Name: Aditya Agrawal
Entry No.: 2021AM10198

Date

Page No.

APL747

Homework - 1

Q.1. a) To prove: $X \perp Y, W | Z \Rightarrow X \perp Y | Z$

$X \perp Y, W | Z \Rightarrow P(X,Y,W|Z) = P(X|Z)\, P(Y,W|Z)$ ——— ①

Now,

$$P(X|Y,Z) = \frac{P(X,Y,Z)}{P(Y,Z)}$$

$$= \frac{\sum_W P(X,Y,Z,W)}{\sum_W P(Y,Z,W)}$$

$$= \frac{\sum_W P(X,Y,W|Z)\, P(Z)}{\sum_W P(Y,Z,W)}$$

$$= \frac{\sum_W P(X|Z)\, P(Y,W|Z)\, P(Z)}{\sum_W P(Y,Z,W)} \quad \{\because \text{Using ①}\}$$

$$= P(X|Z) \frac{\sum_W P(Y,W,Z)}{\sum_W P(Y,Z,W)}$$

$\therefore P(X|Y,Z) = P(X|Z)$

Hence, $X \perp Y, W | Z \Rightarrow X \perp Y | Z$

b) To prove: $X \perp Y, W | Z \Rightarrow X \perp Y | Z, W$

$$P(X|Y,W,Z) = \frac{P(X,Y,W,Z)}{P(Y,W,Z)}$$

$$= \frac{P(X,Y,W|Z)\, P(Z)}{P(Y,W,Z)}$$

$$= \frac{P(X|Z)\, P(Y,W|Z)\, P(Z)}{P(Y,W,Z)} \quad \{\because \text{Using ①}\}$$

$$= \frac{P(X|Z)\, P(Y|W,Z)\, P(W|Z) P(Z)}{P(Y|W,Z)\, P(W,Z)}$$

$$= \frac{P(X|Z)\, P(W|Z) P(Z)}{P(W,Z)} = \frac{P(X|W,Z) P(Z)}{P(W,Z)} \{\because \text{Using a)}\}$$

DELTA Notebook

$$= \frac{P(X, W, Z)}{P(W, Z)}$$

$$\therefore P(X|Y, W, Z) = P(X|W, Z)$$

Hence, $X \perp Y, W | Z \Rightarrow X \perp Y | Z, W$

Q2. a) A chi-squared distribution is a special case of a gamma distribution with parameters $\alpha = \frac{n}{2}$ and $\beta = 2$.

$$\therefore f_X(x) = \begin{cases} \dfrac{x^{n/2 - 1} e^{-x/2}}{\Gamma n/2 \; 2^{n/2}} , & x > 0 \\ 0 , & \text{o.w.} \end{cases}$$

Since its pdf only depends on ~~a~~ the ~~single~~ parameter (n), its degree of freedom = n.

Now,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^\infty x \cdot \frac{x^{\frac{n}{2}-1} e^{-x/2}}{\Gamma n/2 \; 2^{n/2}}\, dx$$

$$= \int_0^\infty \frac{x^{n/2} e^{-x/2}}{\Gamma n/2 \; 2^{n/2}}\, dx$$

$$= \frac{1}{\Gamma n/2 \; 2^{n/2}} \int_0^\infty \frac{x^{n/2} e^{-x/2}}{\cancel{\Gamma n/2 \; 2^{n/2}}}\, dx$$

Let $x = 2t \Rightarrow dx = 2dt$

$$\therefore E[X] = \frac{1}{\Gamma n/2 \; 2^{n/2}} \int_0^\infty (2t)^{n/2} e^{-t} (2\,dt)$$

$$= \frac{2^{n/2} \cdot 2}{\Gamma n/2 \; 2^{n/2}} \int_0^\infty t^{n/2} e^{-t}\, dt$$

