Sareh Kouchaki – Section 1
Statistical-Modeling-II

Exercise 1.1:

We pick 3 balls from an urn with r red and b blue balls. We pick a random ball and note its color, return the ball, plus another ball of the same color. It may not be obvious that the sequence A=r, B=b, C=b has the same probability as the sequence A=b, B=b, C=r since the individual probabilities of picking the red ball first or last are completely different: r/(r+b) when it is the first ball versus r/(r+b+2) when it is the last ball (since two blue balls were added in the meantime). So, the observations are not iid. Writing down the equations makes it clear that the two sequences are exchangeable due to the equal probability but not iid.

p (A=b, B=b, C=r) = $\frac{b}{r+b}\,\frac{b+1}{r+b+1}\,\frac{r}{r+b+2}$

p (A=r, B=b, C=b) = $\frac{r}{r+b}\,\frac{b}{r+b+1}\,\frac{b+1}{r+b+2}$

Exercise 1.2: We want to show that:

$$p(\sum_{i=1}^{N} x_i = s \mid \sum_{i=1}^{M} x_i = t) = \frac{p(\sum_{i=1}^{N} x_i = s\ ,\ \sum_{i=1}^{M} x_i = t)}{p(\sum_{i=1}^{M} x_i = t)}$$

$x_i \rightarrow \{0,1\}$

There are $\binom{N}{S}$ possibilities to select s observations from the set of first N observations for the first group. For each of these possibilities, there are $\binom{M-N}{t-s}$ possibilities to select t-s observations for the first group from the remaining M-N observations.

$$\frac{p(\sum_{i=1}^{N} x_i = s, \sum_{i=1}^{M} x_i = t)}{p(\sum_{i=1}^{M} x_i = t)} = \frac{\binom{N}{S}\binom{M-N}{t-s} p(x_1=1,x_2=1,\dots,x_t=1,x_{t+1}=0,\dots,x_M=0)}{\binom{M}{t} p(x_1=1,x_2=1,\dots,x_t=1,x_{t+1}=0,\dots,x_M=0)} = = \frac{\binom{N}{S}\binom{M-N}{t-s}}{\binom{M}{t}} = \frac{\binom{t}{S}\binom{M-t}{N-s}}{\binom{M}{N}}$$

Solution on the blackboard:

$$p\,(\sum_{i=1}^{N} X_i = s) = \sum_{t=s}^{M} p(\sum_{i=1}^{N} X_i = s \mid \sum_{i=1}^{M} X_i = t)\, p(\sum_{i=1}^{M} X_i = t)$$

N<M

$$p(X_1,\dots,X_M \mid \sum_{i=1}^{M} X_i = t) = \begin{cases} \frac{1}{\binom{M}{t}} & \sum X_i = t \\ 0 & Otherwise \end{cases}$$

$X_1,\dots,X_M$: $\quad t\ 1's\ and\ M-t\ 0's$

Exchangeability → p(1 0 1) = p(0 1 1)

$$p(\sum_{i=1}^{N} X_i = s \mid \sum_{i=1}^{M} X_i \ = t) \ = \sum_{\sum_{i=1}^{N} x_i = s, \sum_{i=1}^{M} x_i = t} p(X_1, \ldots, X_M \mid \sum_{i=1}^{M} X_i = t\ ) \ = \binom{N}{s}\binom{M-N}{t-s}\frac{1}{\binom{M}{t}}$$

$$= \binom{t}{s}\binom{M-t}{N-s}\frac{1}{\binom{M}{N}}$$

Exercise 1.3:

$$\mathrm{p}(\sum_{i=1}^{N} \mathrm{X_i} = \mathrm{S}) \ = \binom{\mathrm{N}}{\mathrm{S}} \sum_{t=s}^{M-N+S} \frac{(\mathrm{t})_{\mathrm{S}}(\mathrm{M}-\mathrm{t})_{\mathrm{N-S}}}{(\mathrm{M})_{\mathrm{N}}}\, \mathrm{p}(\sum_{i=1}^{M} \mathrm{X_i} = \mathrm{t}) = \int_0^1 \frac{(\mathrm{t})_{\mathrm{S}}(\mathrm{M}-\mathrm{t})_{\mathrm{N-S}}}{(\mathrm{M})_{\mathrm{N}}}\, dF_M(\theta)$$

