By Li Xiaoran
www.sauron.online
xiaoranli@daum.net

Linear algebra-『线性代数』同济
Matrix analysis-『矩阵论』程云鹏
Calculus-『高等数学上下』同济
Statistics- Prof.Yeh Benson 台大

Linear algebra

Eigenvalues and eigenvectors

$$Ax = \lambda x \ then\ \lambda = \text{Eigenvalues};\ x = \text{Eigenvectors}$$

Proof: $|A - \lambda E| = 0$

$Ax - \lambda x = 0$

$(A - \lambda E)x = 0$

1)$cram\ rule\ if\ Ax = 0\ \exists\ |A - \lambda E| = 0$

2)$Ax = b$

$if\ R(A) < R(Ab)\ then\ No\ soution$

$if\ R(A) = R(Ab) = \#(independent\ variable)\ then\ One\ soution$

$if\ R(A) = R(Ab) < \#(independent\ variable)\ then\ Infinite\ soution$

$(A - \lambda E)x = 0 \in Infinite\ soution$

$\therefore\ A - \lambda E < \#(independent\ variable)$

$\therefore |A - \lambda E| = 0$

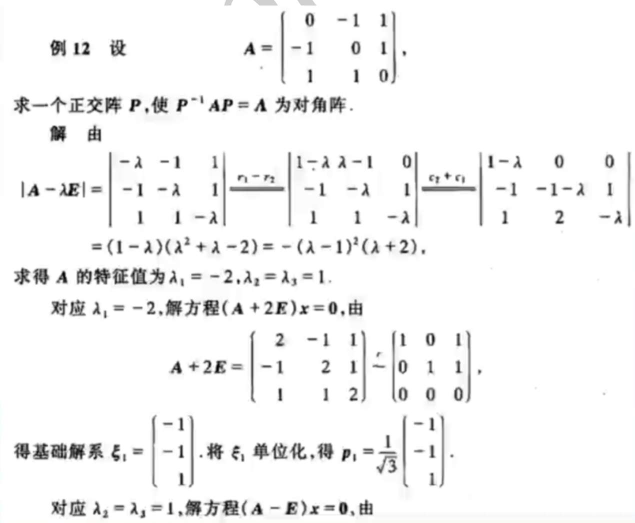
Diagonalization of Symmetric Matrix

0) $\forall matrix\ A\ if\ B \in Right\ triangular\ matrix\ then:\ \ A = P^{-1}BP$

1) $if\ A \in Symmetric\ Matrix\ then\ \Lambda \in \text{Diagonalization matrix}\ then: P^{-1}AP = P^{H}AP = \Lambda$

2) $if\ P \in Unitary\ matrix\ then:\ P^{-1}P = PP^{-1} = \text{E}$

Orthogonalization(P114)



例 12 设 $A = \begin{pmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$,

求一个正交阵 P,使 $P^{-1}AP = \Lambda$ 为对角阵.

解 由

$$|A - \lambda E| = \begin{vmatrix} -\lambda & -1 & 1 \\ -1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} \xlongequal{r_1 - r_2} \begin{vmatrix} 1-\lambda & \lambda-1 & 0 \\ -1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} \xlongequal{c_2 + c_1} \begin{vmatrix} 1-\lambda & 0 & 0 \\ -1 & -1-\lambda & 1 \\ 1 & 2 & -\lambda \end{vmatrix}$$

$$= (1-\lambda)(\lambda^2 + \lambda - 2) = -(\lambda - 1)^2(\lambda + 2),$$

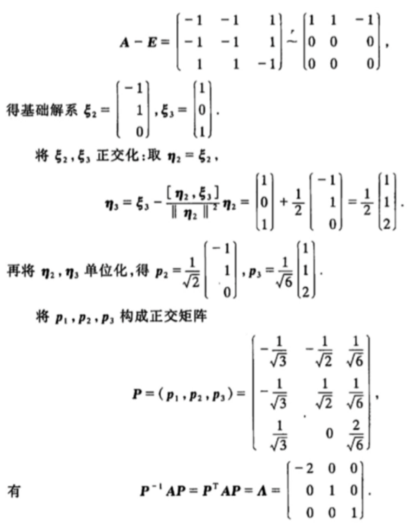
求得 A 的特征值为 $\lambda_1 = -2, \lambda_2 = \lambda_3 = 1$.

对应 $\lambda_1 = -2$,解方程 $(A + 2E)x = 0$,由

$$A + 2E = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \overset{r}{\sim} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

得基础解系 $\xi_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$. 将 ξ_1 单位化,得 $p_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$.

