Preliminaries

a.) $\omega \sim beta(a,b)$ $\quad X_i \sim bernoulli(\omega)$

$$p(w|X_{1:N}) \propto p(X_1,\dots,X_N|w)\,p(w) \qquad \text{Let } n=\sum_1^{} X_i$$

$$= \binom{N}{n} w^n (1-w)^{N-n} \frac{\Gamma(\alpha+b)}{\Gamma(\alpha)\Gamma(b)} w^{a-1} (1-w)^{b-1}$$

$$= \binom{N}{n}\frac{\Gamma(\alpha+b)}{\Gamma(\alpha)\Gamma(b)} w^{a+n-1} (1-w)^{b+N-n-1} \quad \leftarrow \text{beta}(a+n, b+N-n) \text{ kernel}$$

so $\boxed{p(w|X_{1:N}) \sim Beta\left(a+\sum_{i=1}^{n} X_i,\ b+N-\sum_{i=1}^{n} X_i\right)}$

b.) $X_1 \sim Ga(a_1,1)$ $\quad X_2 \sim Ga(a_2,1)$

$y_1 = \frac{X_1}{X_1+X_2}$ in $[0,1]$ $\qquad y_2 = X_1+X_2$ in $\mathbb{R}^+$ $\qquad$ I think this is one-to-one

$$X_1 = y_1(X_1+X_2) = y_1 y_2$$

$$X_2 = y_2 - X_1 = y_2 - y_1y_2 = y_2(1-y_1)$$

$$J = \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{bmatrix}$$

$$|J| = y_2(1-y_1) - y_1(-y_2) = y_2 - y_1y_2 + y_1y_2 = y_2$$

$|y_2| = y_2$ bc X_1, X_2 positive

$$f(y_1,y_2) = f_{X_1}(y_1y_2)\, f_{X_2}(y_2(1-y_1))\, y_2$$

$$= \frac{1}{\Gamma(a_1)}(y_1y_2)^{a_1-1}\exp(-y_1y_2)\cdot\frac{1}{\Gamma(a_2)}[y_2(1-y_1)]^{a_2-1}\exp(-y_2+y_1y_2)\, y_2$$

$$= \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} y_2^{a_1-1+a_2-1+1} (1-y_1)^{a_2-1}\exp(-y_2)$$

$$= \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} y_2^{a_1+a_2-1} (1-y_1)^{a_2-1}\exp(-y_2)$$

Marginal distribution of y_1:

$$p(y_1) = \int_0^\infty \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} y_2^{a_1+a_2-1} (1-y_1)^{a_2-1}\exp(-y_2)\,dy_2$$

$$= \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1}\int_0^\infty y_2^{a_1+a_2-1}\exp(-y_2)\,dy_2 \quad \leftarrow \text{gamma}(a_1+a_2, 1) \text{ kernel}$$

$$= \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} \quad \leftarrow \boxed{\text{beta}(a_1,a_2) \text{ distribution}}$$

Marginal distribution of y_2:

$$\int_0^1 \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1} y_2^{a_1+a_2-1} (1-y_1)^{a_2-1} \exp(-y_2)\, dy_1$$

$$= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_2^{a_1+a_2-1} \exp(-y_2) \int_0^1 y_1^{a_1-1}(1-y_1)^{a_2-1} dy_1 \quad \leftarrow \text{beta}(a_1,a_2) \text{ kernel}$$

$$= \frac{1}{\Gamma(a_1+a_2)} y_2^{a_1+a_2-1} \exp(-y_2) \rightarrow \boxed{\text{Gamma}(a_1+a_2, 1) \text{ distribution}}$$

To simulate beta(a_1,a_2) random variables, generate $X_1 \sim ga(a_1,1)$ and $X_2 \sim ga(a_2,1)$ and compute $\frac{X_1}{X_1+X_2}$.

