# Stastics

NISHANT
ABHANGI

**Bayes' Theorem**: For two events A and B,

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

where $P(x)$ denotes probability of event $x$. and $P(x|y)$ denotes conditional probability of A given B.

Comments: Bayes' rule tells us how to invert conditional probabilities. In practice, $P(A)$ is often computed using law of total probability.

$$\boxed{P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

if the sample space $\Omega$ is divided into n disjoint events $B_1, B_2, \dots B_n$.

**Poisson Distribution**: Let $\Omega = \{0, 1, 2, 3, 4, \dots\}$

For any $K \in \Omega$ if $\boxed{P(K) = e^{-\lambda}\frac{\lambda^K}{K!}}$ Then the random variable follows a poisson distribution with parameter $\boxed{\lambda}$.

**Normal Distribution**: Parameters: $\mu, \sigma$, Range: $(-\infty, \infty)$

Probability density function: $\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$

Cumulative distribution function $F(x)$ has no formula

$\boxed{\text{Mean}: \mu, \quad \text{Variance}: \sigma^2}$. If $\mu = 0, \sigma = 1$, we get the standard normal distribution: $\boxed{\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}$

N. Normal probabilities: $\boxed{P(-1 \le z \le 1) \approx 0.68, P(-2 \le z \le 2) \approx 0.95 \quad P(-3 \le z \le 3) \approx 0.99}$

Notation for normal distribution: $N(\mu, \sigma)$.

**Central Limit Theorem:** Suppose $X_1, X_2, \ldots X_n$ are independent and identically distributed random variables each having mean $\mu$ and standard deviation $\sigma$. Let

$$\bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}, \quad Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

Then the cumulative ~~density~~ distribution function of $Z_n$ converges to $\Phi(z)$:

$$\boxed{\lim_{n \to \infty} F_{Z_n}(z) = \Phi(z)}$$

(cdf of $\phi(z)$)

$\left(\Phi(z) = P(Z \leq z)\right.$ for standard normal distribution$\left.\right)$

**Correlation:** The correlation coefficient between $X$ and $Y$ is:

$$\boxed{Cor(X, Y) = \rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}}$$

where $Cov(X, Y)$ is the covariance of $X$ and $Y$

$$\boxed{\begin{aligned} Cov(X, Y) &= E\left((X - \mu_X)(Y - \mu_Y)\right) \\ &= E(XY) - \mu_X \mu_Y \end{aligned}}$$

where $E$ denotes expectation value.

**Properties:**

- $\rho$ is the covariance of $\frac{X - \mu_X}{\sigma_X}$ and $\frac{Y - \mu_Y}{\sigma_Y}$
- $\rho$ is dimensionless
- $\boxed{-1 \leq \rho \leq 1}$. Furthermore,

  $\rho = +1 \iff Y = aX + b$ with $a > 0$

  $\rho = -1 \iff Y = aX + b$ with $a < 0$
- $\rho$ measures the linear relationship between $X$ & $Y$.

**Maximum Likelihood Estimates:** Given the data, the maximum likelihood estimate (MLE) for the parameter $p$ is the value of $p$ that maximises the likelihood function $P(\text{data} \mid p)$. If the data is drawn from a normal distribution $N(\mu, \sigma)$, then MLE of $\mu$ is the mean of the data and MLE of $\sigma$ is the variance of the data.

**Bayesian Classifier**: Let $\vec{x} = (x_1, x_2, \dots x_n)$ ~~be~~ represent a data, with $n$ independent variables, ~~by Bayes~~ and let there be $K$ possible hypothesis $\{C_1, C_2, \dots C_K$. Then the Bayes Classifier takes the hypothesis that is most probable, i.e.

$$\boxed{\hat{y} = \underset{K \in \{1, \dots K\}}{\operatorname{argmax}} \; P(C_K) \prod_{i=1}^{n} P(x_i | C_K)}$$

. It is based on the ~~for~~ Bayes theorem:

$$P(\text{hypothesis} \mid \text{data}) = \frac{P(\text{data} \mid \text{hypothesis}) \, P(\text{hypothesis})}{P(\text{data})}$$

i.e., $$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

**Regression (Linear)**: When we fit a line to bivariate data, it is called simple linear regression. The word linear refers to the linear algebraic equations for the unknown parameters $\beta_i$, for example we can also fit polynomials by linear regression.

**Fitting a line using least squares**:

Data: $(x_i, y_i)$. Goal: Find the line $y = \beta_1 x + \beta_0$ such that $\sum_i (y_i - \beta_1 x_i - \beta_0)^2$ is minimized.

We find $$\boxed{\beta_1 = \frac{s_{xy}}{s_{xx}} \qquad \beta_0 = \bar{y} - \beta_1 \bar{x}}$$

$$\boxed{\bar{x} = \frac{1}{n} \sum x_i, \quad \bar{y} = \frac{1}{n} \sum y_i \quad s_{xx} = \frac{\sum (x_i - \bar{x})^2}{n-1}, \; s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}}$$

Note that the response variable $y$ regresses closer to its mean than the predictor variable $x$, i.e.

$$\boxed{\frac{|y-\bar{y}|}{\sqrt{S_{yy}}} \leq \frac{|x-\bar{x}|}{\sqrt{S_{xx}}}}$$

Measuring the fit: The coefficient of determination $R^2$

$$\boxed{R^2 = \frac{TSS - RSS}{TSS}} \qquad \boxed{0 \leq R^2 \leq 1}$$

TSS (Total sum of squares) $= \sum (y_i - \bar{y})^2$

RSS (Residual sum of squares) $= \sum (y_i - \beta_0 - \beta_1 x_i)^2$

A value close to 1 indicates a good fit while a value close to 0 indicates a poor fit.