Probability and Probability Distribution

Probability and Probability Distribution

- Discrete Probability Distribution
- Continuous Probability Distribution

- Binomial Probability Distribution
- Poisson Probability Distribution
- Normal Probability distribution

Random Variable

- A random variable is a variable whose value is a numerical outcome of a random phenomenon
- Usually denoted by X, Y or Z.
- Can be
  - Discrete - a random variable that has finite or countable infinite possible values
    - Example: the number of days that it rains yearly
  - Continuous - a random variable that has an (continuous) interval for its set of possible values
    - Example: amount of preparation time for the SAT

Discrete Probability distribution

- Binomial Probability Distribution
- Bernoulli Probability Distribution
- Poisson Probability Distribution
- Negative Binomial Probability Distribution
- Hyper Geometric Probability Distribution
- Geometric Probability Distribution

Continuous Probability Distribution

- Uniform Probability Distribution
- Normal or Gaussian Probability Distribution
- Inverse Gaussian Probability Distribution
- Gamma Probability Distribution
- Inverse Gamma Probability Distribution
- Beta Probability Distribution

Probability distribution

The probability distribution for a random variable X gives

- the possible values for X, and
- the probabilities associated with each possible value (i.e., the likelihood that the values will occur)

The methods used to specify discrete prob. distributions are similar to (but slightly different from) those used to specify continuous prob. distributions.

Probability Mass Function

- $f(x)$ - Probability mass function for a discrete random variable X having possible values $x_1, x_2, \cdots$
- $f(x_i) = \Pr(X = x_i)$ is the probability that X has the value x_i
- Properties
  - $0 \leq f(x_i) \leq 1$
  - $\sum_i f(x_i) = f(x_1) + f(x_2) + \cdots = 1$
- $f(x_i)$ can be displayed as a table or as a mathematical function

- Expected Value of X or (population) mean

$$\mu = E(X) = \sum_{i=1}^{R} x_i \Pr(X = x_i) = \sum_{i=1}^{R} x_i f(x_i),$$

where the sum is over R possible values. R may be finite or infinite.

(Population) variance

$$\begin{aligned}\sigma^2 &= Var(X) \\ &= \sum_{i=1}^{R} (x_i - \mu)^2 \Pr(X = x_i) \\ &= \sum_{i=1}^{R} x_i^2 \Pr(X = x_i) - \mu^2\end{aligned}$$

Represents the spread, relative to the expected value, of all values with positive probability

The standard deviation of X, denoted by σ, is the square root of its variance.

Probability Density function

$f(x)$ - Probability density function for a continuous random variable X

Properties

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P[a \leq X \leq b] = \int_a^b f(x)dx$

Important Notes

- $P[a \leq X \leq a] = \int_a^a f(x)dx = 0$
  This implies that $P[X = a] = 0$
- $P[a \leq X \leq b] = P[a < X < b]$

Probability Density function

Summarizations for continuous prob. distributions

- Mean or Expected Value of X

$$\mu = EX = \int_{-\infty}^{\infty} x f(x) dx$$

- Variance

$$\begin{aligned} \sigma^2 &= VarX \\ &= \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - (EX)^2 \end{aligned}$$

Binomial Probability Distribution

- Structure
  - Two possible outcomes: Success (S) and Failure (F).
  - Repeat the situation n times (i.e., there are n trials).
  - The "probability of success," p, is constant on each trial.
  - The trials are independent.

Binomial Probability Distribution

- Let X = the number of S's in n independent trials. (X has values $x = 0, 1, 2, \cdots, n$)
- Then X has a binomial distribution with parameters n and p.
- The binomial probability mass function is

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \cdots, n$$

- Expected Value: $\mu = E(X) = np$
- Variance: $\sigma^2 = Var(X) = np(1-p)$

- Bernoulli Trials
- Relationship between Binomial and Bernoulli distribution

Question

Example: (Moore p.306) Each child born to a particular set of parents has probability 0.25 of having blood type O. If these parents have 5 children, what is the probability that exactly 2 of them have type O blood?

Let X= the number of boys

$$\Pr(X = 2) = f(2) = \binom{5}{2}(.25)^2(.75)^3 = .2637$$

What is the expected number of children with type O blood?

$\mu = 5(.25) = 1.25$

What is the probability of at least 2 children with type O blood?

$$
\begin{aligned}
\Pr(X \geq 2) &= \sum_{k=2}^{5} \binom{5}{k} (.25)^k (.75)^{5-k} \\
&= 1 - \sum_{k=0}^{1} \binom{5}{k} (.25)^k (.75)^{5-k} \\
&= .3671875
\end{aligned}
$$

What is the expected number of children with type O blood?

