Probablity

Sharique Nawaz

Probablity Vs Statistics

- Probability – Predict the likelihood of a future event
- Statistics – Analyze the past events

- Probability – What will happen in a given ideal world?
- Statistics – How ideal is the world?

Probablity Vs Statistics

Probability is the basis of inferential statistics.

Probability - Applications

8 National Vital Statistics Reports, Vol. 54, No. 14, April 19, 2006

Table 1. Life table for the total population: United States, 2003

Age	Probability of dying between ages x to x+1 $q(_x)$	Number surviving to age x $l(_x)$	Number dying between ages x to x+1 $d(_x)$
0–1	0.006865	100,000	687
1–2	0.000469	99,313	47
2–3	0.000337	99,267	33
3–4	0.000254	99,233	25
4–5	0.000194	99,208	19
5–6	0.000177	99,189	18
6–7	0.000160	99,171	16

Insurance industry uses probabilities in actuarial tables for setting premiums and coverages.

Assigning Probabilities

Classical Method – *A priori* or Theoretical

Probability can be determined prior to conducting any experiment.

$$P(E) = \frac{\#\ of\ outcomes\ in\ which\ the\ event\ occurs}{total\ possible\ \#\ of\ outcomes}$$

Example: Tossing of a fair die

Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Probability can be determined post conducting a thought experiment.

$$P(E) = \frac{\#\ of\ times\ an\ event\ occurred}{total\ \#\ of\ opportunities\ for\ the\ event\ to\ have\ occurred}$$

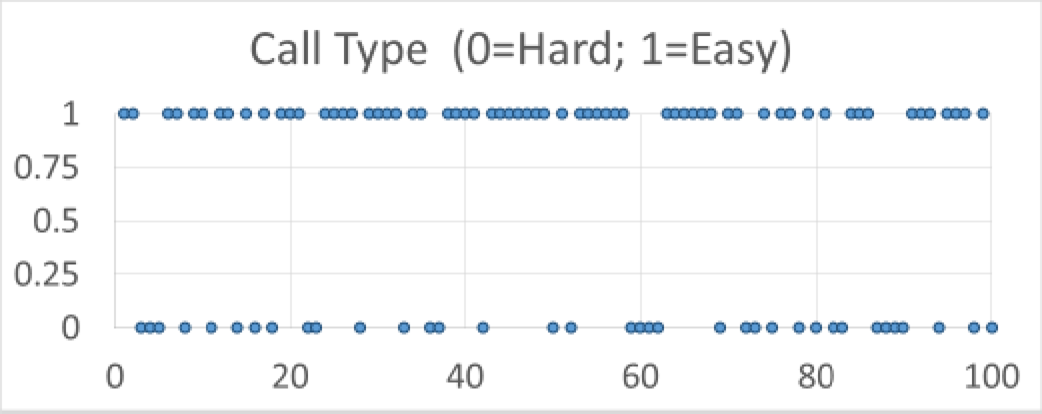
Example: Tossing of a weighted die…well!, even a fair die. The larger the number of experiments, the better the approximation.

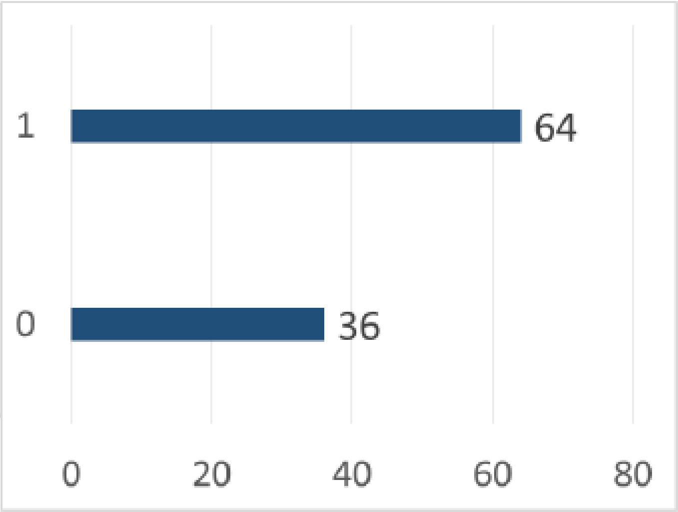
This is the most used method in statistical inference.

Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

100 calls handled by an agent at a call centre

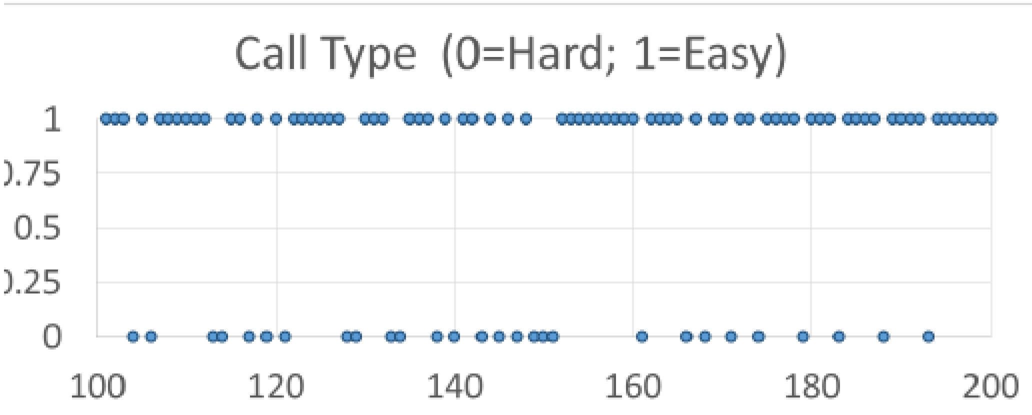


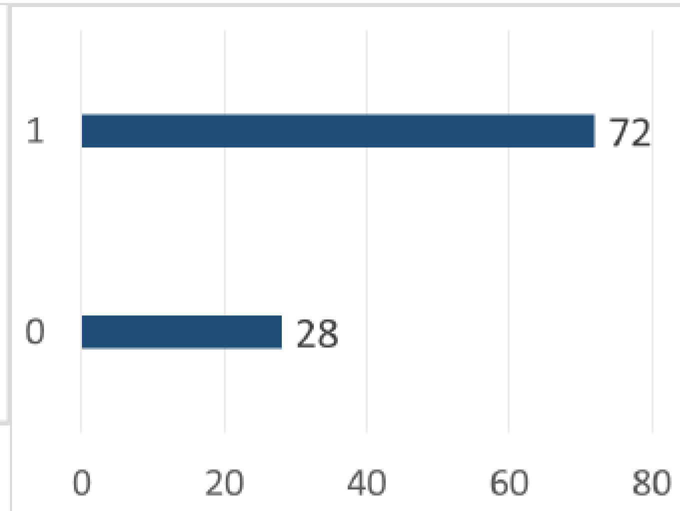


Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Next 100 calls handled by an agent at a call centre