**$(t)_s = t(t-1)\ldots(t-s+1)$

** $\theta \ = \frac{\sum_{i=1}^{M} X_i}{M} \ = \frac{t}{M}$

By, replacing t with Mθ:

$$\int_0^1 \frac{(\mathrm{M}\theta)_{\mathrm{S}}(\mathrm{M}(1-\theta))_{\mathrm{N-S}}}{(\mathrm{M})_{\mathrm{N}}}\, dF_M(\theta)$$

$$= \int_0^1 \frac{M\theta!}{(M\theta - s)!} * \frac{[M(1-\theta)]!}{[M(1-\theta) - (N-s)]!} * \frac{(M-N)!}{M!}\, dF_M(\theta)$$

If $M \to \infty$

$$\boxed{(M\theta)^s} \qquad \boxed{(M(1-\theta))^{N-s}}$$

$$= \int_0^1 \frac{M^N(\theta^s)(1-\theta)^{N-s}}{M^N} dF_M(\theta) \ = \int_0^1 \theta^s(1-\theta)^{N-s} dF_M(\theta$$

Exercise 1.4:

Poisson distribution:

$$p(x|\lambda) \ = \frac{\lambda^x e^{-\lambda}}{x!}$$

Exponential family form:

$$p(x|\lambda) \ = \frac{exp(xlog(\lambda) \ - \ \lambda)}{x!}$$

$\eta = log(\lambda)$

T(x) = x

$A(\eta) = \lambda$

h(x) = $\frac{1}{x!}$

If we have n independent samples:

$$p(x_1, x_2, ..., x_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

Exponential family form:

$$p(x_1, x_2, ..., x_n|\lambda) = \frac{1}{\prod_{i=1}^{n} x_i!} exp\ (\sum_{i=1}^{n} x_i\ log\ (\lambda) - n\lambda)$$

$\eta = log(\lambda)$

T(x) = $\sum_{i=1}^{n} x_i$

$A(\eta) = n\lambda$

h(x) = $\frac{1}{\prod_{i=1}^{n} x_i!}$

Exercise 1.5:

Gamma distribution

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

joint pdf for n observations:

$$p(x_1, x_2, ..., x_n|\alpha, \beta) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} x_i{}^{\alpha-1} e^{-\beta x_i} = (\frac{\beta^\alpha}{\Gamma(\alpha)})^n (\prod_{i=1}^{n} x_i)^{\alpha-1} e^{-\beta \sum_{i=1}^{n} x_i}$$

Exponential family form:

$$p(x_1, x_2, ..., x_n|\alpha, \beta) = exp(log((\frac{\beta^\alpha}{\Gamma(\alpha)})^n) + ((\alpha - 1) log(\prod_{i=1}^{n} x_i)) - \beta \sum_{i=1}^{n} x_i)$$

Natural Parameter or $\eta 1 = \alpha - 1$

Sufficient statistic1 or $T_1(x) = log(\prod_{i=1}^{n} x_i) = \sum_{i=1}^{n} log(x_i)$

Natural Parameter or $\eta 2 = -\beta$

Sufficient statistic1 or $T_2(x) = \sum_{i=1}^{n} x_i$

** point:

The $\sum_{i=1}^{n} log(x_i)$ is not the function of $\sum_{i=1}^{n} x_i$.

Any function of sufficient statistics is sufficient statistics. So, in this case the sufficient statistics is the minimum statistics.