$\mu = 5(.25) = 1.25$

What is the probability of at least 2 children with type O blood?

$$\begin{aligned}\Pr(X \geq 2) &= \sum_{k=2}^{5} \binom{5}{k} (.25)^k (.75)^{5-k} \\ &= 1 - \sum_{k=0}^{1} \binom{5}{k} (.25)^k (.75)^{5-k} \\ &= .3671875\end{aligned}$$

Poisson Distribution

Definition

The **Poisson distribution** is a discrete probability distribution that applies to occurrences of some event *over a specified interval*. The random variable x is the number of occurrences of the event in an interval. The interval can be time, distance, area, volume, or some similar unit.

Formula

$$P(x) = \frac{\mu^{x} \cdot e^{-\mu}}{x!} \quad \text{where } e \approx 2.71828$$

The Poisson Distribution

The Poisson distribution is defined by:

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}$$

Where *f(x)* is the probability of x occurrences in an interval

μ is the expected value or mean value of occurrences within an interval

e is the natural logarithm. *e* = 2.71828

Properties of Poisson distribution

1. Poisson distribution is a distribution of discrete random variable.
2. In Poisson distribution mean=variance=m. Hence its standard deviation is $\sqrt{m}$ .This is the acid test to be applied to any data which might appear to conform to Poisson distribution.
3. The sum of any finite of independent Poisson variates is itself number a Poisson variate, with mean equal to the sum of the means of those variates taken separately.

Question

- A life insurance salesman sells on the average 3 life insurance policies per week. Use Poisson's law to calculate the probability that in a given week he will sell

a. Some policies
b. 2 or more policies but less than 5 policies.
c. Assuming that there are 5 working days per week, what is the probability that in a given day he will sell one policy?

Solution

Here, $\mu = 3$

(a) "Some policies" means "1 or more policies". We can work this out by finding 1 minus the "zero policies" probability:

$$P(X > 0) = 1 - P(x_0)$$

Now $P(X) = \dfrac{e^{-\mu}\mu^x}{x!}$ so $P(x_0) = \dfrac{e^{-3}3^0}{0!} = 4.9787 \times 10^{-2}$

Therefore the probability of 1 or more policies is given by:

$$\text{Probability} = P(X \geq 0)$$

$$= 1 - P(x_0)$$

$$= 1 - 4.9787 \times 10^{-2}$$

$$= 0.95021$$

$= 0.95021$

(b) The probability of selling 2 or more, but less than 5 policies is:

$P(2 \leq X < 5)$

$$= P(x_2) + P(x_3) + P(x_4)$$

$$= \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!} + \frac{e^{-3}3^4}{4!}$$

$$= 0.61611$$

(c) Average number of policies sold per day: $\frac{3}{5} = 0.6$

So on a given day, $P(X) = \frac{e^{-0.6}(0.6)^1}{1!} = 0.32929$

Normal Probability distribution

- Most widely used continuous distribution
- Also known as the Gaussian distribution
- Symmetric