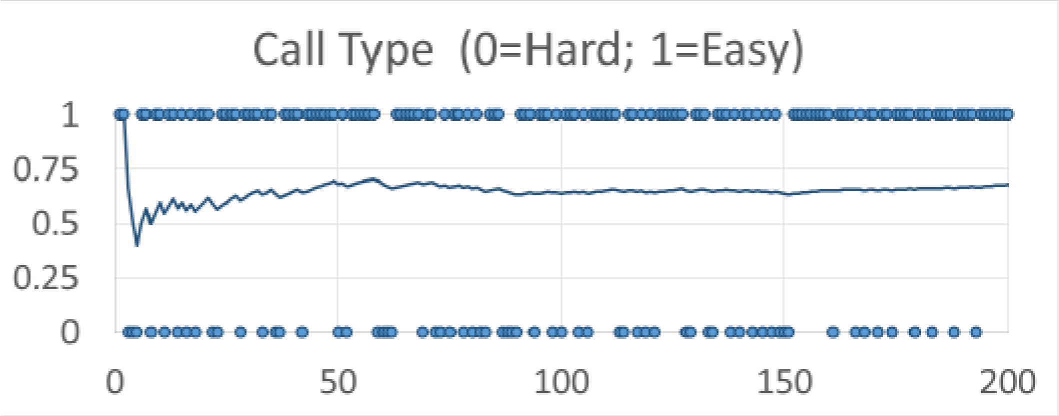


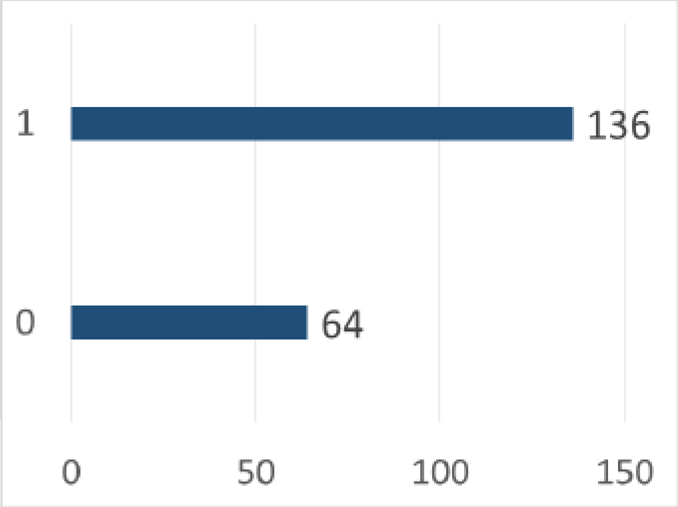


Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Averages over the long run





P(easy) = 0.7

Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Probability of having a monthly income of 1000 BHD is 10/23 = 0.43

INCOME(BHD)	FREQUENCY
100	10
345	1
1000	10
9833	2

Assigning Probabilities

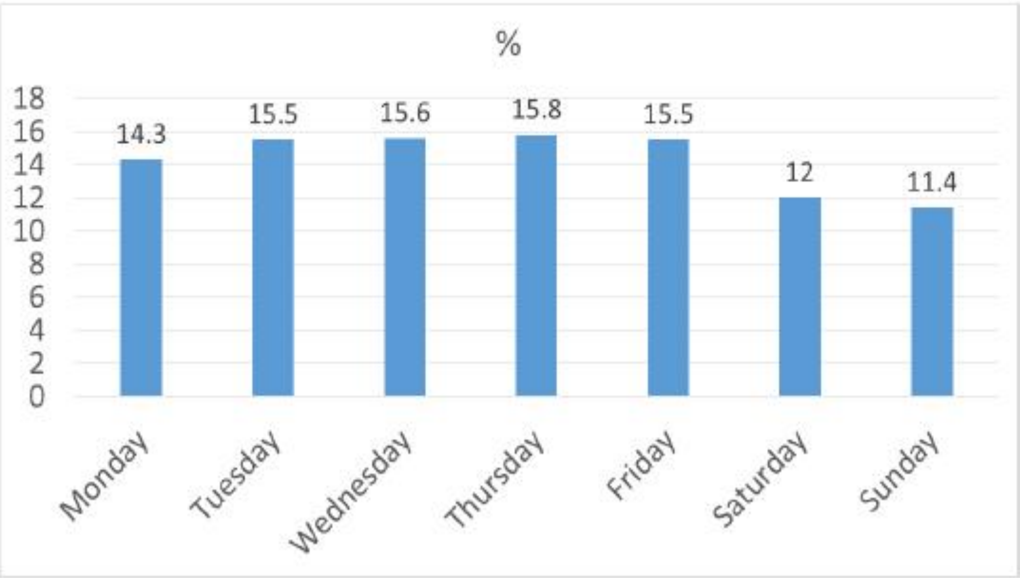
Subjective Method

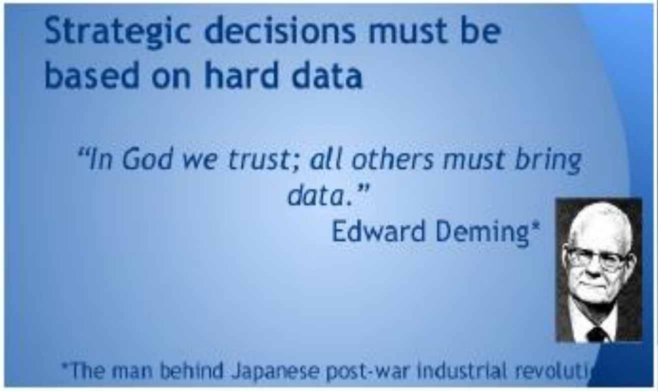
Based on feelings, insights, knowledge, etc. of a person.