c) $\theta \sim N(m, v^2)$ $\quad X_i \sim N(\theta, \sigma^2)$

$$p(\theta \mid X_1,\dots,X_N) \propto p(X_1,\dots,X_N \mid \theta)\, p(\theta)$$

$$= \left(\prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{X_i-\theta}{\sigma}\right)^2\right\}\right)\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\exp\left\{\frac{-1}{2}\left(\frac{\theta-m}{v}\right)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{N+1} \exp\left\{\frac{-1}{2}\left[\left(\frac{\theta-m}{v}\right)^2 + \sum_{i=1}^{N}\left(\frac{X_i-\theta}{\sigma}\right)^2\right]\right\}$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}(X_i-\theta)^2$$

$$= \frac{1}{\sigma^2}\sum(X_i^2 - 2\theta X_i + \theta^2)$$

$$= \frac{n\theta^2}{\sigma^2} - \frac{2\theta n\bar{X}}{\sigma^2} + \frac{1}{\sigma^2}\sum X_i^2$$

$$\propto \exp\left\{\frac{-1}{2}\left[\frac{\theta^2}{v^2} - \frac{2m}{v^2}\theta + \frac{n\theta^2}{\sigma^2} - \frac{2\theta n\bar{X}}{\sigma^2}\right]\right\}$$

complete the square

$$\left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)\theta^2 + \left(\frac{-2m}{v^2} - \frac{2n\bar{X}}{\sigma^2}\right)\theta$$

$$= \underbrace{\left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)}_{a}\theta^2 + \underbrace{\left(\frac{-2m}{v^2} - \frac{2n\bar{X}}{\sigma^2}\right)}_{b}\theta + \left(\frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2$$

$$= \left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)\left(\theta - \left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)^{-1}\left(\frac{m}{v^2}+\frac{n\bar{X}}{\sigma^2}\right)\right)^2 + c$$

$$\propto \exp\left\{\frac{-1}{2}\left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)\left(\theta - \left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)^{-1}\left(\frac{m}{v^2}+\frac{n\bar{X}}{\sigma^2}\right)\right)^2\right.$$

This is the kernel of the $\boxed{\text{Normal}(\mu_N, \sigma_N^2) \text{ w/ } \sigma_N^2 = \left(\frac{1}{v^2}+\frac{n}{\sigma^2}\right)^{-1}, \mu_N = \sigma_N^2\left(\frac{m}{v^2}+\frac{n\bar{X}}{\sigma^2}\right)}$

d.) $\omega \sim \text{gamma}(a,b)$

$$p(\omega \mid x_1,\dots,x_N) \propto \frac{b^a}{\Gamma(a)}\, \omega^{a-1} \exp(-b\omega) \prod_{i=1}^{N} \left(\frac{\omega}{2\pi}\right)^{1/2} \exp\left\{\frac{-\omega}{2}(x_i-\theta)^2\right\}$$

$$\propto \omega^{a-1}\exp(-b\omega)\, \omega^{N/2} \exp\left\{\frac{-\omega}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}$$

$$= \omega^{a+\frac{N}{2}-1} \exp\left\{-\left(b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right)\omega\right\}$$

the posterior is $\boxed{\text{Gamma}\left(a+\frac{N}{2},\ b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right)}$

$\omega = \frac{1}{\sigma^2}$ $\quad \sigma^2 = \frac{1}{\omega}$ using change of variable:

$$p(\sigma^2 \mid x_1,\dots,x_N) = p_\omega\left(\frac{1}{\sigma^2}\right)\cdot\left|\frac{-1}{(\sigma^2)^2}\right|$$

$$= \left(\frac{1}{\sigma^2}\right)^{a+\frac{N}{2}-1} \exp\left\{-\left(b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right)\left(\frac{1}{\sigma^2}\right)\right\}\cdot\left(\frac{1}{\sigma^2}\right)^2$$

$$= (\sigma^2)^{-\left(a+\frac{N}{2}\right)-1} \exp\left\{-\left(b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right)\left(\frac{1}{\sigma^2}\right)\right\}$$

$$\to \boxed{IG\left(a+\frac{N}{2},\ b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right)}$$

e.) $x_i \sim N(\theta, \sigma_i^2)$ $\quad \theta \sim N(m, v^2)$ $\quad$ (I used v^2 as the variance rather than v out of habit.)

