Uniform distⁿ

X taking values $x_1, \ldots, x_n$.

$P(X = x_i) = \frac{1}{n} \quad \forall i$

Poisson Distⁿ

$\lambda > 0.$

$$f(x) = e^{-\lambda} \frac{\lambda^{x}}{x!}, \quad x = 0, 1, 2, \ldots$$

$$E(X) = \sum_{x=1}^{\infty} x \, e^{-\lambda} \frac{\lambda^{x}}{x!}$$

$$= \lambda \sum_{x=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^{y}}{y!} = \lambda.$$

$$\mathrm{Var}(X) = \sum_{x=1}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2$$

$$= \lambda \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} - \lambda^2$$

$$= \lambda \sum_{y=0}^{\infty} (y+1) \frac{e^{-\lambda}\lambda^{y}}{y!} - \lambda^2$$

$$= \lambda \left\{ \sum_{y=0}^{\infty} y \frac{e^{-\lambda}\lambda^{y}}{y!} + \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} \right\} - \lambda^2$$

$$= \cancel{\lambda^2} + \lambda - \cancel{\lambda^2}$$

$$= \lambda$$

Def^n Given two r.v. X and Y, the covariance of X and Y, denoted by $\mathrm{Cov}(X, Y)$ is given by

$$\mathrm{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Thm Let X, Y be r.vs & a, b constants. Then

$$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X,Y).$$

Pf. Denote $E(X) = \mu_X$ & $E(Y) = \mu_Y$.

$$E(aX+bY) = a\mu_X + b\mu_Y.$$

$$\mathrm{Var}(aX+bY) = E(aX+bY)^2 - (a\mu_X + b\mu_Y)^2$$
$$= E(a^2X^2 + b^2Y^2 + 2abXY) - (a^2\mu_X^2 + b^2\mu_Y^2 + 2ab\mu_X\mu_Y)$$
$$= a^2\left(E(X^2) - \mu_X^2\right) + b^2\left(E(Y^2) - \mu_Y^2\right).$$
$$+ 2ab\left(E(XY) - E(X).E(Y)\right)$$
$$= a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y).$$

Ex. Generalise this to more than 2 r.vs.

Ex. Let X be a binomial r.v. with parameters n and p.

$$X_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } q = 1-p \end{cases}$$

$$X = \sum_{i=1}^{n} X_i \; , \quad E(X) = \sum_{i=1}^{n} E(X_i) = np.$$

$$\mathrm{Var}(X_i) = 1^2 \cdot p + 0 \cdot q - p^2.$$

$$= p(1-p) = pq.$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = npq.$$

Suppose there are n points $(x_1, y_1), (x_2, y_2) \ldots\ldots (x_n, y_n)$ on the plane. Given a +ve integer m, which polyⁿ of degree m is "closest" to these points. Let the desired polyⁿ be

$$p(x) = a + \sum_{i=1}^{m} b_i x^i$$

The polyⁿ that is the "best" is the polyⁿ whose coeffⁿˢ minimizes the following

$$\mathcal{L} = \sum_{i=1}^{n} \left(y_i - p(x_i) \right)^2$$

The polyⁿ that is the "best"

is the polyⁿ whose coeff^ns

minimizes the following

$$\mathcal{L} = \sum_{i=1}^{n} \left(y_i - p(x_i) \right)^2$$

Thm 1. Given n points $(x_1, y_1), \ldots (x_n, y_n)$. the st. line that minimizes $\mathcal{L}$.

has gradient

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}$$

and y-intercept

$$a = \bar{y} - b\bar{x},$$

Where $\bar{x} = \frac{1}{n}\sum x_i$

$\bar{y} = \frac{1}{n}\sum y_i$

Pf Since (a,b) minimizes $\mathcal{L}$, we must have $\dfrac{\partial \mathcal{L}}{\partial b} = 0$

& $\dfrac{\partial \mathcal{L}}{\partial a} = 0$.

We have

$$\mathcal{L} = \sum_{i=1}^{n} \left(y_i - (a + b x_i) \right)^2$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{n} 2 \left(y_i - a - b x_i \right) \left(-x_i \right).$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \quad \text{implies}$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (1)$$

$$na + bn\bar{x} = n\bar{y}$$

$$a + b\bar{x} = \bar{y} \implies \boxed{a = \bar{y} - b\bar{x}}$$

from (1)

$$(\bar{y} - b\bar{x})n\bar{x} + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$b\left[\sum x_i^2 - n\bar{x}^2\right] = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$\therefore\ b = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \Big/ \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

Exponential Regression.

Suppose X and Y are related by an exponential $f^{\underline{n}}$ of the form

$$\boxed{y = ae^{bx}.}$$

$\therefore \log y = \log a + bx.$

Hence the linear relation between x & $\log y$ that is the best fit is given by

$\log y = \log a + bx$, where

$$b = \frac{\sum_{i=1}^{n} x_i \log y_i - \bar{x} \cdot \sum_{i=1}^{n} \log y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\log a = \frac{1}{n}\sum_{i=1}^{n} \log y_i - b\bar{x}$$

Log regression.

$$y = a x^{b}.$$

$$\log y = \log a + b \log x.$$

$$\log y \rightarrow Y \qquad \log x \rightarrow X.$$

Logistic regression.

$$y = \frac{L}{1+e^{a+bx}}, \quad L, a, b \text{ are const.}$$

$$1 + e^{a+bx} = \frac{L}{y}.$$

$$e^{a+bx} = \frac{L-y}{y}.$$

$$a + bx = \log\left(\frac{L-y}{y}\right)$$

$$\Big\downarrow$$

$$Y$$

Correlation coeffient

Let X, Y be r.vs with means μ_X, μ_Y

& S.Ds. σ_x, σ_y respectively

Put $\quad X^* = \dfrac{X - \mu_X}{\sigma_x}$

$$Y^* = \frac{Y - \mu_Y}{\sigma^*}$$

We define the correlation coeff^{n} ρ by

$$\rho = \operatorname{Cov}(X^*, Y^*) = E\left(\frac{X-\mu_X}{\sigma_X} \cdot \frac{Y-\mu_Y}{\sigma_Y} \right)$$

$$- E\left(\frac{X-\mu_X}{\sigma_X} \right) E\left(\frac{Y-\mu_Y}{\sigma_Y} \right)$$

<u>Prop.</u> $|\rho| \leq 1$.

Pf $0 \le E(X^* + Y^*)^2$

$= E(X^{*2}) + E(Y^{*2}) + 2E(X^* \cdot Y^*).$

$= E(X^{*2}) + E(Y^{*2}) + 2\rho + 2E(X^*) \cdot E(Y^*).$

$x \in \mathbb{R}^n$ $\quad x_i = x_i' - \mu$

$y \in \mathbb{R}^n$

$$0 \le \left(\sum_{i=1}^{n} x_i y_i \right)^2 \le \left(\sum_{i=1}^{n} x_i^2 \right) \cdot \left(\sum y_i^2 \right)$$

$X_i - \mu_{X_i}$