SDS 383D Exercises 1: Preliminaries

Preston Biro

November 2018

1 Bayesian Inference in Simple Conjugate Families

A

Sol:

$$
\begin{aligned}
p(w|X_1,\dots,X_N) &\propto p(w)p(X_1,\dots,X_N|w) \\
p(X_1,\dots,X_N|w) &= \prod_{i=1}^{N} w^{X_i}(1-w)^{1-X_i} \\
p(w) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1}(1-w)^{b-1} \\
\implies p(w|X_1,\dots,X_N) &\propto w^{a-1}(1-w)^{b-1}\prod_{i=1}^{N} w^{X_i}(1-w)^{1-X_i} \\
&= w^{a-1}(1-w)^{b-1} w^{\sum_{i=1}^{N} X_i}(1-w)^{n-\sum_{i=1}^{N} X_i} \\
&= w^{a+n\bar{X}-1}(1-w)^{b+n(1-\bar{X})-1} \\
\implies w|X_{1:N} &\sim Beta(a+n\bar{X}, b+n(1-\bar{X}))
\end{aligned}
$$

■

B

Sol:

$$
\begin{aligned}
y_1 &= \frac{x_1}{x_1 + x_2} \\
y_2 &= x_1 + x_2 \\
y_1 &= \frac{x_1}{y_2} \\
\implies x_1 &= y_1 y_2 \\
y_2 &= y_1 y_2 + x_2 \\
\implies x_2 &= y_2(1 - y_1) \\
f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(y_1 y_2, y_2(1 - y_1))|J| \\
|J| &= \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_2}{\partial y_1}\frac{\partial x_1}{\partial y_2} \\
&= y_2(1 - y_1) + y_2 y_1 = y_2 \\
\implies f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(y_1 y_2, y_2(1 - y_1))|J| \\
&= f_{X_1}(y_1 y_2) f_{X_2}(y_2(1 - y_1)) y_2 \\
&= \frac{(y_1 y_2)^{a_1 - 1}}{\Gamma(a_1)} e^{-y_1 y_2} \frac{(y_2(1 - y_1))^{a_2 - 1}}{\Gamma(a_2)} e^{-y_2(1 - y_1)} y_2 \\
f_{Y_1,Y_2}(y_1, y_2) &= \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1} y_2^{a_1 + a_2 - 1}}{\Gamma(a_1)\Gamma(a_2)} e^{-y_2} \\
p(y_1) &= \int f_{Y_1,Y_2}(y_1, y_2) dy_2 \\
&= \int \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1} y_2^{a_1 + a_2 - 1}}{\Gamma(a_1)\Gamma(a_2)} e^{-y_2} dy_2 \\
&= \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1}\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \int \frac{y_2^{a_1 + a_2 - 1}}{\Gamma(a_1 + a_2)} e^{-y_2} dy_2 \\
&= \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1}}{Beta(a_1, a_2)} \int Ga(a_1 + a_2, 1) dy_2 \\
p(y_1) &= \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1}}{Beta(a_1, a_2)} \\
\implies Y_1 &\sim Beta(p = y_1, a_1, a_2) \\
p(y_2) &= \int f_{Y_1,Y_2}(y_1, y_2) dy_1 \\
&= \int \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1} y_2^{a_1 + a_2 - 1}}{\Gamma(a_1)\Gamma(a_2)} e^{-y_2} dy_1 \\
&= \frac{y_2^{a_1 + a_2 - 1} e^{-y_1}}{\Gamma(a_1 + a_2 - 1)} \int \frac{y_1^{a_1 - 1}(1 - y_1)^{a_2 - 1}}{Beta(a_1, a_2)} dy_1 \\
p(y_2) &= \frac{y^{a_1 + a_2 - 1} e^{-y_2}}{\Gamma(a_1 + a_2 - 1)} \\
\implies Y_2 &\sim Ga(a_1 + a_2, 1)
\end{aligned}
$$

