a medical procedure.
  - A customer buying behavior.
- One Decision Making Process: Collect Data, Model the Phenomenon, Extrapolate and make decisions.

What is probability? And Why?

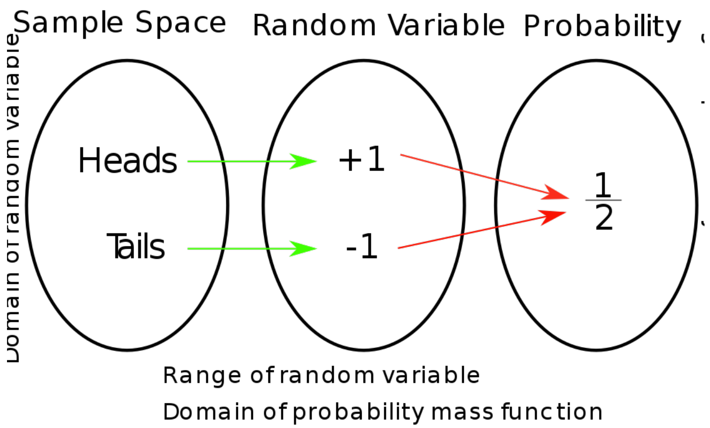
- **Probability** is a measure quantifying the likelihood that events will occur. … Probability quantifies as a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur.
- A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes ("heads" and "tails") are both equally probable; the probability of "heads" equals the probability of "tails"; and since no other outcomes are possible, the probability of either "heads" or "tails" is 1/2 (which could also be written as 0.5 or 50%).
- These concepts have been given an axiomatic mathematical formalization in probability theory, which is used widely in such areas of study as mathematics, statistics, finance, gambling, science (in particular physics), artificial intelligence/machine learning, computer science, game theory, and philosophy to, for example, draw inferences about the expected frequency of events. Probability theory is also used to describe the underlying mechanics and regularities of complex systems.
  - https://en.wikipedia.org/wiki/Probability (2019)

What is probability? And Why?



Random Variables

- Example:
- In [probability and statistics](), a **random variable**, **random quantity**, **aleatory variable**, or **stochastic variable** is described informally as a [variable whose values depend]() on [outcomes]() of a [random]() phenomenon.[1] The formal mathematical treatment of random variables is a topic in [probability theory](). In that context, a random variable is understood as a [measurable function]() defined on a [probability space]() whose outcomes are typically real numbers. (https://en.wikipedia.org/wiki/Random_variable)
- Help to express our thoughts mathematically
- https://towardsdatascience.com/understanding-random-variable-a618a2e99b93
- https://www.youtube.com/watch?v=kPRA0W1kECg



Random Variables

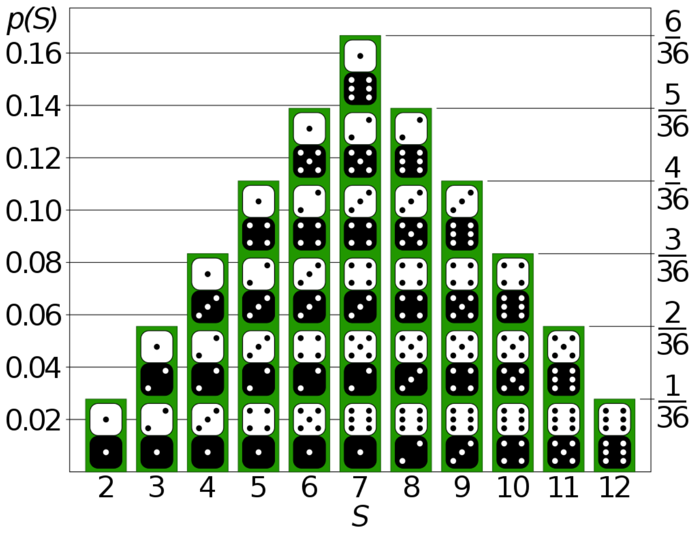
- Find an example (in your field)!

Probability Distributions

- In probability theory and statistics, a **probability distribution** is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment. In more technical terms, the probability distribution is a description of a random phenomenon in terms of the probabilities of events. For instance, if the random variable X is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of X would take the value 0.5 for X = heads, and 0.5 for X = tails (assuming the coin is fair). Examples of random phenomena can include the results of an experiment or survey. (https://en.wikipedia.org/wiki/Probability_distribution)

Probability Distributions (Dice)

- The probability mass function (pmf) $p(S)$ specifies the probability distribution for the sum S of counts from two dice. For example, the figure shows that $p(11) = 2/36 = 1/18$. The pmf allows the computation of probabilities of events such as $P(S > 9) = 1/12 + 1/18 + 1/36 = 1/6$, and all other probabilities in the distribution.
- https://en.wikipedia.org/wiki/Probability_distribution



Probability Distributions – Discrete Variables & Mass Functions

- A **discrete probability distribution** is a probability distribution that can take on a countable number of values.[1] For the probabilities to add up to 1, they have to decline to zero fast enough. For example, if $P(X = n) = \frac{1}{2^n}$ $for\ n = 1,2, \ldots$ the sum of probabilities would be $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1$.
- In probability and statistics, a **probability mass function** (**PMF**) is a function that gives the probability that a discrete random variable is exactly equal to some value.[1] The probability mass function is often the primary means of defining a discrete probability distribution, and such functions exist for either scalar or multivariate random variables whose domain is discrete.