$$p(\theta \mid x_1,\dots,x_N) \propto p(x_1,\dots,x_N \mid \theta)\, p(\theta)$$

$$= \left(\prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{x_i-\theta}{\sigma_i}\right)^2\right\}\right) \frac{1}{v\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{\theta-m}{v}\right)^2\right\}$$

Lemma:

$\exp\left\{\frac{-1}{2}(-2b\theta + A\theta^2)\right\}$

$N(A^{-1}b, A^{-1})$

$$\propto \exp\left\{\frac{-1}{2}\left[\sum_{i=1}^{N}\left(\frac{x_i-\theta}{\sigma_i}\right)^2 + \left(\frac{\theta-m}{v}\right)^2\right]\right\}$$

$$\frac{1}{\sigma_i^2}(\cancel{x_i^2} - 2x_i\theta + \theta^2) + \frac{1}{v^2}(\theta^2 - 2m\theta + \cancel{m^2})$$

$$\left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)\theta^2 - 2\sum_{i=1}^{N}\left(\frac{x_i}{\sigma_i^2}\right)\theta + \frac{1}{v^2}\theta^2 - \frac{2m\theta}{v^2}$$

$$\propto \exp\left\{\frac{-1}{2}\left[\left(\frac{1}{v^2}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)\theta^2 - 2\left(\frac{m}{v^2}+\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}\right)\theta\right]\right\}$$

$$= \exp\left\{\frac{-1}{2}\left(\frac{1}{v^2}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)\left[\theta^2 - 2\left(\frac{1}{v^2}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)^{-1}\left(\frac{m}{v^2}+\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}\right)\theta\right]\right\}$$ complete the square

$$\propto \exp\left\{\frac{-1}{2}\left(\frac{1}{v^2}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)\left[\theta - \left(\frac{1}{v^2}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)^{-1}\left(\frac{m}{v^2}+\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}\right)\right]^2\right\}$$

the posterior mean: $\boxed{\underbrace{\left(\frac{1}{v^2}+\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)^{-1}}_{\text{sum of all weights}}\left(\underbrace{\frac{m}{v^2}}_{\text{prior mean}}+\sum_{i=1}^{N}\underbrace{\frac{\overbrace{x_i}^{\text{observation}}}{\sigma_i^2}}_{\text{weight}}\right)}$

each observation mean weighted by reciprocal of its variance.

f.) $X|\omega \sim N(m, \omega^{-1})$ $\omega \sim gamma(\frac{a}{2}, \frac{b}{2})$

$$p(x) = \int_0^\infty \frac{1}{\sqrt{\omega^{-1}}\sqrt{2\pi}} \exp\left\{\frac{-1}{2}\left(\frac{x-m}{\sqrt{\omega^{-1}}}\right)^2\right\} \frac{(b/2)^{a/2}}{\Gamma(\frac{\alpha}{2})} \omega^{\frac{a}{2}-1} \exp\left(\frac{-b}{2}\omega\right) d\omega$$

$$= \frac{1}{\sqrt{2\pi}} \frac{(b/2)^{a/2}}{\Gamma(\frac{\alpha}{2})} \int_0^\infty \omega^{\frac{1}{2}+\frac{a}{2}-1} \exp\left\{\frac{-1}{2}\omega (x-m)^2 - \frac{1}{2}b\omega\right\} d\omega$$

$$= \frac{1}{\sqrt{2\pi}} \frac{(b/2)^{a/2}}{\Gamma(\frac{\alpha}{2})} \int_0^\infty \omega^{\frac{1}{2}+\frac{a}{2}-1} \exp\left\{\frac{-1}{2}[(x-m)^2+b]\,\omega\right\} d\omega$$

Gamma$(\frac{1}{2}+\frac{a}{2}, \frac{1}{2}[(x-m)^2+b])$ kernel

$$= \frac{1}{\sqrt{2\pi}} \frac{(b/2)^{a/2}}{\Gamma(\frac{\alpha}{2})} \cdot \frac{\Gamma(\frac{1}{2}+\frac{\alpha}{2})}{[\frac{1}{2}(x-m)^2+\frac{b}{2}]^{\frac{\alpha+1}{2}}} = \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\frac{\alpha}{2})} \cdot \frac{(b/2)^{a/2}}{\sqrt{2\pi}} \left[\frac{(x-m)^2}{b}+1\right]^{-\frac{(\alpha+1)}{2}} \left(\frac{b}{2}\right)^{\frac{\alpha+1}{2}}$$

$$= \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{b\pi}\,\Gamma(\frac{\alpha}{2})} \left[\frac{(x-m)^2}{b}+1\right]^{-\frac{(\alpha+1)}{2}}$$