■

C

Sol:

$$
\begin{aligned}
\theta &\sim N(m, v) \\
x_i &\sim N(\theta, \sigma^2) \\
p(\theta|x_1, \ldots, x_N) &\propto p(\theta)p(x_1, \ldots, x_N|\theta) \\
&\propto e^{\frac{-(\theta-m)^2}{2v}} \prod_{i=1}^{N} e^{\frac{-(x_i-\theta)^2}{2\sigma^2}} \\
&= e^{-\frac{(\theta-m)^2}{2v} - \sum_{i=1}^{N} \frac{(x_i-\theta)^2}{2\sigma^2}} \\
&= e^{-\frac{1}{2}(\frac{(\theta-m)^2}{v} + n\frac{(\theta-\bar{x})^2}{\sigma^2} + \sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{\sigma^2})} \\
&\propto e^{-\frac{1}{2}(\frac{\theta^2-2m\theta}{v} + \frac{n\theta^2-2n\theta\bar{x}}{\sigma^2})} \\
p(\theta|x_1, \ldots, x_N) &\propto e^{-\frac{(\frac{1}{v}+\frac{n}{\sigma^2})}{2}(\theta-(\frac{m}{v}+\frac{n\bar{x}}{\sigma^2})(\frac{1}{v}+\frac{n}{\sigma^2})^{-1})^2} \\
\theta|x_{1:N} &\sim N((\frac{m}{v} + \frac{n\bar{x}}{\sigma^2})(\frac{1}{v} + \frac{n}{\sigma^2})^{-1}, (\frac{1}{v} + \frac{n}{\sigma^2})^{-1})
\end{aligned}
$$

■

D

Sol:

$$
\begin{aligned}
w &\sim Ga(a, b) \\
p(w|x_1, \ldots, x_N) &\propto p(w)p(x_1, \ldots, x_N|w) \\
&\propto w^{a-1}e^{-bw} \prod_{i=1}^{N} (w)^{1/2} e^{-\frac{w}{2}(x_i-\theta)^2} \\
&= w^{a-1}e^{-bw}w^{N/2}e^{-\frac{w}{2}\sum(x_i-\theta)^2} \\
p(w|x_1, \ldots, x_N) &\propto w^{\frac{N}{2}+a-1}e^{-w(b+\frac{\sum(x_i-\theta)^2}{2})} \\
w|x_{1:N} &\sim Ga(\frac{N}{2} + a, b + \frac{\sum(x_i-\theta)^2}{2}) \\
p(\sigma^2|x_1, \ldots, x_N) &\propto (\frac{1}{\sigma^2})^{\frac{N}{2}+a-1}e^{-(\frac{1}{\sigma^2})(b+\frac{\sum(x_i-\theta)^2}{2})} \\
\sigma^2|x_{1:N} &\sim IG(\frac{N}{2} + a, b + \frac{\sum(x_i-\theta)^2}{2})
\end{aligned}
$$

■

E

Sol:

$$
\begin{aligned}
x_i &\sim N(\theta, \sigma_i^2) \\
\theta &\sim N(m, v) \\
p(\theta|x_1, \ldots, x_N) &\propto p(\theta)p(x_{1:N}|\theta) \\
&= \frac{1}{\sqrt{2\pi v}} e^{-\frac{(\theta-m)^2}{2v}} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}} \\
&= \frac{1}{\sqrt{2\pi v}} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(\theta-m)^2}{2v} - \sum_{i=1}^{N} \frac{(x_i-\theta)^2}{2\sigma_i^2}} \\
&\propto e^{-\frac{1}{2}(\frac{\theta^2-2\theta m+m^2}{v} + \sum \frac{\theta^2-2\theta x_i+x_i^2}{\sigma_i^2})} \\
&\propto e^{-\frac{1}{2}\theta^2(\frac{1}{v}+\sum \frac{1}{\sigma_i^2}) - 2\theta(\frac{m}{v}+\sum \frac{x_i}{\sigma_i^2})} \\
&\propto e^{-\frac{1}{2(\frac{1}{v}+\sum \frac{1}{\sigma_i^2})^{-1}}(\theta-(\frac{m}{v}+\sum \frac{x_i}{\sigma_i^2})(\frac{1}{v}+\sum \frac{1}{\sigma_i^2})^{-1})^2} \\
\theta|x &\sim N((\frac{m}{v} + \sum \frac{x_i}{\sigma_i^2})(\frac{1}{v} + \sum \frac{1}{\sigma_i^2})^{-1}, (\frac{1}{v} + \sum \frac{1}{\sigma_i^2})^{-1})
\end{aligned}
$$

