

Unit 1: Basic Probability

1.	$Probability = P(A) = \frac{No. \text{ of } possible \text{ outcomes}}{Total \text{ no. of outcomes}}$
2.	$Probability \text{ of an impossible event} = P(\phi) = 0$
3.	$P(\bar{A}) = 1 - P(A)$ $P(A) + P(\bar{A}) = 1$ $P(A) \leq 1$
4.	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ <ul style="list-style-type: none"> • If A and B are mutually exclusive events (do not occur simultaneously) then $P(A \cap B) = 0 \therefore P(A \cup B) = P(A) + P(B)$ • If A and independent events then $P(A \cap B) = P(A) P(B)$ $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$ <ul style="list-style-type: none"> • If A, B and C are mutually exclusive events (do not occur simultaneously) then $P(A \cap B) = 0, P(B \cap C) = 0, P(C \cap A) = 0$ and $P(A \cap B \cap C) = 0$ • If A and independent events then $P(A \cap B \cap C) = P(A) P(B) P(C)$ $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ $P(A \cap \bar{B}) = P(A) - P(A \cap B)$ <p>De Morgan's Laws:</p> $P(\bar{A} \cap \bar{B}) = P(\bar{A} \cup \bar{B})$ $P(\bar{A} \cup \bar{B}) = P(\bar{A} \cap \bar{B})$ $P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$ $P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$
5.	<p>Conditional Probability:</p> $P(A/B) = \frac{P(A \cap B)}{P(B)} \therefore P(A \cap B) = P(A/B) P(B)$ $P(B/A) = \frac{P(A \cap B)}{P(A)} \therefore P(A \cap B) = P(B/A) P(A)$
6.	<p>Bayes' Theorem:</p> <p>A_1, A_2, \dots, A_n are mutually exclusive and exhaustive events and B is an event that occurs in combination with any one of the events A_1, A_2, \dots, A_n then</p> $P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^n P(A_i)P(B/A_i)}$
7.	<p>Properties of Probability Mass Function (One Dimensional Discrete):</p> <p>$p(x_i) \geq 0$ for all values of i</p> $\sum_{i=1}^{\infty} p(x_i) = 1$

8.	<p>Properties of Probability Density Function (One Dimensional Continuous):</p> $f(x) \geq 0 \quad -\infty < x < \infty$ $\int_{-\infty}^{\infty} f(x) dx = 1$ $P(a < x < b) = \int_a^b f(x) dx$
9.	<p>Properties of Joint Probability Mass Function (Two Dimensional Discrete):</p> $p_{XY}(x_i, y_j) \geq 0 \text{ for all values of } i \text{ and } j$ $\sum_{j=1}^n \sum_{i=1}^m p_{XY}(x_i, y_j) = 1$ <p>Marginal Probability $p_X(x_i) = \sum_{j=1}^m p(x_i, y_j)$</p> <p>Marginal Probability $p_Y(y_i) = \sum_{i=1}^n p(x_i, y_j)$</p> <p>Conditional Probability $p_{X/Y} = P(X = x/Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$</p> <p>Conditional Probability $p_{Y/X} = P(Y = y/X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$</p> <p>Necessary and sufficient condition for X and Y to be independent is $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$</p>
10.	<p>Properties of Joint Probability Density Function (Two Dimensional Continuous):</p> $f(x, y) \geq 0 \text{ for all } x, y$ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$ <p>Marginal Probability $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$</p> <p>Marginal Probability $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$</p> <p>Conditional Probability $f(x/y) = \frac{f(x, y)}{f_Y(y)}$</p> <p>Conditional Probability $f(y/x) = \frac{f(x, y)}{f_X(x)}$</p> <p>Necessary and sufficient condition for X and Y to be independent is $f(x, y) = f_X(x) f_Y(y)$</p>