Formal definition [edit]

Suppose that $X : S \mapsto A$ for $A \subseteq \mathbb{R}$ is a discrete random variable defined on a sample space S. Then the probability mass function $f_X : A \mapsto [0, 1]$ for X is defined as[3]

$$\boxed{f_X(x) = P(X = x) = P(\{s \in S : X(s) = x\})}$$

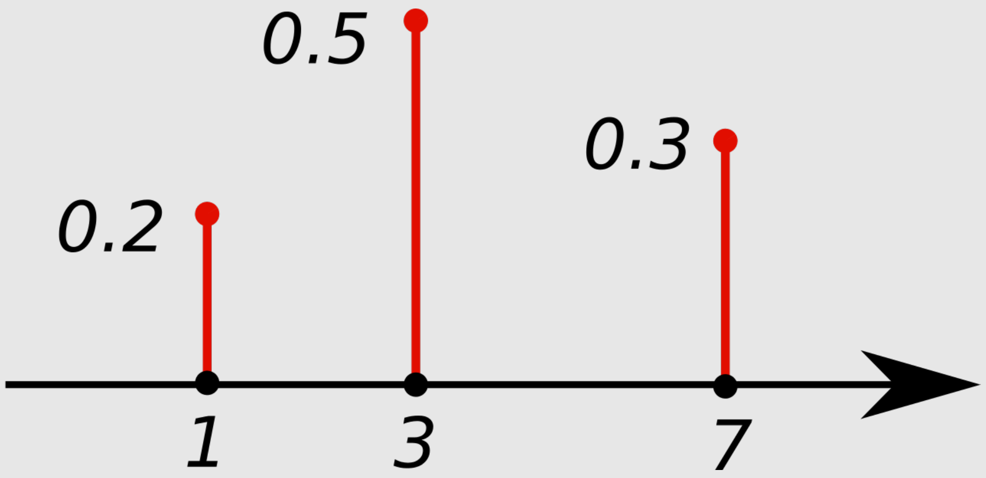
Thinking of probability as mass helps to avoid mistakes since the physical mass is conserved as is the total probability for all hypothetical outcomes x:

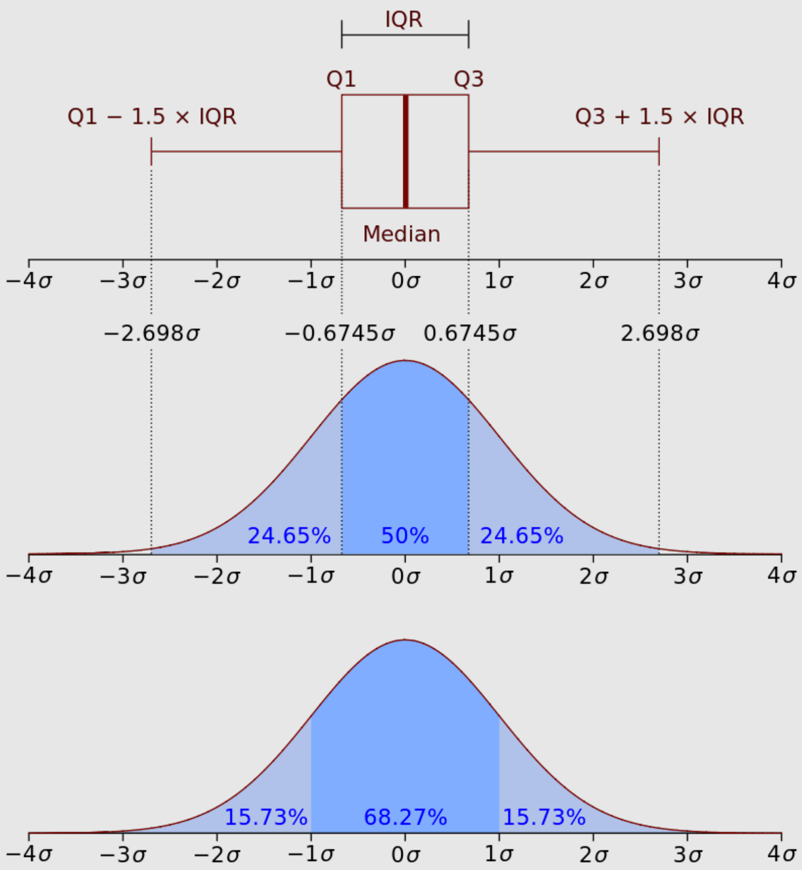
$$\sum_{x \in A} f_X(x) = 1$$

Probability Distributions – Continuous Variables & Density Functions

- A **continuous probability distribution** is a probability distribution with a cumulative distribution function that is absolutely continuous. Equivalently, it is a probability distribution on the real numbers that is absolutely continuous with respect to Lebesgue measure. Such distributions can be represented by their probability density functions. If the distribution of X is continuous, then X is called a **continuous random variable**. There are many examples of continuous probability distributions: normal, uniform, chi-squared, and others.
- Formally, if X is a continuous random variable, then it has a probability density function $f(x)$, and therefore its probability of falling into a given interval, say $[a, b]$, is given by the integral $P[a \le X \le b] = \int_a^b f(x)dx$
- https://en.wikipedia.org/wiki/Probability_distribution

Probability Distributions – What is what?





Probability Distributions

- Examples?
- https://en.wikipedia.org/wiki/List_of_probability_distributions

Marginal Probability

- Sometimes we know the probability distribution over a set of variables and we want to know the probability distribution over just a subset of them. The probability distribution over the subset is known as the marginal probability distribution. (https://www.deeplearningbook.org/contents/prob.html)
- For example, suppose we have discrete random variables x and y, and we know P (x, y). We can find P (x) with the sum rule:
- $\forall x \in x$, $P(x = x) = \sum_y P(x = x, y = y)$
- For continues variables we need the integration:
- $p(x) = \int p(x, y)dy$

Y \ X	x_1	x_2	x_3	x_4	$p_Y(y)$ ↓
y_1	4/32	2/32	1/32	1/32	**8/32**
y_2	3/32	6/32	3/32	3/32	**15/32**
y_3	9/32	0	0	0	**9/32**
$p_X(x)$ →	**16/32**	**8/32**	**4/32**	**4/32**	**32/32**

Joint and marginal distributions of a pair of discrete random variables, *X* and *Y*, having nonzero mutual information *I*(*X*; *Y*). The values of the joint distribution are in the 3×4 rectangle; the values of the marginal distributions are along the right and bottom margins.

