1 Problem 1

a.

$$\begin{aligned} p(w|x_1,\ldots,x_N) &\propto p(x_1,\ldots,x_N|w)p(w) \\ &\propto [\prod_{i=1}^{N} w^{x_i}(1-w)^{1-x_i}]w^{a-1}(1-w)^{b-1} \\ &= w^{\sum_{i=1}^{N} x_i+a-1}(1-w)^{n-\sum_{i=1}^{N} x_i+b-1} \\ &\propto \text{Beta}(\sum_{i=1}^{N} x_i + a, n - \sum_{i=1}^{N} x_i + b) \end{aligned} \tag{1}$$

b. $y_1 = \frac{x_1}{x_1+x_2}, \quad y_2 = x_1 + x_2 \to x_1 = y_1 y_2, \quad x_2 = y_2(1-y_1)$

$$\begin{aligned} |J| &= | \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{bmatrix} | \\ &= y_2 - y_2 y_1 + y_1 y_2 \\ &= y_2 \end{aligned} \tag{2}$$

$$\begin{aligned} f_{y_1,y_2}(y_1,y_2) &= f_{x_1,x_2}(x_1,x_2)|J| \\ &= \frac{b^{a_1}}{\Gamma(a_1)} x_1^{a_1-1} e^{-bx_1} \frac{b^{a_2}}{\Gamma(a_2)} x_2^{a_2-1} e^{-bx_2} y_2 \\ &= \frac{b^{a_1}}{\Gamma(a_1)} (y_1 y_2)^{a_1-1} e^{-b(y_1 y_2)} \frac{b^{a_2}}{\Gamma(a_2)} (y_2(1-y_1))_2^{a_2-1} e^{-b(y_2(1-y_1))} y_2 \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} y_2^{a_1+a_2-1} e^{-by_2} \end{aligned} \tag{3}$$

$$\begin{aligned} f_{y_1}(y_1) &= \int_0^\infty f_{y_1,y_2}(y_1,y_2) dy_2 \\ &= \int_0^\infty \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} y_2^{a_1+a_2-1} e^{-by_2} dy_2 \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} \int_0^\infty y_2^{a_1+a_2-1} e^{-by_2} dy_2 \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} \frac{\Gamma(a_1+a_2)}{b^{a_1+a_2}} \\ &= \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} \\ &= \text{Beta}(a_1,a_2) \end{aligned} \tag{4}$$

$$
\begin{aligned}
f_{y_2}(y_2) &= \int_0^\infty f_{y_1,y_2}(y_1,y_2)dy_1 \\
&= \int_0^\infty \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} y_2^{a_1+a_2-1} e^{-by_2} dy_1 \\
&= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_2^{a_1+a_2-1} e^{-by_2} \int_0^\infty y_1^{a_1-1}(1-y_1)^{a_2-1} dy_1 \\
&= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} y_2^{a_1+a_2-1} e^{-by_2} \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)} \\
&= \frac{b^{a_1+a_2}}{\Gamma(a_1+a_2)} y_2^{a_1+a_2-1} e^{-by_2} \\
&= \text{Gamma}(a_1+a_2, b)
\end{aligned} \tag{5}
$$

In our case $b = 1$, so we have $\text{Gamma}(a_1 + a_2, 1)$.

c.

$$
\begin{aligned}
p(\theta|x_1,\ldots,x_N) &\propto p(x_1,\ldots,x_N|\theta)p(\theta) \\
&\propto [\prod_{i=1}^N \exp(-\frac{(x_i-\theta)^2}{2\sigma^2})] \times \exp(-\frac{(\theta-m)^2}{2v}) \\
&\propto \exp(-\frac{\sum_{i=1}^N (x_i-\theta)^2}{2\sigma^2} - \frac{(\theta-m)^2}{2v}) \\
&\propto \exp(-\frac{1}{2\sigma^2 v}(v\sum_{i=1}^N (x_i-\theta)^2 + \sigma^2(\theta-m)^2)) \\
&= \exp(-\frac{1}{2\sigma^2 v}(v\sum_{i=1}^N (x_i^2 + 2x_i\theta + \theta^2) + \sigma^2(\theta^2 - 2\theta m + m^2))) \\
&\propto \exp(-\frac{1}{2\sigma^2 v}(-2\theta v n\bar{x} + n\theta^2 v + \sigma^2\theta^2 - 2\theta\sigma^2 m)) \\
&= \exp(-\frac{1}{2\sigma^2 v}[(\sigma^2 + nv)\theta^2 - 2\theta(vn\bar{x} + \sigma^2 m)]) \\
&= \exp(-\frac{(\sigma^2+nv)}{2\sigma^2 v}[\theta^2 - 2\theta(\frac{vn\bar{x}+\sigma^2 m}{\sigma^2+nv})]) \\
&\propto = \exp(-\frac{(\sigma^2+nv)}{2\sigma^2 v}(\theta - \frac{vn\bar{x}+\sigma^2 m}{\sigma^2+nv})^2) \\
&\propto \mathcal{N}(\tilde{m}, \tilde{v})
\end{aligned} \tag{6}
$$

