#################################### [**Probability**] ######################################

**Definition of probability**:

* **Classic, exact, objective probability** – ex. play card, dies, I have 4 options, choose 1 which is 25%, need all option to be equally the same
* **Empirical, relative frequency probability** – ex. Look at the historical data (trials in the past) to estimate probability based on portion of many, many trials, 4 out of 10000 died each year, what probability?
* **Subjective probability** – ex. (If no historical data available) assess the probability with guts and without data, how likely a new product will be popular?

**Definition and Rules:**

* **Random Experiment** – Anything that has uncertain outcome is a **random experiment**, weather next Monday, stock price, etc.
  + **Basic** **outcomes –** result from experiment, like 1,3,5 from a die
  + **Sample Space –** Union of all Basic outcomes = **S**
  + **Events –** Subsets of sample space = **A, B, C, …**
* **Probability General Definition –** That’s the chance or the likelihood that an uncertain event will occur, value between 1 and 0, or 0% and 100%. It is also [**Event** / **Sample Space**] = **P(Event)**
* **Fundamental Probability Rules –** 
  + **Rule 1: The probability of an outcome is the sample space is 1. P(S) = 1**
  + **Rule 2: For any event A, the probability of A is between 0 and 1, 0 <= P(A) <= 1**
  + **Rule 3: For disjoint events A and B, P(A union B) = P(A) + P(B)**
* **Complement Rule –** A**n** complement of event A or a set A are all the elements in S that are not in A:  **P(Ac) = 1 – P(A)**
* **General Additional Rule –For joint events A and B, P(A union B) = P(A or B) = P(A) + P(B) – P(A intersect B)**

**Introduction to Independence:**

* **Example –** Roll two faire dies, probability of the sum-up values?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **…** |
| **1/(6\*6)** | **2/(6\*6)** | **3/(6\*6)** | **4/(6\*6)** | **5/(6\*6)** | **6/(6\*6)** | **5/(6\*6)** | **4/(6\*6)** | **3/(6\*6)** |  |

* **1 and 1 – one possible combination / 6 \* 6 total combinations = 1/(6\*6)**
* **There are 11 outcomes to the sample space but each has different probability to occur**
* **So, re-think the sample space – 6\*6 = 36 outcomes**

**A different way to calculate this –**

* First die independent to the second die **(Independent): P(A intersect B) = P(A) \* P(B)**
* **1 and 1 – 1/6 \* 1/6 = 1/(6\*6)**

**Subjective Probabilities:**

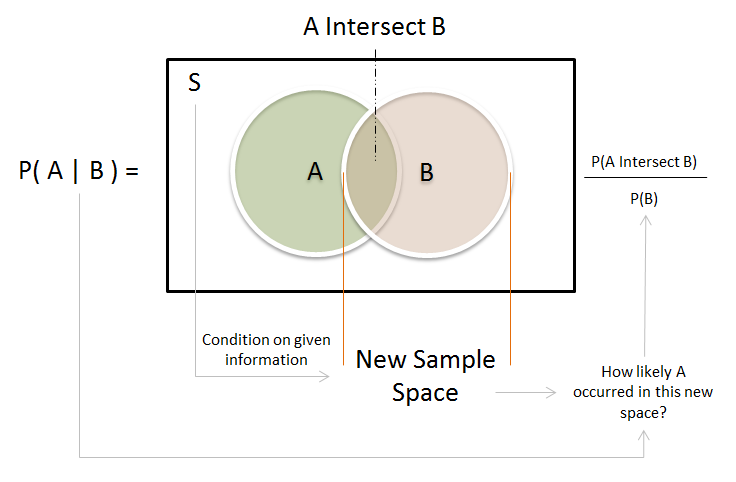
* **Still have to obtain the General Rules, like 0-100%, intersect smaller than either side, etc.**