In contrast to:

- In probability theory and statistics, 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. This contrasts with a conditional distribution, which gives the probabilities contingent upon the values of the other variables.

Conditional Probability

- In many cases, we are interested in the probability of some event, given that some other event has happened. This is called a conditional probability. We denote the conditional probability that y=$\mathcal{y}$ given x=$\mathcal{x}$ as P(y=$\mathcal{y}$|x= $\mathcal{x}$). This conditional probability can be computed with the formula

$$P(y = \mathcal{y}|\mathrm{x} = \mathcal{x}) = \frac{P(y = \mathcal{y}, x = \mathcal{x})}{P(x = \mathcal{x})}$$

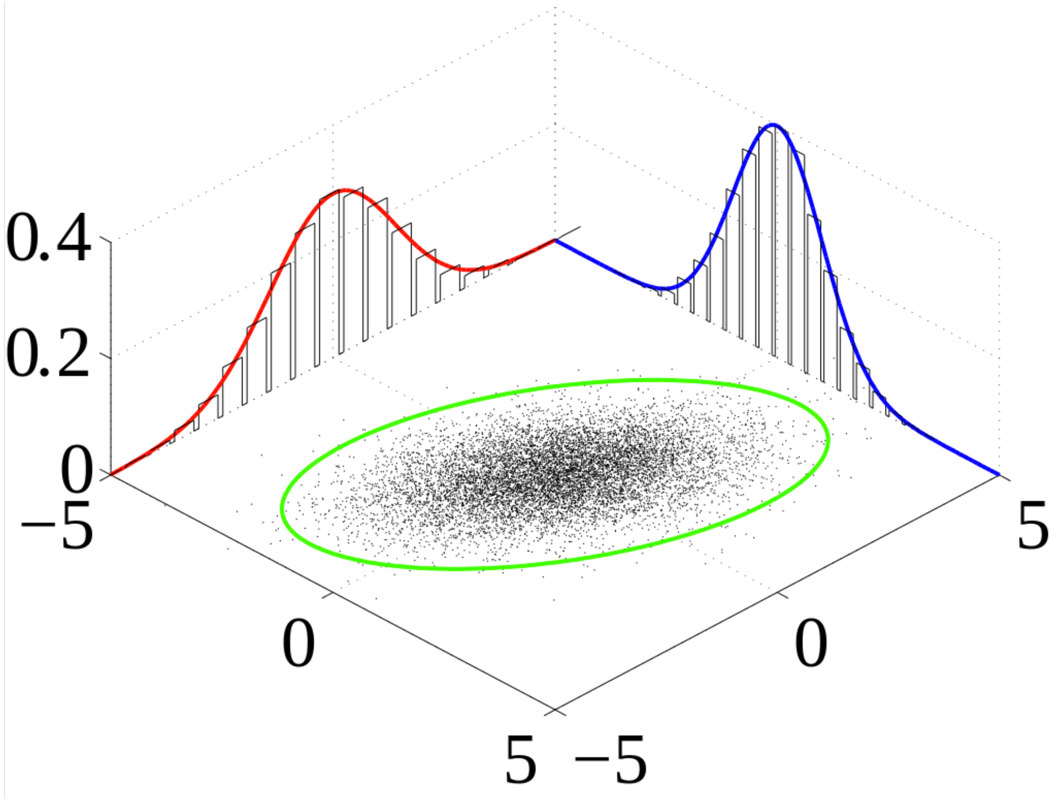
For: $P(x = \mathcal{x}) > 0$

Example [edit]

Consider the roll of a fair die and let $X = 1$ if the number is even (i.e. 2, 4, or 6) and $X = 0$ otherwise. Furthermore, let $Y = 1$ if the number is prime (i.e. 2, 3, or 5) and $Y = 0$ otherwise.

	1	2	3	4	5	6
X	0	1	0	1	0	1
Y	0	1	1	0	1	0

Then the unconditional probability that $X = 1$ is 3/6 = 1/2 (since there are six possible rolls of the die, of which three are even), whereas the probability that $X = 1$ conditional on $Y = 1$ is 1/3 (since there are three possible prime number rolls—2, 3, and 5—of which one is even).



JOINT DISTRIBUTION

Given random variables $X, Y, \ldots,$ that are defined on a probability space, the **joint probability distribution** for $X, Y, \ldots$ is a probability distribution that gives the probability that each of $X, Y, \ldots$ falls in any particular range or discrete set of values specified for that variable. In the case of only two random variables, this is called a **bivariate distribution**, but the concept generalizes to any number of random variables, giving a **multivariate distribution**.

Chain Rule of CP

- For two events: A & B
  - $P(A \cap B) = P(A|B) * P(B)$
- Example?
- More than two events $A_1, \ldots, A_n$:
  - $P(A_n \cap \ldots \cap A_1) = P(A_n | A_{n-1} \cap \ldots \cap A_1) * P(A_{n-1} \cap \ldots \cap A_1)$
  - Also:

$$P(A_n \cap \ldots \cap A_1) = \prod_{k=1}^{n} P\left(A_k \middle| \cap_{j=1}^{k-1} A_j\right)$$

- Other writing:

$$P(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(n)}) = P(\mathrm{x}^{(1)}) \Pi_{i=2}^{n} P(\mathrm{x}^{(i)} \mid \mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(i-1)})$$

Exercise – Chain Rule:

- With 4 events (n=4), the chain rule is:
  - ?????

Move me!!!!

Independence and Conditional Independence

- Independence:
  - Two events are independent(A⊥B), **if and only if** their joint probability equals the product of their probabilities:
  - $P(A \cap B) = P(A)P(B)$

WHY??