■

F

Sol:

$$
\begin{aligned}
p(x,w) &= p(x|w)p(w) \\
&= \frac{w^{1/2}}{\sqrt{2\pi}} e^{-\frac{(x-m)^2 w}{2}} \frac{(b/2)^{a/2}}{\Gamma(a/2)} w^{a/2-1} e^{-bw/2} \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} w^{\frac{a+1}{2}-1} e^{\frac{w}{2}(b+(x-m)^2)} \\
p(x) &= \int p(x,w)dw \\
&= \int_0^\infty \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} w^{\frac{a+1}{2}-1} e^{\frac{w}{2}(b+(x-m)^2)} dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \int_0^\infty w^{\frac{a+1}{2}-1} e^{\frac{w}{2}(b+(x-m)^2)} dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \frac{\Gamma(\frac{a+1}{2})}{(\frac{b+(x-m)^2}{2})^{\frac{a+1}{2}}} \int_0^\infty \frac{(\frac{b+(x-m)^2}{2})^{\frac{a+1}{2}}}{\Gamma(\frac{a+1}{2})} w^{\frac{a+1}{2}-1} e^{\frac{w}{2}(b+(x-m)^2)} dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \frac{\Gamma(\frac{a+1}{2})}{(\frac{b+(x-m)^2}{2})^{\frac{a+1}{2}}} \int_0^\infty Ga(\frac{a+1}{2}, \frac{b+(x-m)^2}{2}) dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \frac{\Gamma(\frac{a+1}{2})}{(\frac{b+(x-m)^2}{2})^{\frac{a+1}{2}}} \\
&= \frac{(\frac{b}{2})^{\frac{a}{2}}\Gamma(\frac{a+1}{2})}{\sqrt{2\pi}\Gamma(\frac{a}{2})} (\frac{b+(x-m)^2}{2})^{-\frac{a+1}{2}} \\
&= \frac{(\frac{b}{2})^{\frac{a}{2}}\Gamma(\frac{a+1}{2})}{\sqrt{2\pi}\Gamma(\frac{a}{2})} (1+\frac{(x-m)^2}{b})^{-\frac{a+1}{2}} (\frac{b}{2})^{-\frac{a+1}{2}} \\
&= \frac{\sqrt{a}}{\sqrt{a}} \frac{\Gamma(\frac{a+1}{2})}{\sqrt{b\pi}\Gamma(\frac{a}{2})} (\frac{a+\frac{a(x-m)^2}{b}}{a})^{-\frac{a+1}{2}} \\
&= \frac{\Gamma(\frac{a+1}{2})}{\sqrt{a\pi}\sqrt{\frac{b}{a}}\Gamma(\frac{a}{2})} (\frac{a+\frac{(x-m)^2}{\sqrt{\frac{b}{a}}}}{a})^{-\frac{a+1}{2}}
\end{aligned}
$$

Which is by definition, a Student t distribution with scale $\sqrt{\frac{b}{a}}$, mean m, and d = a degrees of freedom. ■

2 Multivariate Normal Distribution

A

Sol: Part a)

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

Part b) Using $E[Ax+b] = AE[x] + b = A\mu + b$:

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

■

B

We know the pdf and moment generating function of an independent standard normal distribution are $f_{Z_i}(z_i) = \frac{1}{\sqrt{2\pi}} e^{z_i^2/2}$ and $M_{z_i}(t) = e^{t^2/2}$ respectively. Thus since each of the z_i's are independent, we have:

$$\begin{aligned} f_Z(z_1,\ldots,z_p) &= f_{Z_1}(z_1)f_{Z_2}(z_2)\ldots f_{Z_p}(z_p) \\ &= \frac{1}{\sqrt{2\pi}} e^{z_1^2/2} \frac{1}{\sqrt{2\pi}} e^{z_2^2/2} \cdots \frac{1}{\sqrt{2\pi}} e^{z_p^2/2} \\ &= (\frac{1}{\sqrt{2\pi}})^p e^{\frac{1}{2}\sum_{i=1}^n z_i^2} \\ &= (\frac{1}{\sqrt{2\pi}})^p e^{\frac{z^Tz}{2}} \end{aligned}$$

Also:

$$\begin{aligned} M_Z(t) &= \prod_{i=1}^p E[e^{tz_i}] \\ &= \prod_{i=1}^p e^{t^2/2} \\ &= e^{t^Tt/2} \end{aligned}$$

C

Sol: If x is multivariate normal, then we can assume that for all not identically zero vectors a, $z = a^T x$ and the univariate normal moment generating function $M_z(t) = E[e^{tz}] = e^{\mu t + \sigma^2 t^2/2}$, where μ is the mean and σ^2 is the variance of z. Thus:

$$\begin{aligned} E[z] = E[a^T x] &= a^T E[x] = a^T \mu \\ Var[z] = Var[a^T x] &= a^T Var[x] a = a^T \Sigma a \\ M_z(t) &= E[e^{t^T a^T x}] \\ &= E[e^{\sum_{i=1}^p t_i a_i x_i}] \\ &= \prod_{i=1}^p E[e^{t_i a_i x_i}] \end{aligned}$$