Also,

$$\Gamma \alpha = \int_0^\infty t^{\alpha - 1} e^{-t}\, dt \quad \text{———} \; ①$$

$$\therefore E[X] = \frac{2}{\Gamma n/2} \times \Gamma \tfrac{n}{2}+1 \quad \{\text{Using } ①\}$$

$$= \frac{2 \times \frac{n}{2} \, \Gamma n/2}{\Gamma n/2} \quad \{\Gamma \alpha+1 = \alpha \Gamma \alpha\}$$

$$\therefore E[X] = n$$

$$E[X^2] = \int_0^\infty x^2 \frac{x^{\frac{n}{2}-1} e^{-x/2}}{\Gamma n/2 \; 2^{n/2}} dx$$

$$= \frac{1}{\Gamma n/2 \; 2^{n/2}} \int_0^\infty x^{\frac{n}{2}+1} e^{-x/2} dx$$

Let $x = 2t \Rightarrow dx = 2dt$

$$E[X^2] = \frac{1}{\Gamma n/2 \; 2^{n/2}} \int_0^\infty (2t)^{\frac{n}{2}+1} e^{-t} (2dt)$$

$$= \frac{2 \cdot 2^{\frac{n}{2}+1}}{\Gamma n/2 \; 2^{n/2}} \int_0^\infty t^{\frac{n}{2}+1} e^{-t} dt$$

$$= \frac{4 \; \Gamma \frac{n}{2}+2}{\Gamma n/2} \quad \{\text{Using } ①\}$$

$$= 4\left(\frac{n}{2}+1\right)\left(\frac{n}{2}\right) \frac{\Gamma n/2}{\Gamma n/2} \quad \{\Gamma \alpha+1 = \alpha \Gamma \alpha\}$$

$$= n(n+2)$$

$$\therefore Var[X] = E[X^2] - (E[X])^2 = n^2 + 2n - n^2 = 2n$$

Therefore, the variance of a chi-squared distributed random variable ($2n$) equals twice its degrees of freedom (n).

b) The formula for Kurtosis of a R.V. X with mean μ and standard deviation σ is given by:

$$K = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

Now, $X \sim N(\mu, \sigma)$

$$\therefore K = \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^4 \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

Let $\frac{x-\mu}{\sigma} = z \Rightarrow dx = \sigma dz$

Date

Page No.

$$K = \int_{-\infty}^{\infty} \frac{z^4 e^{-z^2/2}}{\sqrt{2\pi}} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^4 e^{-z^2/2} dz$$

$$= \sqrt{\frac{2}{\pi}} \left[\int_0^{\infty} z^3 \left(z e^{-z^2/2} \right) dz \right]$$

$$= \sqrt{\frac{2}{\pi}} \left[-z^3 e^{-z^2/2} \Big|_0^{\infty} + 3 \int_0^{\infty} z^2 e^{-z^2/2} dz \right]$$

$$= \sqrt{\frac{2}{\pi}} \left[3 \int_0^{\infty} z^2 e^{-z^2/2} dz \right]$$

We know that,

$$\sqrt{2\pi} = \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \Rightarrow \sqrt{\frac{\pi}{2}} = \int_0^{\infty} x^2 e^{-x^2/2} dx$$

$$\therefore K = 3 \sqrt{\frac{2}{\pi}} \times \sqrt{\frac{\pi}{2}}$$

$$= 3$$

For ~~a~~ $\mu = 0$: $K = E\left[\left(\frac{X}{\sigma}\right)^4\right]$

$$\text{or, } 3 = \frac{1}{\sigma^4} E[X^4]$$

$$\therefore E[X^4] = 3\sigma^4$$

Q.3. Theorem: Let X be a continuous R.V. with pdf $f(x)$. Let $y = g(x)$ be differentiable & strictly increasing or strictly decreasing function, i.e., $g'(x) > 0$ or $g'(x) < 0$ $\forall$ $x \in [a, b]$. Then, $Y = g(X)$ is a continuous R.V. with pdf given by

$$h(y) = \begin{cases} f(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \alpha < y < \beta \\ 0 & \text{O.W.} \end{cases}$$

$\alpha = \min\{g(a), g(b)\}$

$\beta = \max\{g(b), g(a)\}$

Date

Page No.