$\rightarrow$ t-distribution with b degrees of freedom, center m

scale: $\frac{b}{a}$

Multivariate Normal

a.) $\mathrm{Cov}(X) = E[(X-\mu)(X-\mu)'] = E[(X-\mu)(X'-\mu')]$ because $(A+B)' = A'+B'$

$= E[XX' - X\mu' - \mu X' + \mu\mu']$ distribute

$= E(XX') - E(X)\mu' - \mu E(X') + \mu\mu'$ linearity of expectation

$= E(XX') - \mu\mu' - \mu\mu' + \mu\mu'$ bc $E(X)=\mu$, $E(X')=\mu'$

$= E(XX') - \mu\mu'$

$\mathrm{Cov}(AX+b) = E[((AX+\cancel{b}) - (A\mu+\cancel{b}))((AX+\cancel{b}) - A\mu+\cancel{b}))']$ $E(AX+b) = AE(X)+b$ $= A\mu+b$

$= E[(AX-A\mu)(AX-A\mu)']$

$= E[A(X-\mu)(A(X-\mu))']$

$= E[A(X-\mu)(X-\mu)'A']$ bc $(AB)' = B'A'$

$= A\underbrace{E[(X-\mu)(X-\mu)']}_{\mathrm{cov}(X)}A'$ linearity of expectation

$= A\,\mathrm{Cov}(X)A'$

b.) $Z_i \sim N(0,1)$

$f_Z(z) = \prod_{i=1}^{p} f_{Z_i}(z_i) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}z_i^2\} = (2\pi)^{-p/2} \exp\{-\frac{1}{2}\sum_{i=1}^{p} z_i^2\}$

$= \boxed{(2\pi)^{-p/2} \exp\{-\frac{1}{2}z'z\}}$

$E[\exp\{t'Z\}] = E[\exp\{t_1Z_1 + \dots + t_pZ_p\}] = E[\prod_{i=1}^{p}\exp(t_iZ_i)] = \prod_{i=1}^{p} \underbrace{E[\exp(t_iZ_i)]}_{M_{Z_i}(t_i)} = \prod_{i=1}^{p} e^{t_i^2/2}$

due to independence

$= e^{\frac{1}{2}\sum_{i=1}^{p} t_i^2} = \boxed{e^{\frac{1}{2}t't}}$

c.) Prove that for any $a \neq 0$, $Z = a'X \sim N(m, \sigma^2) \Leftrightarrow$ mgf of X $E[\exp(t'X)] = \exp\{t'\mu + \frac{1}{2}t'\Sigma t\}$
$X \sim N(\mu, \Sigma)$

First we find the mean, variance, and mgf of $Z = a'X$.

$E[a'X] = a'EX = a'\mu$ $\quad \mathrm{Var}(a'X) = a'\mathrm{Var}(X)a = a'\Sigma a$

For a univariate normal, say $Y \sim N(m, \sigma^2)$

$$M_Y(t) = E[\exp\{tY\}] = \int_{-\infty}^{\infty} e^{yt} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-m}{\sigma}\right)^2} dy$$

Let $u = \frac{y-m}{\sigma}$ then $du = \frac{1}{\sigma} dy$ and $y = \sigma u + m$

then $$M_Y(t) = \int_{-\infty}^{\infty} e^{(\sigma u + m)t} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \sigma \, du$$

$$= e^{mt} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\sigma u t - \frac{1}{2}u^2} du$$

↳ equal to $E[\exp(\sigma t U)]$ where $U \sim N(0,1)$

$$= e^{mt} e^{\frac{1}{2}(\sigma t)^2}$$

$p=1$ case of mgf from part (b)

$$= e^{mt + \frac{1}{2}\sigma^2 t^2}$$

So for $Y \sim N(m, \sigma^2)$, $M_Y(t) = e^{mt + \frac{1}{2}\sigma^2 t^2}$.