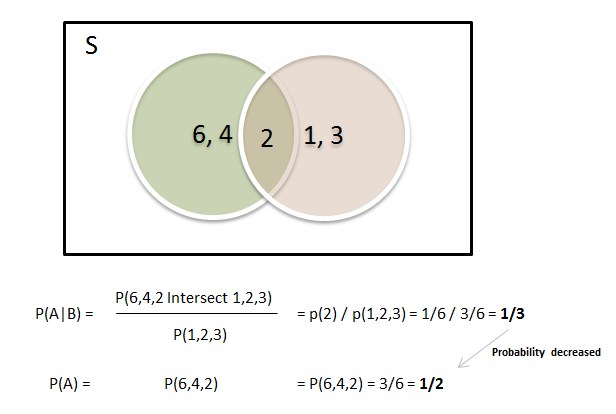
**Empirical Probabilities:**

* **Estimate probability based on the data of real world**
* **Leading Digits: the first digit of a number – 235, 64, 9 = 2, 6, 9**
  + **How often [1-9] shows up in real-world leading digits? Distribution?**
  + **Benford’s Law – P(d) = log10 (1 + 1/d), d – [1-9] which p(1) = 30%, p(2) = 17.6%, …**
  + **Application –** evaluation large dataset of taxpayer’s amount and see if fit this law

**Introduction to Conditional Probabilities:**

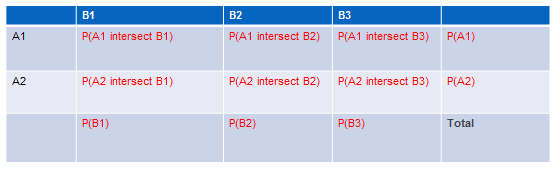
* **Example –** A visit to Switzerland, what is the probability that the person is younger than 20, what is the probability that person speaks French?
  + **General (Empirical) a person – P(Q1) = 21%, P(Q2) = 20.3%**
  + **Given person live in location G – P(Q1|G) = 17%, P(Q2|G) = 3%**
  + **Given person live in location J – P(Q1|J) = 32%, P(Q2|J) = 26%**
* **General Idea:**
  + The probability of event A occurring is **P(A)**
  + Information update**: Event B occurred**
  + What is the probability of Event A occurring given that Event B occurred **P(A|B)**
* **Diagram: Conditional Probability –**

****

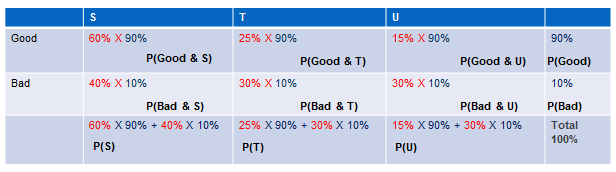
****

**Independence:**

* **If P(A) != P(A|B), Event A dependent on Event B**
* **If P(A) == P(A|B), Event A independent to Event B**
* **General Multiplication Rules:** 
  + **Because Conditional Probability:**
    - **P(A|B) = P(A intersect B) / P(B)**
    - **P(B|A) = P(A intersect B) / P(A)**
  + **Multiplication Rules** *for dependent Events***:** 
    - **P(A intersect B) = P(A|B) X P(B)**
    - **P(A intersect B) = P(B|A) X P(A)**
    - **Example –** *A proportion of a proportion in population***:**
      * **P(B) = 17%** population in Swifts
      * **P(A|B)** = **19%** of Swifts citizeneat ffts
      * So**: P(A intersect B) = 17% \* 19% = 3.45%** population is in Swifts and eat ffts
  + **Multiplication Rules** *for Independent Events***:** 
    - **P(A intersect B) = P(A) X P(B)**
    - **P(A intersect B) = P(B) X P(A)**
    - **Example –** *Rolling three fair Dices and get three ones in a row:*
      * **P(A) = 1/6, P(B) = 1/6, P(C) = 1/6**
      * **P(A intersect B intersect C) = 1/6 ^ 3 = 1/216**
* **Table Probability:**

****

* **Bayes Rule:** 
  + **Example -** 
    - **Suppliers = S, T, U**
    - **P(Good) = 90%, P(Bad) = 10%**
    - **P(S|Good) = 60%, P(S|Bad) = 40%**
    - **P(T|Good) = 25%, P(T|Bad) = 30%**
    - **P(U|Good) = 15%, P(U|Bad) = 30%**
    - **Qs:** Which of my supplier delivers the best parts? Highest **-> P(Good|Suppliers)**

****

* + - **Since: Based on General Multiplication Rule (Above):**
      * **P(A intersect B) = P(A|B) X P(B)**
      * **P(A intersect B) = P(B|A) X P(A)**
    - **So: P(A|B) X P(B) = P(B|A) X P(A)**
    - **Then: P(A|B) = P(B|A) X P(A) / P(B) (Bayes Rule)**
    - **P(Good|S) = P(S|Good) X P(Good) / P(S)**
    - **P(Good|T) = P(T|Good) X P(Good) / P(T)**
    - **P(Good|U) = P(U|Good) X P(Good) / P(U)**