For a gamma distribution, $G(a,b)$:

$$f_X(x) = \begin{cases} \dfrac{b^a x^{a-1} e^{-bx}}{\Gamma a} & , x > 0 \\ 0 & , \text{O.W.} \end{cases}$$

$\therefore$ $Y = g(X) = \dfrac{1}{X}$

For inverse of g: $X = \dfrac{1}{Y} \Rightarrow Y = \dfrac{1}{X}$

$\therefore g^{-1}(y) = \dfrac{1}{y} \Rightarrow \dfrac{d}{dy} g^{-1}(y) = \dfrac{d}{dy}(1/y) = -\dfrac{1}{y^2}$

$$\therefore h_Y(y) = \begin{cases} \dfrac{b^a}{\Gamma a} \dfrac{1}{y^{a-1}} e^{-b/y} \dfrac{1}{y^2} & , \alpha < y < \beta \\ 0 & , \text{O.W.} \end{cases}$$

$g(a) = g(0) = \infty$; $g(b) = g(\infty) = 0$

$\therefore \alpha = 0$; $\beta = \infty$

$$\therefore h_Y(y) = \begin{cases} \dfrac{b^a}{\Gamma a} \dfrac{e^{-b/y}}{y^{a+1}} & , y > 0 \\ 0 & , \text{O.W.} \end{cases}$$

$$\frac{p_X(x)}{p_Y(y)} = \frac{b^a x^{a-1} e^{-bx}}{\Gamma a \left(\dfrac{b^a e^{-b/y}}{\Gamma a \, y^{a+1}} \right)} = \frac{x^{a-1} e^{-bx + b/y}}{y^{a+1}}$$

For $y = \dfrac{1}{x}$: $\dfrac{p_X(x)}{p_Y(y)} = x^{a-1} \cdot x^{a+1} e^{-bx+bx} = x^2$

$$\therefore p_Y(y) = \frac{1}{x^2} p_X(x)$$

Q.4. $X \sim \mathcal{N}(30, 4.5)$

Now,

a) $P(X < 25) = P\left(\frac{X-30}{4.5} < \frac{25-30}{4.5}\right)$ $\{Z = X-30/4.5\}$

$= P\left(Z < -\frac{10}{9}\right) = \Phi(-10/9) = 1 - \Phi(10/9)$ $\{\because \Phi(x) + \Phi(-x) = 1\}$

$= 1 - 0.8686$ {Using a standard CDF table}

$= 0.13134$

b) $P(30 \le X < 40) = P\left(\frac{30-30}{4.5} \le \frac{X-30}{4.5} < \frac{40-30}{4.5}\right)$

$= P\left(0 \le Z < \frac{10}{4.5}\right) = P(0 \le Z < 2.22)$

$= \Phi(2.22) - \Phi(0)$

$= 0.9868 - 0.5$

$= 0.4868$

Q.5. $p(x) \sim \mathcal{N}(x \mid \mu, \sigma^2)$; $q(x) \sim \mathcal{N}(x \mid m, s^2)$

$KL(p \| q) = -\sum_x p(x) \log \frac{q(x)}{p(x)}$ (Discrete ~~distributi~~ R.V.)

$= -\int_{-\infty}^{\infty} p(x) \log \frac{q(x)}{p(x)} dx$ (Cont. R.V.)

$= -\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \log\left[\frac{\sigma}{s} \exp\left(\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x-m}{s}\right)^2\right)\right] dx$

$= -\log\frac{\sigma}{s} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx}_{\int_{-\infty}^{\infty} f_X(x)\,dx = 1} - \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \left(\frac{x-\mu}{\sigma}\right)^2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$

$- \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{(x^2 - 2mx + m^2)}{s^2} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$

$= \left[\frac{1}{2} \log \frac{s^2}{\sigma^2}\right] - \frac{1}{2} \underbrace{\int_{-\infty}^{\infty} \frac{z^2 e^{-z^2/2}}{\sqrt{2\pi}} dz}_{z = \frac{x-\mu}{\sigma}} - \frac{1}{2s^2}\left[\underbrace{\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx}_{E[x^2] = \sigma^2 + \mu^2} - 2m \int_{-\infty}^{\infty} x \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx\right.$