What is the probability of rain tomorrow?

Assigning Probabilities

What is the probability of a baby being born on a Sunday?



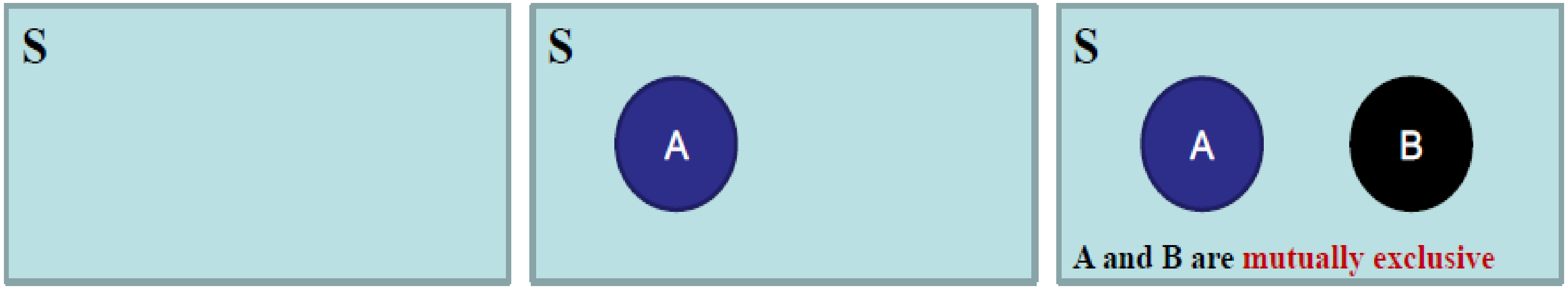


Probability - Terminology

Sample Space – Set of all possible outcomes, denoted S.

Event – A subset of the sample space.

Probability - Rules



$$P(S) = 1$$

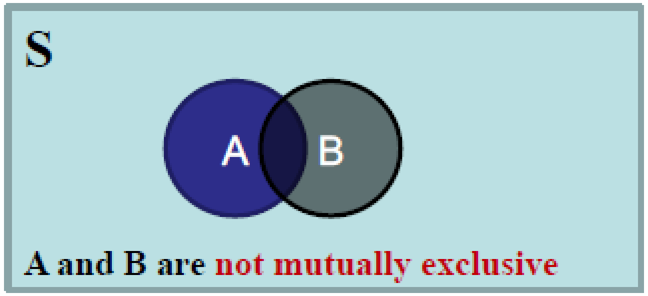
$$0 \leq P(A) \leq 1$$

$$P(A \; or \; B) = P(A) + P(B)$$

Area of the rectangle denotes sample space, and since probability is associated with area, it cannot be negative.

Mutually Exclusive – If event A happens, event B cannot.

Probability - Rules



$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Example

Event A – Customers who default on loans

Event B – Customers who are High Net Worth Individuals

Probability - Rules

Independent Events – Outcome of event B is not dependent on the outcome of event A.

Probability of customer B defaulting on the loan is not dependent on default (or otherwise) by customer A.

$$P(A\ and\ B) = P(A) * P(B)$$

If the probability of getting an *easy* call is 0.7, what is the probability that the next 3 calls will be *easy*?

$$P(easy_1\ and\ easy_2\ and\ easy_3) = 0.7^3 = 0.343$$

Probability - Question

A basketball team is down by 2 points with only a few seconds remaining in the game. Given that:

- Chance of making a 2-point shot to tie the game = 50%
- Chance of winning in overtime = 50%
- Chance of making a 3-point shot to win the game = 30%

What should the coach do: go for 2-point or 3-point shot?

What are the assumptions, if any?

Probability - Types

Contingency table summarizing 2 variables, *Loan Default* and *Age*:

		Age			
		Young	**Middle-aged**	**Old**	**Total**
Loan Default	**No**	10,503	27,368	259	38,130
	Yes	3,586	4,851	120	8,557
	Total	14,089	32,219	379	46,687

Probability - Types

Convert it into probabilities:

		Age			
		Young	**Middle-aged**	**Old**	**Total**
Loan Default	**No**	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

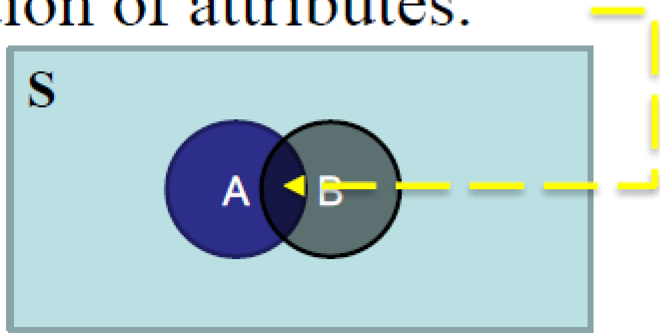
Probability - Types

Joint Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Probability describing a combination of attributes.