We recognize the kernel as that of a normal distribution. We can identify the parameters by looking at the line prior to the last line.

$\tilde{m} = (\frac{vn\bar{x} + \sigma^2 m}{\sigma^2 + nv})$

$\tilde{v} = \frac{\sigma^2 v}{\sigma^2 + nv}$

d.

$$\begin{aligned}
p(w|x_1,\ldots,x_N) &\propto p(x_1,\ldots,x_N|w)p(w) \\
&\propto [\prod_{i=1}^{N} w^{1/2}\exp(-\frac{w}{2}(x_i-\theta)^2)]w^{a-1}\exp(-bw) \\
&= w^{\frac{N}{2}+a-1}\exp(-w(\frac{\sum_{i=1}^{N}(x_i-\theta)^2}{2}+b)) \\
&\propto \text{Gamma}(\frac{N}{2}+a, \frac{\sum_{i=1}^{N}(x_i-\theta)^2}{2}+b)
\end{aligned} \tag{7}$$

Claim: If $X \sim \text{Gamma}(a,b), Y = \frac{1}{X} \sim \text{InvGamma}(a,b)$.
$Y = \frac{1}{X} \rightarrow X = \frac{1}{Y}$
$|\frac{dx}{dy}| = |-\frac{1}{y^2}| = \frac{1}{y^2}$

$$\begin{aligned}
f_y(y) &= f_x(x)|\frac{dx}{dy}| \\
&= \frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)y^{-2} \\
&= \frac{b^a}{\Gamma(a)}y^{-a+1-2}\exp(-\frac{b}{y}) \\
&= \frac{b^a}{\Gamma(a)}(y)^{-a-1}\exp(-\frac{b}{y}) \\
&= \text{InvGamma}(a,b)
\end{aligned} \tag{8}$$

Thus, $p(\sigma^2|x_1,\ldots,x_N) = \text{InvGamma}(\frac{N}{2}+a, \frac{\sum_{i=1}^{N}(x_i-\theta)^2}{2}+b)$.

e.

$$\begin{aligned}
p(\theta|x_1,\ldots,x_N) &\propto p(x_1,\ldots,x_N|\theta)p(\theta) \\
&\propto (\prod_{i=1}^{N}\exp(-\frac{(x_i-\theta)^2}{2\sigma_i^2}))(\exp(-\frac{(\theta-m)^2}{2v})) \\
&= \exp(-\frac{\sum_{i=1}^{N}(x_i-\theta)^2}{2\sigma_i^2} - \frac{(\theta-m)^2}{2v}) \\
&= \exp(-\frac{1}{2}(\sum_{i=1}^{N}(\frac{x_i^2-2x_i\theta+\theta^2}{\sigma_i^2}) + (\frac{\theta^2-2\theta m+m^2}{v}))) \\
&\propto \exp(-\frac{1}{2}(\theta^2(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}) - 2\theta(\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}+\frac{m}{v}))) \\
&= \exp(-\frac{1}{2}(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v})(\theta^2-2\theta(\frac{\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}+\frac{m}{v}}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v}}))) \\
&\propto \mathcal{N}(\tilde{m},\tilde{v})
\end{aligned} \tag{9}$$

$$\tilde{v} = \frac{1}{(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}+\frac{1}{v})} \tag{10}$$

$$\tilde{m} = \tilde{v}(\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \frac{m}{v}) \tag{11}$$

$$= \tilde{v} \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \frac{\tilde{v}}{v} m \tag{12}$$

f.