对应 $\lambda_2 = \lambda_3 = 1$,解方程 $(A - E)x = 0$,由



$$A - E = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \overset{r}{\sim} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

得基础解系 $\xi_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \xi_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

将 ξ_2, ξ_3 正交化:取 $\eta_2 = \xi_2$,

$$\eta_3 = \xi_3 - \frac{[\eta_2, \xi_3]}{\|\eta_2\|^2}\eta_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

再将 η_2, η_3 单位化,得 $p_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, p_3 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

将 p_1, p_2, p_3 构成正交矩阵

$$P = (p_1, p_2, p_3) = \begin{pmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix},$$

有 $$P^{-1}AP = P^{T}AP = \Lambda = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Quadratic matrix : $\sum_{i,j=1}^{n} a_{ij}x_i x_j$

xiaoranli@daum.net

$$f = a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n + \cdots + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2 = \sum_{i,j=1}^{n} a_{ij}x_ix_j.$$

$$= (x_1, x_2, \cdots, x_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

例如,二次型 $f = x^2 - 3z^2 - 4xy + yz$ 用矩阵记号写出来,就是

$$f = (x, y, z)\begin{pmatrix} 1 & -2 & 0 \\ -2 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & -3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Matrix analysis

Singular value decomposition (SVD)

1) $if\ \exists\ A^HA\ then\ \lambda \geq 0$

Proof: $A^HA \in Positive\ semi-definite$

$$x^HA^HAx = (Ax)^HAx$$
$$s.t.\ Ax = z$$
$$z^Hz = \left||z|\right|^2 \geq 0$$

2) $rank(A^HA) = rankA$

$if\ Ax = 0$

$then\ \#(rank(A)) + \#(variable\ space\ of\ A) = \#(independent\ variable)$

Proof: $\#(variable\ space\ of\ A) = \#(variable\ space\ of\ A^HA)$

$$\forall Ax = 0 \Longrightarrow A^HAx = 0$$
$$\forall A^HAx = 0 \Longrightarrow x^HA^HAx = 0$$
$$\Longrightarrow (Ax)^HAx = 0$$
$$\Longrightarrow Ax = 0$$

3) $if A = 0\ then\ \ A^HA = 0$

4) $if\ U, V \in Unitary\ matrix\ then\ U^HAV = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$

Proof: $\boldsymbol{U^HAV} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}$ **or** $\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{V^H}$

$if\ V \in Unitary\ matrix\ then\ V^H(A^HA)V = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$

$s.t. V = [V_1, V_2\,]$

$$A^HA[V_1, V_2\,] = [V_1, V_2\,]\begin{bmatrix} \Sigma^2 & O \\ O & O \end{bmatrix} \Longrightarrow [A^HAV_1, A^HAV_2] = [V_1\Sigma^2, 0]$$
$$A^HAV_1 = V_1\Sigma^2$$
$$A^HAV_2 = 0$$

$\therefore$

$$V_1^HA^HAV_1 = \Sigma^2$$
$$(AV_1\Sigma^{-1})^H(AV_1\Sigma^{-1}) = I_r$$

$\therefore$

$$A^HAV_2 = 0 \Longrightarrow (A^HV_2^H)AV_2 = 0 \Longrightarrow AV_2 = 0$$

$if\ U_1 = AV_1\Sigma^{-1}\ \ \exists\ U_1^HU_1 = I_r$

$s.t. U = [U_1, U_2]$

xiaoranli@daum.net

$$U_1^H U_1 = I_r$$
$$U_2^H U_1 = 0$$

∴

$$U^H AV = U^H[AV_1, AV_2]$$
$$= \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix} [U_1\Sigma, O] = \begin{bmatrix} U_1^H U_1 \Sigma & O \\ U_2^H U_1 \Sigma & O \end{bmatrix} = \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix}$$

例 4.14 求矩阵 $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ 的奇异值分解.

解 $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ 的特征值是 $\lambda_1 = 3$，$\lambda_2 = 1$，$\lambda_3 = 0$，对应的特征向量依次为

$$\boldsymbol{\xi}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{\xi}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\xi}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

于是可得

$$\text{rank}\boldsymbol{A} = 2, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

且使得式(4.4.5)成立的正交矩阵为

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

计算

$$\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix}$$

构造

$$\boldsymbol{U}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{U} = [\boldsymbol{U}_1 \vdots \boldsymbol{U}_2] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

则 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{V}^T$$

xiaoranli

Image Compression

$A = U \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} V^H$

$A = \Sigma_1 U_1 V_1^T + \Sigma_2 U_2 V_2^T + \cdots + \Sigma_n U_n V_n^T$

Matrix multiplication acceleration

$A = U \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} V^H$

$s.t.\ A_{200*100} = U_{200*100} \Lambda_{100*100} V_{100*100}^H \overset{SVD}{\Longrightarrow} A_{200*100} = U_{200*10} \Lambda_{10*10} V_{10*100}^H$

$20000th \overset{SVD}{\Longrightarrow} 2000 + 1000 = 3000th$

Multiple linear regression

$$\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$\text{s.t. } \min\|Xa - Y\|^2 = J$

$proof : \dfrac{\partial J}{\partial a} = 0$

$\dfrac{\partial J}{\partial a} = \|Xa - Y\|^2$
$= (Xa - Y)^T(Xa - Y)$
$= (a^TX^T - Y^T)(Xa - Y)$
$= a^TX^TXa - a^TX^TY - Y^TXa + Y^TY$
$= a^TX^TXa - (X^TY)^Ta - Y^TXa + \left||Y|\right|^2$
$= a^TX^TXa - 2Y^TXa + ||Y||^2$
$= (2X^TX)a - 2(Y^TX)^T$
$= X^TXa - X^Ty$