Exercise 1.6: Show that the moment generating function for the sufficient statistic of an arbitrary exponential family random variable with natural parameter η can be written as:

$$M_{T(x)}(s) = exp\, A\,(\eta + s) - A(\eta)$$

Suppose X is distributed as:

$$p(x|\eta) = h(x) exp(\eta T(x) - A(\eta))$$

$$M_{T(x)}(s) = E[\,(e^{sT(x)})|\eta] = \int e^{sT(x)} h(x) e^{(\eta T(x) - A(\eta))} \mathrm{dx} = \int h(x) e^{((s+\eta)T(x) - A(\eta))} \mathrm{dx}$$

$$= e^{A(s+\eta)-A(\eta)} \int h(x) e^{((s+\eta)T(x) - A(s+\eta))} \mathrm{dx} = e^{A(s+\eta)-A(\eta)} * 1 = e^{A(s+\eta)-A(\eta)}$$

Pdf: p (x| $\eta + s$)

Exercise 1.7:

Part a) the moment generating function:

p(x|λ) $= \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0,1,2,\ldots$

$M_x(t) = e^{\lambda e^t - \lambda}$ $where\ x = 0,1,2,\ldots and\ \lambda \geq 0$

E(x) = $M'_x(t) = \lambda e^t e^{\lambda e^t - \lambda} = \lambda e^{\lambda e^t - \lambda + t}$ $at\ t = 0\ is\ equal\ to\ \lambda$

$E(x^2) = M''_x(t) = \lambda e^t e^{\lambda e^t - \lambda} + \lambda e^t \lambda e^t e^{\lambda e^t - \lambda}$ $at\ t = 0\ is\ equal\ to\ \lambda^2 + \lambda$

var(x) $=$ E$(x^2) - [E(x)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Part b) The cumulant generating function:

Exponential family form: $\frac{1}{x!}\ e^{xlog\ \lambda - \lambda}$

h(x) $= \frac{1}{x!}$ $\quad t(x) = x \quad \eta = log\lambda \quad A(\eta) = exp(\eta)$

$C_x(s) = A(\eta + s) - A(\eta) = \exp(\eta + s) - exp(\eta) = \exp(\log\lambda + s) - exp(log\lambda) = \lambda exp(s) - \lambda = \lambda(e^s - 1)$

Mean $= \frac{d}{ds}\ C_x(s)|_{s=0} = \lambda e^s|_{s=0} = \lambda$

Variance $= \frac{d^2}{ds^2}\ C_x(s)|_{s=0} = \lambda e^s|_{s=0} = \lambda$

Exercise 1.8:

All exponential families have exponential likelihood, so, they have exponential posterior.

$x_1, x_2, \ldots, x_n \sim Normal(\mu, \sigma^2)\ with\ known\ \sigma^2$

Likelihood: $p(x_1, x_2, \ldots, x_n|\mu) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\ exp(-\frac{1}{2\sigma^2}\sum(x_i - \mu)^2)$

Posterior: $\mathrm{p}(\mu|x_1, x_2, \ldots, x_n) = p(x_1, x_2, \ldots, x_n|\mu) * P(\mu|\mu_0, \sigma_0^2)$

$= \exp(-\frac{1}{2\sigma^2}\sum(x_i - \mu)^2 - \frac{1}{2\sigma_0{}^2}(\mu - \mu_0)^2) = \exp(-\frac{1}{2\sigma^2}\sum(x_i{}^2 - 2x_i\mu + \mu^2) - \frac{1}{2\sigma_0{}^2}(\mu_0{}^2 - 2\mu_0\mu + \mu^2) = exp(-\frac{1}{2}\ [\frac{\sigma_0{}^2(\sum(x_i{}^2 - 2x_i\mu + \mu^2)) + \sigma^2(\mu_0{}^2 - 2\mu_0\mu + \mu^2)}{\sigma^2\sigma_0{}^2}]) =$
$exp(-\frac{1}{2}[\frac{-2\sigma_0{}^2 n\bar{x}\mu + \sigma_0^2 n\mu^2 + \sigma^2\mu^2 - 2\sigma^2\mu_0\mu}{\sigma^2\sigma_0{}^2}]) = exp(-\frac{1}{2}\ [\frac{(n\sigma_0^2 + \sigma^2)\mu^2 - 2(\sigma^2\mu_0 + \sigma_0{}^2 n\bar{x})\mu}{\sigma^2\sigma_0{}^2}])$