$$\begin{aligned}
p(x) &= \int_0^\infty p(x|w)p(w)dw \\
&= \int_0^\infty \frac{w^{1/2}}{\sqrt{2\pi}} \exp(-w\frac{(x-m)^2}{2})\frac{(b/2)^{a/2}}{\Gamma(a/2)} w^{a/2-1}\exp(-w\frac{b}{2})dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \int_0^\infty w^{\frac{a+1}{2}-1} \exp(-w(\frac{b+(x-m)^2}{2}))dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \frac{\Gamma(\frac{a+1}{2})}{(\frac{b+(x-m)^2}{2})^{\frac{a+1}{2}}} \\
&= \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})\sqrt{\pi}\sqrt{b}} \frac{1}{(1+\frac{(x-m)^2}{b})^{\frac{a+1}{2}}} \\
&= \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})\sqrt{\pi a}\sqrt{\frac{b}{a}}} \frac{1}{(1+\frac{a(x-m)^2}{ba})^{\frac{a+1}{2}}}
\end{aligned} \tag{13}$$

2 Problem 2

a.

$$\begin{aligned}
Cov(x) &= E[(x-\mu)(x-\mu)^T] \\
&= E[xx^T - x\mu^T - \mu x^T + \mu\mu^T] \\
&= E[xx^T] - E[x]\mu^T - \mu E[x^T] + \mu\mu^T \\
&= E[xx^T] - \mu\mu^T
\end{aligned} \tag{14}$$

The third line is justified because expectation is linear and $\mu = E[x]$ and $\mu^T = E[x]^T = E[x^T]$.

$$\begin{aligned}
Cov(Ax+b) &= E[(Ax+b)(Ax+b)^T] - E[Ax+b]E[Ax+b]^T \\
&= E[Axx^TA^T + Axb^T + bx^TA^T + bb^T] - [A\mu + b][\mu^TA^T + b^T] \\
&= AE[xx^T]A^T + AE[x]b^T + bE[x^T]A^T + bb^T - A\mu\mu^TA^T - A\mu b^T - b\mu^TA^T - bb^T \\
&= AE[xx^T]A^T - A\mu\mu^TA^T \\
&= A[E[xx^T] - \mu\mu^T]A^T \\
&= ACov(x)A^T
\end{aligned} \tag{15}$$

The first line we use the property that we proved in the first part. Again, expectation is linear and $\mu = E[x]$ and $\mu^T = E[x]^T = E[x^T]$ so the result follows.

b.

$$
\begin{aligned}
p(z_1, \ldots, z_p) &= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z_i^2}{2}) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{\sum_{i=1}^{p} z_i^2}{2}) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{z^T z}{2})
\end{aligned} \tag{16}
$$

The first line is justified because of z_i are independent and identically distributed $\mathcal{N}(0, 1)$. In the last equality we re-write sum square as inner product of the vector z.

$$
\begin{aligned}
M_z(t) &= E[\exp(z^T t)] \\
&= E[\exp(\sum_{i=1}^{p} z_i t_i)] \\
&= E[\prod_{i=1}^{p} \exp(z_i t_i)] \\
&= \prod_{i=1}^{p} E[\exp(z_i t_i)] \\
&= \prod_{i=1}^{p} \exp(\frac{t_i^2}{2}) \\
&= \exp(\frac{\sum_{i=1}^{p} t_i^2}{2}) \\
&= \exp(\frac{t^T t}{2})
\end{aligned} \tag{17}
$$

In the fourth equality, we use the independent assumption. In the next line, we apply the definition of univariate mgf of a normal distribution.

Alternative solution.

$$
\begin{aligned}
M_z(t) &= E[\exp(z^T t)] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(z^T t) \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{z^T z}{2}) dz_1 \ldots dz_p \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{1}{2}(z^T z - 2z^T t)) dz_1 \ldots dz_p \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{1}{2}(z^T z - 2z^T t + t^T t)) \exp(\frac{t^T t}{2}) dz_1 \ldots dz_p \\
&= \exp(\frac{t^T t}{2}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{1}{2}((z - t)^T (z - t)) dz_1 \ldots dz_p \\
&= \exp(\frac{t^T t}{2})
\end{aligned} \tag{18}
$$

In the last line, the integral integrates to 1 because we recognize the integral as that the integral of $\mathcal{N}(t, I)$ over the whole space.

c. We will first prove the if direction.