$X^TXa - X^Ty = 0$
$X^TXa = X^TY$
$|A| \neq 0 \propto Fullrank \propto Linearly\ independent \propto Invertible$
1)$N > n$
$if\ N = 5, n = 3\ (X^TX)_{3*3} \in Pseudo\ inverse\ Matrix$
$a = (X^TX)^{-1}x^TY$
2) $N < n$
$if\ N = 3, n = 5\ (X^TX)_{5*5} \in Irrevertible\ Matrix$
$R(X^TX) = R(X) \leq 3$
$Proof :$

$$a^T(X^TX)a = (Xa)^T(Xa) \geq 0 \rightarrow \lambda_i \geq 0$$

Add R2

$$J = \left||Xa - Y|\right|^2 + \lambda\left||a|\right|^2$$

$$\frac{\partial J}{\partial a} = X^TXa - X^Ty + \lambda a = 0$$

$$(X^TX + \lambda I) = X^TY$$

$Proof : X^TX + \lambda I \in Invertible\ Matrix$

$$a^T(X^TX + \lambda I)a = (Xa)^T(Xa) + \lambda a^Ta > 0 \rightarrow \lambda_i > 0$$

$a = (X^TX + \lambda I)^{-1}X^TY$**(Ridge regression)**

$$X^TX = P^{-1}\begin{pmatrix} \lambda_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \lambda_n \end{pmatrix}P$$

$|X^TX| = \lambda_1\lambda_2 \dots \lambda_n$

Calculus

Mean value theorem

$$\frac{f(a) - f(b)}{a - b} = f'(\xi)$$

Cauchy Mean value theorem

xiaoranli@daum.net

$$\frac{f(a)-f(b)}{F(a)-F(b)}=\frac{f'(\xi)}{F'(\xi)}$$

Law of Robida

$Proof:\ \lim_{x\to x_0}\frac{f(x)}{g(x)}=\lim_{x\to x_0}\frac{f'(x)}{g'(x)}$

$s.t.\, f(x_0)=g(x_0)=0$

$$\begin{aligned}\lim_{x\to x_0}\frac{f(x)}{g(x)}&=\lim_{x\to x_0}\frac{f(x)-f(x_0)}{g(x)-g(x_0)}\\&=\lim_{x\to x_0}\frac{(x-x_0)f'(\xi)}{(x-x_0)g'(\xi)}\\&=\lim_{x\to x_0}\frac{f'(\xi)}{g'(\xi)}\end{aligned}$$

Taylor series

$$\sum_{x=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f^2(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^n(x_0)}{n!}(x-x_0)^n+R_n(x)$$

Peano:

$$R_n(x)=o((x-x_0)^n).$$

Lagranga:

$$R_n(x)=\frac{f^{n+1}(x_0)}{(n+1)!}(x-x_0)^{n+1}$$

Maclaurin series:

$$x_0=0$$

Concavity and convexity of function

$$f\left(\frac{x_1+x_2}{2}\right)\leftrightarrow\frac{f(x_1+x_2)}{2}$$
$$f^2(x)\leftrightarrow 0$$

Stationary point

1) $f'(x)=0$

2)

$$if\ x\in(x_0-\delta,x_0), f'(x)>0\ \&\ x\in(x_0,\delta+x_0), f'(x)<0\ then\ maximum=f(x_0)$$
$$if\ x\in(x_0-\delta,x_0), f'(x)<0\ \&\ x\in(x_0,\delta+x_0), f'(x)>0\ then\ minimum=f(x_0)$$

3) $if\ f'(x_0)=0$

$$if\ f^2(x_0)<0\ then\ maximum=f(x_0)$$
$$if\ f^2(x_0)>0\ then\ minimum=f(x_0)$$

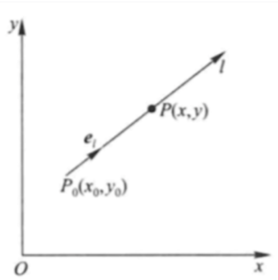
Proof:

xiaorani@daum.net

$$f(x) = f(x_0) + \frac{f^2(x_0)}{2}(x - x_0)^2 + \cdots + R_n(x)$$

$$if\ f^2(x_0) < 0\ then\ \frac{f^2(x_0)}{2}(x - x_0)^2 \le f(x)$$

Directional derivative:



$$\begin{cases} x = x_0 + tcos\alpha \\ y = y_0 + tcos\beta \end{cases}; (t \ge 0)$$

$$\left.\frac{\partial f}{\partial l}\right|_{x_0,y_o} = \lim_{t\to 0}\frac{f(x_0 + tcos\alpha, y_0 + tcos\beta) - f(x_0, y_0)}{t}$$

$$= \lim_{t\to 0}\frac{f(x_0 + tcos\alpha, y_0 + tcos\beta) - f(x_0, y_0 + tcos\beta) + f(x_0, y_0 + tcos\beta) - f(x_0, y_0)}{tcos\alpha}cos\alpha$$

$$= f_x(x_0, y_0)cos\alpha + f_y(x_0, y_0)cos\beta$$

Gradient :