$(\Rightarrow)$ Let $X \sim N(\mu, \Sigma)$. Assume $Z = a'X$ for $a \neq 0$ is normally distributed. We show that $E[\exp\{t'X\}] = \exp(t'\mu + \frac{1}{2}t'\Sigma t)$.

$E(Z) = E(a'X) = a'E(X) = a'\mu$

$Var(Z) = a' cov(X) a = a'\Sigma a$

Then $M_Z(t) = \exp\{a'\mu t + \frac{1}{2}a'\Sigma a t^2\} = \exp\{\mu' a t + (at)'\Sigma (at)\}$

Now let $s = at$. Then

?

$(\Leftarrow)$ We show that $M_X(t) = \exp\{t'\mu + \frac{1}{2}t'\Sigma t\} \Rightarrow Z = a'X$ is normal for all $a \neq 0$.

$$M_Z(t) = E[\exp\{t Z\}] = E[\exp\{ta'X\}]$$
$$= M_X(ta')$$
$$= \exp\{(ta')\mu + \tfrac{1}{2}(ta')\Sigma (ta')\}$$
$$= \exp\{t\underbrace{(a'\mu)}_{E(Z)} + \tfrac{1}{2}\underbrace{(a'\Sigma a)}_{Var(Z)} t^2\}$$

By uniqueness of MGFs, Z is normal.

$X \sim N(\mu, \Sigma) \quad \forall a \neq 0 \quad a'X \sim N(a'\mu, a'\Sigma a)$

$$E[\exp\{ta'X\}] = \exp\{ta'\mu + \tfrac{1}{2}t^2 a'\Sigma a$$

$$= \exp\{ \underbrace{(ta)}_{\tilde{t}}'\mu + \tfrac{1}{2}(ta)'\Sigma(ta)\}$$

$$\text{MGF of } X = E[\exp\{\tilde{t}'X\}] = \exp\{\tilde{t}'\mu + \tfrac{1}{2}\tilde{t}'\Sigma\tilde{t}\}$$

d.) $Z \sim N(0, I_p)$ $X = LZ + \mu$ Prove that X is MVN.

We have that $M_Z(t) = E[\exp\{t'Z\}]$.

from part (b.)

Then $M_X(t) = E[\exp\{t'(LZ+\mu)\}] = e^{t'\mu} E[\exp\{t'LZ\}] = e^{t'\mu} M_Z((t'L)') = e^{t'\mu} M_Z(L't) = e^{t'\mu} e^{\frac{1}{2}(L't)'(L't)} = e^{t'\mu + \frac{1}{2}t'LL't}$

$= e^{t'\mu + \frac{1}{2}t'\Sigma t}$

$E[LZ+\mu] = LE(Z) + \mu = 0 + \mu = \mu$

$Cov(LZ+\mu) = L\,Cov(Z)L' = LL'$ (from part (a.))

Therefore X is MVN, using part (c.)

e.) $X \sim MVN(\mu, \Sigma)$. Prove that X is affine transformation of standard normals.

Σ is symmetric because it is a covariance matrix, thus by spectral decomposition we can write $\Sigma = PDP'$ where P is orthonormal and D is diagonal. So $\Sigma = PD^{1/2}D^{1/2}P' = (PD^{1/2})(PD^{1/2})'$ Let $A = PD^{1/2}$.