If x has mgf of the form $E(\exp(t^T x)) = \exp(t^T \mu + \frac{t^T \Sigma t}{2})$, then $x \sim \mathcal{N}(\mu, \Sigma)$

Let $a \in \mathbb{R}^p, a \neq 0$, we will show that $s = a^T x$ is a univariate normal distribution so x follows a multivariate normal distribution since every linear combination is a univariate normal.

$$\begin{aligned} M_s(t) &= E(\exp(ts)) \\ &= E(\exp(t(a^T x))) \\ &= E(\exp((ta)^T x)) \\ &= M_x(ta) \\ &= \exp\{(ta)^T \mu + \frac{(ta)^T \Sigma (ta)}{2}\} \\ &= \exp\{t(a^T \mu) + \frac{t^2 a^T \Sigma a}{2}\} \end{aligned} \tag{19}$$

We recognize the mgf as that of a univariate normal and we can identify mean and variance. $s \sim \mathcal{N}(a^T \mu, a^T \Sigma a)$.

Now, we will prove the only if direction. Suppose $x \sim \mathcal{N}(\mu, \Sigma)$, then x has mgf $E(\exp(t^T x)) = \exp(t^T \mu + \frac{t^T \Sigma t}{2})$.

Let t be a non-zero vector $s = t^T x$ is a univariate normal. $s \sim \mathcal{N}(t^T \mu, t^T \Sigma t)$ by exercise A.

$$\begin{aligned} M_x(t) &= E(\exp(t^T x(1))) \\ &= M_s(1) \\ &= \exp(t^T \mu(1) + \frac{(1^2) t^T \Sigma t}{2}) \\ &= \exp(t^T \mu + \frac{t^T \Sigma t}{2}) \end{aligned} \tag{20}$$

In the first line, we use the fact that $s = t^T x$ is univariate normal and use the mgf of a univariate normal evaluated at 1.

If $t = 0$, $M_x(0) = E(\exp(0^T x)) = 1 = \exp(0^T \mu + \frac{0 \Sigma 0}{2})$.

Thus, $M_x(t) = \exp(t^T \mu + \frac{t^T \Sigma t}{2})$ for all t.

d. We will first prove the if direction. Suppose x is written as an affine transformation of a standard multivariate normal i.e. $x = Lz + \mu$, then x is multivariate normal.

$$\begin{aligned} M_x(t) &= E(\exp(t^T x)) \\ &= \exp(t^T \mu) E(\exp(t^T L z)) \\ &= \exp(t^T \mu) E(\exp((L^T t)^T z)) \\ &= \exp(t^T \mu) \exp(\frac{(L^T t)^T (L^T t)}{2}) \\ &= \exp\{t^T \mu + \frac{t^T L L^T t}{2}\} \end{aligned} \tag{21}$$

We recognize this as the mgf we found in the previous problem. Thus, $x \sim \mathcal{N}(\mu, LL^T)$.

e. Supose $x \sim \mathcal{N}(\mu, \Sigma)$. We can apply Cholesky decomposition to the covariance matrix Σ. $\Sigma = LL^T$ since we assume Σ is psd. Now, define $x = Lz + \mu$. Then, by the previous part, $x \sim \mathcal{N}(\mu, LL^T = \Sigma)$ by construction.

f. Suppose $x \sim \mathcal{N}(\mu, \Sigma)$, x can be written as an affine transformation $x = Lz + \mu$. The inverse transformation is $z = L^{-1}(x - \mu)$. Using the hint, the jacobian of this transformation is $|J| = |L^{-1}| = \frac{1}{|L|}$.

$$|\Sigma| = |LL^T| = |L||L^T| = |L|^2$$

Thus, $|J| = \frac{1}{|L|} = \frac{1}{|\Sigma|^{1/2}}$.

Apply change of variable theorem,

$$\begin{aligned}
f_x(x) &= f_z(z)|J| \\
&= f_z(L^{-1}(x-\mu))\frac{1}{|\Sigma|^{1/2}} \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{1/2}} \exp\{-\frac{(L^{-1}(x-\mu))^T(L^{-1}(x-\mu))}{2}\} \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{1/2}} \exp\{-\frac{(x-\mu)^T(L^T)^{-1}(L^{-1})(x-\mu)}{2}\} \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{1/2}} \exp\{-\frac{(x-\mu)^T(LL^T)^{-1}(x-\mu)}{2}\} \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{1/2}} \exp\{-\frac{(x-\mu)^T(\Sigma)^{-1}(x-\mu)}{2}\}
\end{aligned} \tag{22}$$

g.