与方向导数有关联的一个概念是函数的梯度. 在二元函数的情形，设函数 $f(x,y)$ 在平面区域 D 内具有一阶连续偏导数，则对于每一点 $P_0(x_0,y_0) \in D$，都可定出一个向量

$$f_x(x_0,y_0)\boldsymbol{i} + f_y(x_0,y_0)\boldsymbol{j},$$

这向量称为函数 $f(x,y)$ 在点 $P_0(x_0,y_0)$ 的梯度，记作 $\mathbf{grad}\ f(x_0,y_0)$ 或 $\nabla f(x_0,y_0)$，即

$$\mathbf{grad}\ f(x_0,y_0) = \nabla f(x_0,y_0) = f_x(x_0,y_0)\boldsymbol{i} + f_y(x_0,y_0)\boldsymbol{j}.$$

如果函数 $f(x,y)$ 在点 $P_0(x_0,y_0)$ 可微分，$\boldsymbol{e}_l = (\cos\alpha, \cos\beta)$ 是与方向 l 同向的单位向量，那么

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = f_x(x_0,y_0)\cos\alpha + f_y(x_0,y_0)\cos\beta$$
$$= \mathbf{grad}\ f(x_0,y_0)\cdot\boldsymbol{e}_l = |\mathbf{grad}\ f(x_0,y_0)|\cos\theta,$$

Multiple Taylor series

$$\begin{aligned} f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) \\ + f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y \\ + f_{xx}(x_0, y_0)\Delta x^2 + 2f_{xy}(x_0, y_0)\Delta x\Delta y + f_{yy}(x_0, y_0)\Delta y^2 \end{aligned}$$

@daum.net

$$
\begin{aligned}
&= f(x_0, y_0) \\
&+ (f_x, f_y)\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix} \\
&+ (\Delta x, \Delta y)\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy}\end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}
\end{aligned}
$$

Multiple Hessian matrix

$$
\begin{aligned}
f(x_1, x_2, \dots, x_n) &= f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) \\
&= f(x_1, x_2, \dots, x_n) \\
&+ (\Delta x_1, \Delta x_2, \dots, \Delta x_n)\begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix} \\
&+ (\Delta x_1, \Delta x_2, \dots, \Delta x_n)\begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1 x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n x_n} \end{pmatrix}\begin{pmatrix}\Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n\end{pmatrix} \\
&= f(x_1, x_2, \dots, x_n) + \Delta x^T \nabla f + \frac{\Delta x^T H \Delta x}{2!}
\end{aligned}
$$

Multiple stationary point

定义 设函数 $z=f(x,y)$ 的定义域为 D，$P_0(x_0,y_0)$ 为 D 的内点. 若存在 P_0 的某个邻域 $U(P_0)\subset D$，使得对于该邻域内异于 P_0 的任何点 (x,y)，都有

$$f(x,y) < f(x_0,y_0),$$

则称函数 $f(x,y)$ 在点 (x_0,y_0) 有极大值 $f(x_0,y_0)$，点 (x_0,y_0) 称为函数 $f(x,y)$ 的极大值点；若对于该邻域内异于 P_0 的任何点 (x,y)，都有

$$f(x,y) > f(x_0,y_0),$$

则称函数 $f(x,y)$ 在点 (x_0,y_0) 有极小值 $f(x_0,y_0)$，点 (x_0,y_0) 称为函数 $f(x,y)$ 的极小值点. 极大值与极小值统称为极值. 使得函数取得极值的点称为极值点.

定理 1（必要条件） 设函数 $z=f(x,y)$ 在点 (x_0,y_0) 具有偏导数，且在点 (x_0,y_0) 处有极值，则有

$$f_x(x_0,y_0)=0,\ f_y(x_0,y_0)=0.$$

定理 2（充分条件） 设函数 $z=f(x,y)$ 在点 (x_0,y_0) 的某邻域内连续且有一阶及二阶连续偏导数，又 $f_x(x_0,y_0)=0, f_y(x_0,y_0)=0$，令

$$f_{xx}(x_0,y_0)=A,\ f_{xy}(x_0,y_0)=B,\ f_{yy}(x_0,y_0)=C,$$

则 $f(x,y)$ 在 (x_0,y_0) 处是否取得极值的条件如下：

（1）$AC-B^2>0$ 时具有极值，且当 $A<0$ 时有极大值，当 $A>0$ 时有极小值；

（2）$AC-B^2<0$ 时没有极值；

（3）$AC-B^2=0$ 时可能有极值，也可能没有极值，还需另作讨论.