↳ λ_i's all ≥ 0 bc PSD

Let Z be n-dimensional, same as X. Then consider the random vector

$$AZ + \mu$$

We know that this random vector is multivariate normal. Then we also have

$$E(AZ+\mu) = AE(Z) + \mu = \mu$$

$$Cov(AZ+\mu) = A\,Cov(Z)A' = AA' = (PD^{1/2})(PD^{1/2})' = PD^{1/2}D^{1/2}P' = PDP' = \Sigma$$

which means $AZ + \mu \overset{d}{=} X$.

f.) $f_Z(z) = \frac{1}{(2\pi)^{k/2}} \exp\{-\frac{1}{2}z'z\}$

$X \sim N(\mu, \Sigma)$ $p(X) = C\exp\{-\frac{1}{2}Q(x-\mu)\}$

Multivariate $X = AZ + \mu$ for some A such that $\Sigma = AA'$ as shown above. $A = PD^{1/2}$ then $A^{-1} = (D^{1/2})^{-1}P^{-1} = D^{1/2}P$

Using our transformations of random variables, we have that (assuming A full rank)

↳ true bc P is orthonormal

$$f_X(x) = f_Z(A^{-1}(x-\mu))\,|A^{-1}|$$

$$= \frac{1}{(2\pi)^{k/2}} \exp\{-\frac{1}{2}(A^{-1}(x-\mu))'(A^{-1}(x-\mu))\}\,|A^{-1}|$$

$|A^{-1}| = \frac{1}{|A|}$

$$= \frac{1}{(2\pi)^{k/2}|A|} \exp\{-\frac{1}{2}(x-\mu)'(A^{-1})'A^{-1}(x-\mu)\}$$

$(A^{-1})'A^{-1} = (D^{1/2}P)'(D^{1/2}P) = P'DP = \Sigma^{-1}$

$$= \frac{1}{(2\pi)^{k/2}|\Sigma|^{-1/2}} \exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}$$

← bc $\Sigma = AA^T$

g.) $X_1 \sim N(\mu_1, \Sigma_1)$ $X_2 \sim N(\mu_2, \Sigma_2)$ $y = AX_1 + BX_2$.

$AX_1 \sim N(A\mu_1, A\Sigma_1 A')$ $M_{AX_1}(t) = \exp\{t'A\mu_1 + \frac{1}{2}t'A\Sigma_1A't\}$

$BX_2 \sim N(B\mu_2, B\Sigma_2 B')$ $M_{BX_2}(t) = \exp\{t'B\mu_2 + \frac{1}{2}t'B\Sigma_2B't\}$

$M_{AX_1+BX_2}(t) = M_{AX_1}(t)M_{BX_2}(t) = \exp\{t'(A\mu_1 + B\mu_2) + \frac{1}{2}t'(A\Sigma_1A' + B\Sigma_2B')t\}$

$\Rightarrow \boxed{AX_1 + BX_2 \sim MVN(A\mu_1 + B\mu_2,\ A\Sigma_1A' + B\Sigma_2B')}$

Conditionals and Marginals

a) $X_1 = \underbrace{\begin{bmatrix} I_k & \vdots & \underset{n-k}{0} \end{bmatrix}}_{M} X \qquad X \sim N(\mu, \Sigma_1)$

then X_1 is MVN

$E(X_1) = M E(X) = M \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mu_1$

$Cov(X_1) = M \Sigma_1 M^T = \begin{bmatrix} I_k \mid 0 \end{bmatrix} \begin{bmatrix} \Sigma_{1\,11} & \Sigma_{1\,12} \\ \Sigma_{1\,21} & \Sigma_{1\,22} \end{bmatrix} \begin{bmatrix} I_k \\ \hline 0 \end{bmatrix} = \Sigma_{1\,11}$

so $\boxed{X_1 \sim N(\mu_1, \Sigma_{1\,11})}$

b) If a matrix can be partitioned into 4 blocks, it can be inverted blockwise, as follows (from Wikipedia):

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1} BD^{-1} \\ -(D - CA^{-1}B)^{-1} CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

Since we have

$$\Omega = \begin{bmatrix} \Sigma_{1\,11} & \Sigma_{1\,12} \\ \Sigma_{1\,21} & \Sigma_{1\,22} \end{bmatrix}^{-1}$$

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

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

$\Omega_{21} = \Omega_{12}'$

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

c) $p(X_1, X_2) = \det(2\pi\Sigma_1)^{-1/2} \exp\{-\frac{1}{2}(x-\mu)' \Sigma_1^{-1} (x-\mu)\}$

$= \det(2\pi\Sigma_1)^{-1/2} \exp\{-\frac{1}{2} \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}' \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}\}$