$\left. + m^2 \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx\right.$

Date ________
Page No. ________

$$= \frac{1}{2}\log\left(\frac{s^2}{\sigma^2}\right) - \frac{1}{2} - \frac{1}{2s^2}\left[\sigma^2 + \mu^2 - 2m\mu + m^2\right]$$

$$= \frac{1}{2}\left[\ln\frac{s^2}{\sigma^2} + \frac{\sigma^2+\mu^2-2m\mu+m^2}{s^2} - 1\right]$$

Q6. Given, the multivariate Gaussian distribution of random vector $\underline{x}$:

$$p(\underline{x}\,|\,\underline{\mu},\underline{\underline{\Sigma}}) = \frac{1}{(2\pi)^{n/2}|\underline{\underline{\Sigma}}|^{-1/2}}\exp\left(\frac{-1}{2}(\underline{x}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}-\underline{\mu})\right)$$

Now,

Suppose the number of trials of the experiment $= N$,

According MLE, the estimates of $\underline{\mu}$ and $\underline{\underline{\Sigma}}$ are given by:

$$\underline{\mu}^* = \arg\max_{\underline{\mu}} P(\mathcal{D}\,|\,\underline{\mu})$$

$$\Sigma^* = \arg\max_{\Sigma} P(\mathcal{D}\,|\,\underline{\underline{\Sigma}})$$

where, $\mathcal{D}$ is the data collected after N trials. $\mathcal{D} = \{\underline{x}_1, \underline{x}_2, \ldots \underline{x}_N\}$

Since all the trials are independent, we can write:

$$P(\mathcal{D}\,|\,\underline{\mu},\underline{\underline{\Sigma}}) = \prod_{i=1}^{N}\frac{1}{(2\pi)^{n/2}|\underline{\underline{\Sigma}}|^{1/2}}\exp\left(\frac{-1}{2}(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$$

Maximising $\log P(\mathcal{D}\,|\,\underline{\mu},\underline{\underline{\Sigma}}) \Rightarrow$ Maximising $P(\mathcal{D}\,|\,\underline{\mu},\underline{\underline{\Sigma}})$, hence, taking log:

$$\log P(\mathcal{D}\,|\,\underline{\mu},\underline{\underline{\Sigma}}) = \log\prod_{i=1}^{N}\frac{1}{(2\pi)^{n/2}|\underline{\underline{\Sigma}}|^{1/2}}\exp\left(\frac{-1}{2}(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$$

$$= \sum_{i=1}^{N}\left(\frac{-n}{2}\log(2\pi) - \frac{1}{2}\log|\underline{\underline{\Sigma}}| - \frac{1}{2}(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$$

$$= -\frac{Nn}{2}\log(2\pi) - \frac{N}{2}\log|\underline{\underline{\Sigma}}| - \frac{1}{2}\sum_{i=1}^{N}(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})$$

Date________
Page No.________

a) Estimating $\underline{\mu}^*$:

$\rightarrow \dfrac{\partial P(\mathcal{D}|\underline{\mu},\underline{\underline{\Sigma}})}{\partial \underline{\mu}} = 0$ {Condition for maxima}

or, $0 = \dfrac{\partial}{\partial \underline{\mu}}\left(-\dfrac{Nn}{2}\log(2\pi)\right) + \dfrac{\partial}{\partial \underline{\mu}}\left(-\dfrac{N}{2}\log|\underline{\underline{\Sigma}}|\right) + \dfrac{\partial}{\partial \underline{\mu}}\left(\dfrac{-1}{2}\sum_{i=1}^{N}(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$

or, $0 = 0 + 0 - \dfrac{1}{2}\sum_{i=1}^{N}\dfrac{\partial}{\partial \underline{\mu}}\left((\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$

or, $0 = \sum_{i=1}^{N}\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})$ {Using the identity: $\dfrac{\partial \omega^T A\omega}{\partial \omega} = 2A\omega$ for symmetric A}