$$\begin{aligned}
M_y(t) &= E[\exp(t^T(Ax_1 + Bx_2))] \\
&= E[\exp(t^TAx_1) + \exp(t^TBx_2)] \\
&= E[\exp((A^Tt)^Tx_1)]E[\exp((B^Tt)^Tx_2)] \\
&= M_{x_1}(A^Tt)M_{x_2}(B^Tt) \\
&= \exp\{\mu_1^T(A^Tt) + \frac{(A^Tt)^T\Sigma_1(A^Tt)}{2}\}\exp\{\mu_2^T(B^Tt) + \frac{(B^Tt)^T\Sigma_2(B^Tt)}{2}\} \\
&= \exp\{(A\mu_1)^Tt + (B\mu_2)^Tt + \frac{(t^TA)\Sigma_1(A^Tt)}{2} + \frac{(t^TB)\Sigma_2(B^Tt)}{2}\} \\
&= \exp\{((A\mu_1) + (B\mu_2))^Tt + \frac{t^T(A\Sigma_1A^T + B\Sigma_2B^T)t}{2}\}
\end{aligned} \tag{23}$$

The third equality is justified by independence of x_1 and x_2. By uniqueness of mgf, we recognize this as the mgf of a multivariate normal. Thus, $y \sim \mathcal{N}(A\mu_1 + B\mu_2, A\Sigma_1A^T + B\Sigma_2B^T)$.

(A)

$$x_{p\times 1} = (x_1 \; x_2)^T$$

since we know x_1 is $k\times 1$,

we want to find a transformation

that picks out elements of x_1 to

get the marginal distribution.

$$A_{k\times p} = \begin{bmatrix} I_{k\times k} & O_{k\times (p-k)} \end{bmatrix}$$

$$x_1 = Ax \sim N(A\mu, A\Sigma A^T).$$

$$A\mu = \begin{bmatrix} I_{k\times k} & O_{k\times (p-k)} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 I_{k\times k} + O_{k\times (p-k)}\mu_2 \end{bmatrix} = \mu_1$$

$$A\Sigma A^T = \begin{bmatrix} I_{k\times k} & O_{k\times (p-k)} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_{k\times k}^T \\ O_{k\times (p-k)}^T \end{bmatrix}$$

$$= I_{k\times k}\Sigma_{11}I_{k\times k}^T + I_{k\times k}\Sigma_{12}O_{k\times (p-k)}^T + O_{k\times (p-k)}\Sigma_{21}I_{k\times k}^T + O_{k\times (p-k)}\Sigma_{22}O_{k\times (p-k)}^T$$

$$= \Sigma_{11}$$

(B)

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{bmatrix}$$

$$= \begin{pmatrix} A & -A\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}A & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}A\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix}$$

$$A = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$$

(C)

$$x - \mu = [x_1 - \mu_1 \quad x_2 - \mu_2]^T$$

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{bmatrix}$$

$$p(x_1 | x_2) \propto p(x_1, x_2)$$

$$\propto e^{-\underbrace{\frac{(x-\mu)^T \Omega (x-\mu)}{2}}}$$

$$(x-\mu)^T \Omega (x-\mu) = [x_1 - \mu_1 \quad x_2 - \mu_2]^T \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$= (x_1-\mu_1)^T \Omega_{11} (x_1 - \mu_1) + (x_1 - \mu_1)^T \Omega_{12} (x_2 - \mu_2) + (x_2 - \mu_2)^T \Omega_{12}^T (x_1 - \mu_1) +$$

$(x_2-\mu_2)^T \Omega_{22}(x_2-\mu_2)$

As a function of x_1 the exponent is proportional to

$$x_1^T\Omega_{11}x_1 - 2x_1^T\Omega_{11}\mu_1 + x_1^T\Omega_{12}x_2 - x_1^T\Omega_{12}\mu_2 + x_2^T\Omega_{12}^T x_1 - \mu_2^T\Omega_{12}^T x_1$$

$$= x_1^T\Omega_{11}x_1 - 2x_1^T\Omega_{11}\mu_1 + 2x_1^T\Omega_{12}x_2 - 2x_1^T\Omega_{12}\mu_2$$

$$= x_1^T\Omega_{11}x_1 - 2x_1^T(\Omega_{11}\mu_1 - \Omega_{12}x_2 + \Omega_{12}\mu_2)$$

