Probability Cheatsheet

Yahya Almardeny
almardeny@gmail.com

June 2019

1 Overview

Probability provides measures for reasoning the likelihood that an event will occur in an experiment.

2 Terminology

1. **Experiment**: Procedure that yields one of the possible outcomes. E.g. Tossing a coin.

2. **Sample Space S**: A set of all possible outcomes of an experiment. E.g. S = {Head, Tail} when tossing a fair coin.

3. **Event E**: Set of outcomes of an experiment. Note that if the Event contains only one outcome, it is then called an elementary event. E.g. The event that coin is Head.

3 Random Variable R.V

A variable that its possible values are the outcomes of an experiment. It has two types:

1. **Discrete R.V**: Take one countable number of distinct values. E.g 1, 2, 3, 4

2. **Continuous R.V**: Take infinite number of possible values (usually measurements). E.g. all interval values in the rage [1 - 2].

4 Probability of Events

1. **Marginal probability**: Simply $P(A)$ which means "unconditional probability". And for each outcome s, it should satisfy: $0 \leq P(s) \leq 1$ and $\Sigma P(s_i) = 1$.

2. **Addition Rule**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ where $\cup$ means OR, $\cap$ means AND.

3. **Joint (Compound) Probability Rule**:

   (a) Independent Events: $P(A \cap B) = P(A) \times P(B)$.

   (b) Dependent Events: $P(A \cap B) = P(A) \times P(B|A)$. where $P(B|A)$ read: Probability of B given A had happened.

4. **Conditional Probability**: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

5. **Bayes Theorem**: $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$. It gives the probability of A based on prior knowledge of B.

5 Probability Distributions

1. **Probability Mass Function (PMF)**: Gives the probability a *discrete* R.V X takes on the value x: $P(X = x)$.
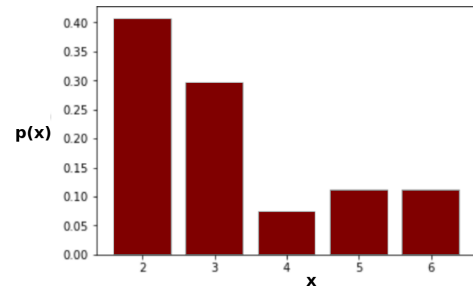
2. **Probability Density Function (PDF)**: Gives the probability a *continuous* R.V X takes on the value x: $P(X = x)$ which means how dense the probability of X near x, and it will be the definite Integrals between two points.
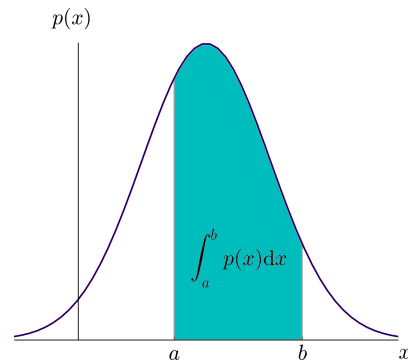
3. **Cumulative Density Function (CDF)**: Gives the probability a R.V X is less than or equals the value x: $P(X \leq x)$. It is cumulative because it adds the probabilities up to x.

   (a) For Discrete R.V: It is the summation of previous probabilities $\Sigma p(x_i)$.

   (b) For Continuous R.V: It is the integral $\int_{-\infty}^{x} p(x)dx$

6 Expected Value of R.V

Intuitively, it is the long-run average value of repetitions of the same experiment it represents (i.e. what outcome to expect on long run)

1. **For a Discrete R.V**: It is the average of R.V values based on their associated probabilities $E(X) = \Sigma(x_i \times p(x_i))$.

2. **For a Continuous R.V**: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ where $f(x)$ is probability density function.

7 Combining Random Variables

If we have two independent random variables X and Y, we can add or subtract to get the total or the difference of their means (expected values) and variances: $T = X \pm Y \Rightarrow \mu_T = \mu_X \pm \mu_Y \ \ and \ \ \sigma_T^2 = \sigma_X^2 \pm \sigma_Y^2$.

8 The Law of Large Numbers

Let X be R.V of a population where its expected value (i.e Mean) is $E(X)$. Let $\overline{X_n} = \frac{x_1 + x_2 + ... x_n}{n}$ be the mean of n samples $\in X$. As $n \to \infty : \overline{X_n} \to E(X)$. Which means as we do more experiments, the outcomes become closer and closer to the outcomes of the probability theory.

9 Statistical Distributions

[1]

1. **Binomial Distribution**: How many success in finite number of trials:

   - Made up of *independent* trials.
   - Each trial has one of two *discrete* outcomes (success or failure).
   - *Fixed* number of trials
   - Probability of success in each trial is *constant*.
   - E.g. X = number of heads after 10 flips of a coin.

---

[1]Included because it is related directly to probability

- *10% Rule*: We assume independence if the sample $\leq 10\%$ of the population.
- PMF: $P(X = x) = \binom{k}{n} p^x (1-p)^{n-x}$ which means the probability of k successes in n trials.
- $E(X) = \mu_X = n.p$ where p is the probability of success per trial.
- $\sigma_X = \sqrt{var(X)} = \sqrt{n.p(1-p)}$.
- The more trials we do, the more Binomial becomes Normal distribution.