Unit 2: Some Special Probability Distributions

<p>1.</p> <p>Binomial Distribution:</p> $P(X = x) = nCx p^x q^{n-x}, x = 0, 1, 2, \dots, n$ <p>$n = \text{no. of trials}$</p> <p>$p = \text{probability of success}$</p> <p>$q = 1 - p = \text{probability of failure}$</p> <p>$\text{mean} = np$</p> <p>$\text{variance} = npq$</p> <p>$SD = \sqrt{npq}$</p>	<p>2.</p> <p>Poisson Distribution:</p> <p><i>Used when n is very large and p is very small.</i></p> $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots, n$ <p>$\lambda = np$</p> <p>$\text{mean} = np = \lambda$</p> <p>$\text{variance} = \lambda$</p> <p>$SD = \sqrt{\lambda}$</p>
<p>3.</p> <p>Normal Distribution:</p> <p><i>Take a new random variable $Z = \frac{X - \mu}{\sigma}$</i></p> <p><i>then use Z – table to find probability</i></p>	
<p>4.</p> <p>Exponential Distribution:</p> $f(x) = \lambda e^{-\lambda x}, x > 0$ <p>$\text{mean} = \frac{1}{\lambda}$</p> <p>$\text{variance} = \frac{1}{\lambda^2}$</p> <p>$SD = \frac{1}{\lambda}$</p>	
<p>5.</p> <p>Gamma Distribution:</p> $f(x) = \frac{\lambda^r}{\Gamma r} x^{r-1} e^{-\lambda x}, x > 0$ <p>$\text{mean} = \frac{r}{\lambda}$</p> <p>$\text{variance} = \frac{r}{\lambda^2}$</p> <p>$SD = \frac{\sqrt{r}}{\lambda}$</p>	

Unit 3: Basic Statistics

Note: If data is given in the form of class then $x_i = \text{middle value of class}$

<p>1.</p> <p>Arithmetic Mean \bar{x} or $\mu = \frac{\sum_{i=1}^n x_i}{n}$</p> <p>→ when frequency is given, Weighted Arithmetic Mean \bar{x} or $\mu = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$</p> <p>Expectation $E(X) = \text{Mean} = \sum_{i=1}^{\infty} x_i p(x_i)$</p> <p>→ Important Results: $E(X + k) = E(X) + k$ $E(aX \pm b) = aE(X) \pm b$ $E(X + Y) = E(X) + E(Y)$ $E(XY) = E(X)E(Y)$ if X and Y are independent random variables.</p>	<p>For median first arrange observations in ascending order.</p> <p>→ If no. of observations = $n = \text{odd}$</p> <p>Median $M = \left(\frac{n+1}{2}\right)^{\text{th}}$ observation</p> <p>→ If no. of observations = $n = \text{even}$</p> <p>Median $M = \frac{\left(\frac{n}{2}\right)^{\text{th}} \text{observation} + \left(\frac{n}{2}+1\right)^{\text{th}} \text{observation}}{2}$</p> <p>→ when frequency is given, Weighted Median $M = l + \frac{\left(\frac{N}{2}\right) - m}{f} \times c$</p> <p>$l = \text{lower limit of median class}$ $m = \text{cumulative frequency upto median class}$ $f = \text{frequency of median class}$ $c = \text{class length of median class}$</p> <p>Median class is a class corresponding to cumulative frequency which is lowest greater than or equal to $\frac{N}{2}$.</p>
<p>3.</p> <p>Mode $Z = \text{Observation which occurs maximum time.}$</p> <p>→ when frequency is given, Weighted Mode $Z = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$</p> <p>$l = \text{lower limit of modal class}$ $f_1 = \text{frequency of modal class}$ $f_0 = \text{frequency of class just above modal class}$ $f_2 = \text{frequency of class just below modal class}$ $c = \text{class length of modal class}$</p> <p>Modal class is a class with highest frequency.</p>	