  - Short: the events do not effect each other.

Independence and Conditional Independence

- Independence: Example?

Independence and Conditional Independence

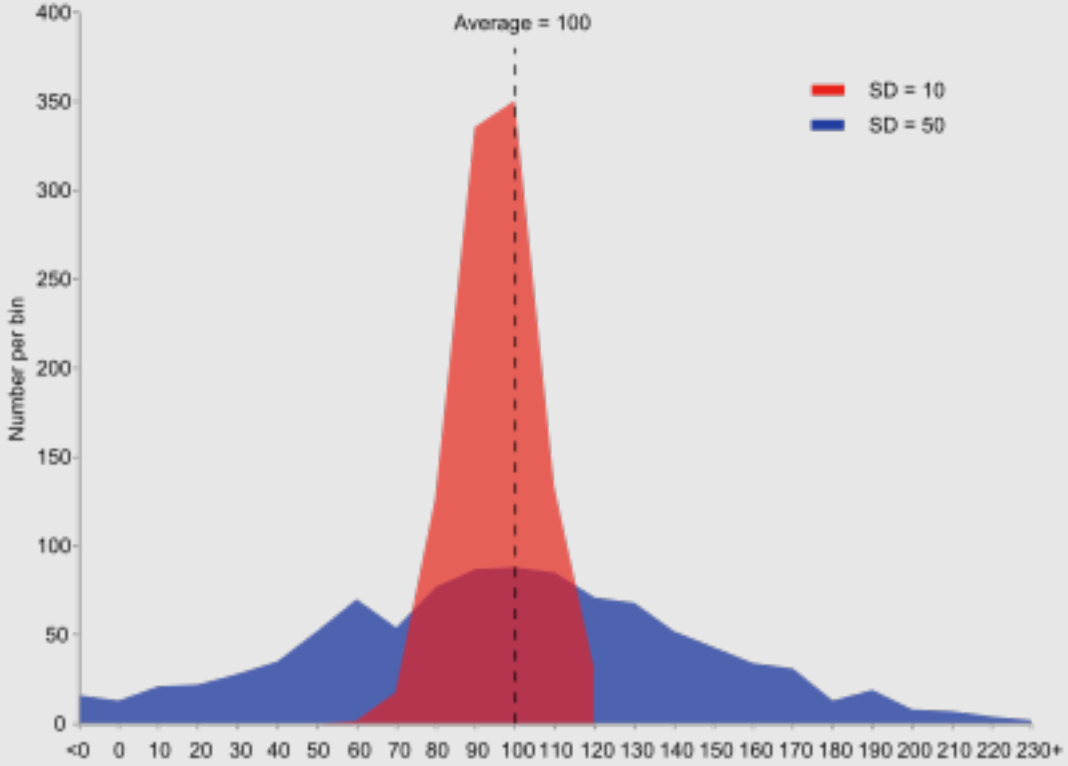
- Conditional independence:
  - A and B are conditionally independent given C, **if and if only,** $P(A \cap B|C) = P(A|C)P(B|C)$
  - $(A \perp B)|C \Leftrightarrow P(A \cap B|C) = P(A|C)P(B|C)$
  - Find an example in your field:
  - https://en.wikipedia.org/wiki/Conditional_independence

Expectation

- Expectation or Expected Value
  - Intuitively, a random variable's expected value represents the average of a large number of independent realizations of the random variable.
- X is a random variable with a finite number of outcomes $(x_1, x_2, \ldots, x_k)$ occurring with probabilities $(p_1, p_2, \ldots, p_k)$.
  - $E[X] = \sum_{i=1}^{k} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k$ for discrete variables
  - For continues variables we compute it with an integral.
  - Example???

Variance

- Variance: The variance of a standard X (function) is the expected value of the squared deviation from the mean $X, \mu = E[X]$:
- $Var(X) = E[(X - \mu)^2]$
- Example?? E.g.: Fair Die
- The $\sqrt{Var(X)}$ is called the "standard deviation"



Covariance

- The covariance gives some sense of how much two values are linearly related to each other, as well as the scale of these variables:
- X and Y (random variables (jointly distributed)): (If the covariance is positive, there is a positive correlation. In fact, correlation coefficients can simply be understood as a normalized version of covariance.)
  - $cov(X, Y) = E[(X - E[X])(Y - E[Y])]$
- Example or need?
- What is a covariance matrix and what for?

Common Probability Distributions

Bernoulli Distribution

- (Yes/No – distribution)
  - $P(X = 1) = p; P(X = 0) = 1 - p = q$

The probability mass function f of this distribution, over possible outcomes k, is

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ q = 1 - p & \text{if } k = 0. \end{cases}$$ [2]

This can also be expressed as

$$f(k;p) = p^k (1-p)^{1-k} \quad \text{for } k \in \{0, 1\}$$

  - $E[X] = p$
  - $Var_x(X) = p(1 - p)$
  - What For????
    - https://en.wikipedia.org/wiki/Bernoulli_process

Multinoulli Distribution & Categorical dis.