P(Yes **and** Young) = 0.077

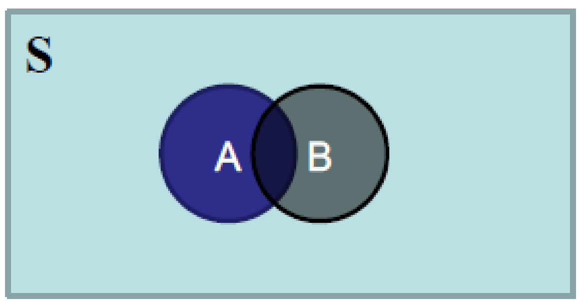


Probability - Types

Union Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

P(Yes **or** Young) = P(Yes) + P (Young) – P(Yes and Young) = 0.184 + 0.302 – 0.077 = 0.409



Probability - Types

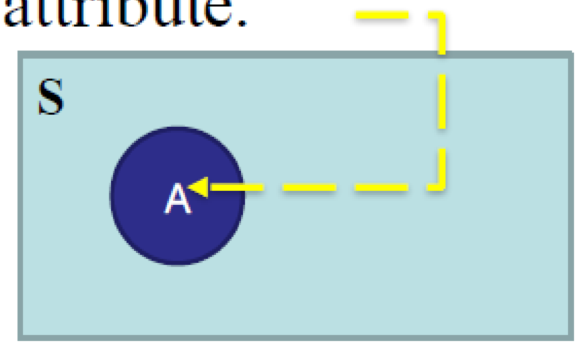
Marginal Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Probability describing a single attribute.

P(No) = 0.816

P(Old) = 0.008



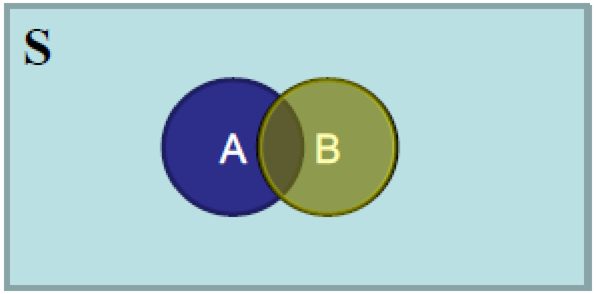
Probability - Types

Conditional Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Probability of A occurring **given that** B has occurred.

The sample space is restricted to a single row or column.
This makes rest of the sample space irrelevant.



Probability - Types

Conditional Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

What is the probability that a person will not default on the loan payment **given** she is middle-aged?

P(No | Middle-Aged) = 0.586/0.690 = 0.85

Note that this is the ratio of **Joint Probability** to **Marginal Probability**, i.e., $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

P(Middle-Aged | No) =

Probability - Types

Conditional Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

What is the probability that a person will not default on the loan payment **given** she is middle-aged?

P(No | Middle-Aged) = 0.586/0.690 = 0.85

Note that this is the ratio of **Joint Probability** to **Marginal Probability**, i.e., $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$

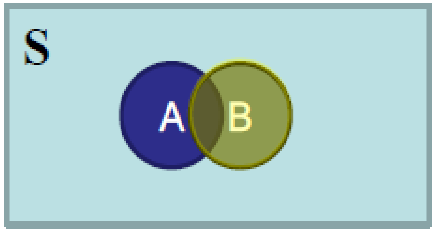
P(Middle-Aged | No) = 0.586/0.816 = 0.72 (<u>Order Matters</u>)

Probability - Types

Conditional Probability – Visualizing using Probability Tables and Venn Diagrams

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	10,503	27,368	259	38,130
	Yes	3,586	4,851	120	8,557
	Total	14,089	32,219	379	46,687

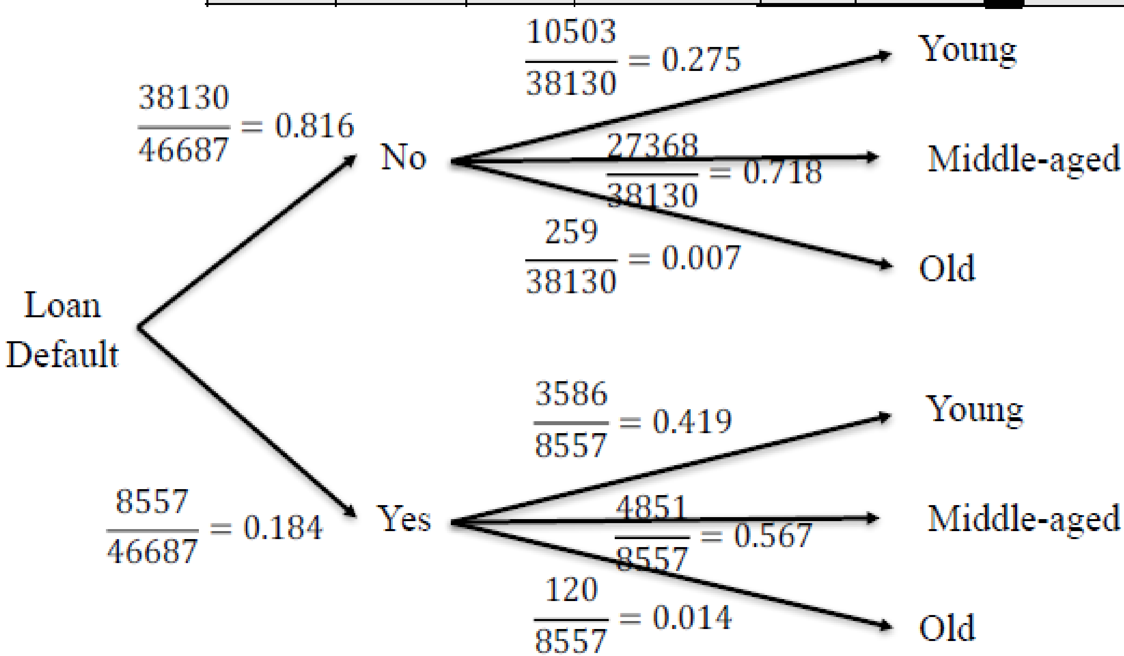
		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000



Probability - Types

Conditional Probability – Visualizing using Probability Trees

		Age (Numbers)			Total	Age (Probabilities)			Total
		Young	Middle-aged	Old	Total	Young	Middle-aged	Old	Total
Loan Default	No	10,503	27,368	259	38,130	0.225	0.586	0.005	0.816
	Yes	3,586	4,851	120	8,557	0.077	0.104	0.003	0.184
	Total	14,089	32,219	379	46,687	0.302	0.690	0.008	1.000