$$= x_1^T\Omega_{11}x_1 - 2x_1^T(\Omega_{11}\mu_1 - \Omega_{12}(x_2-\mu_2))$$

$$\tilde{\Omega} = \Omega_{11}$$

$$= A \quad \text{(from ex. 2)}$$

$$= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$$

$$\tilde{\Sigma} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$\tilde{\mu} = \tilde{\Omega}^{-1}(\Omega_{11}\mu_1 - \Omega_{12}(x_2-\mu_2))$$

$$= \mu_1 - \Omega_{11}^{-1}\Omega_{12}(x_2-\mu_2)$$

$$= \mu_1 + A^{-1}A\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)$$

$$= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)$$

1 Conditional and Marginal

c. Let

$\mathbf{x_1} \in \mathbb{R}^{\mathbf{k}}$

$\mathbf{x_2} \in \mathbb{R}^{\mathbf{p-k}}$

$\mathbf{x} = [\mathbf{x_1} \quad \mathbf{x_2}]^{\mathbf{T}}$

$$\begin{aligned}
f(\mathbf{x_1}|\mathbf{x_2}) &\propto f(\mathbf{x_1}, \mathbf{x_2}) \\
&\propto \exp\{(\mathbf{x} - \boldsymbol{\mu})^{\mathbf{T}}\boldsymbol{\Omega}(\mathbf{x} - \boldsymbol{\mu})\} \\
&\propto \exp\{\begin{bmatrix}(\mathbf{x_1} - \boldsymbol{\mu_1})^{\mathbf{T}} & (\mathbf{x_2} - \boldsymbol{\mu_2})^{\mathbf{T}}\end{bmatrix} \begin{bmatrix}\Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{bmatrix} \begin{bmatrix}(\mathbf{x_1} - \boldsymbol{\mu_1}) \\ (\mathbf{x_2} - \boldsymbol{\mu_2})\end{bmatrix}\} \\
&= \exp\{(\mathbf{x_1} - \boldsymbol{\mu_1})^{\mathbf{T}}\boldsymbol{\Omega_{11}}(\mathbf{x_1} - \boldsymbol{\mu_1}) + (\mathbf{x_1} - \boldsymbol{\mu_1})^{\mathbf{T}}\boldsymbol{\Omega_{12}}(\mathbf{x_2} - \boldsymbol{\mu_2}) \\
&+ (\mathbf{x_2} - \boldsymbol{\mu_2})^{\mathbf{T}}\boldsymbol{\Omega_{12}^{\mathbf{T}}}(\mathbf{x_1} - \boldsymbol{\mu_1}) + (\mathbf{x_2} - \boldsymbol{\mu_2})^{\mathbf{T}}\boldsymbol{\Omega_{22}}(\mathbf{x_2} - \boldsymbol{\mu_2})^{\mathbf{T}}\} \\
&\propto \exp\{\mathbf{x_1^T}\boldsymbol{\Omega_{11}}\mathbf{x_1} - \mathbf{2x_1^T}\boldsymbol{\Omega_{11}}\boldsymbol{\mu_1} + \mathbf{2x_1^T}\boldsymbol{\Omega_{12}}\mathbf{x_2} - \mathbf{2x_1^T}\boldsymbol{\Omega_{12}}\boldsymbol{\mu_2}\} \\
&\propto \exp\{\mathbf{x_1^T}\boldsymbol{\Omega_{11}}\mathbf{x_1} - \mathbf{2x_1^T}[\boldsymbol{\Omega_{11}}\boldsymbol{\mu_1} - \boldsymbol{\Omega_{12}}[\mathbf{x_2} - \boldsymbol{\mu_2}]]\} \\
&\propto \exp\{(\mathbf{x_1} - \tilde{\boldsymbol{\mu}}_{\mathbf{1}})^{\mathbf{T}}\tilde{\boldsymbol{\Omega}}(\mathbf{x_1} - \tilde{\boldsymbol{\mu}}_{\mathbf{1}})\}
\end{aligned} \tag{1}$$

$$\begin{aligned}
\tilde{\Omega} &= \Omega_{11} \\
&= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})
\end{aligned} \tag{2}$$

$$\tilde{\Omega}\tilde{\boldsymbol{\mu}} = [\Omega_{11}\boldsymbol{\mu}_1 - \Omega_{12}[\mathbf{x_2} - \boldsymbol{\mu_2}]]$$

$$\begin{aligned}
\tilde{\boldsymbol{\mu}} &= (\Omega_{11})^{-1}[\Omega_{11}\boldsymbol{\mu}_1 - \Omega_{12}[\mathbf{x_2} - \boldsymbol{\mu_2}]] \\
&= \boldsymbol{\mu}_1 - \Omega_{11}^{-1}\Omega_{12}(\mathbf{x_2} - \boldsymbol{\mu_2}) \\
&= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x_2} - \boldsymbol{\mu_2})
\end{aligned} \tag{3}$$