**Dividing the numerator and operator by $(n\sigma_0^2 + \sigma^2)$:

$$exp(-\frac{1}{2}\ [\frac{\mu^2 - 2\frac{(\sigma^2\mu_0 + \sigma_0{}^2 n\bar{x})}{(n\sigma_0^2 + \sigma^2)}\mu}{\frac{\sigma^2\sigma_0{}^2}{(n\sigma_0^2 + \sigma^2)}}]) \propto exp(-\frac{1}{2}\ [\frac{(\mu - \frac{(\sigma^2\mu_0 + \sigma_0{}^2 n\bar{x})}{(n\sigma_0^2 + \sigma^2)})^2}{\frac{\sigma^2\sigma_0{}^2}{(n\sigma_0^2 + \sigma^2)}}])$$

This shows that μ is normally distributed with mean of $\frac{(\sigma^2\mu_0 + \sigma_0{}^2 n\bar{x})}{(n\sigma_0^2 + \sigma^2)}$ and variance of $\frac{\sigma^2\sigma_0{}^2}{(n\sigma_0^2 + \sigma^2)}$.

Exercise 1.9: Derive the posterior distribution for ω.

$$f(x_i|\mu,\omega) \;=\; \sqrt{\frac{\omega}{2\pi}}\; exp\{-\frac{\omega}{2}(x_i-\mu)^2\}$$

Likelihood function: $\prod_{i=1}^{n}\sqrt{\frac{\omega}{2\pi}}\; exp\{-\frac{\omega}{2}(x_i-\mu)^2\} \;=\; (\frac{\omega}{2\pi})^{\frac{n}{2}}\, exp^{\sum_{i=1}^{n}-\frac{\omega}{2}(x_i-\mu)^2} \;=$
$(\frac{\omega}{2\pi})^{\frac{n}{2}}\; exp^{-\frac{\omega}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$

Gamma prior: $p(\omega) \;=\; \frac{\beta^{\alpha}}{\Gamma(\alpha)}\,\omega^{\alpha-1}e^{-\beta\omega}$

Posterior:

$$\left(\frac{\omega}{2\pi}\right)^{\frac{n}{2}} exp^{-\frac{\omega}{2\pi}\sum_{i=1}^{n}(x_i-\mu)^2} * \frac{\beta^{\alpha}}{\Gamma(\alpha)}\,\omega^{\alpha-1}e^{-\beta\omega} \propto \omega^{\alpha+\frac{n}{2}-1} exp\left(-\omega\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta\right)\right)$$

$Gamma(\alpha+\frac{n}{2},\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)$

Exercise 1.10: Show that the marginal distribution of x is given by a Student's t distribution.

Marginal distribution is obtained through the following:

$$f(x) \;=\; \int f(x,\omega)\,d\omega \;=\; \int \sqrt{\frac{1}{2\pi}}\;\frac{\beta^{\alpha}}{\Gamma(\alpha)}\,\omega^{\alpha+\frac{1}{2}-1}\,e^{-\omega(\beta+(x-\mu)^2)}d\omega$$

Kernel of Gamma dist.

$$= \frac{1}{\sqrt{2\pi}}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{(\beta+\frac{1}{2}(x-\mu)^2)^{\alpha+\frac{1}{2}}}\int\frac{(\beta+\frac{1}{2}(x-\mu)^2)^{\alpha+\frac{1}{2}}}{\Gamma\left(\alpha+\frac{1}{2}\right)}\,\omega^{\alpha+\frac{1}{2}-1}\,e^{-\omega(\beta+(x-\mu)^2)}d\omega$$

$$= \frac{1}{\sqrt{2\pi}}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{(\beta+\frac{1}{2}(x-\mu)^2)^{\alpha+\frac{1}{2}}} \;=\; \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\sqrt{2\pi}\beta}(1+\frac{x^2}{2\beta})^{-\frac{1}{2}-\alpha}$$

**$\mu = 0$

** t distribution format

Exercise 1.11:

a)