2. **Bernoulli Distribution**: Simply a Binomial Distribution with just *one* trial.

3. **Geometric Distribution**: How many trials until success?. It is similar to Binomial dist. but here we do not have fixed number of trials because we do not know ahead how many trials until we get the desired outcome:

   - PMF: $P(X = x) = (1-p)^{k-1} p$.
   - $E(X) = \mu_X = \frac{1}{p}$.
   - $\sigma_X = \sqrt{var(X)} = \sqrt{\frac{1-p}{p^2}}$.

4. **Hypergeometric Distribution**: It is a *discrete* probability distribution that describes the probability of k successes in n draws, *without* replacement, from a finite population of size N that contains exactly K objects. In contrast, the binomial distribution describes the probability of k successes in n draws *with* replacement.

5. **Poisson Distribution**: It is a *discrete* probability distribution that predicts *rare* events that are independent of one another and occur with a known constant rate λ. It is usually used instead of Binomial dist. if the number of trials is very high (fixed interval of time) and the probability (occurrence) of each is relatively low:

   - PMF: $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$.
   - Mean is $E(X) = \lambda$. and Variance is λ.

6. **Normal Distribution (a.k.a Gaussian)**: It is a *continuous* probability distribution that has the notation $X \sim \mathcal{N}(\mu, \sigma^2)$. It is symmetric and has a bell shape where most of the values

are near to the mean and follows the 68–95–99.7 rule (a.k.a empirical rule). This rule states that 68.27%, 95.45% and 99.73% of the values lie within one, two and three standard deviations of the mean, respectively:

- PDF: $P(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$.
- Mean is $E(X) = \mu_X$. and Variance is σ_X^2.

7. **Log-Normal Distribution**: It is a *continuous* probability distribution of a random variable whose *logarithm is normally distributed*.Because the values in a log-normal distribution are positive, they create a right-skewed curve which is important in determining which distribution is appropriate to use in investment decision-making. One of the most common applications of log-normal distributions is in the analysis of stock prices. A log-normal distribution can be translated to a normal distribution and vice versa using associated logarithmic calculations.

8. **Uniform Distribution**: It is a probability distribution that has *constant* probability.

  (a) *Discrete Uniform Dist.*: It takes one of a finite n possible values:
    - PMF: $P(X = x) = \frac{1}{n}$
    - $E(X) = \mu_X = \frac{(a+b)}{2}$
    - $\sigma_X^2 = \frac{(b-a+1)^2 - 1}{12}$
  (b) *Continuous Uniform Dist.*: It takes values within a specified range a to b:
    - PDF:
    $$P(X = x) = \begin{cases} \frac{1}{b-a} \ if \ x \in [a,b] \\ 0 \ otherwise \end{cases}$$
    - $E(X) = \mu_X = \frac{1}{2}(a + b)$
    - $\sigma_X^2 = \frac{1}{12}(b - a)^2$

9. **Gamma Distribution**: It is a family of a *continuous*, *positive-only*, *unimodal* distributions that encode the time required for *alpha* events to occur in a Poisson process with mean arrival time of *beta*. These distributions are useful in real-life where something has a natural minimum of 0. For example, it is commonly used

in finance, for elapsed times, or during Poisson processes. In other words, it is a generalization of both the exponential and chi-squared distributions.

10. **Exponential Distribution**: Special case of Gamma Distribution. It is a *continuous* analogue of the geometric distribution and often used to model the time elapsed between events. Formally, it is the probability distribution of the time between events in a process where events occur continuously and independently at a *constant* average rate (i.e. Poisson process):

- PDF: $P(X = x) = \lambda e^{-\lambda x}$
- $E(X) = \mu_X = \lambda^{-1} = \beta$
- $\sigma_X^2 = \lambda^{-2} = \beta^2$

11. **Chi-Square Distribution**: It is a special case of the Gamma distribution. It is a *continuous* distribution of a sum of the squares of k independent standard normal deviates (where k is known as degree of freedom). A standard normal deviate is a *random sample* from the standard normal distribution.

- PDF: $P(X = x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{K}{2})} x^{\frac{k}{2}-1} e^{\frac{-x}{2}}$ where Γ is the gamma function
- $E(X) = \mu_X = k$
- $\sigma_X^2 = 2k$

12. **Student's t-Distribution**: It is a probability distribution used to estimate population parameters when the sample size is small and standard deviation is unknown. It is mainly used by applying Student's t-test for assessing the statistical significance of the difference between two sample means, and in linear regression analysis.

- PDF:
$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$$
where ν is the number of degrees of freedom and Γ is the gamma function.
- $E(X) = \mu_X = 0 \ for \ \nu > 1$, otherwise undefined.
- $\sigma_X^2 = \frac{\nu}{\nu-2} \ for \ \nu > 2, \ \infty \ for \ 1 < \nu \leq 2$, otherwise undefined.