We can see that the mean of $\mathbf{x_1}|\mathbf{x_2}$ is a linear function of $\mathbf{x_2}$.
If x_1 and x_2 are both scalar, we get a regression line.

$$\begin{aligned}
E(x_1|x_2) &= \mu_1 + \frac{Cov(x_1, x_2)}{\sigma_2^2}(x_2 - \mu_2) \\
&= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)
\end{aligned} \tag{4}$$

If x_1 is a scalar and $\mathbf{x_2}$ is a $p-1$ dimension vector, we can read off the regression coefficient from the precision matrix. The regression coefficient of x_1 on the other variables are $\frac{\Omega_{1j}}{\Omega_{11}}$ $j = 2, \ldots, p$. Here, Ω_{ij} denote the i th j th entry of the precision matrix.
The regression equation is
$E(x_1|x_{2:p}) = \mu_1 - \sum_{j=2}^{p} \frac{\Omega_{1j}}{\Omega_{11}}(x_j - \mu_j)$

(A)

$\boxed{\text{Least Square}}$

$$f(\beta) = (Y - X\beta)^T (Y - X\beta)$$

$$\underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}}\ f(\beta)$$

$$f(\beta) = Y^TY - 2Y^TX\beta + \beta^TX^TX\beta$$

$$= Y^TY - 2(X^TY)^T\beta + \beta^T(X^TX)\beta$$

$$\nabla f = -2(X^TY) + (X^TX + X^TX)\beta$$

$$= -2(X^TY) + 2(X^TX)\beta$$

$$\nabla f = 0 \rightarrow (X^TX)\beta = X^TY$$

$$\beta = (X^TX)^{-1}X^TY$$

$\boxed{\text{MLE}}$

$$f(\beta) = \prod_{i=1}^{n} p(y_i \mid \beta, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y_i - \beta x_i)^2}{2\sigma^2}}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{\sum_{i=1}^{n}(y_i - \beta x_i)^2}{2\sigma^2}}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{(Y - X\beta)^T (Y - X\beta)}{2\sigma^2}}$$

since e^{-x} is a decreasing function,

maximizing e^{-x} is the same as

minimizing $(Y - X\beta)^T (Y - X\beta)$ which

has the same solution as least square.

$$\boxed{\text{Method of Moment}}$$

Moment condition

$$E(X^T \varepsilon) = X^T E(\varepsilon) = 0$$ on average the X and residual are uncorrelated.

$$\mathrm{cov}(x_i, \varepsilon) = \frac{1}{n}(x_i - \bar{x}_i)^T(\varepsilon - \bar{\varepsilon})$$

$$= \frac{1}{n} x_i^T \varepsilon \qquad \forall i = 1,2,\ldots,p$$

Assume $\bar{x}_i = 0$, $\bar{\varepsilon} = 0$

writing in matrix form we obtain.

$$\varepsilon = \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{n \times 1} \qquad X^T = \underbrace{\begin{bmatrix} — & x_1^T & — \\ & \vdots & \\ — & x_p^T & — \end{bmatrix}}_{p \times n}$$

$$X = \underbrace{\begin{bmatrix} | & & | \\ x_1 & \cdots & x_p \\ | & & | \end{bmatrix}}_{n \times p}$$

$$X^T \varepsilon = \begin{bmatrix} x_1^T \varepsilon \\ x_2^T \varepsilon \\ \vdots \\ x_p^T \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$X^T(Y - X\beta) = 0$$

$$X^T Y - X^T X \beta = 0$$

$$X^T X \beta = X^T Y$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

(B)