4.	<p>Standard deviation $\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$</p> <p>→ when frequency is given, Standard deviation $\sigma = \sqrt{\frac{\sum_{i=1}^n f_i(x_i - \bar{x})^2}{\sum_{i=1}^n f_i}}$</p> <p>Standard deviation $\sigma = \sqrt{E(X^2) - [E(X)]^2}$</p>
5.	<p>Variance is square of standard deviation.</p> <p>Variance $\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$</p> <p>→ when frequency is given, Variance $\sigma^2 = \frac{\sum_{i=1}^n f_i(x_i - \bar{x})^2}{\sum_{i=1}^n f_i}$</p> <p>Variance $V(X) = \sigma^2 = E(X^2) - [E(X)]^2$</p> <p>→ Important Results:</p> <p>$V(X + k) = V(X)$</p> <p>$V(aX \pm b) = a^2V(X)$</p> <p>$V(k) = 0$</p> <p>$V(kX) = k^2V(X)$</p>
6.	<p>Mean deviation $M.D. = \frac{\sum_{i=1}^n x_i - \bar{x} }{N}$</p> <p>→ when frequency is given, Mean deviation $M.D. = \frac{\sum_{i=1}^n f_i x_i - \bar{x} }{\sum_{i=1}^n f_i}$</p>
7.	<p>Moments:</p> <p>Central Moments (Moments about actual mean) $\mu_r = \frac{\sum_{i=1}^n f_i(x_i - \bar{x})^r}{\sum_{i=1}^n f_i}$</p> <p>Raw Moments (Moments about arbitrary mean a) $\mu'_r = \frac{\sum_{i=1}^n f_i(x_i - a)^r}{\sum_{i=1}^n f_i}$</p> <p>Moments about origin $v_r = \frac{\sum_{i=1}^n f_i x_i^r}{\sum_{i=1}^n f_i}$</p> <p>$\mu_1 = \mu'_1 - \mu_1' = 0$ $\mu_2 = \mu'_2 - (\mu'_1)^2$ $\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$ $\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4$</p> <p>$v_1 = a + \mu'_1$ $v_2 = \mu_2 + (v_1)^2$ $v_3 = \mu_3 + 3v_1v_2 - 2(v_1)^3$ $v_4 = \mu_4 + 4v_1v_3 - 6(v_1)^2v_2 + 3(v_1)^4$</p>
8.	<p>Karl Pearson's Coeffiecient of Skewness $S_k = \frac{Mean - Mode}{Standard Deviation}$</p> <p>If the mode is ill-defined then $S_k = \frac{3(Means - Median)}{Standard Deviation}$</p>

	<p>S_k lies between -1 and 1. For a positively skewed distribution, $S_k > 0$ For a negatively skewed distribution, $S_k < 0$ For a symmetrical distribution, $S_k = 0$</p> <p>Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ For a positively skewed distribution, $\mu_3 > 0$ For a negatively skewed distribution, $\mu_3 < 0$ For a symmetrical distribution, $\mu_3 = 0$</p>
9.	<p>Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$ For a Leptokurtic, $\beta_2 > 3$ For a Platykurtic, $\beta_2 < 3$ For a Mesokurtic, $\beta_2 = 3$</p>
10.	<p>Correlation:</p> <p>Karl Pearson's Correlation Coefficient $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$</p> <p>Spearman's Rank Correlation Coefficient $r = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$</p> <p>$d$ = Difference between ranks R_1 and R_2 given by two judges N = No. of pairs (contestants)</p>
11.	<p>Regression:</p> <p>Method of least squares: $\rightarrow y = ax + b$ (Regression line of y on x) $\sum y = a \sum x + nb$ $\sum xy = a \sum x^2 + b \sum x$</p> <p>$\rightarrow x = ay + b$ (Regression line of x on y) $\sum x = a \sum y + nb$ $\sum xy = a \sum y^2 + b \sum y$</p> <p>Regression Coefficients:</p> <p>$b_{yx} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$</p> <p>$b_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$</p> <p>Regression line of y on x: $y - \bar{y} = b_{yx}(x - \bar{x})$</p> <p>Regression line of x on y: $x - \bar{x} = b_{xy}(y - \bar{y})$</p> <p>Correlation coefficient $r = \sqrt{b_{xy} \times b_{yx}}$</p> <p>If $b_{xy} > 0, b_{yx} > 0$ then $r > 0$ and if $b_{xy} < 0, b_{yx} < 0$ then $r < 0$</p>