**Random Variables: *(Usually use capital number – XYZ)***

***Probability distribution(PDF): P(X=x)***

***Probability Mass Function -> Discrete Random Variable***

***All single value add up to 1, all positive value***

***Probability Density Function -> Continuous Random Variable***

***All area AUC is 1, all positive values***

***Cumulative Probability distribution(CDF): F(x)***

***Cumulatively add up probability by single values or value range, add up to 1***

* **Discrete Random Variables**
  + **Definition:** Discrete random variables can take on a list of possible values (“finitely many” or “countable infinitely many” values)
  + **Expected Value *(Measure of summarizing):*** 
    - **Why we need? –** Hard to understand random variable so need to summarize
    - **Definition:** Mean (or Expected Value) of a discrete random variable X is the probability-weighted sum of all possible values.
    - **u = E(X) = x1\*P(x1) + x2\*P(x2) + … + xk\*P(xk)**
    - **Interpretation**: If sample from X many, many times, you will expect the average value you got is the **E(X)**
    - **Example – Rolette Game**

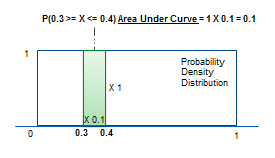
**Win or lost, bid RED – Random variable X [1 win, -1 lost]**

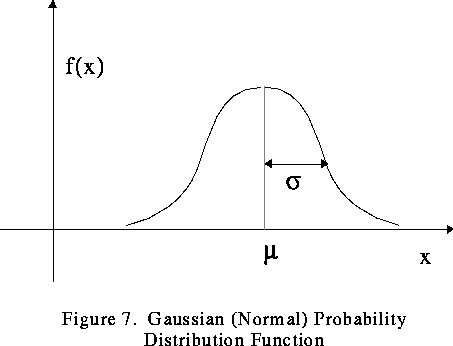
**P(X=1) = 18/37 = 0.486 [18 out of 37 are RED]**

**P(X=-1) = 19/37 = 0.513 [19 out of 37 are BLACK]**

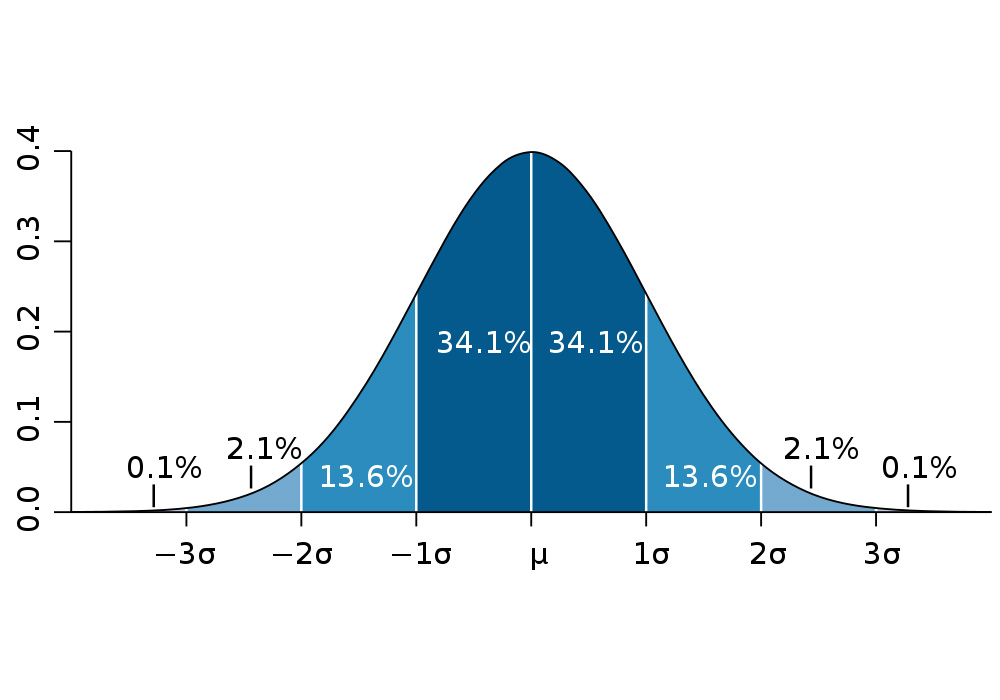
**E(X) = 1 X 18/37 + (-1) X 19/37 = -0.027027 [On average you will lose money] & house advantage**