Since the vector a is not identically zero and t can take any value, write $\mathbf{t} = ta$. Thus:

$$\begin{aligned} M_x(t) &= \prod_{i=1}^p E[e^{t_i a_i x_i}] \\ &= \prod_{i=1}^p E[e^{\mathbf{t}_i x_i}] \\ &= \prod_{i=1}^p e^{\mathbf{t}_i \mu_i + \mathbf{t}^2 \sigma_i^2/2} \\ &= e^{\mathbf{t}^T \mu + \mathbf{t}^T \Sigma \mathbf{t}/2} \end{aligned}$$

And since we know that moment generating function are unique, it follows that if a random variable has this moment generating function, then the random variable follows a multivariate normal distribution. ■

D

Sol:

$$\begin{aligned} M_x(t) &= E[e^{t^T x}] \\ &= E[e^{t^T (Lz + \mu)}] \\ &= E[e^{t^T \mu} e^{t^T L z}] \\ &= e^{t^T \mu} E[e^{t^T L z}] \\ &= e^{t^T \mu} e^{t^T L L^T t/2} \\ &= e^{t^T \mu + t^T \Sigma t/2} \end{aligned}$$

Thus the mean vector is μ and covariance matrix is $\Sigma = LL^T$. ■

E

E) Claim: if $X \sim MVN(\mu, \Sigma)$ then $\exists A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^{n \times 1}$ s.t. $X \sim Az + b$ where $Z \sim N(0, I)$

Since Σ is positive definite, we can write

$\Sigma = LL^T$ using a cholesky Transformation.

Let $A = L$ & $b = \mu \Rightarrow X = Lz + \mu$, $z = L^{-1}(x - \mu)$

$\Rightarrow f_X(x) = f_Z(L^{-1}(x-\mu)) \left| \frac{\partial f^2}{\partial z} \right|$

$\frac{\partial f^2}{\partial z} = |L^{-1}| = \sqrt{|L^{-1}|^2} = \sqrt{|L^{-1}||L^{-1}|} = |\Sigma|^{\frac{1}{2}}$

$\Rightarrow = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(L^{-1}(x-\mu))^T (L^{-1}(x-\mu))} |L^{-1}|$

$= \frac{1}{(2\pi)^{n/2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma (x-\mu)}$

F, G

F] Want to show $p(x) = Ce^{\frac{-Q(x-\mu)}{2}}$

We have shown $p(x) = \frac{1}{(2\pi)^{n/2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma (x-\mu)}$

$\Rightarrow C = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}$, $Q = \Sigma$

G] If $X_1 \sim N(\mu_1, \Sigma_1)$ & $X_2 \sim N(\mu_2, \Sigma_2)$, we have shown we can write

$$X_1 = L_1 z_1 + \mu_1 \quad \& \quad X_2 = L_2 z_2 + \mu_2$$

$$\Rightarrow y = AX_1 + BX_2 = A(L_1 z_1 + \mu_1) + B(L_2 z_2 + \mu_2)$$

$$= AL_1 z_1 + A\mu_1 + BL_2 z_2 + B\mu_2$$

$$= \begin{pmatrix} AL_1 & BL_2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + A\mu_1 + B\mu_2 \Rightarrow \text{MVN}$$

Thus we need $E[Y]$ & $Var[Y]$

$$E[Y] = A\mu_1 + B\mu_2$$

$$Var[Y] = A\, Var(X_1) A^T + B\, Var(X_2) B^T$$

$$= A\Sigma_1 A^T + B\Sigma_2 B^T$$

Conditionals and Marginals

A

A)

$X \sim N(\mu, \Sigma)$ $\quad \mu = (\mu_1, \mu_2)^T \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

Using our previous results, we know there is an affine transformation of X such that

$$X_1 = AX + b$$

Such that $X_1 \sim N(\mu', \Sigma')$

Choosing $A = \begin{bmatrix} I_k & O_{k \times p-k} \end{bmatrix}$

& $b = O_k$ makes

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

Since $E[X_1] = A E[X_0] = A\mu = \mu_1$

& $Var(X_1) = A Var(X) A^T = A \Sigma A^T = \Sigma_{11}$

B

B) Since $\Omega = \Sigma^{-1}$, we have $\Sigma\Omega = I$

$$\Rightarrow \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = I$$

1) $\Sigma_{11}\Omega_{11} + \Sigma_{12}\Omega_{21} = I$; 2) $\Sigma_{21}\Omega_{11} + \Sigma_{22}\Omega_{21} = 0$