Unit 4: Applied Statistics: Test of Hypothesis

Large Samples: $n > 30$

		Critical Value Z_α			Level of Significance (α)								
		1%		5%		10%							
1.	<i>Two tailed test</i>	$ Z_\alpha = 2.58$		$ Z_\alpha = 1.96$		$ Z_\alpha = 1.645$							
	<i>Right/Left tailed test</i>	$ Z_\alpha = 2.33$		$ Z_\alpha = 1.645$		$ Z_\alpha = 1.28$							
2.	Test of Significance for Single Proportion												
	$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$												
	$p = \text{Sample Proportion}$												
	$n = \text{Sample size}$												
3.	$P = \text{Total population proportion}$												
	$Q = 1 - P$												
	<i>when P and Q are not known, p and q are used.</i>												
	Confidence limits:												
4.	$95\% \text{ confidence limits} = p \pm 1.96 \sqrt{\frac{PQ}{n}}$												
	$99\% \text{ confidence limits} = p \pm 2.58 \sqrt{\frac{PQ}{n}}$												
	Test of Significance for Difference of Proportions												
	$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$												
5.	<i>when P_1 and P_2 are not known, p_1, p_2, q_1 and q_2 are used.</i>												
	Confidence limits:												
	$95\% \text{ confidence limits} = (p_1 - p_2) \pm 1.96 \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$												
	$99\% \text{ confidence limits} = (p_1 - p_2) \pm 2.58 \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$												
6.	Test of Significance for Single Mean												
	$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$												
	$\bar{x} = \text{sample mean}$												
	$\mu = \text{population mean}$												

	<p>When Standard deviation of population σ is not known</p> $Z = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$ <p>s = sample standard deviation</p> <p>Confidence limits:</p> $95\% \text{ confidence limits} = \bar{x} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}} \right)$ $99\% \text{ confidence limits} = \bar{x} \pm 2.58 \left(\frac{\sigma}{\sqrt{n}} \right)$
5.	<p>Test of Significance for Difference of Means</p> $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$ <p>When Standard deviations of population σ_1 and σ_2 are not known</p> $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$ <p>Confidence limits:</p> $95\% \text{ confidence limits} = (\bar{x}_1 - \bar{x}_2) \pm 1.96 \sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$ $99\% \text{ confidence limits} = (\bar{x}_1 - \bar{x}_2) \pm 2.58 \sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$
6.	<p>Test of Significance for Difference of S.D.</p> $Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}\right)}}$ <p>When Standard deviations of population σ_1 and σ_2 are not known</p> $Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}\right)}}$
7.	<p>Small Samples: $n \leq 30$</p> <p><i>t – Test: Test of Significance for Single Mean</i></p> $t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n-1}}\right)} \text{ where } s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \text{ with degree of freedom (df) } v = n - 1$