Since, $\underline{\underline{\Sigma}}$ is invertible, it is positive definite and hence,
$\sum_{i=1}^{N}(\underline{x}^{(i)}-\underline{\mu}) = 0$

or, $\sum_{i=1}^{N}\underline{x}^{(i)} - N\underline{\mu} = 0$

$\therefore \underline{\mu}^* = \dfrac{\sum_{i=1}^{N}\underline{x}^{(i)}}{N}$

b) Estimating $\underline{\underline{\Sigma}}^*$:

$\dfrac{\partial P(\mathcal{D}|\underline{\mu},\underline{\underline{\Sigma}})}{\partial \underline{\underline{\Sigma}}} = 0$

or, $0 = \dfrac{\partial}{\partial \underline{\underline{\Sigma}}}\left(\dfrac{-Nn}{2}\log(2\pi)\right) + \dfrac{\partial}{\partial \underline{\underline{\Sigma}}}\left(-\dfrac{N}{2}\log|\underline{\underline{\Sigma}}|\right) + \dfrac{\partial}{\partial \underline{\underline{\Sigma}}}\left(\dfrac{-1}{2}\sum_{i=1}^{N}(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$

or, $0 = 0 - \dfrac{N}{2}\dfrac{\partial}{\partial \underline{\underline{\Sigma}}}(\log|\underline{\underline{\Sigma}}|) - \dfrac{1}{2}\dfrac{\partial}{\partial \underline{\underline{\Sigma}}}\left(\sum(\underline{x}^{(i)}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}^{(i)}-\underline{\mu})\right)$

Using identities : ① $\dfrac{\partial}{\partial A}\log|A| = (A^{-1})^T = (A^T)^{-1}$

② $x^TAx = tr(x^TAx) = tr(xx^TA)$

③ $\dfrac{\partial}{\partial A}tr(AB) = B^T$

Date
Page No.

$$\text{or, } 0 = -N(\Sigma^T)^{-1} - \sum_{i=1}^{N} \frac{\partial}{\partial \Sigma} tr\left[(x^{(i)} - \mu)(x^{(i)} - \mu)^T \Sigma^{-1} \right]$$

$$\text{or, } N \Sigma^{-1} = - \sum_{i=1}^{N} \left\{ \frac{\partial \left(tr\left[(x^{(i)} - \mu)(x^{(i)} - \mu)^T \Sigma^{-1} \right] \right)}{\partial \Sigma} \cancel{\frac{\partial \Sigma^{-1}}{\partial \Sigma}} \right)$$

$$\text{or, } N \Sigma^{-1} = - \frac{\partial\, tr}{\partial \Sigma} \sum_{i=1}^{N} \left[(x^{(i)} - \mu)(x^{(i)} - \mu)^T \right] \Sigma^T \cancel{(-\Sigma^{-2})} \qquad \cancel{\left\{ \frac{\partial A^{-1}}{\partial A} = -A^{-2} \right\}}$$

$$\text{or, } N \Sigma^{-1} = \left(\Sigma^{-1} \sum_{i=1}^{N} (x^{(i)} - \mu)(x^{(i)} - \mu)^T \Sigma^{-1} \right)^T$$

$$\text{Using } \frac{\partial\, tr[AX^{-1}B]}{\partial \Sigma} = -(X^{-1}BAX^{-1})^T \quad \{B = I\}$$

$$\text{or, } N \Sigma^{-1} = \Sigma^{-1} \sum_{i=1}^{N} (x^{(i)} - \mu)(x^{(i)} - \mu)^T \Sigma^{-1}$$

$$\therefore \Sigma^{*} = \frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \mu^{*})(x^{(i)} - \mu^{*})^T$$

Q.7. Given,

$$\underline{X} = \begin{pmatrix} \underline{X}_a \\ \underline{X}_b \end{pmatrix}, \quad \underline{\mu} = \begin{pmatrix} \underline{\mu}_a \\ \underline{\mu}_b \end{pmatrix}, \quad \underline{\underline{\Sigma}} = \begin{bmatrix} \underline{\underline{\Sigma}}_{aa} & \underline{\underline{\Sigma}}_{ab} \\ \underline{\underline{\Sigma}}_{ba} & \underline{\underline{\Sigma}}_{bb} \end{bmatrix}$$