- Special case of the "multinomial dis."
- The multinoulli, or categorical, distribution is a distribution over a single discrete variable with k different states, where k is finite.
- Vector $p \in [0,1]^{k-1}$, p_i gives the probability
- k>0, number of categories
- Often over "objects" (no need for the var, or E[x])
- Example??
- https://www.statisticshowto.datasciencecentral.com/multinomial-distribution/

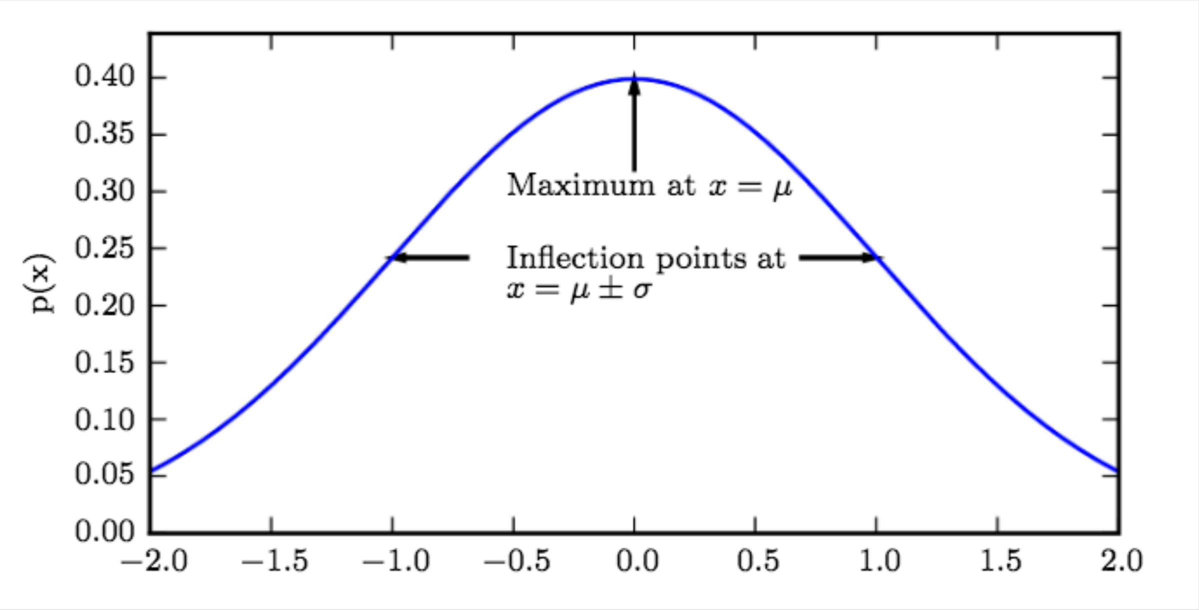
Gaussian Distribution

- Also known as: normal distribution.

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

- $\mu = E[X], \sigma^2 = Var(X), \sigma = SD$

- central limit theorem
- multivariate normal distribution
- What for???



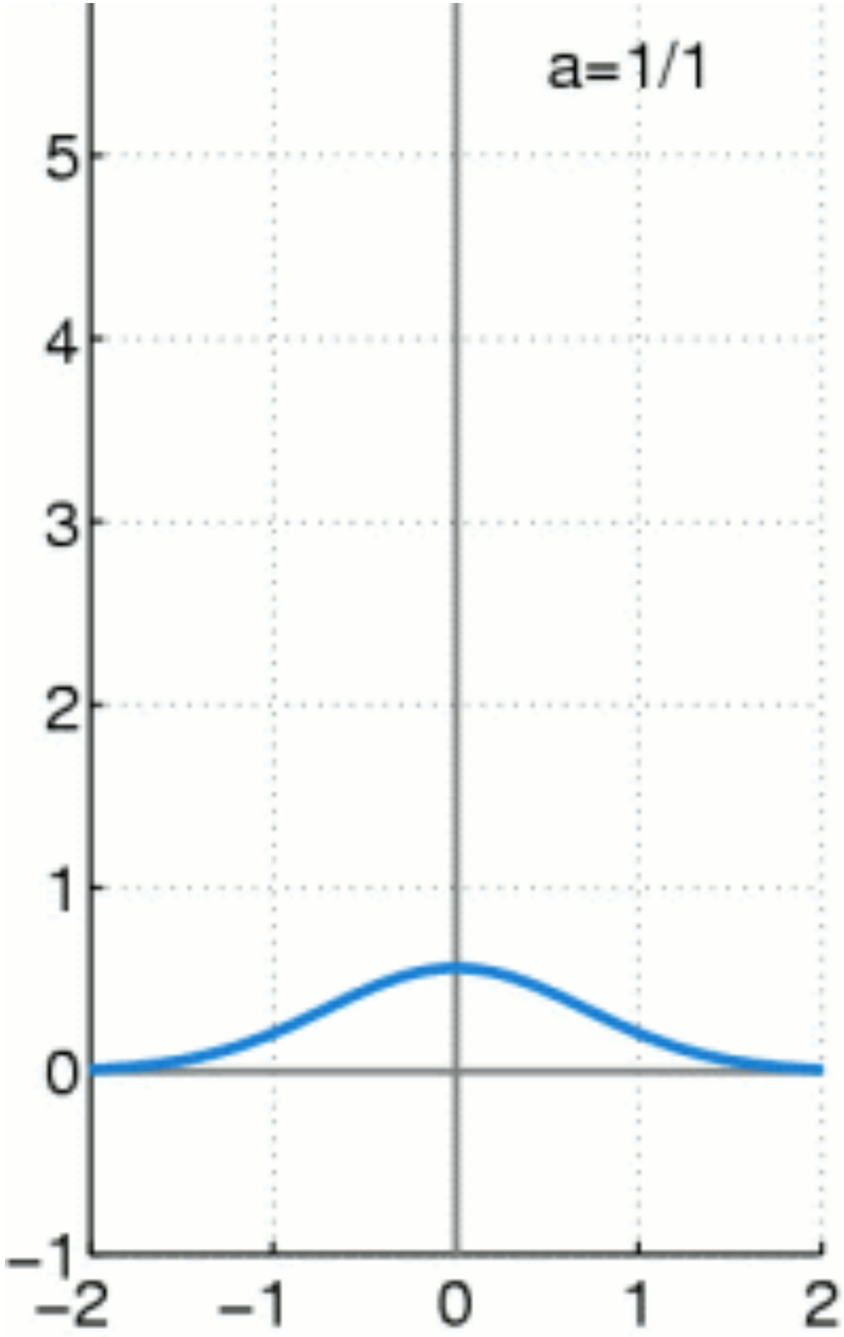
Laplace Distribution aka. Exponential Dis.