3) $\Sigma_{11}\Omega_{12} + \Sigma_{12}\Omega_{22} = 0$, 4) $\Sigma_{21}\Omega_{12} + \Sigma_{22}\Omega_{22} = I$

1) $\Rightarrow$ $\Omega_{11} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Sigma_{12}\Omega_{21}$

2) $\Rightarrow$ $\Omega_{21} = -\Sigma_{22}^{-1}\Sigma_{21}\Omega_{11}$

combining $\Rightarrow$ $\Omega_{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Omega_{11}$

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

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

$$\Rightarrow \boxed{\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)^{-1} = \Omega_{11}}$$

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

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

$$\Omega_{22} = \Sigma_{22}^{-1} - \Sigma_{22}^{-1}\Sigma_{21}\Omega_{12}$$

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

C

c) $p(X_1|X_2) = \frac{p(X_1, X_2)}{p(X_2)}$

$p(X_1, X_2) \sim N(\mu, \Sigma)$, $p(X_2) \sim N(\mu_2, \Sigma_{22})$

$\log(p(X_1|X_2)) \propto \log\left(\frac{p(X_1, X_2)}{p(X_2)}\right)$

$= \frac{1}{2}\left[(X_1 - \mu_1)^T \Omega_{11}(X_1 - \mu_1) + 2(X_1 - \mu_1)^T \Omega_{12}(X_2 - \mu_2) + (X_2 - \mu_2)^T \Omega_{22}(X_2 - \mu_2) - (X_2 - \mu_2)^T \Sigma_{22}^{-1}(X_2 - \mu_2)\right]$

$\Rightarrow X_1 - \mu_1 | X_2 \sim N(\Omega_{11}^{-1}\Omega_{12}(X_2 - \mu_2), \Omega_{11}^{-1})$

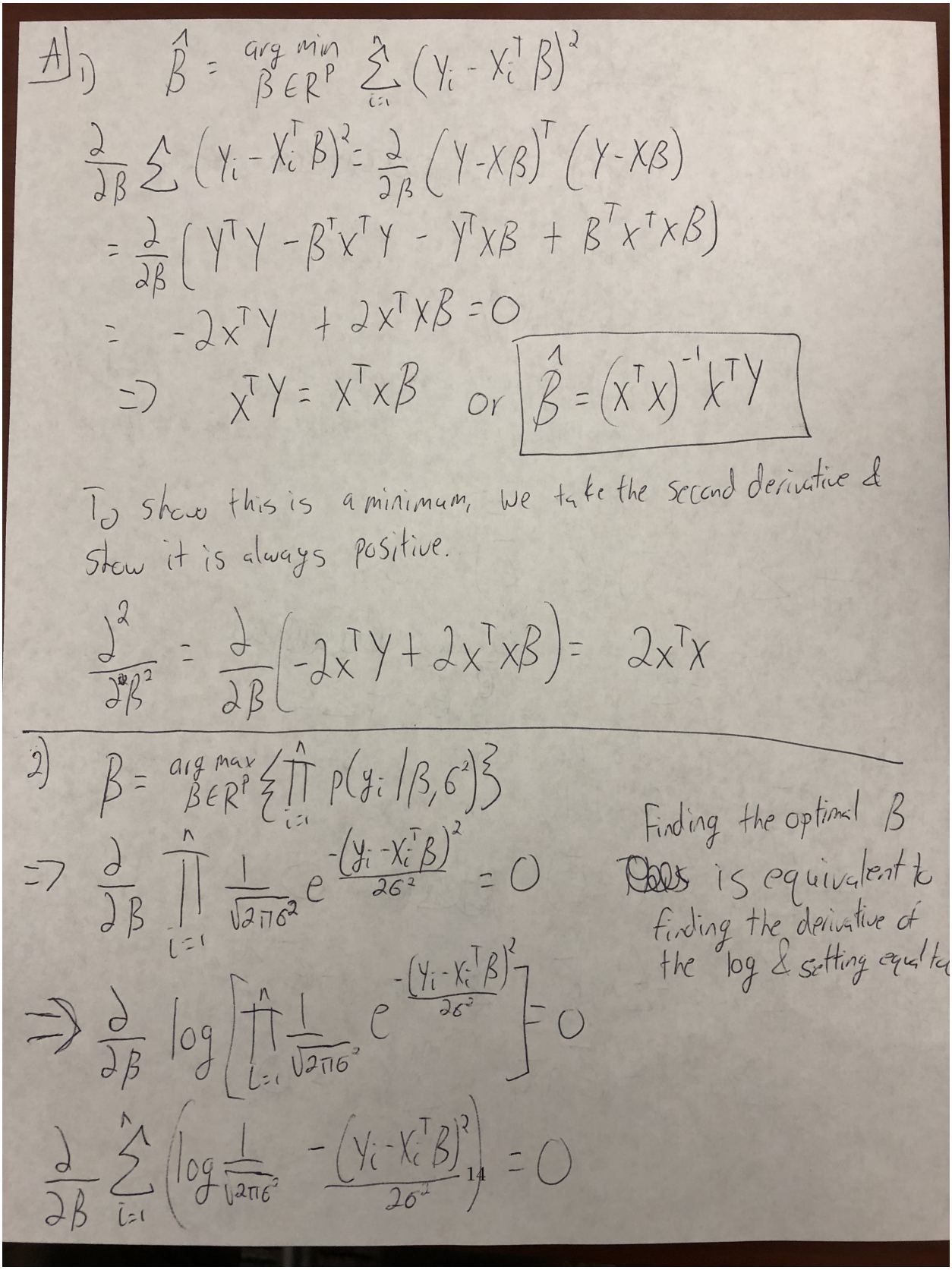
$\Rightarrow X_1 | X_2 \sim N(\mu_1 + \Omega_{11}^{-1}\Omega_{12}(X_2 - \mu_2), \Omega_{11}^{-1})$