Find

- P(Old and Yes)
- P(Yes and Old)
- P(Old)
- P(Yes)
- P(Old | Yes)
- P(Yes | Old)
- P(Young | No)

Probability - Types

Attention Check

Identify the type of probability in each of the below cases:

1. P(Old and Yes)
2. P(Yes and Old)
3. P(Old)
4. P(Yes)
5. P(Old | Yes)
6. P(Yes | Old)
7. P(Young | No)
8. P(Middle-aged or No)
9. P(Old or Young)

		Age (Probabilities)			
		Young	**Middle-aged**	**Old**	**Total**
Loan Default	**No**	0.225	0.586	0.005	**0.816**
	Yes	0.077	0.104	0.003	**0.184**
	Total	**0.302**	**0.690**	**0.008**	**1.000**

Probability - Types

Attention Check

Identify the type of probability in each of the below cases:

1. P(Old and Yes)
2. P(Yes and Old)
3. P(Old)
4. P(Yes)
5. P(Old | Yes)
6. P(Yes | Old)
7. P(Young | No)
8. P(Middle-aged or No)
9. P(Old or Young)

		Age (Probabilities)			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	**0.816**
	Yes	0.077	0.104	0.003	**0.184**
	Total	**0.302**	**0.690**	**0.008**	**1.000**

1 and 2: **Joint**; 3 and 4: **Marginal**; 5, 6 and 7: **Conditional**; 8 and 9: **Union**

Probability - Types

Conditional Probability

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)} \Rightarrow P(A\ and\ B) = P(B) * P(A|B)$$

Similarly

$$P(B|A) = \frac{P(A\ and\ B)}{P(A)} \Rightarrow P(A\ and\ B) = P(A) * P(B|A)$$

Equating, we get

$$P(A|B) * P(B) = P(A) * P(B|A)$$

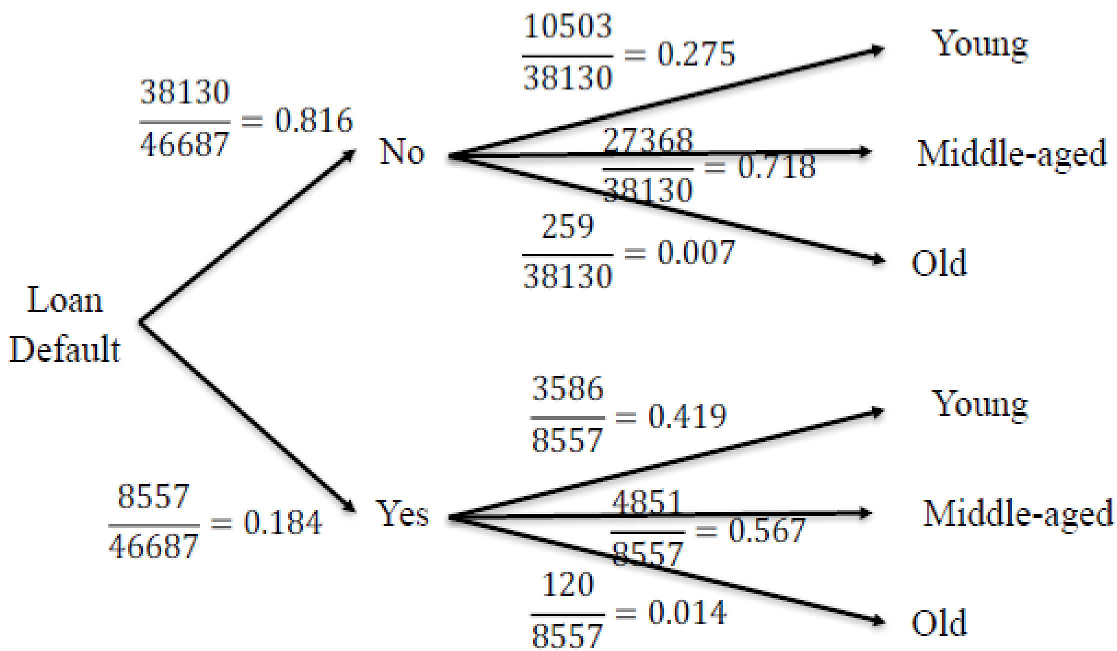
$$\therefore P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

Probability - Types

Conditional Probability – Visualizing using Probability Trees

		Age (Probabilities)			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

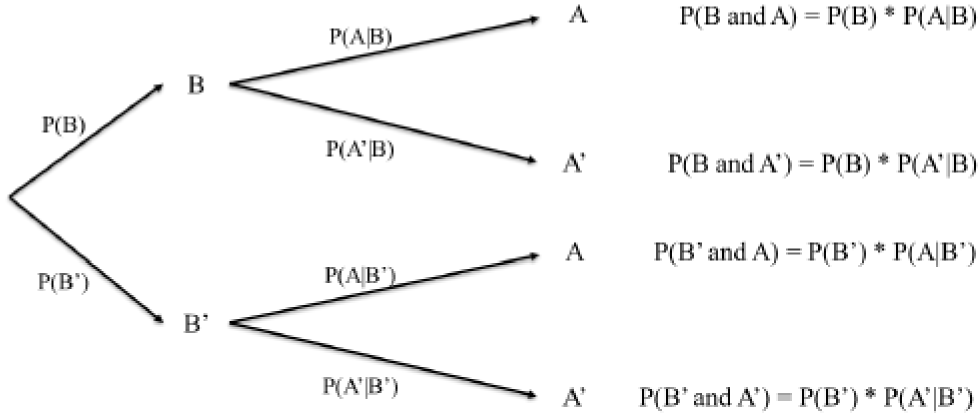
$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$



Now find P(Yes | Old)

Probability - Types

Conditional Probability -> Bayes' Theorem



$$P(B|A) = \frac{P(B) * P(A|B)}{P(A)} = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|not\ B) * P(not\ B)}$$

Note B' means "not B"

Bayes' Theorem

Bayes' Theorem allows you to find reverse probabilities, and to allow **revision of original probabilities** with new information.