$\sum \;=\; E[(x-\mu)(x-\mu)^T] \;=\; E[xx^T-2\mu x^T+\mu\mu^T] \;=\; E(xx^T) \;-\; 2\mu E(x^T)+E(\mu\mu^T) \;=$
$E(xx^T) \;-\; 2\mu\mu^T+E(\mu\mu^T) \;=\; E(\,xx^T)-\mu\mu^T$

b) $Cov(Ax+b) \;=\; E[(Ax+b \;-\; E(Ax+b))(Ax+b \;-\; E(Ax+b))^T] \;=\; E[(Ax - A\mu)(Ax-A\mu)^T] \;=\; E[A(x-\mu)(x-\mu)^T A^T] = \; AE[(x-\mu)(x-\mu)^T]\, A^T \;=\; A\Sigma A^T$

Exercise 1.12:

Moment Generation function for a univariate normal with mean m and variance υ^2:

$$M_X(t) \;=\; E[e^{tx}] \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\upsilon} exp(-\frac{(x\text{-m})^2}{2\upsilon^2}) exp(tx)dx \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\upsilon} exp(-\frac{x^2\text{-2mx-2}\upsilon^2\text{tx+m}^2}{2\upsilon^2})dx \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\upsilon} exp(-\frac{(\text{x-(m+}\upsilon^2\text{t))}^2\text{-(m+}\upsilon^2\text{t)}^2\text{+m}^2}{2\upsilon^2})dx \;=\; \exp\{\frac{\text{-(m+}\upsilon^2\text{t)}^2\text{+m}^2}{2\upsilon^2}\} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\upsilon} exp(-\frac{(\text{x-(m+}\upsilon^2\text{t))}^2}{2\upsilon^2})dx}_{\text{Normal }(\text{m}+\upsilon^2\text{t},\upsilon^2)} \;=\; \exp\{\frac{\text{-(m+}\upsilon^2\text{t)}^2\text{+m}^2}{2\upsilon^2}\} \;=\; \exp\{\text{mt}+\frac{\upsilon^2\text{t}^2}{2}\}$$

$$M_Z(t) \;=\; E[e^{t^T z}] \;=\; E[exp(\textstyle\sum_{i=1}^{n} t_i z)] = E[\prod_{i=1}^{n} exp(\,t_i z)] \;=\; exp(\frac{1}{2}\sum_{i=1}^{n} {t_i}^2) \;=\; exp(\frac{1}{2} t^T t)$$

PDF:

$$p(x) \;=\; \frac{1}{\sqrt{2\pi}}\; exp\{-\frac{x^2}{2}\}$$

$$p(x) \;=\; \prod_{i=1}^{n} p(x_i) \;=\; \frac{1}{(2\pi)^{\frac{n}{2}}} exp\{-\frac{\sum_{i=1}^{n} {x_i}^2}{2}\}$$

Exercise 1.13:

$$\forall\, a \quad M_{a^T x}(s) \;=\; E(e^{a^T x s}) \;=\; E(e^{(sa^T)x}) \;=\; e^{sa^T\mu + \frac{1}{2} sa^T \Sigma a s}$$

$$**\mu^* = a^T\mu \qquad \sigma^* = \sqrt{a^T\Sigma a}$$

$$=\; e^{s\mu^* + \frac{1}{2}s^2{\sigma^*}^2}$$

Exercise 1.14:

$$x \;=\; DZ + \mu$$

$$E(x) \;=\; E(DZ+\mu) \;=\; DE(Z) \;+\; \mu \;=\; \mu$$

$$\Sigma \;=\; cov(x) \;=\; cov(DZ+\mu) \;=\; cov(DZ) \;=\; Dcov(Z)D^T \;=\; DD^T$$

Is equal to I, because it is standard.