	<p>Confidence limits:</p> <p>95% confidence limits = $\bar{x} \pm t_{0.05} \left(\frac{s}{\sqrt{n-1}} \right)$</p> <p>99% confidence limits = $\bar{x} \pm t_{0.01} \left(\frac{s}{\sqrt{n-1}} \right)$</p>
	<p>t – Test: Test of Significance for Difference of Means</p> <p>$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ where } s = \sqrt{\frac{\sum(x - \bar{x})^2 + \sum(y - \bar{y})^2}{n_1 + n_2 - 2}}$ or $s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$</p> <p>with degree of freedom (df) $v = n_1 + n_2 - 2$</p>
8.	<p>Confidence limits:</p> <p>95% confidence limits = $(\bar{x} - \bar{y}) \pm t_{0.05} \left(\frac{1}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right)$</p> <p>99% confidence limits = $(\bar{x} - \bar{y}) \pm t_{0.01} \left(\frac{1}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right)$</p>
9.	<p>t – Test: Test of Significance for Correlation Coefficients</p> <p>$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$ with degree of freedom (df) $v = n - 2$</p>
10.	<p>Snedecor's F – Test for Ratio of Variances</p> <p>$F = \frac{S_1^2}{S_2^2}$ where $S_1^2 > S_2^2$</p> <p>$S_1^2 = \frac{\sum(x - \bar{x})^2}{n_1 - 1}$ and $S_2^2 = \frac{\sum(y - \bar{y})^2}{n_2 - 1}$</p> <p>with numerator df $v_1 = n_1 - 1$ and denominator df $v_2 = n_2 - 1$</p> <p>if s_1 and s_2 (Sample SD) are given then $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ and $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$</p>
11.	<p>Chi – Square Test: Goodness of Fit</p> <p>$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$ with degree of freedom (df) $v = n - 1$</p> <p>Chi – Square Test for Independence of Attributes</p> <p>$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$ with df $v = (\text{no. of rows} - 1)(\text{no. of columns} - 1)$</p> <p>$f_e = \frac{(\text{Row total}) \times (\text{Column total})}{\text{Total frequency}} = \frac{(A_i)(B_j)}{N}$</p>

Unit 5: Curve Fitting by Numerical Method

<p>1.</p> <p>Linear Approximation:</p> <ul style="list-style-type: none"> • Equations for the best fitting straight line $y = a + bx$ $\Sigma xy = a\Sigma x + b\Sigma x^2$ $\Sigma y = na + b\Sigma x$ • Equations for the best fitting straight line $y = ax + b$ $\Sigma xy = b\Sigma x + a\Sigma x^2$ $\Sigma y = nb + a\Sigma x$
<p>2.</p> <p>Least Square Approximation:</p> <ul style="list-style-type: none"> • Equations for the best fitting parabola of second degree $y = ax^2 + bx + c$ $\Sigma x^2y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2$ $\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$ $\Sigma y = a\Sigma x^2 + b\Sigma x + nc$ • Equations for the best fitting parabola of second degree $y = a + bx + cx^2$ $\Sigma x^2y = c\Sigma x^4 + b\Sigma x^3 + a\Sigma x^2$ $\Sigma xy = c\Sigma x^3 + b\Sigma x^2 + a\Sigma x$ $\Sigma y = c\Sigma x^2 + b\Sigma x + na$
<p>3.</p> <p>Non-polynomial Approximation OR Non-linear Regression</p> <p>$y = ae^{bx}$ yields</p> <ul style="list-style-type: none"> • Taking Logarithm on both sides $\log y = \log a + bx$ • Denoting $\log y = Y$ and $\log a = A$, the above equation becomes • $Y = A + bx$ which is a straight line. • From above equation A, b can be found & consequently $a = \text{Antilog } A$ can be calculated. <p>$y = ax^b$ yields</p> <ul style="list-style-type: none"> • Taking Logarithm on both sides $\log y = \log a + b \log x$ • Denoting $\log y = Y, \log a = A$ and $\log x = X$ the above equation becomes • $Y = A + bX$ which is a straight line. • From above equation A, b can be found & consequently $a = \text{Antilog } A$ can be calculated. <p>$y = ab^x$ yields</p> <ul style="list-style-type: none"> • Taking Logarithm on both sides $\log y = \log a + x \log b$ • Denoting $\log y = Y, \log a = A$ and $\log b = B$ the above equation becomes • $Y = A + Bx$ which is a straight line. • From above equation A, B can be found & consequently $a = \text{Antilog } A$ and $b = \text{Antilog } B$ can be calculated.