$= N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Omega_{11}^{-1})$

Can be considered regressing X_1 on X_2 by rotating and scaling the elements by the covariance elements by the variance of X_2.

Multiple Regression: Three Classical Principles for Inference

A, B



$$\Rightarrow \quad -\frac{\partial}{\partial \beta}\frac{(Y-X\beta)^2}{2\sigma^2} = 0$$

or $-\frac{\partial}{\partial \beta}(Y-X\beta)^2 = 0$

Thus, by the same argument as before,

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

However since this derivative was negative & we showed the last value was a minimum, this $\hat{\beta}$ must represent a maximum.

---

3) $Cov(X_i, \epsilon_i) = 0 \Rightarrow$

$$\frac{1}{n-1}\sum (x_{ij} - \bar{x}_j)(\epsilon_i - \bar{\epsilon}) = 0$$

where $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ & $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i$

If we center each of the predictors about its mean, we can force $\bar{x}_j = 0$ for $j = 1, \ldots, p$. Thus

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(\epsilon_i - \bar{\epsilon}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij}\epsilon_i - \sum x_{ij}\bar{\epsilon} = \sum_{i=1}^{n} x_{ij}\epsilon_i - \bar{\epsilon}\sum_{i=1}^{n} x_{ij}$$

$$= \sum_{i=1}^{n} x_{ij}\epsilon_i - \bar{\epsilon} n \bar{x}_j = \sum_{i=1}^{n} x_i \epsilon_i = 0$$

Thus we can write this as

$$X^T \epsilon = 0$$

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

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

Thus we have the same 3 estimators

2] $\hat{\beta} = \underset{\beta \in R^p}{\arg\min} \left\{ \sum_{i=1}^{n} w_i (Y_i - X_i^T \beta)^2 \right\}$

We will write this problem in vector/matrix form by saying

$$W = \begin{bmatrix} w_1 & & 0 \\ & w_2 & \\ 0 & & \ddots \\ & & & w_n \end{bmatrix}$$ Thus

$$\hat{\beta} = \underset{\beta \in R^p}{\arg\min} (Y - X\beta)^T W (Y - X\beta)$$

$$\Rightarrow \frac{\partial}{\partial \beta} (-\beta^T X^T + Y^T) W (Y - X\beta) = 0$$

$$= \frac{\partial}{\partial \beta} (Y^T WY - Y^T WX\beta - \beta^T X^T WY + \beta^T X^T WX\beta) = 0$$

$$= -2X^T WY + 2X^T WX\beta = 0$$

$$\Rightarrow \boxed{\hat{\beta} = (X^T WX)^{-1} X^T WY}$$

To show this is a minimum, we will derive again & show it is convex.