Case – Clinical trials

Epidemiologists claim that probability of breast cancer among Caucasian women in their mid-50s is 0.005. An established test identified people who had breast cancer and those that were healthy. A new mammography test in clinical trials has a probability of 0.85 for detecting cancer correctly. In women without breast cancer, it has a chance of 0.925 for a negative result. If a 55-year-old Caucasian woman tests positive for breast cancer, what is the probability that she in fact has breast cancer?

Bayes' Theorem

Case – Clinical trials

P(Cancer) = 0.005

P(Test positive | Cancer) = 0.85 (*aka* Prior Probability)

P(Test negative | No cancer) = 0.925

P(Cancer | Test positive) = ? (*aka* Posterior or Revised Probability)

$$P(Cancer|Test +) = \frac{P(Cancer) * P(Test + |Cancer)}{P(Test + |Cancer) * P(Cancer) + P(Test + |No\ cancer) * P(No\ cancer)}$$

$$= \frac{0.005 * 0.85}{0.85 * 0.005 + 0.075 * 0.995} = \frac{0.00425}{0.078875} = 0.054$$

Homework

Draw a Probability Table and a Probability Tree for the above case.

A slight detour

HOW GOOD IS YOUR CLASSIFICATION?

Confusion Matrix

Spam filtering		Predicted		Total
		Positive	Negative	
Actual	Positive	952	526	1478
	Negative	167	3025	3192
Total		1119	3551	4670

		Predicted		
		Positive	Negative	
Actual	Positive	True +ve	False –ve	Recall/Sensitivity/True Positive Rate (Minimize False –ve)
	Negative	False +ve	True –ve	Specificity/True Negative Rate (Minimize False +ve)
		Precision		Accuracy, F_1 score

Confusion Matrix

Spam filtering		Predicted		Total
		Positive	Negative	
Actual	Positive	952	526	1478
	Negative	167	3025	3192
Total		1119	3551	4670

$$Recall\ (Sensitivity) = \frac{952}{1478} = 0.644$$

$$Precision = \frac{952}{1119} = 0.851$$

$$Accuracy = \frac{952 + 3025}{952 + 3025 + 526 + 167} = \frac{3977}{4670} = 0.852$$

$$Specificity = \frac{3025}{3025 + 167} = \frac{3025}{3192} = 0.948$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2 * 0.851 * 0.644}{0.851 + 0.644} = \frac{1.096}{1.495} = 0.733$$

Which measure(s) is/are more important?

Case Study - Which is the Important measure ?

A Cancer report says that out of 100 patients visiting the hospital 95 patients are non cancerous. If the model is biased and classifies everyone as non cancerous which still means the model is 95% accurate.

Consider

1 - Denotes patients having cancer

0 - Not having cancer

Is this Good model ?

Confusion Matrix

Breast cancer detection		Predicted		Total
		Positive	Negative	
Actual	Positive	852	126	978
	Negative	67	1025	1092
Total		919	1151	2070

$$Recall\ (Sensitivity) = \frac{852}{978} = 0.871$$

$$Precision = \frac{852}{919} = 0.927$$

$$Accuracy = \frac{852 + 1025}{852 + 1025 + 126 + 67} = \frac{1877}{2070} = 0.907$$

$$Specificity = \frac{1025}{1025 + 67} = \frac{1025}{1092} = 0.939$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2 * 0.871 * 0.927}{0.871 + 0.927} = \frac{1.615}{1.798} = 0.898$$

Which measure(s) is/are more important?

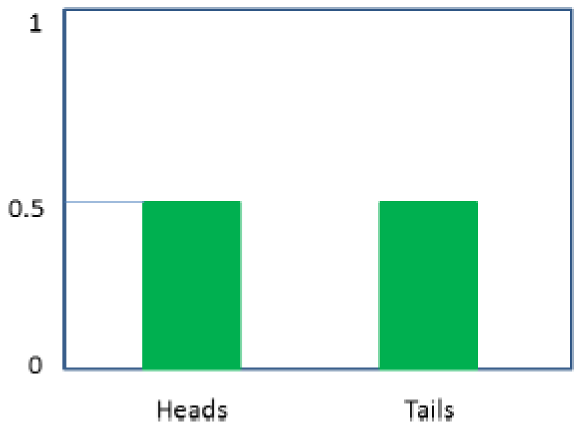
Analyzing attributes

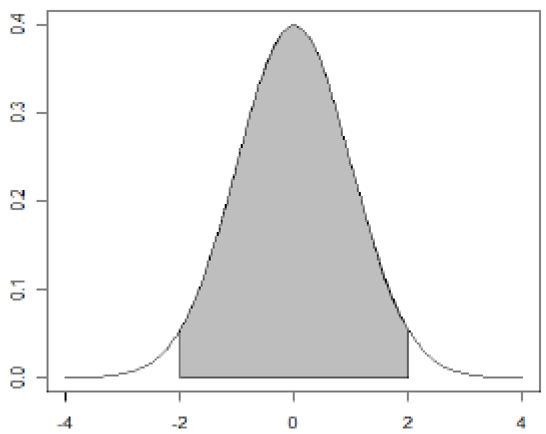
PROBABILITY DISTRIBUTIONS

Random variable

- A variable that can take multiple values with different probabilities.
- The mathematical function describing these possible values along with their associated probabilities is called a probability distribution.

Discrete and Continuous





Countable

Measurable

Can any function be a probability distribution?

Discrete Distributions	Continuous Distributions
Probability that X can take a specific value x is $P(X = x) = p(x)$.	Probability that X is between two points a and b is $P(a \leq X \leq b) = \int_a^b f(x)dx$.
It is non-negative for all real x.	It is non-negative for all real x.
The sum of $p(x)$ over all possible values of x is 1, i.e., $\sum p(x) = 1$.	$\int_{-\infty}^{\infty} f(x)dx = 1$
Probability Mass Function	**Probability Density Function**

Histogram

A series of contiguous rectangles that represent the frequency of data in given class intervals.

How many class intervals?

