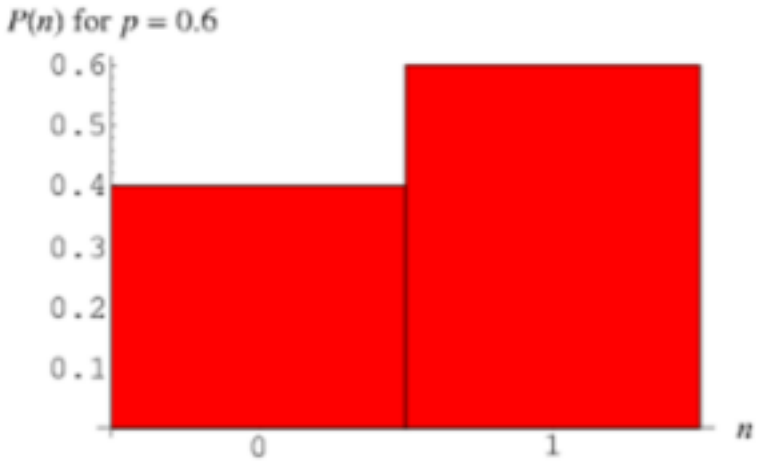
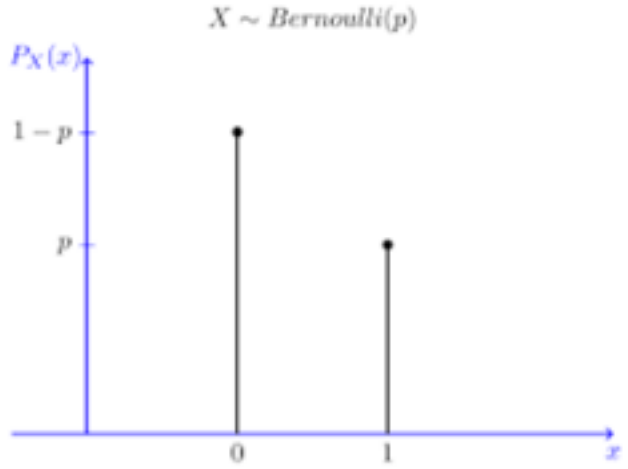
Danbury AI
Deep Learning Study Group
Notes on Section 3.9
Doug Cowles

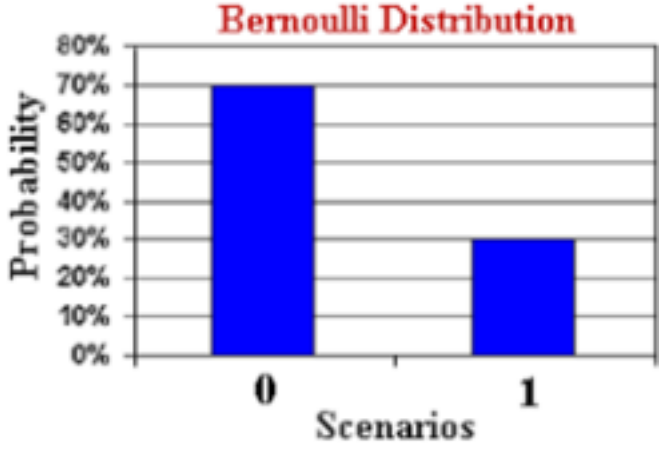
Bernoulli Distribution

- The simplest distribution, for yes-no questions: coin flips, Male/Female, etc.,
- One happens with probability p, the other with 1-p. – described in book as 3.16 & 3.17 with letter phi.
- 3.18 P(X = x) refers to the probability that the random variable X is equal to a particular value, denoted by x. As an example, P(X = 1) refers to the probability that the random variable X is equal to 1.
- 3.19 is the Expected Value
- 3.20 is the Variance

Examples of Bernoulli Distributions







Multinoulli Distribution

- Essentially, a form of a binomial distribution where instead of 2 outcomes, there are K outcomes – like rolling a die.
- The **multinomial** distribution is when there are **multiple** identical independent trials where each trial has *k* possible outcomes. The categorical (multinoulli) distribution is when there is **only one** such trial.
- Best explained here: https://www.statlect.com/probability-distributions/multinomial-distribution
- A Multinoulli Distribution is not the same thing as a multinomial distribution. Explained here: https://geekyisawesome.blogspot.com/2016/12/bernoulli-vs-binomial-vs-multinoulli-vs.html

Gaussian Distribution

- Also known as “bell curve” or “Normal” distribution
- The formula is interesting but few people worry about it 3.21
- The **normal distribution** is **important** because of the Central limit theorem. In simple terms, if you have many independent variables that may be generated by all kinds of **distributions**, assuming that nothing too crazy happens, the aggregate of those variables will tend toward a **normal distribution**.
- It’s sometimes stated that many things in nature just seem to be normally distributed but I hear that challenged more and more. In other words, test scores, heights, foot sizes, density of leaves under a tree. There appear to be slight variations on the normal distribution being brought up more frequently.

Gaussian Disribution part 2

- Normal or Gaussian Distribution
- 68.2% within 1 Standard Deviation
- Another 95.4% within 2 Standard Deviations
- 99.6% within 3 Standard Deviations
- Often called the 68-95-99.7 rule.

Central Limit Theorem

- The Central Limit Theorem (*CLT* for short) basically says that for non-normal data, the distribution of the sample means has an approximate normal distribution, no matter what the distribution of the original data looks like, as long as the sample size is large enough (usually at least 30) and all samples have the same size. And it doesn't just apply to the sample mean; the CLT is also true for other sample statistics, such as the sample proportion. Because statisticians know so much about the normal distribution, these analyses are much easier.
- Great explanation here - https://www.quora.com/What-is-the-importance-of-the-central-limit-theorem if you see the answer by Wisnu Subekti, K-12 math teacher

Exponential Distribution

- Consider how long glass vases last before being broken. Glass vases do not age, so it is just as likely to break at 10 years old, as 80 years old. This is unlike a human, whose probability goes up at 80, compared to 10.
- It is “memoryless” - it is the probability **distribution** that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. (Examples of a Poisson process - # of pieces of mail that arrive at your mailbox.. Cars at a tollbooth – Assuming no obvious external factors like rush hour)
- Great link here - https://www.statlect.com/probability-distributions/exponential-distribution

Laplace Distribution

- Haven’t figured this out yet. Video here: https://www.youtube.com/watch?v=1wsig_TGrtg I haven’t gotten through it yet.

Dirac Delta Function

- Explained very well here: https://www.khanacademy.org/math/differential-equations/laplace-transform/properties-of-laplace-transform/v/dirac-delta-function
- It is a function that is zero at every point but one, and whose integral is 1. It is a “generalized function”
- The book describing the function as that which is defined in terms of its properties when integrated seems a good definition.
- The concepts of “empirical distribution” and “empirical frequency” are still unclear to me at this time.

Mixture Distributions

- On the face of it, seems like obvious concept. Distributions that are a composite of other distributions.
- Significant for finding “latent” variables.
- Gaussian Mixture Models https://www.youtube.com/watch?v=Rkl30Fr2S38