$$\Rightarrow \frac{\partial^2}{\partial \beta^2} = \frac{\partial}{\partial \beta}\left(-2X^T W Y + 2X^T W X \beta\right)$$

$$= 2X^T W X$$

---

2) $$\hat{\beta} = \underset{\beta \in R^p}{\arg\max}\left(\prod_{i=1}^{n} p(y_i \mid \beta, \sigma_i^2)\right)$$

$$\Rightarrow \frac{\partial}{\partial \beta} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(Y_i - X_i^T\beta)^2}{2\sigma_i^2}} = 0$$

This is Equivalent in the log

$$\Rightarrow \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \underbrace{\log \frac{1}{\sqrt{2\pi\sigma_i^2}}}_{\text{Constant in } \beta} - \frac{(Y_i - X_i^T\beta)^2}{2\sigma^2} = 0$$

$$\Rightarrow \frac{\partial}{\partial \beta} - \sum \frac{(Y_i - X_i^T\beta)^2}{2\sigma^2} \quad \text{[illegible]}$$

If we say $w_i = \frac{1}{\sigma_i^2}$ & use the same W matrix as before, we have (Precision matrix)

$$\Rightarrow -\frac{\partial}{\partial \beta}(Y - X\beta)^T W (Y - X\beta) = 0$$