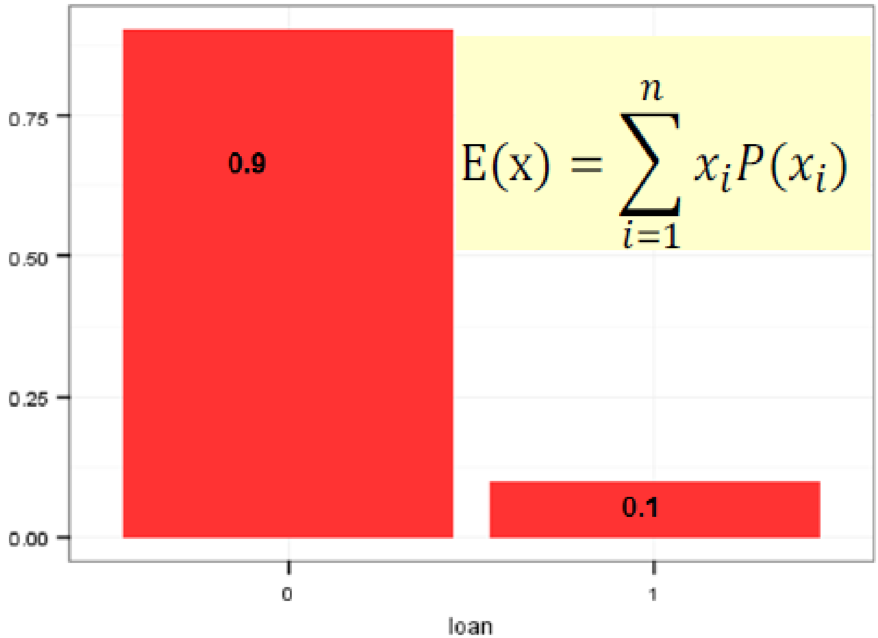
Rule of thumb: 5-15 (not too many and not too few)

Freedman-Diaconis rule:

$$No. of\ bins = \frac{(max - min)}{2 * IQR * n^{\frac{-1}{3}}},$$

$$where\ the\ denominator\ is\ the\ bin - width$$

Expectation: Discrete



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

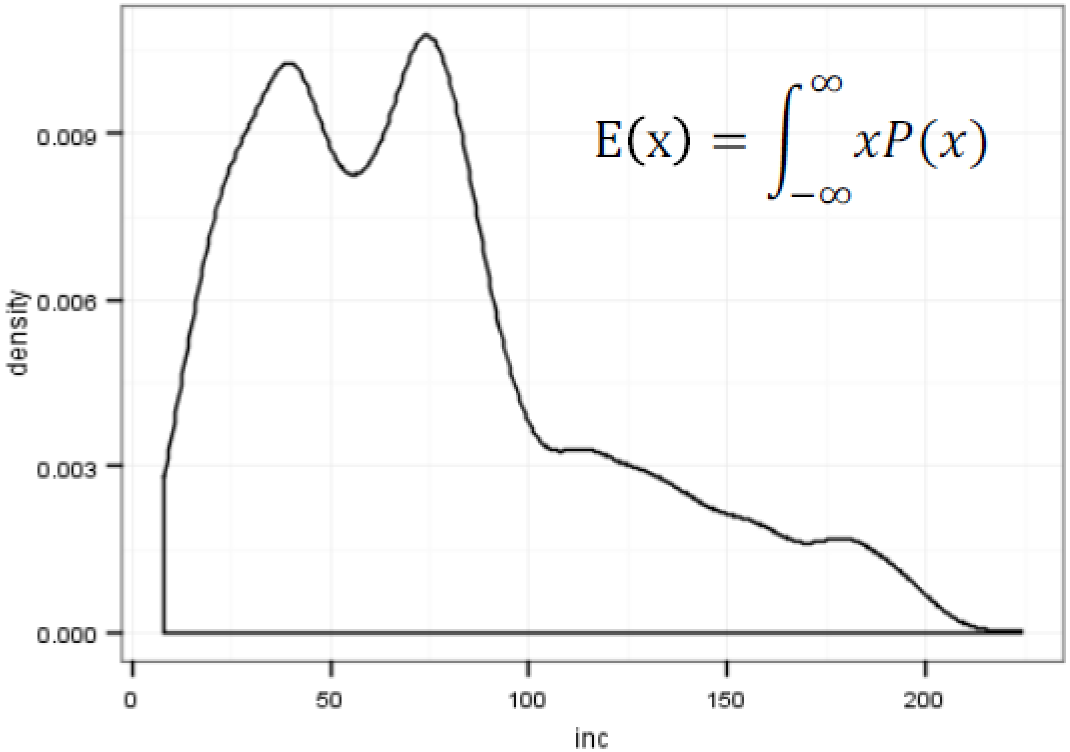
Recall anything like this?

Salary (BHD)	100	345	1000	9833
Frequency, f	10	1	10	2
Probability	0.43	0.04	0.43	0.09

$$\text{Mean, } \mu = \frac{\Sigma x}{n} = \frac{\Sigma f x}{\Sigma f} = \frac{100X10+345X1+1000X10+9833X2}{10+1+10+2} = 1348$$

$$\text{Expectation, } E(X) = 100 * 0.43 + 345 * 0.04 + 1000 * 0.43 + 9833 * 0.09 = 1348$$

Expectation: Continuous



Describing a Distribution – Summary of Moments

Measure	Formula	Description
Mean (μ)	$E(X)$	Measures the centre of the distribution of X
Variance (σ^2)	$E[(X-\mu)^2]$	Measures the spread of the distribution of X about the mean

Simplifying the Formula

$E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2]$

$= E[X^2] - 2\mu E[X] + \mu^2$ (we get this from previous formula as m is just a number)

$= E[X^2] - 2\mu^2 + \mu^2$

$= E[X^2] - \mu^2 = E[X^2] - [E(X)]^2$

Expectation Properties

$E(X+Y) = E(X) + E(Y)$ e.g., Playing a game each on 2 slot machines with different probabilities of winning. This is called Independent Observation.

$E(aX+b) = aE(X)+E(b) = aE(X) + b$ e.g., values x have been changed. This is called Linear Transformation.