Proof:

$$\begin{aligned} f(x,y) &= f(x_0,y_0)+\nabla^T f(x,y)\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}+\frac{1}{2}(x-x_0,y-y_0)\begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} \\ &= f(x_0,y_0)+\frac{1}{2}(\Delta x,\Delta y)\begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \end{aligned}$$

$if : H = Positive\ definite\ then\ u^T M u \geq 0$

$$\exists\, f(x,y) > f(x_0,y_0)$$

$else$:

$$\exists\, f(x,y) < f(x_0,y_0)$$

Proof：$\boldsymbol{Positive\ definite}$

$$\begin{vmatrix} A-\lambda & B \\ B & C-\lambda \end{vmatrix} = (A-\lambda)(C-\lambda)-B^2$$

$$=\lambda^2-(A+C)\lambda+AC-B^2=0$$

$Vieta\ formulas : if\ ax^2+bx+c=0\ then\ x_1+x_2=-\dfrac{b}{a}\ ; x_1x_2=\dfrac{c}{a}$

has a Stationary point: $\begin{cases} Positive\ definite(minimum): \lambda_1>\lambda_2>0 \Longrightarrow \begin{cases} A+C>0 \\ AC-B^2>0 \end{cases} \\ else(maximum): \lambda_1<\lambda_2<0 \Longrightarrow \begin{cases} A+C<0 \\ AC-B^2>0 \end{cases} \end{cases}$

No Stationary point: $C-B^2<0$

Matrix Derivative

1)$f(x)=Ax$

$$\frac{\partial f(x)}{\partial x^T}=\frac{\partial(Ax)}{\partial x^T}=A$$

2)$f(x)=Ax$

$$\frac{\partial f(x)}{\partial x}=\frac{\partial(x^TAx)}{\partial x}=Ax+A^Tx$$

3)$f(x)=Ax$

$$\frac{\partial a^Tx}{\partial x}=\frac{\partial x^Ta}{\partial x}=a$$

4)$f(x)=Ax$

$$\frac{\partial x^TAy}{\partial x}=Ay$$

$$\frac{\partial x^TAy}{\partial x}=xy^T$$

xiaoranli@daum.net

Proof:

$$\frac{\partial\left(tr(ZZ^T)\right)}{\partial Z} = \frac{\partial\left(tr(Z^TZ)\right)}{\partial Z}$$
$$= \sum_{i=1}^{m} C_{ii}$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n} Z_{ij}Z_{ji}^T$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{n} Z_{ij}^2$$
$$= \frac{\partial\left(\sum_{i=1}^{m}\sum_{j=1}^{n} Z_{ij}^2\right)}{\partial Z_{m*n}}$$
$$= 2Z_{m*n}$$

$$\frac{\partial tr(A)}{\partial A} = I_{n*n}$$

$$\frac{\partial |Z|}{\partial Z} = (Z^*)\hat{}T = |Z|(Z^{-1})^*$$

$$Z^*Z = ZZ^* = |Z|I$$

Statistics

1. **Choose with Replacement**

$$n \times n \times n \times \cdots \times n = n^k$$

2. **Permutation**

$$n \times n-1 \times n-2 \times \cdots \times n-(k-1) = \frac{n!}{(n-k)!}$$

3. **Combination**

$$\frac{\text{Permutation}}{\text{k!}} = \binom{n}{k}$$

Binomial theorem:

$$(x+y)^n = \sum\nolimits_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

4. **Independence**

$$P(A \cap B) = P(A) \cdot P(B)$$

From the Bayesian point of view

$$P(A|B) = P(A)$$
$$\frac{P(A \cap B)}{P(B)} = P(A)$$
$$P(A \cup B) = P(A) \cdot P(B)$$

5. **CDF(Cumulative Distribution Function)**

For any random R.V(Random Variable) **X**, we define its CDF as:

$$F_X(x) \stackrel{\text{def}}{=} P(X \leq x)$$

xiaorani@daum.net

The purpose is to calculate the probability of being in a certain range.
When the R.V is discrete:

$$F_X(x^+) = F_X(x)$$
$$F_X(x^-) = F_X(x) - P(X = x)$$

When the R.V is continuous:

$$F_X(x^-) = F_X(x) = F_X(x^+)$$

Common nature:

$$F_X(-\infty) = P(X \le -\infty) = 0$$
$$F_X(\infty) = P(X \le \infty) = 1$$
$$0 \le F_X(x) \le 1$$

6. **PMF(Probability Mass Function)**

For any random R.V **X**, we define its PDF as:

$$p_X(x) \stackrel{\text{def}}{=} P(X = x)$$

Ex:PMF vs CDF

$$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n)$$
$$P_X(x) = F_X(x^+) - F_X(x^-)$$

7. **PDF(Probability Density function)**

$$PDF: f_X(x) = \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x)}{\Delta x}$$
$$= \lim_{\Delta \to 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x}$$
$$= F_X'(x)$$

CDF vs PDF:

$$CDF\ F_X(x) \xrightarrow{\Delta} PDF\ f_X(x)$$

CDF s.t.

$$f_X(x) = F_X'(x)$$
$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$
$$P(a \le X \le b) = \int_a^b f_X(x)dx$$
$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$
$$f_X(x) \ge 0$$

8. **Expectation**

Law of Large Numbers

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

Mean:

$$\sum_{x=1}^{n} \frac{x \cdot N_x}{N}$$

so:

$$\lim_{N\to\infty} mean = \sum_{x=1}^{n} x \cdot \frac{N_x}{N}$$
$$= \sum_{x=1}^{n} x \cdot P(A)$$