- In the context of deep learning, we often want to have a probability distribution with a sharp point at x = 0
- $\mu = Mean; b > 0\ (scale)$
- What for???

A random variable has a $\mathrm{Laplace}(\mu, b)$ distribution if its probability density function is

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b} \right)$$

$$= \frac{1}{2b} \begin{cases} \exp\left(-\frac{\mu - x}{b}\right) & \text{if } x < \mu \\ \exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases}$$



Dirac Dis.

- Dirac Delta Function as a “Limit”
- to specify that all the mass in a probability distribution clusters around a single point.
- $p(x) = \delta(x - \mu)$
- Aka: generalized function
- We can think of the Dirac delta function as being the limit point of a series of functions that put less and less density on all points other than zero.

By defining $p(x)$ to be δ shifted by $-\mu$ we obtain an infinitely narrow and infinitely high peak of probability density where $x = \mu$.

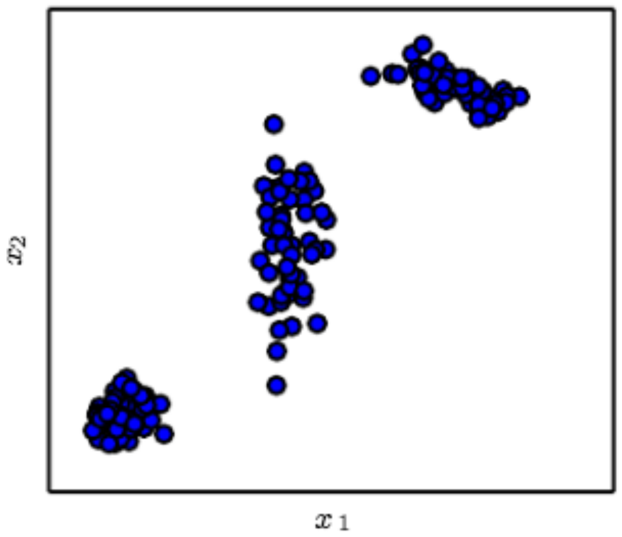
A common use of the Dirac delta distribution is as a component of an **empirical distribution**,

$$\hat{p}(\boldsymbol{x}) = \frac{1}{m} \sum_{i=1}^{m} \delta(\boldsymbol{x} - \boldsymbol{x}^{(i)}) \tag{3.28}$$

which puts probability mass $\frac{1}{m}$ on each of the m points $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$, forming a given data set or collection of samples. The Dirac delta distribution is only necessary to define the empirical distribution over continuous variables. For discrete variables, the situation is simpler: an empirical distribution can be conceptualized as a multinoulli distribution, with a probability associated with each possible input value that is simply equal to the **empirical frequency** of that value in the training set.

Empirical Dis. & Dirac

- What for??



Mixtures of Dis.

- $Px) = \sum_i P(c = i)P(x|c = i)$
- P(c) is the multinoulli distribution over component identities
- E.g.: Empirical Dis. & Gaussian Mixture Model

A very powerful and common type of mixture model is the **Gaussian mixture model**, in which the components $p(\mathbf{x} \mid \mathrm{c} = i)$ are Gaussians. Each component has a separately parametrized mean $\boldsymbol{\mu}^{(i)}$ and covariance $\boldsymbol{\Sigma}^{(i)}$. Some mixtures can have more constraints. For example, the covariances could be shared across components via the constraint $\boldsymbol{\Sigma}^{(i)} = \boldsymbol{\Sigma}, \forall i$. As with a single Gaussian distribution, the mixture of Gaussians might constrain the covariance matrix for each component to be diagonal or isotropic.

Bayes´ Rule

- $P(A|B) = \frac{P(B\,|A)P(A)}{P(B)}$
- Often we know P(B|A), but want to find P(A|B)
- Example:??

Technical Details of Continuous Variables

- Measure Theory
- Measure Zero
- Almost Everywhere
- https://www.deeplearningbook.org/contents/prob.html#pf11

Information Theory

- Units of "nats", "bits", "Shannons"
- https://www.deeplearningbook.org/contents/prob.html#pf11

The basic intuition behind information theory is that learning that an unlikely event has occurred is more informative than learning that a likely event has occurred. A message saying "the sun rose this morning" is so uninformative as to be unnecessary to send, but a message saying "there was a solar eclipse this morning" is very informative.

We would like to quantify information in a way that formalizes this intuition.

- Likely events should have low information content, and in the extreme case, events that are guaranteed to happen should have no information content whatsoever.
- Less likely events should have higher information content.
- Independent events should have additive information. For example, finding out that a tossed coin has come up as heads twice should convey twice as much information as finding out that a tossed coin has come up as heads once.

Structured Probabilistic Models

- $P(a, b, c) = p(a)p(b|a)p(b|c)$
- Suppose that a influences the value of b, and b influences the value of c, but that a and c are independent given b.
- structured probabilistic model, or graphical model
  - Directed & Undirected

Directed:

Directed models use graphs with directed edges, and they represent factorizations into conditional probability distributions, as in the example above. Specifically, a directed model contains one factor for every random variable x_i in the distribution, and that factor consists of the conditional distribution over x_i given the parents of x_i, denoted $Pa_{\mathcal{G}}(\mathrm{x}_i)$:

$$p(\mathbf{x}) = \prod_i p\left(\mathrm{x}_i \mid Pa_{\mathcal{G}}(\mathrm{x}_i)\right). \tag{3.53}$$

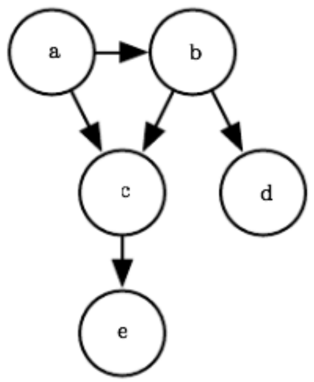


Figure 3.7: A directed graphical model over random variables a, b, c, d and e. This graph corresponds to probability distributions that can be factored as

$$p(\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}) = p(\mathrm{a})p(\mathrm{b} \mid \mathrm{a})p(\mathrm{c} \mid \mathrm{a}, \mathrm{b})p(\mathrm{d} \mid \mathrm{b})p(\mathrm{e} \mid \mathrm{c}). \tag{3.54}$$

Undirected:

- Clique: C^i

$$p(\mathbf{x}) = \frac{1}{Z} \prod_i \phi^{(i)} \left(\mathcal{C}^{(i)} \right).$$

- ϕ are factors
- (just functions, no probability dis.)