Let $\quad W = \mathrm{diag}(w_1, \ldots, w_n)$

$$\varepsilon = (y - X\beta)$$

$$\sum_{i=1}^{n} w_i (y_i - x_i^T \beta) = \varepsilon^T W \varepsilon = (y - X\beta)^T W (y - X\beta)$$

$$= y^T W y - 2 y^T W X \beta + \beta^T X^T W X \beta$$

$$= y^T W y - 2 (X^T W y)^T \beta + \beta^T X^T W X \beta$$

$$\nabla \varepsilon^T W \varepsilon = -2(X^T W y) + 2(X^T W X)\beta = 0$$

$$\longrightarrow \quad (X^T W X)\beta = X^T W y$$

$$\hat{\beta} = (X^T W X)^{-1} X^T W y$$

$$L(\beta) = p(y_1, \ldots, y_n \mid \beta, \sigma_i^2)$$

$$= \prod_{i=1}^{n} p(y_i \mid \beta, \sigma_i^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma_i^2}} e^{-\frac{(y_i - x_i^T \beta)^2}{2\sigma_i^2}}$$

$$= \frac{1}{(2\pi\sigma_i^2)^{\frac{n}{2}}} e^{-\sum_{i=1}^{n} \frac{(y_i - x_i^T \beta)^2}{2\sigma_i^2}} \qquad \Lambda = \begin{bmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n^2} \end{bmatrix}$$

$$= C \, e^{-\frac{(y - X\beta)^T \Lambda (y - X\beta)}{2}}$$

maximize $L(\beta)$ is equivalent to minimizing $(Y-X\beta)^T \Lambda (Y-X\beta)$, which is the same problem as least square but with $W=\Lambda$.

Thus the solution is

$$\hat{\beta} = (X^T \Lambda X)^{-1} X^T \Lambda Y$$

Thus, individual precision becomes the weight.

Quantifying Uncertainty

(A)

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$y \sim N(X\beta, \sigma^2 I)$$

since $\hat{\beta} = (X^T X)^{-1} X^T y$ is a linear transformation of y, from previous exercise, $\hat{\beta}$ is also a multivariate normal.

$$\cdot E(\hat{\beta}) = E((X^T X)^{-1} X^T y) = (X^T X)^{-1} X^T X \beta$$

$$= \beta$$

$$var(\hat{\beta}) = var((X^T X)^{-1} X^T y)$$

$$= (X^T X)^{-1} X^T var(y) X (X^T X)^{-1}$$

$$= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}$$

(A)

$$f(\hat{\theta}) = j^T \hat{\theta} \qquad j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ \vdots \end{bmatrix}$$

$$= \hat{\theta}_1 + \hat{\theta}_2$$

Since this is a linear transformation,

$f(\hat{\theta})$ also follows normal distribution.

$$E(f(\hat{\theta})) = E(j^T \hat{\theta}) = j^T E(\hat{\theta})$$

$$= j^T \theta$$

$$= \theta_1 + \theta_2$$

$$var(f(\hat{\theta})) = var(j^T \hat{\theta})$$

$$= j^T var(\hat{\theta}) j$$

$$= \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \hat{\theta}_{11}^2 & \cdots & \hat{\sigma}_{1p}^2 \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1}^2 & \cdots & \hat{\sigma}_{pp}^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \hat{\Sigma}_{11} + \hat{\Sigma}_{22} + \hat{\Sigma}_{12} + \hat{\Sigma}_{21}$$

$$= \hat{\Sigma}_{11} + \hat{\Sigma}_{22} + 2\hat{\Sigma}_{12}$$

For p components

$$j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\mathrm{var}(j^T\hat{\theta}) = j^T \mathrm{var}(\hat{\theta}) j$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1p} \\ \vdots & & \\ \hat{\sigma}_{p1} & \cdots & \hat{\sigma}_{np} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \sum_{j=1}^{p} \sum_{i=1}^{p} (\Sigma_{ij})$$

(B)

$$f(\hat{\theta}) \approx f(\theta) + (\nabla f(\theta))^T (\hat{\theta} - \theta)$$

$$\mathrm{var}(f(\hat{\theta})) = \mathrm{var}[f(\theta) + (\nabla f(\theta))^T (\hat{\theta} - \theta)]$$

$$= \mathrm{var}[(\nabla f(\theta))^T (\hat{\theta} - \theta)]$$

$$= [\nabla f(\theta)]^T \mathrm{var}(\hat{\theta} - \theta) [\nabla f(\theta)]$$

3 main approximations

1. Taylor expansion
2. Gradient
3. Nuisance parameter