For discrete R.V, we define the Expectation as:

$$E[X] = \mu_X = \sum_{x=-\infty}^{\infty} x \cdot P_X(x)$$

The nature of Expectation：

$$E[\alpha] = \alpha$$
$$E[\alpha \cdot g(X)] = \alpha \cdot E[g(X)]$$
$$E[\alpha \cdot g(X) + \beta \cdot h(X)]X = \alpha \cdot E[g(X)] + \beta \cdot E[h(X)]$$

9. **Variance**

$$\sigma_X^2 = E[(X-\mu_X)^2] = \sum_{x=-\infty}^{\infty} (X-\mu_X)^2 \cdot P_X(x)$$
$$= E[X^2 - 2\mu_X \cdot X + \mu_X^2]$$
$$= E[X^2] + E[-2\mu_X X] + E[\mu_X^2]$$
$$= E[X^2] - 2\mu_X \cdot E[X] + \mu_X^2$$
$$= E[X^2] - \mu_X^2$$

10. **Standard Deviation**

$$\sigma_X = \sqrt{Variance}$$

11. **Bernoulli Distribution X~Bernoulli(p)**

Suppose:The probability of success is **P**, do an experiment, X represents the number of successes.

Expectation&Variance

$$\mu_X = 1 \cdot p + 0 \cdot (1-p)$$
$$= p$$
$$\sigma_X^2 = E[X^2] - \mu_X^2$$
$$= \sum_{x=0}^{1} x^2 \cdot p_X(x) - \mu_X^2$$
$$= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2$$
$$= p(1-p)$$

PMF:

$$p_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & therwise \end{cases}$$

CDF:

$$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n)$$
$$= \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

xiaoranli@daum.net

12. **Binomial Distribution X~Binomial(n,p)**
    Suppose:The probability of success is **P**, do **N** experiments, and X represents the number of successes.
    Expectation&Variance

$$\mu_X = np$$
$$\sigma_X^2 = np(1-p)$$

PMF:

$$\begin{aligned} P_X(x) &= P(X = x) \\ &= \binom{n}{x} p^x (1-p)^{n-x} \end{aligned}$$

CDF:

$$\begin{aligned} F_X(x) &= \sum_{m=-\infty}^{\lfloor x \rfloor} P_X(m) \\ &= \sum_{m=-\infty}^{\lfloor x \rfloor} \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m} \end{aligned}$$

13. **Uniform Distribution X~Unif(a,b)**
    Suppose:X=a,a+1⋯,b
    Expectation&Variance

$$\mu_X = \frac{a+b}{2}$$
$$\sigma_{X(\text{discrete})}^2 = \frac{1}{12}(b-a)(b-a+2)$$
$$\sigma_{X(\text{continuous})}^2 = \frac{1}{12}(b-a)^2$$

PMF:

$$p_X(x) = \begin{cases} \dfrac{1}{b-a+1} & x = a, a+1, \dots b \\ 0 & otherwise \end{cases}$$

CDF(discrete):

$$\begin{aligned} F_X(x) &= \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n) \\ &= \begin{cases} 0 & x < a \\ \dfrac{\lfloor x \rfloor - a + 1}{b-a+1} & a \le x < b \\ 1 & x \ge b \end{cases} \end{aligned}$$

CDF(continuous):

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u) du \\ &= \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases} \end{aligned}$$

14. **Geometric Distribution X~Grometric(p)**
    Suppose: If the probability of success of the experiment is **P**, try until it succeeds.

xiaorani@daum.net

Expectation&Variance

$$\begin{aligned}\mu_X &= \sum_{x=0}^{\infty} x \cdot p_X(x) \\ &= \sum_{x=0}^{\infty} x \cdot (1-p)^{x-1} \cdot p \\ &= \frac{1}{p} \\ \sigma_X^2 &= E[X^2] - \mu_X^2 \\ &= \frac{1-p}{p^2}\end{aligned}$$

PMF：

$$p_X(x) = \begin{cases} (1-p)^x - 1 \cdot p & x = 1,2,3 \ldots \\ 0 & otherwise \end{cases}$$

CDF:

$$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n)$$

$$= \begin{cases} x \geq 1: & \sum_{n=1}^{\lfloor x \rfloor} (1-p)^{n-1} p = p \cdot \dfrac{1-(1-p)^{\lfloor x \rfloor}}{1-(1-p)} \\ x < 1: & 0 \end{cases}$$

$$= \begin{cases} 1-(1-p)^{\lfloor x \rfloor} & x \geq 1 \\ 0 & otherwise \end{cases}$$

15. **Pascal Distribution X~Pascal(k,p)**

Suppose :If the experiment success rate is P, try until the kth success.

Expectation&Variance

$$\mu_X = \frac{k}{p}$$

$$\sigma_X^2 = \frac{k(1-p)}{p^2}$$

PMF:

$$p_X(x) = \begin{cases} \binom{x-1}{k-1} (1-p)^{x-k} \cdot p^k & x = k, k+1, \ldots \\ 0 & otherwise \end{cases}$$

CDF:Pascal as Negative Binomial

$$\begin{aligned}F_X(x) &= P(X \leq x) \\ &= P(Y \geq k), Y \sim BIN(x,p)\end{aligned}$$

16. **Poisson Distribution Poi(λT)**

Suppose:It is known that something happens at a rate of **λ** times per unit time. The time unit is **T**. X is the total number of occurrences during the observation period.