$= \det(2\pi\Sigma_1)^{-1/2} \exp\{-\frac{1}{2}[(X_1-\mu_1)'\Omega_{11}(X_1-\mu_1) - (X_2-\mu_2)'\Omega_{21}(X_1-\mu_1) - (X_1-\mu_1)'\Omega_{12}(X_2-\mu_2) + (X_2-\mu_2)'\Omega_{22}(X_2-\mu_2)]\}$

Now we try to identify $p(X_1|X_2)$ by dropping unrelated terms:

$p(X_1|X_2) \propto \exp\{-\frac{1}{2}[(X_1-\mu_1)'\Omega_{11}(X_1-\mu_1) - 2(X_2-\mu_2)'\Omega_{12}(X_1-\mu_1)]\}$

complete the square: $x'Mx - 2b'x = (x - M^{-1}b)'M(x - M^{-1}b) - b'M^{-1}b$

Here let $x = X_1 - \mu_1$

$M = \Omega_{11}$

$b' = (X_2-\mu_2)'\Omega_{12} \Rightarrow b = \Omega_{21}(X_2-\mu_2)$

$= \exp\{-\frac{1}{2}[((X_1-\mu_1) - \Omega_{11}^{-1}\Omega_{21}(X_2-\mu_2))' \Omega_{11}^{-1} ((X_1-\mu_1) - \Omega_{11}^{-1}\Omega_{21}(X_2-\mu_2)) - \cancel{c}]\}$

then $\boxed{p(X_1|X_2) \sim N(\mu_1 + \Omega_{11}^{-1}\Omega_{21}(X_2-\mu_2), \Omega_{11}^{-1})}$ where $\Omega_{11}^{-1} = \Sigma_{1\,11} - \Sigma_{1\,12}\Sigma_{1\,22}^{-1}\Sigma_{1\,21}$ as above.

Multiple Regression

a) Least squares: $\hat{\beta} = \underset{\beta}{\arg\min}\ (Y - X\beta)'(Y - X\beta)$

$$= \underset{\beta}{\arg\min}\ Y'Y - Y'X\beta - (X\beta)'Y + \beta'X'X\beta$$

$$= \underset{\beta}{\arg\min}\ -2\beta'X'Y + \beta'X'X\beta$$

Using gradients,

$$-2X'Y + (X'X + (X'X)')\hat{\beta}_{LS} = 0$$

$$X'X\hat{\beta}_{LS} = X'Y$$

$$\hat{\beta}_{LS} = (X'X)^{-1}X'Y \quad \text{assuming } X \text{ full rank}$$

MLE under Gaussianity

$$\hat{\beta} = \underset{\beta}{\arg\max} \prod_{i=1}^{n} p(y_i \mid \beta, \sigma^2) \qquad y_i \sim N(x_i'\beta, \sigma^2)$$

$$= \underset{\beta}{\arg\max} \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right\}$$

$$= \underset{\beta}{\arg\max}\ -\frac{1}{2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2$$

$$= \underset{\beta}{\arg\min} \sum_{i=1}^{n}(y_i - x_i'\beta)^2 = \underset{\beta}{\arg\min}\ (Y - X\beta)'(Y - X\beta)$$

which is equivalent to the $\hat{\beta}_{LSE}$ formulation.

For method of moments: choose $\hat{\beta}$ such that

$$Cov(\varepsilon, x_p) = 0 \quad \text{or}$$

$$\sum_{i=1}^{n}(e_i - \bar{e})(x_{ij} - \bar{x}_j) = 0 \quad \text{(sample covariance)}$$

$$= \sum_{i=1}^{n}(e_i x_{ij} - e_i\bar{x}_j - \bar{e}x_{ij} + \bar{e}\bar{x}_j)$$

$$= \sum_{i=1}^{n} e_i x_{ij} - n\bar{e}\bar{x}_j - n\bar{e}\bar{x}_j + n\bar{e}\bar{x}_j$$

$$= \sum_{i=1}^{n} e_i x_{ij} + n\bar{e}\bar{x}_j \quad \text{we could center } x_p\text{'s wlog so } \bar{x}_j = 0.$$

so this is $\sum_{i=1}^{n} e_i x_{ij} = 0$

so we have $\forall j = 1, \dots, p$ that $\sum_{i=1}^{n} e_i x_{ij} = 0$.