Exercise 1.15:

$p(z) = \frac{1}{(2\pi)^{\frac{n}{2}}} exp\{-\sum_i \frac{z_i^2}{2}\}$

$x = u(z) = Dz + \mu$

$z = v(x) = D^{-1}(x-\mu)$

$f_z(z) = f_x(x)|v^T(x)|$

$$\boxed{|\Sigma|^{\frac{-1}{2}}}$$

$$f_z(z) = \frac{1}{(2\pi)^{\frac{n}{2}}} exp\{-\frac{z^T z}{2}\} \rightarrow \frac{1}{(2\pi)^{\frac{n}{2}}}|D^{-1}| exp\{-\frac{(x-\mu)D^{-1^T}D^{-1}(x-\mu)}{2}\}$$

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

$**v = D^{-1},$

Exercise 1.16:

Solution 1:

$p(\left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(\left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right] - \left[\begin{smallmatrix}\mu_1\\\mu_2\end{smallmatrix}\right])^T \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}(\left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right] - \left[\begin{smallmatrix}\mu_1\\\mu_2\end{smallmatrix}\right]))$

Integrate out over $x_2$to find the marginal of x_1

$p(x_1) = \int p(\left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]; \mu, \Sigma)\, dx_2$

$= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\int exp(-\frac{1}{2}((x_1-\mu_1)^T\Sigma_{11}(x_1-\mu_1) + (x_2-\mu_2)^T\Sigma_{22}(x_2-\mu_2) +$
$2(x_2-\mu_2)^T\Sigma_{21}(x_1-\mu_1))dx_2 = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\int exp(-\frac{1}{2}((x_1-\mu_1)^T\Sigma_{11}(x_1-\mu_1) +$
$(x_2-\mu_2)^T\Sigma_{22}(x_2-\mu_2) + 2(x_2-\mu_2)^T\Sigma_{22}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1-\mu_1)) + (x_1 -$
$\mu_1)^T\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{22}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1-\mu_1) - (x_1-\mu_1)^T\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{22}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1-\mu_1))dx_2 =$
$\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}((x_1-\mu_1)^T\Sigma_{11}(x_1-\mu_1) - (x_1-\mu_1)^T\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{22}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1-\mu_1)))$

$* \int exp(-\frac{1}{2}((x_2-\mu_2)^T\Sigma_{22}(x_2-\mu_2) + 2(x_2-\mu_2)^T\Sigma_{21}(x_1-\mu_1)+)) + (x_1 -$
$\mu_1)^T\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{22}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1-\mu_1))dx_2=$

$\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}((x_1-\mu_1)^T\Sigma_{11}(x_1-\mu_1) - (x_1-\mu_1)^T\Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1 -$
$\mu_1))) \int exp(-\frac{1}{2}((x_2-\mu_2+\Sigma_{22}{}^{-1}\Sigma_{21}(x_1-\mu_1))^T\Sigma_{22}(x_2-\mu_2+\Sigma_{22}{}^{-1}\Sigma_{21}(x_1 -$

$\mu_1))))dx_2 = \frac{(2\pi)^{\frac{n_2}{2}}|\Sigma_{22}{}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \; exp(-\frac{1}{2}((x_1 - \mu_1)^T \Sigma_{11}(x_1 - \mu_1) - (x_1 - \mu_1)^T \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21}(x_1 - \mu_1))) =$

$\frac{(2\pi)^{\frac{n_2}{2}}|\Sigma_{22}{}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \; exp(-\frac{1}{2}((x_1 - \mu_1)^T(\Sigma_{11} - \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21})(x_1 - \mu_1)))$

Solution 2 (Blackboard):

$x_1 = (I \quad 0)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

E$(x_1) = E[(I \quad 0)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}] = \mu_1$ (use transformation & combination)

$Cov(x_1) =$ cov $[(I \quad 0)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}] = (I \quad 0)cov(x)\begin{bmatrix} I \\ 0 \end{bmatrix} = (I \quad 0)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}{}^T & \Sigma_{22} \end{pmatrix}\begin{bmatrix} I \\ 0 \end{bmatrix} = \Sigma_{11}$

If we have p number of x:

PDF: f$(x_1) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_{11}|^{\frac{1}{2}}} \; e^{-\frac{1}{2}(x_1-\mu_1)^T\Sigma_{11}{}^{-1}(x_1-\mu_1)}$