Expectation&Variance Poi(α)

$$\begin{aligned}\mu_X &= \sum_{x=0}^{\infty} x \cdot p_X(x) \\ &= \sum_{x=0}^{\infty} x \cdot \frac{\alpha^x}{x!} e^{-\alpha}\end{aligned}$$

xiaoranli@daum.net

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

$$\because \sum_{x=0}^{\infty} \frac{\alpha^x}{x!} e^{-a} = 1 \quad \therefore = \alpha \sum_{x=1}^{\infty} \frac{\alpha^{x-1}}{(x-1)!} e^{-\alpha}$$

$$Let: x' = x - 1 \quad = \alpha \sum_{x'=x-1=1-1=0}^{\infty-1=\infty} \frac{\alpha^{x'}}{x'!} e^{-\alpha}$$

$$= \alpha \cdot 1$$

$$= \alpha$$

$$\sigma_X^2 = \alpha$$

PMF:

$$p_X(x) = P(X = x) = e^{-\lambda T} \cdot \frac{(\lambda T)^x}{x!}$$

$$※\mu = \lambda T, X\sim POI(\mu) \Longrightarrow P_X(x) = e^{-\mu} \cdot \frac{\mu^x}{x!}$$

CDF:

$$F_X(x) = \sum_{n=-\infty}^{\lfloor x \rfloor} p_X(n)$$

$$= \begin{cases} \sum_{n=-\infty}^{\lfloor x \rfloor} e^{-\mu} \cdot \frac{\mu^n}{n!} & x = 1,2,\ldots \\ 0 & otherwise \end{cases}$$

17. **Exponential Distribution X~Exponential(λ)**
    Expectation&Variance

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$= \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx$$

$$= -\int_0^{\infty} -x \cdot \lambda e^{-\lambda x} dx$$

$$= -\int_0^{\infty} x \cdot e^{-\lambda x} d(-\lambda x)$$

$$= -\int_0^{\infty} x d e^{-\lambda x}$$

$$= -\left[x e^{-\lambda x} \Big|_0^{\infty} - \int_0^{\infty} e^{-\lambda x} dx \right]$$

$$= -\left[0 - 0 - \int_0^{\infty} e^{-\lambda x} dx \right]$$

$$= \int_0^{\infty} e^{-\lambda x} dx$$

$$= \frac{1}{\lambda} \int_0^{\infty} \lambda e^{-\lambda x} dx$$

$$\sigma_X^2 = \frac{1}{\lambda^2}$$

PDF:

xiaoranli@daum.net

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

CDF:

$$\begin{aligned} &if\ x \geq 0: \\ &F_X(x) = \int_{-\infty}^{x} f_X(u)du \\ &\quad\quad = \int_{-\infty}^{x} \lambda e^{-\lambda u}du \\ &\quad\quad = -\int_{-\infty}^{x} e^{-\lambda u}d(-\lambda u) \\ &\quad\quad = -\left[e^{-\lambda u}\right]_0^x \\ &\quad\quad = 1 - e^{-\lambda u} \\ &if\ x < 0: F_X(x) = 0 \end{aligned}$$

18. **Erlang Distribution X~Erlang(n, λ)**
Expectation&Variance

$$\mu_X = \frac{n}{\lambda}$$

$$\sigma_X^2 = \frac{n}{\lambda^2}$$

PDF:
Gamma Distribution

$$f_X(x) = \begin{cases} \frac{1}{(n-1)!}\lambda^n x^{n-1} e\hat{}(-\lambda x) & x \geq 0 \\ 0 & otherwise \end{cases}$$

Nth convolution

$$f_X(x) = (\lambda e^{-\lambda x}) * (\lambda e^{-\lambda x}) * \cdots * (\lambda e^{-\lambda x})$$

CDF:

$$F_X(x) = \begin{cases} 1 - \sum_{k=1}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

The distribution of time required in each level: **X~Exponential(λ)**
The distribution of time required in **n** levels: **X~Erlang(n, λ)**

19. **Normal/Gaussian Distribution X~Gaussian(μ,σ)/X~(μ,σ^2)**
Expectation&Variance

$$\mu_X = \mu$$
$$\sigma_X^2 = \sigma^2$$

PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Standard Normal Distribution Z~N(0,1)
PDF:

xiaoranli@daum.net

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

CDF: (https://en.wikipedia.org/wiki/Standard_normal_table)

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

X~N(μ,σ^2) to Z~N(0,1):

$$F_X(x) = \Phi(\frac{x-\mu}{\sigma})$$

PROOF:

Let:

$$\omega = \frac{x-\mu}{\sigma}$$

$$\begin{aligned} P\left(\frac{x-\mu}{\sigma} \le z\right) &= P(X \le \mu + \sigma z) \\ &= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{\omega^2}{2}} d\omega \\ &= \Phi(z) \end{aligned}$$