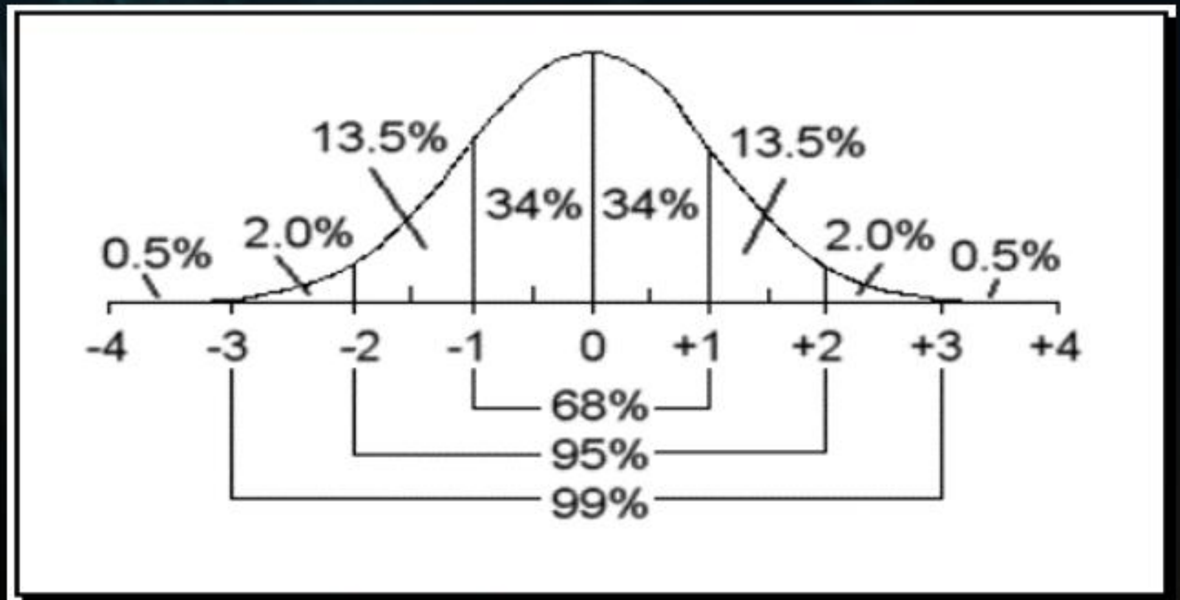
Normal Distribution

- Normal distribution: the standard deviation indicates precisely how the scores are distributed. Empirical rule:
  - About 68% of all scores lie within one standard deviation of the mean. In another word, roughly two thirds of the scores lie between one standard deviation on either side of the mean.

Normal distribution

- About 95% of all scores lie within two standard deviation of the mean (Normal scores: close to the mean).
- About 99.7% of all scores lie within three standard deviation of the mean.

Standard normal distribution or Z distribution



Standard normal distribution or Z distribution

- A normal distribution with mean = 0, and standard deviation = 1.
- A Z score is a value on the x-axis of a standard normal distribution

Normal Probability distribution

- Probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

- $EX = \mu$
- $VarX = \sigma^2$
- Notation: $X \sim N(\mu, \sigma^2)$
  means that X is normally distributed with mean μ and variance σ^2.

Standard Normal Probability distribution

- A normal distribution with mean 0 and variance 1 is called a standard normal distribution.
- Standard normal probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-x^2}{2}\right]$$

- Standard normal cumulative probability function
  Let $Z \sim N(0,1)$

$$\Phi(z) = P(Z \leq z)$$

- Symmetry property

$$\Phi(-z) = 1 - \Phi(z)$$

Standardization of a Normal Random Variable

- Suppose $X \sim N(\mu, \sigma^2)$ and let $Z = \frac{X-\mu}{\sigma}$. Then $Z \sim N(0,1)$.
- If $X \sim N(\mu, \sigma^2)$, what is $P(a < X < b)$?
  - Form equivalent probability in terms of Z :

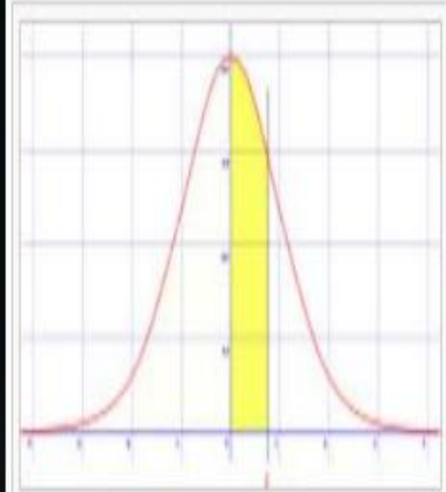
$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

  - Use standard normal tables to compute latter probability.

• Normal table/z table

Cumulative from mean (0 to Z) [edit source | edit beta]

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	1.33
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586	
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535	
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409	
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173	
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793	
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240	
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490	
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524	
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327	
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891	
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214	
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298	
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147	
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774	
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189	
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408	
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449	
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327	



This table gives a probability that a statistic is between 0 (mean) and Z.

Question

- Suppose the distribution of heights of young women are normally distributed with $\mu = 64$ and $\sigma^2 = 2.7^2$ What is the probability that a randomly selected young woman will have a height between 60 and 70 inches?

$$
\begin{aligned}
\Pr(60 < X < 70) &= \Pr\left(\frac{60-64}{2.7} < Z < \frac{70-64}{2.7}\right) \\
&= \Pr(-1.48 < Z < 2.22) \\
&= \Phi(2.22) - \Phi(-1.48) \\
&= .9868 - .0694 \\
&= .9174
\end{aligned}
$$