Σ_{aa} & Σ_{bb} are symmetric & $\Sigma_{ba} = \Sigma_{ab}^T$

$$\Sigma^{-1} = \Lambda = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$, Λ_{aa} & Λ_{bb} are symmetric, $\Lambda_{ba} = \Lambda_{ab}^T$

Now,

We can find $p(\underline{x}_a | \underline{x}_b)$ by using the joint distribution $p(\underline{x}) = p(\underline{x}_a, \underline{x}_b)$ & fixing $\underline{x}_b$ to be the already observed value.

We know that,

$$p(\underline{x}) = \frac{-1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})$$

$$= \frac{-1}{2}(\underline{x}_a - \underline{\mu}_a)^T \underline{\underline{\Lambda}}_{aa} (\underline{x}_a - \underline{\mu}_a) - \frac{1}{2}(\underline{x}_a - \underline{\mu}_a)^T \underline{\underline{\Lambda}}_{ab} (\underline{x}_b - \underline{\mu}_b)$$

$$- \frac{1}{2}(\underline{x}_b - \underline{\mu}_b)^T \underline{\underline{\Lambda}}_{ba} (\underline{x}_a - \underline{\mu}_a) - \frac{1}{2}(\underline{x}_b - \underline{\mu}_b)^T \Lambda_{bb} (\underline{x}_b - \underline{\mu}_b)$$

Date

Page No.

This is a standard method of using the partitioning of $\underline{X}$, $\underline{\mu}$ & $\underline{\underline{\Sigma}}^{-1}$.

Since this term is quadratic in ~~nature~~ x_a, $p(\underline{x}_a|\underline{x}_b)$ will also be Gaussian.

Now,

For a general Gaussian distribution $\mathcal{N}(\underline{X}|\mu,\underline{\underline{\Sigma}})$,

$$-\frac{1}{2}(\underline{x}-\underline{\mu})^T\underline{\underline{\Sigma}}^{-1}(\underline{x}-\underline{\mu}) = -\frac{1}{2}\underline{x}^T\underline{\underline{\Sigma}}^{-1}\underline{x} + \underline{x}^T\Sigma^{-1}\underline{\mu} + \text{const.}$$

Here, the constant is independent of $\underline{x}$.

Comparing these 2 eqns while treating $\underline{x}_b$ as a constant (since it is an observed vector), we compare the corresponding matrices of coefficients for the quadratic & the linear term.

Quadratic terms:

$$-\frac{1}{2}\underline{x}_a^T\underline{\underline{\Lambda}}_{aa}\underline{x}_a = -\frac{1}{2}\underline{x}^T\underline{\underline{\Sigma}}^{-1}\underline{x} \Rightarrow \underline{\underline{\Sigma}}^{-1} = \Lambda_{aa} \Rightarrow \underline{\underline{\Sigma}} = \Lambda_{aa}^{-1}$$

$$\therefore \underline{\underline{\Sigma}}_{a|b} = \underline{\underline{\Lambda}}_{aa}^{-1} \text{ for } p(\underline{x}_a|\underline{x}_b)$$

Linear Terms:

$$\underline{x}_a^T\{\underline{\underline{\Lambda}}_{aa}\underline{\mu}_a - \underline{\underline{\Lambda}}_{ab}(\underline{x}_b-\underline{\mu}_b)\} = \underline{x}^T\underline{\underline{\Sigma}}^{-1}\underline{\mu}$$

$$\therefore \underline{\mu}_{a|b} = \underline{\underline{\Sigma}}_{a|b}\{\underline{\underline{\Lambda}}_{aa}\underline{\mu}_a - \underline{\underline{\Lambda}}_{ab}(\underline{x}_b-\underline{\mu}_b)\}$$
$$= \underline{\mu}_a - \underline{\underline{\Lambda}}_{aa}^{-1}\Lambda_{ab}(\underline{x}_b-\underline{\mu}_b)$$

Hence, Proved.