If I have a portfolio of 30% TCS, 50% Wipro and 20% Ranbaxy stocks, the expected return of my portfolio is

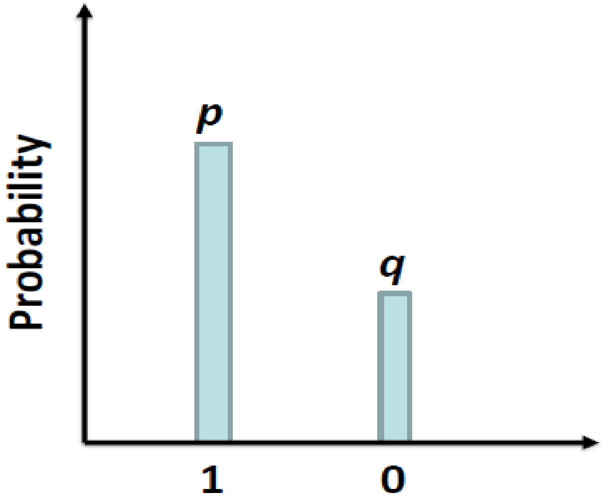
E(Portfolio) = 0.3 E(TCS) + 0.5 E(Wipro) + 0.2 E (Ranbaxy)

SOME COMMON DISTRIBUTIONS

Bernoulli

There are two possibilities (loan taker or non-taker) with probability *p* of success and *1-p* of failure

- Expectation: *p*
- Variance: *p(1-p)* or *pq*, where *q=1-p*



Geometric Distribution

Number of independent and identical Bernoulli trials needed to get ONE success, e.g., number of people I need to call for the first person to accept the loan.

Geometric Distribution

PMF*, $P(X = r) = q^{r-1}p$ — *(r-1)* failures followed by ONE success.

$P(X > r) = q^r$ — Probability you will need more than *r* trials to get the first success.

CDF**, $P(X \leq r) = 1 - q^r$ — Probability you will need *r* trials or less to get your first success.

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{q}{p^2}$$

* Probability Mass Function   ** Cumulative Distribution Function

Geometric Distribution

- You run a series of independent trials.
- There can be either a success or a failure for each trial, and the probability of success is the same for each trial.
- The main thing you are interested in is how many trials are needed in order to get the first successful outcome.

Binomial Distribution

If there are two possibilities with probability *p* for success and *q* for failure, and if we perform *n* trials, the probability that we see *r* successes is

PMF, $\mathrm{P}(\mathrm{X} = \mathrm{r}) = C_r^n p^r q^{n-r}$

CDF, $\mathrm{P}(\mathrm{X} \leq \mathrm{r}) = \sum_{i=0}^{r} C_i^n p^i q^{n-i}$

Binomial Distribution

$$E(X) = np \qquad\qquad Var(X) = npq$$

When to use?

- You run a series of independent trials.
- There can be either a success or a failure for each trial, and the probability of success is the same for each trial.
- There are a finite number of trials, and you are interested in the number of successes or failures.

Poisson Distribution

Probability of getting 15 customers requesting for loans in a given day given on average we see 10 customers

$\lambda = 10 \; and \; r = 15$

PMF, $P(X = r) = \frac{e^{-\lambda}\lambda^{r}}{r!}$

CDF, $P(X \leq r) = e^{-\lambda} \sum_{i=0}^{r} \frac{\lambda^{i}}{i!}$

Poisson Distribution

$E(X) = \lambda$ Can be equated to *np* of Binomial if *n* is large (>50) and *p* is small (<0.1)

$Var(X) = \lambda$ Can be equated to *npq* of Binomial in the above situation.

When to use?

- Individual events occur at random and independently in a given interval (time or space).
- You know the mean number of occurrences, λ, in the interval or the rate of occurrences, and it is finite.

Poisson Distribution

The probability that no customer will visit the store in one day

$$P(X=0) = \frac{e^{-\lambda}\lambda^{0}}{0!} = e^{-\lambda}$$

Probability that she will not have a customer for *n* days

$$e^{-n\lambda}$$

Exponential Distribution

Probability that a customer will visit in *n* days: $1 - e^{-n\lambda}$

$$CDF = 1 - e^{-n\lambda}, n \geq 0$$
$$PDF = \lambda e^{-n\lambda}, n \geq 0$$

Distributions

Geometric:	For estimating number of attempts before first success
Binomial:	For estimating number of successes in *n* attempts
Poisson:	For estimating *n* number of events in a given time period when on average we see *m* events
Exponential:	Time between events

Probability Distributions

Here are a few scenarios. Identify the distribution and calculate expectation, variance and the required probabilities.

Q1. A man is bowling. The probability of him knocking all the pins over is 0.3. If he has 10 shots, what is the probability he will knock all the pins over less than 3 times?

Q2. On average, 1 bus stops at a certain point every 15 minutes. What is the probability that no buses will turn up in a single 15 minute interval?

Q3. 20% of cereal packets contain a free toy. What is the probability you will need to open fewer than 4 cereal packets before finding your first toy?

Probability Distributions

Solutions

A man is bowling. The probability of him knocking all the pins over is 0.3. If he has 10 shots, what is the probability he will knock all the pins over less than 3 times?

X ~ B(10,0.3); n=10, p=0.3, q=1-0.3=0.7, r=0, 1, 2 (< 3)

E(X) = np = 3

Var(X) = npq = 2.1

$$P(X = r) = {}^{n}C_{r}p^{r}q^{n-r}$$

P(X=0) = 0.028; P(X=1) = 0.121; P(X=2) = 0.233

∴ P(X<3) = 0.028 + 0.121 + 0.233 = 0.382

Probability Distributions

Solutions

On average, 1 bus stops at a certain point every 15 minutes. What is the probability that no buses will turn up in a single 15 minute interval?

X ~ Po(1); λ=1, r=0

E(X) = λ = 1

Var(X) = λ = 1

$$P(X = r) = \frac{e^{-\lambda}\lambda^{r}}{r!}$$

P(X=0) = 0.368