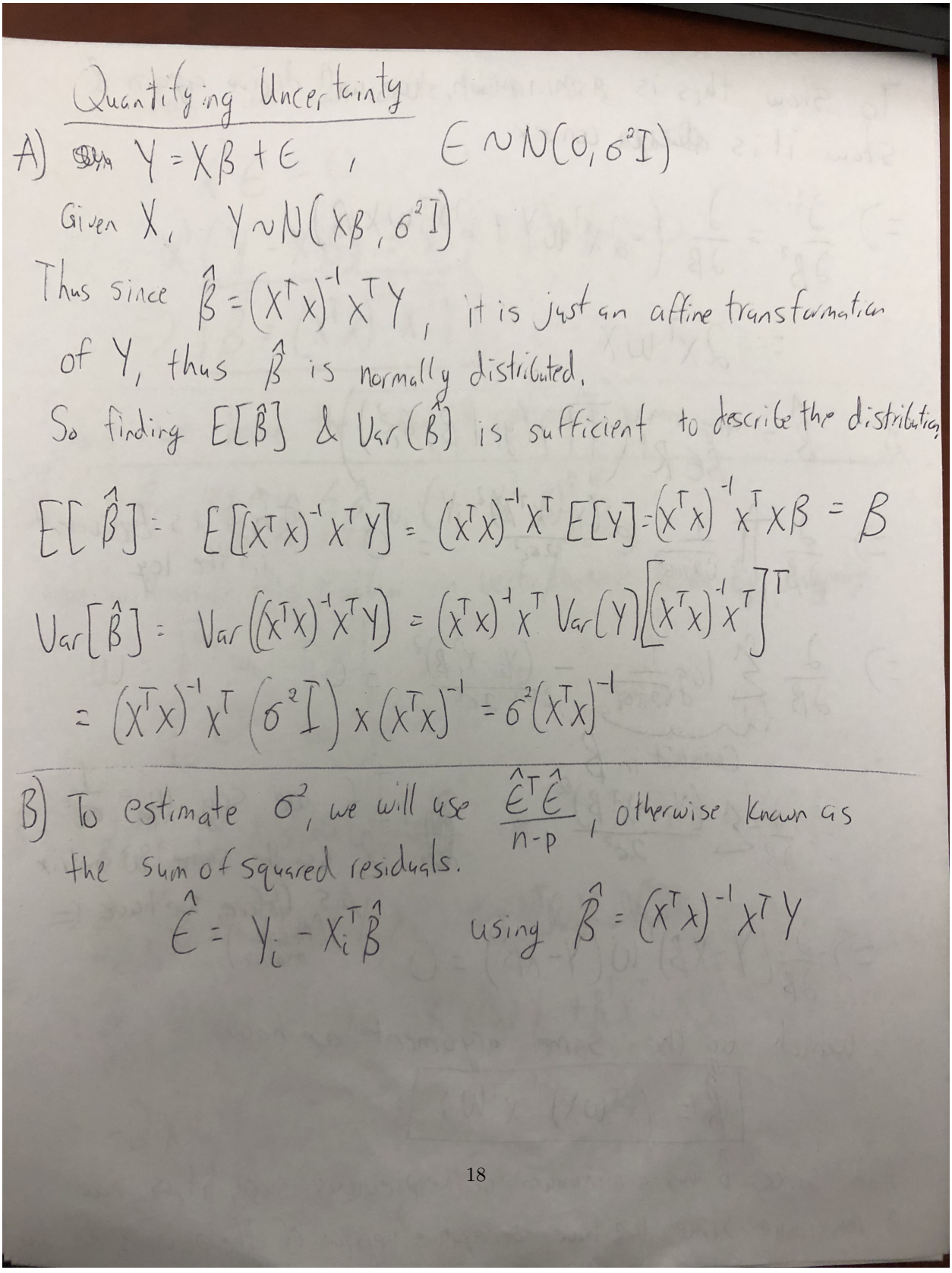
which by the same argument we have

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

And since $\hat{\beta}$ was a minimum in the previous case, it is now a maximum since we have a negative version of the previous derivative.

Quantifying Uncertainty: Some Basic Frequentist Ideas

A, B



We received the following output when estimating the standard errors of the β_j's using the above approximations.

```
> # compute the estimator
> betahat = solve(t(x) %*% x) %*% t(x) %*% y
>
> # Fill in the blank
> residual = y - x%*%betahat
> betacov = (t(residual) %*% residual)/(n-p)
>
>
> # Now compare to lm
> # the 'minus 1' notation says not to fit an intercept (we've already hard-coded it as an extra column)
> lm1 = lm(y~x-1)
>
> summary(lm1)

Call:
lm(formula = y ~ x - 1)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2055  -2.7232  -0.2741   3.0891  13.3442

Coefficients:
       Estimate Std. Error t value Pr(>|t|)
x    59.9517553 38.3286940   1.564 0.119421
xV5  -0.0139111  0.0072511  -1.918 0.056527 .
xV6   0.0276862  0.1741433   0.159 0.873847
xV7   0.0808740  0.0237694   3.402 0.000812 ***
xV8   0.1503404  0.0692994   2.169 0.031272 *
xV9   0.5253439  0.1247136   4.212 3.87e-05 ***
xV10 -0.0010052  0.0003944  -2.549 0.011586 *
xV11  0.0049796  0.0147772   0.337 0.736501
xV12 -0.1543882  0.1192917  -1.294 0.197140
xV13 -0.0033951  0.0048963  -0.693 0.488883
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.482 on 193 degrees of freedom
Multiple R-squared:  0.9026,    Adjusted R-squared:  0.8976
F-statistic: 178.9 on 10 and 193 DF,  p-value: < 2.2e-16

> betacovlm = vcov(lm1)
> sqrt(diag(betacovlm))
           x          xV5          xV6          xV7          xV8          xV9         xV10         xV11
3.832869e+01 7.251104e-03 1.741433e-01 2.376936e-02 6.929942e-02 1.247136e-01 3.943697e-04 1.477719e-02
        xV12         xV13
1.192917e-01 4.896254e-03
```

Propagating Uncertainty

A, B

Propagating Uncertainty

A) $f(\theta) = \theta_1 + \theta_2$

$$Var(f(\hat{\theta})) = Var(\hat{\theta}_1 + \hat{\theta}_2) = Var(\hat{\theta}_1) + 2Cov(\hat{\theta}_1, \hat{\theta}_2) + Var(\hat{\theta}_2)$$

Let $f(\theta) = \sum_{i=1}^{p} \theta_i$

$$\Rightarrow Var(f(\hat{\theta})) = Var\left(\sum_{i=1}^{p} \hat{\theta}_i\right)$$

$$= \sum_{i=1}^{p} Var(\hat{\theta}_i) + 2\sum_{i \neq j} Cov(\hat{\theta}_i, \hat{\theta}_j)$$

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

B) Let f be a nonlinear function of $\hat{\theta}_j$'s. Thus, using a 1st-Order Taylor Approx.:

$$f(\hat{\theta}) \approx f(\theta) + (\theta - \hat{\theta}) f'(\hat{\theta})$$

Thus

$$Var(f(\hat{\theta})) \approx Var(f(\theta) + (\theta - \hat{\theta}) f'(\hat{\theta}))$$

$$= f'(\hat{\theta})^2 Var(\hat{\theta}) = \boxed{f'(\hat{\theta})^2 \Sigma}$$

Since we used a ~~1st~~ - order Taylor Approx., the error can be bounded by the second order Taylor approx. term. Thus is bounded by

$$\frac{f''(\hat{\theta})(\theta - \hat{\theta})^2}{2}$$