20. **Gaussian Distribution MLE**

Suppose:

$$\begin{aligned} &data: X = (x_1, x_2, x_3 \cdots x_n)^T_{N\times p} \\ &x_i \in \mathbb{R}^p \\ &x_i: \overset{iid}{\sim} N(\mu, \sigma) \\ &\theta = (\mu, \sigma) \end{aligned}$$

MLE:

$\boldsymbol{let: p = 1, \theta = (\mu, \sigma^2)}$

$$\begin{aligned} \log P(X|\theta) &= \log \prod_{i=1}^{N} p(x_i|\theta) \\ &= \sum_{i=1}^{N} \log p(x_i|\theta) \\ &= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \sum_{i=1}^{N} \left[\log \frac{1}{\sqrt{2\pi}} + \log \frac{1}{\sigma} - \frac{(x_i-\mu)^2}{2\sigma^2}\right] \end{aligned}$$

MU-MLE:

$$\mu_{MLE} = \underset{\mu}{argmax} \sum_{i=1}^{N} \left[\log \frac{1}{\sqrt{2\pi}} + \log \frac{1}{\sigma} - \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

xiaoranli@daum.net

$$= \underset{\mu}{argmax} \sum_{i=1}^{N} -\frac{(x_i-\mu)^2}{2\sigma^2}$$

$$= \underset{\mu}{argmin} \sum_{i=1}^{N} (x_i-\mu)^2$$

$$\frac{\partial \sum_{i=1}^{N}(x_i-\mu)^2}{\partial\mu} = 0$$

$$\sum_{i=1}^{N} 2\cdot(x_i-\mu)\cdot(-1) = 0$$

$$\sum_{i=1}^{N}(x_i-\mu) = 0$$

$$\sum_{i=1}^{N} x_i \sum_{i=1}^{N} \mu = 0$$

$$\mu_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Unbiased estimate:

$$E[\mu_{MLE}] = \frac{1}{N}\sum_{i=1}^{N} E[x_i]$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu$$

$$= \mu$$

SIGMA-MLE:

$$\sigma_{MLE}^2 = \underset{\sigma}{argmax} \sum_{i=1}^{N}\left[\log\frac{1}{\sqrt{2\pi}} + \log\frac{1}{\sigma} - \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$= \underset{\sigma}{argmax} \sum_{i=1}^{N}\left[-\log\sigma - \frac{1}{2\sigma^2}(x_i-\mu)^2\right]$$

$$\frac{\partial \sum_{i=1}^{N}\left[-\log\sigma - \frac{1}{2\sigma^2}(x_i-\mu)^2\right]}{\partial\sigma} N = 0$$

$$\sum_{i=1}^{N}\left[-\frac{1}{\sigma} + \frac{1}{2}(x_i-\mu)^2\cdot 2\sigma^3\right] = 0$$

$$\sum_{i=1}^{N}\left[-\frac{1}{\sigma} + (x_i-\mu)^2\cdot\sigma^3\right] = 0$$

xiaoranli@daum.net

$$\sum_{i=1}^{N}[-\sigma^2 + (x_i - \mu)^2] = 0$$

$$-\sum_{i=1}^{N}\sigma^2 + \sum_{i=1}^{N}(x_i - \mu)^2 = 0$$

$$\sum_{i=1}^{N}\sigma^2 = \sum_{i=1}^{N}(x_i - \mu)^2$$

$$\sigma_{MLE}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Biased estimate:

$$\sigma_{MLE}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu_{MLE}\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}(x_i^2 - 2x_i \cdot \mu_{MLE} + \mu_{MLE}^2)$$

$$= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \frac{1}{N}\sum_{i=1}^{N}2x_i^2 \cdot \mu_{MLE} + \frac{1}{N}\sum_{i=1}^{N}\mu_{MLE}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\mu_{MLE}^2 + \mu_{MLE}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu_{MLE}^2$$

$$E[\sigma_{MLE}^2] = E\left[\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu_{MLE}^2\right]$$

$$= E\left[\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2\right) - (\mu_{MLE}^2 - \mu^2)\right]$$

$$= E\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2\right) - E(\mu_{MLE}^2 - \mu^2)$$

$$= \frac{1}{N}\sum_{i=1}^{N}E(x_i^2 - \mu^2) - \left(E(\mu_{MLE}^2) - E(\mu^2)\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}(E(x_i^2) - \mu^2) - (E(\mu_{MLE}^2) - \mu^2)$$

xiaoranli@daum.net

$$= \frac{1}{N}\sum_{i=1}^{N}(E(x_i^2) - \mu^2) - \left(E(\mu_{MLE}^2) - E^2(\mu_{MLE})\right)$$

$$\because Var[\mu_{MLE}] = Var\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N^2}\sum_{i=1}^{N} Var(x_i) = \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2 = \frac{\sigma^2}{N}$$

$$\therefore E[\sigma_{MLE}^2] = \frac{1}{N}\sum_{i=1}^{N} Var(x_i) - Var(\mu_{MLE})$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sigma^2 - \frac{1}{N}\sigma^2$$

$$= \sigma^2 - \frac{1}{N}\sigma^2$$

$$= \frac{N-1}{N}\sigma^2$$

xiaoranli@daum.net