* **Variance *(Measuring dispersion)\* cause different distribution may has the same E(X)***
  + - **Why we need? –** Averaging will loss of a lot of information, don’t know how distributed.
    - **Definition:**Variance of a discrete random variable is the probability-weighted sum of all possible squared deviations from the mean.
    - **A^2 = Var(X) = X((X-*u)^2)***
    - ***= (x1-u)^2 \* P(x1) + (x2-u)^2 \*P(x2) + … +(xk-u)^2 \* P(xk)***
    - **However, Variance need to translate to real measure (business)**
    - **Standard Deviation:** SD is a measure (based on the variance) for the average deviation of the values of a random variable X from the mean u
    - **SD(X) = sqrt(Var(X))**
  + **Binomial Distribution *Bino(success#, total trail#, p)****:*
    - **Bernoulli Random Variable: only two outcome [success, failure]**
      * **P(success) = P(B=1) = p**
      * **P(failure) = P(B=0) = (1-p)**
      * **Mean – E(B) = u = 1\*p + 0\*(1-p) = p**
      * **Variance – a^2 = (1-p)^2 X p + (0-p)^2 X (1-p) = p(1-p)**
    - **Definition: Sum of many Bernoulli trials – X = B1 + B2 + … + Bn**
      * **Mean – E(X) = E(B1 + B2 + … + Bn)**
      * **= p + p + … + p**
      * **= np**
      * **Variance – Var(X) = Var(B1 + B2 + … + Bn)**
      * **= p(1-p) + p(1-p) + … + p(1-p)**
      * **= np(1-p)**
    - **Example – taste test of two beers (very similar tatste)**
      * **Less 1/3 people favor beer1?**
      * **Bino(33, 100, 0.5) = 0.0043, so smaller than 0.0043 (Very unlikely)**
      * **More than 2/3 people favor beer1?**
      * **Bino(45,100, 0.5) = 0.135, so 1 – 0.135 = 0.86 (Very likely)**
* **Continues Random Variables** 
  + **Definition:** Continues random variable X can take on a continuum of possible values, many real world are continues random variables since it doesn’t jump from value to value like discrete. Ex. thickness, width. \* Sometime it’s easier to model random variable in continues random variable.
  + **There are infinite values so we can only take range of value rather than specific value**
    - **Discrete:** probability with single number
    - **Continues:** Probability with area under curve (AUC)

****

* + **Uniform Distribution: X on [0,1]**
    - P(0 >= X <= 0.5) = 50%
    - P(0.5 >= X <= 1) = 50%
  + **Normal Distribution**:
    - [](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwjvi7Pq_9vUAhVQID4KHdbPC2gQjRwIBw&url=https://www.phy.ornl.gov/csep/mc/node19.html&psig=AFQjCNEH7nMVY19r3HRxicUOYn68Xmlm7A&ust=1498583235889725)

**Normal(u,a^2) – distribution function**

* + - **Central limit theorem:** Pullsamples from a random variable X with sample size n (same variance), calculate metric from each sample size n, all those calculated metrics tends to form a normal distribution (Regardless what the distribution of the random variable)
    - [](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwj89d2cgNzUAhUGeD4KHZeCDHsQjRwIBw&url=https://commons.wikimedia.org/wiki/File:Normal_Distribution_PDF.svg&psig=AFQjCNEH7nMVY19r3HRxicUOYn68Xmlm7A&ust=1498583235889725)
    - [](http://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwjdnqPKgNzUAhXByT4KHQejBRgQjRwIBw&url=http://www.benlcollins.com/spreadsheets/histograms-normal-distribution/&psig=AFQjCNEH7nMVY19r3HRxicUOYn68Xmlm7A&ust=1498583235889725)
    - **Example – Pillar sink test, average 0.8 sink degree, vary by 0.1 sink degree => Q: if get test result 1, how good?**
      * **1 in Normal(0.8, 0.1) ~ 2% which considered large if build a lot of**
    - **Calculation – different dist with different range has the same p()**
      * **Normal(x, u, a)**
      * **Normal(0.7, 0.8, 0.1) = 0.15866**
      * **Normal(-8, 20, 28) = 0.15866**
      * **Here always: (x – u) / a = [failure]**
  + **Standard Normal Distribution – Any normal can be transform into one**
    - **Definition: It has a mean of 0 and standard deviation of 1 – Normal(0,1)**
    - **Z-score (Standard deviation of SND)**
    - **Can use it for any value of normal distributions**