Thursday, 3 March 2022 11:59

$$P_1 = (x-\mu_x)\cos\alpha + (y-\mu_y)\sin\alpha$$

$$P_2 = -(x-\mu_x)\sin\alpha + y-\mu_y\cos\alpha.$$

$\vec{X} = \{X_1 \cdots X_n\}$

$$\begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \underbrace{\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}}_{R(\alpha) \equiv \text{Rotation Matrix in 2-dimension.}} \begin{pmatrix} x-\mu_x \\ Y-\mu_Y \end{pmatrix}$$

$\vec{Y} = \{Y_1 \cdots Y_n\}$

$\left.\begin{matrix} P_1 \\ P_2 \end{matrix}\right\}$ Principle axes $\quad \tan(2\alpha) = \dfrac{2\,\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}. \leftarrow$

$$\left\{ \sigma_{1,2}^2 = \frac{\sigma_x^2 + \sigma_y^2}{2} \pm \sqrt{\left(\frac{\sigma_x^2 - \sigma_y^2}{2}\right)^2 + \sigma_{xy}^2}. \right. \Rightarrow \underline{\text{Check}}$$

$\sigma_1^2 \to$ variance of P_1 $\quad Cov(P_1, P_2) = 0.$

$\sigma_2^2 \to$ variance of P_2.

Marginal distribution.

$\int dx$  $= m(y).$

$$m(y|I) = \int_{-\infty}^{+\infty} P(x,y|I)\,dx = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left(\frac{-(Y-\mu_Y)^2}{2\sigma_Y^2}\right)$$

where $I = (\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy})$

$m(y|I)$ depends on $N(\mu_y, \sigma_y)$
$\uparrow$

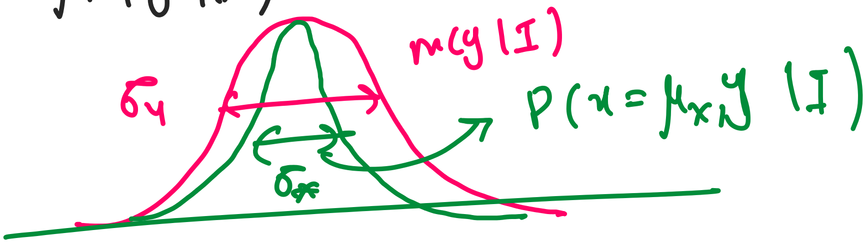
$M(y|I)$ & $p(x,y|I)$

$$P(x=\mu_x, y|I) = \frac{1}{\sigma_x\sqrt{2\pi}}\frac{1}{\sigma_y\sqrt{2\pi}}\exp\left(\frac{-(y-\mu_y)^2}{2\sigma_*^2}\right)$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} N(\mu_y, \sigma_*)$$

where $\sigma_* = \sigma_y\sqrt{1-\rho^2} \le \sigma_y$.

$P(x=\mu_x, y|I)$ is narrower than $m(y|I)$



It means that $M(y|I)$ carries additional uncertainty due to unknown x (marginalized over) x.

Estimates of a Bivariate Gaussian Distribution From Data:

$(\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy})$

$\uparrow$ True values.

(x_i, y_i)

$\swarrow$

$(\bar{x}, \bar{y}, S_x, S_y, S_{xy})$

} Sample values

$$\tan(2\alpha) = \frac{2 S_x S_y}{S_x^2 - S_y^2}\, r \,.$$ } sample values

- working with the real data sets → outliers.

S_x and S_y from the interquartile range

r : Correlation coefficient.

$$r = \frac{V_u - V_w}{V_u + V_w}$$ ← variances of (u, w)

V_u and V_w are the variances of the transformed variables.

$$u = \frac{1}{\sqrt{2}}\left(\frac{x}{\sigma_x} + \frac{y}{\sigma_y}\right) \rightarrow \frac{1}{\sqrt{2}}\left(\frac{x}{S_x} + \frac{y}{S_y}\right)$$

$$w = \frac{1}{\sqrt{2}}\left(\frac{x}{\sigma_x} - \frac{y}{\sigma_y}\right) \rightarrow \frac{1}{\sqrt{2}}\left(\frac{x}{S_x} - \frac{y}{S_y}\right)$$

Estimator for the principal axis angle α.

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \,.$$

Hypothesis testing → r follows t-dist

Multivariate Gaussian Dist

$$\vec{X} = (X_1, X_2, X_3 \cdots, X_n)$$

$$p(x|I) = \frac{1}{(2\pi)^{M/2}\sqrt{\det(C)}} \exp\left(-\frac{1}{2} x^T H x\right)$$

H is a Symmetric Matrix which depends on the inverse of the Covariance Matrix. C^{-1}

C: Covariance Matrix

det: determinant.