To put this concisely into matrix form is to say that

$$X'e = 0 \leftarrow p \times 1 \text{ zero vector}$$

$$X'(Y - Xb) = 0 \quad \text{bc } e = Y - Xb$$

$$X'Y - X'Xb = 0$$

$$\Rightarrow \hat{\beta}_{MOM} = (X'X)^{-1}X'Y$$

Now we find $Var(\hat{\beta}) = Var((X'X)^{-1}X'Y)$

$$= (X'X)^{-1}X'\, Var(Y)\, ((X'X)^{-1}X')' \qquad Var(Y) = \sigma^2 I$$

$$= (X'X)^{-1}X'(\sigma^2 I)X((X'X)^{-1})^T \qquad \text{bc } (X'X)^{-1} = ((X'X)^{-1})^T$$

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

b.) $y \sim N(X\beta, \Sigma)$

$$\hat{\beta} = \underset{\beta}{\arg\max} \; \det(2\pi\Sigma)^{-1/2} \exp\{-\tfrac{1}{2}(Y-X\beta)'\Sigma^{-1}(Y-X\beta)\}$$

$$= \underset{\beta}{\arg\min} \; (Y-X\beta)'\Sigma^{-1}(Y-X\beta)$$

$$= \underset{\beta}{\arg\min} \; \cancel{Y'\Sigma^{-1}Y} - Y'\Sigma^{-1}X\beta - \beta'X'\Sigma^{-1}Y + \beta'X'\Sigma^{-1}X\beta$$

$$= \underset{\beta}{\arg\min} \; -2Y'\Sigma^{-1}X\beta + \beta'X'\Sigma^{-1}X\beta$$

$\Sigma = PDP^{-1}$

$\Sigma^{-1} = (PDP^{-1})^{-1} = PD^{-1}P^{-1}$

$(\Sigma^{-1})^T = (PD^{-1}P^{-1})^T = (P^{-1})^T(D^{-1})^T P^T$

$= PD^{-1}P^{-1}$ so $\Sigma^{-1} = (\Sigma^{-1})^T$

Again take matrix gradients:

$$-2(Y'\Sigma^{-1}X)' + (X'\Sigma^{-1}X + (X'\Sigma^{-1}X)')\hat{\beta} = 0$$

$$-2X'\Sigma^{-1}Y + 2X'\Sigma^{-1}X\hat{\beta} = 0$$

$$X'\Sigma^{-1}X\hat{\beta} = X'\Sigma^{-1}Y$$

$$\Rightarrow \boxed{\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y}$$

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y]$$

$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathrm{Var}(Y)[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}]^T$$

$$= (X'\Sigma^{-1}X)^{-1}X'\cancel{\Sigma^{-1}\Sigma}\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}$$

$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}$$

$$= \boxed{(X'\Sigma^{-1}X)^{-1}}$$

c.) $\hat{\beta}_{MLE} = (X'\Sigma^{-1}X)^{-}X'\Sigma^{-1}Y$ from above.

Note that $\Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2)$ so $\Sigma^{-1} = \mathrm{diag}(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_n^2})$.

$(X'\Sigma X)\hat{\beta} = X'\Sigma Y$

$\hat{\beta} = (X'\Sigma X)^{-}X'\Sigma Y$

weighted least squares: $\underset{\beta}{\min} \sum_{i=1}^{n} w_i(y_i - X_i'\beta)^2$

Some Practical Details

a.) Inversion is much slower

solving system gives you $\frac{2}{3}n^3+2n^2$ flops

direct inversion gives $\frac{14}{3}n^3$ flops

Inversion isn't more stable either — matrix may be ill-conditioned.

solve for $\hat{\beta}$ in $X'WX\hat{\beta}=X'Wy$

Factor $(X'WX)=LU$

solve for z in $Lz=X'Wy$ forward substitution

solve for $\hat{\beta}$ in $U\hat{\beta}=z$ using backward substitution

matlib inv