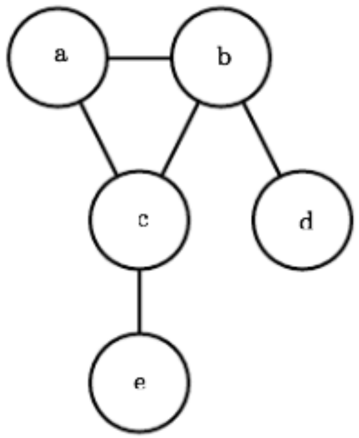


Figure 3.8: An undirected graphical model over random variables a, b, c, d and e. This graph corresponds to probability distributions that can be factored as

$$p(\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}) = \frac{1}{Z} \phi^{(1)}(\mathrm{a}, \mathrm{b}, \mathrm{c}) \phi^{(2)}(\mathrm{b}, \mathrm{d}) \phi^{(3)}(\mathrm{c}, \mathrm{e}). \tag{3.56}$$

Keep in mind that these graphical representations of factorizations are a language for describing probability distributions. They are not mutually exclusive families of probability distributions. Being directed or undirected is not a property of a probability distribution; it is a property of a particular **description** of a probability distribution, but any probability distribution may be described in both ways.

Monte Carlo

- **Monte Carlo methods**, or **Monte Carlo experiments**, are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. The underlying concept is to use randomness to solve problems that might be deterministic in principle. They are often used in physical and mathematical problems and are most useful when it is difficult or impossible to use other approaches. Monte Carlo methods are mainly used in three problem classes:[1] optimization, numerical integration, and generating draws from a probability distribution.
- Monte Carlo methods vary, but tend to follow a particular pattern:
  - Define a domain of possible inputs
  - Generate inputs randomly from a probability distribution over the domain
  - Perform a deterministic computation on the inputs
  - Aggregate the results
- https://en.wikipedia.org/wiki/Monte_Carlo_method

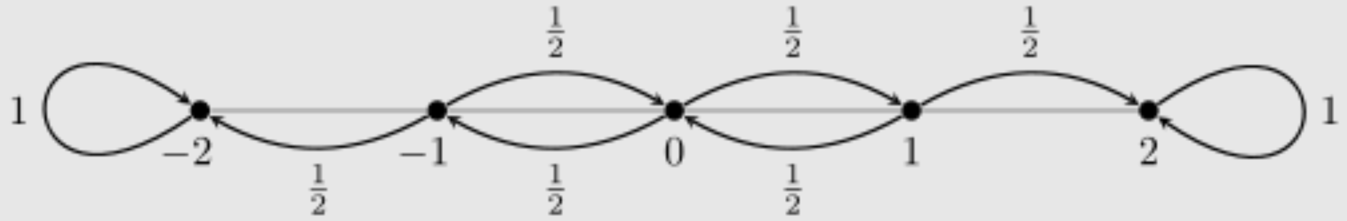
Monte Carlo

- $\langle A \rangle = \sum_{x\in\Omega} P(x)A(x)$, (Descrete)
- $\int_{x\in\Omega} P(x)A(x)d^n x$,

- Examples???

Markov Chains/Models

- A **Markov chain** is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.
- https://en.wikipedia.org/wiki/Markov_chain
- https://towardsdatascience.com/introduction-to-markov-chains-50da3645a50d
- http://web.math.ku.dk/noter/filer/stoknoter.pdf



Definition of a Markov chain

A **stochastic process** in discrete-time is a family, $(X(n))_{n\in\mathbb{N}_0}$, of random variables indexed by the numbers $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The possible values, S, of $X(n)$ are referred to as the **state space** of the process. In this course we consider only stochastic processes with values in a finite or countable state space. The mathematician may then think of a **random variable** , X, on S as a measurable map [1]

$$X: \quad (\Omega, \mathcal{F}) \to (S, \mathcal{P}(S))$$

where $\mathcal{P}(S)$ is the family of all subsets of S.

The distribution of a discrete-time stochastic process [2] with at most countable state space, S, is characterised by the point probabilities

$$P(X(n) = i_n, X(n-1) = i_{n-1}, \ldots, X(0) = i_0)$$

for $i_n, i_{n-1}, \ldots, i_0 \in S$ and $n \in \mathbb{N}_0$. From the definition of elementary conditional probabilities it follows that

$$\begin{aligned} & P(X(n) = i_n, \ldots, X(0) = i_0) \\ = \; & P(X(n) = i_n | X(n-1) = i_{n-1}, \ldots, X(0) = i_0) \\ \times \; & P(X(n-1) = i_{n-1} | X(n-2) = i_{n-2}, \ldots, X(0) = i_0) \\ \times \; & \ldots \\ \times \; & P(X(1) = i_1 | X(0) = i_0) \times P(X(0) = i_0). \end{aligned}$$

This is a general identity that holds for *any* discrete-time stochastic process on a countable state space. In these lecture notes we are only going to discuss the class of Markov chains to be defined below.

MARKOV CHAINS

A sequence of random variables:

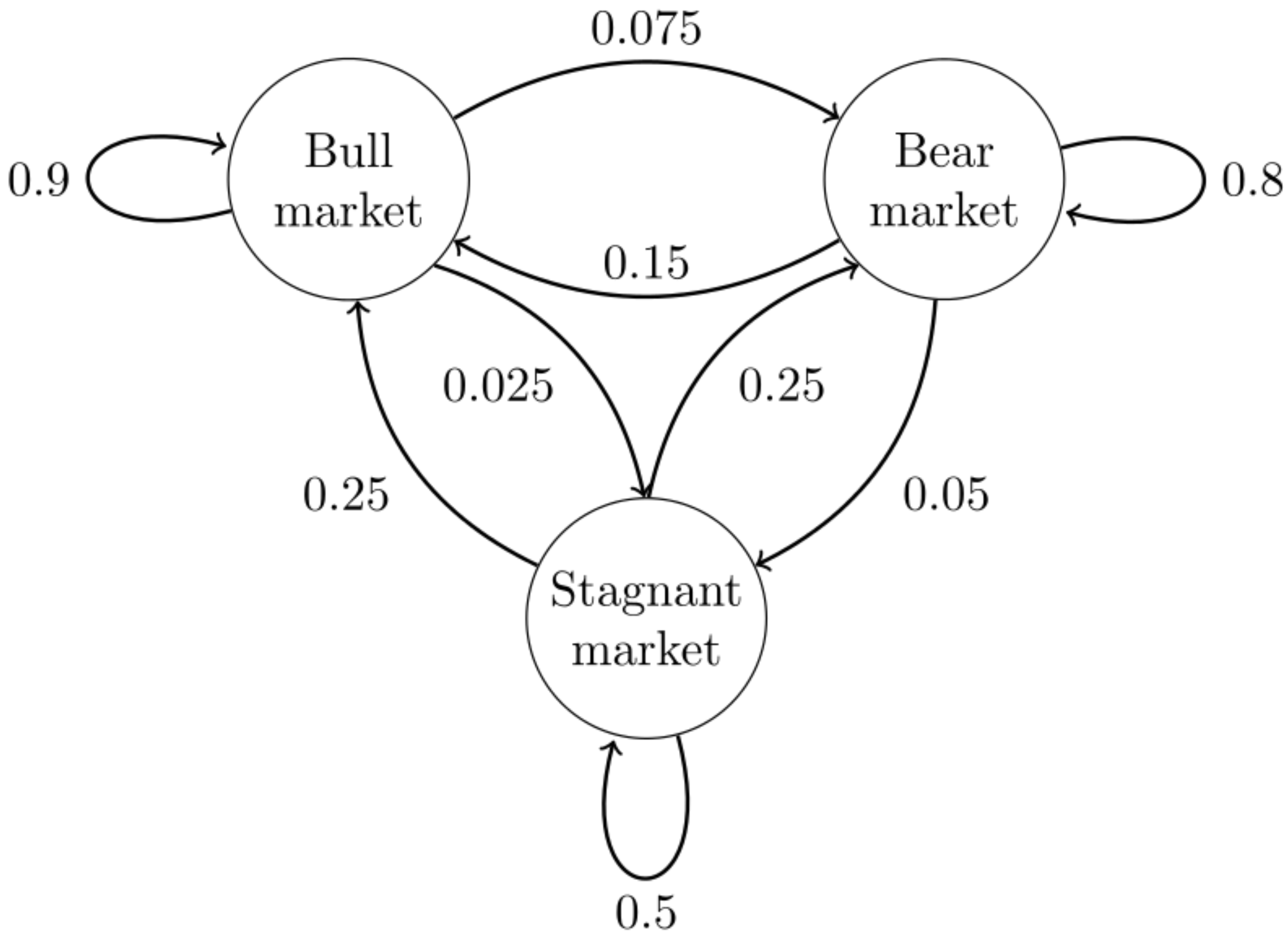
Markov Chains

Discrete-time Markov chain [edit]

A discrete-time Markov chain is a sequence of random variables X_1, X_2, X_3, ... with the Markov property, namely that the probability of moving to the next state depends only on the present state and not on the previous states:

$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n)$, if both conditional probabilities are well defined, that is, if $\Pr(X_1 = x_1, \ldots, X_n = x_n) > 0.$

The possible values of X_i form a countable set S called the state space of the chain.



Markov Chains

- Example.
- How does it work?
- https://en.wikipedia.org/wiki/Markov_chain#Formal_definition

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

The distribution over states can be written as a stochastic row vector x with the relation $x^{(n+1)} = x^{(n)}P$. So if at time n the system is in state $x^{(n)}$, then three time periods later, at time $n + 3$ the distribution is

$$\begin{aligned} x^{(n+3)} &= x^{(n+2)} P = \left(x^{(n+1)} P\right) P \\ &= x^{(n+1)} P^2 = \left(x^{(n)} P\right) P^2 \\ &= x^{(n)} P^3 \end{aligned}$$

In particular, if at time n the system is in state 2 (bear), then at time $n + 3$ the distribution is

$$\begin{aligned} x^{(n+3)} &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}^3 \\ &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.7745 & 0.17875 & 0.04675 \\ 0.3575 & 0.56825 & 0.07425 \\ 0.4675 & 0.37125 & 0.16125 \end{bmatrix} \\ &= \begin{bmatrix} 0.3575 & 0.56825 & 0.07425 \end{bmatrix}. \end{aligned}$$

MARKOV CHAINS

- Own Idea??

Hashing Algorithms

- What does hashing had to do with probability???

Links:

- https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/probability-main-index/
  - Like a dictoinary for DS
- https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-041-probabilistic-systems-analysis-and-applied-probability-spring-2006/lecture-notes/
  - MIT Lecture notes
- https://towardsdatascience.com/introduction-to-markov-chains-50da3645a50d
  - DS
- http://web.math.ku.dk/noter/filer/stoknoter.pdf
  - Markov Chains