Exercise 1.17:

precision matrix: $\Omega = \Sigma^{-1}$

$\Sigma\Sigma^{-1} = \mathrm{I} \rightarrow \Sigma\Omega = \mathrm{I} = \begin{bmatrix} \Sigma_{11}\Omega_{11} + \Sigma_{12}\Omega_{21} & \Sigma_{11}\Omega_{12} + \Sigma_{12}\Omega_{22} \\ \Sigma_{21}\Omega_{11} + \Sigma_{22}\Omega_{21} & \Sigma_{21}\Omega_{12} + \Sigma_{22}\Omega_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$

$\Sigma_{21}\Omega_{11} + \Sigma_{22}\Omega_{21} = 0 \rightarrow \Omega_{21} = \Sigma_{12}{}^{-1}\Sigma_{11}\Omega_{11}$

$\Sigma_{11}\Omega_{12} + \Sigma_{12}\Omega_{22} = 0 \rightarrow \Omega_{12} = \Sigma_{11}{}^{-1}\Sigma_{12}\Omega_{22}$

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

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

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

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

Exercise 1.18:

$p(x_1|x_2) \propto exp\,(-\frac{1}{2}\,(x_1-\mu_1)^T\sum_{11}(x_1-\mu_1)\;-\;(x_1-\mu_1)^T\sum_{12}(x_2-\mu_2)-$
$\frac{1}{2}\,(x_2-\mu_2)^T\sum_{22}(x_2-\mu_2))$

$\propto exp((-\frac{1}{2}\,(x_1-\mu_1)^T\sum_{11}(x_1-\mu_1)\;-\;(x_1-\mu_1)^T\sum_{12}(x_2-\mu_2))$

$\propto\; exp(-\frac{1}{2}\,(x_1-\mu_1)^T\sum_{11}(x_1-\mu_1)-\;(x_1-\mu_1)^T\sum_{11}{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2)\;-\frac{1}{2}\,(x_2-$
$\mu_2)\sum_{12}{\sum_{11}}^{-1}\sum_{11}{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2)+\;\frac{1}{2}\,(x_2-\mu_2)\sum_{12}{\sum_{11}}^{-1}\sum_{11}{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2)$

$$\propto\; exp(-\frac{1}{2}(x_1-\mu_1+{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2)^T\sum_{11}(x_1-\mu_1)+{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2))\;exp(\frac{1}{2}\,(x_2-\mu_2)\sum_{12}{\sum_{11}}^{-1}\sum_{11}{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2))$$

$$\propto\; exp((-\frac{1}{2}(x_1-\mu_1+{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2)^T\sum_{11}(x_1-\mu_1)+{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2))$$

So, we have:

$x_1|x_2 \sim N(\mu_1-{\sum_{11}}^{-1}\sum_{12}(x_2-\mu_2))$

Exercise 1.19:

$y\;=\;\beta x+\epsilon \rightarrow \epsilon\;=\;y-\beta x$

$cov(x_i,\epsilon_i)=E(x\epsilon)-E(x)E(\epsilon)=E\big(x(y-\beta x)\big)-E(x)E(y-\beta x)=E(x^Ty-x^T\beta x)-$
$E(x)\big(E(y)-\beta E(x)\big)=E(x^Ty)-E(x^T\beta x)-E(x)E(y)+\beta E^2(x)\;=\;0\rightarrow E(x^Ty)-$
$E(x)E(y)\;=\;\beta(E(x^Tx)\;-\;E^2(x))\;\rightarrow\;\widehat{\beta_{MM}}\;=\;\frac{E(x^Ty)-E(x)E(y)}{E(x^Tx)\,-\,E^2(x)}\;=\;\frac{\frac{1}{n}\sum_{i=1}^n x_iy_i\,-\,\bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2-\,(\bar{x})^2}\;=$
$\frac{\sum_{i=1}^n x_iy_i\,-\,n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2-\,n(\bar{x})^2}\;=\;\frac{\sum_{i=1}^n(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n(x_i-\bar{x})^2}\;=\;(X^TX)^{-1}X^TY$

Exercise 1.20:

$y\;=\;\beta x+\epsilon\;\;\;\rightarrow\epsilon\;=\;y\;-\;\beta x\,,\;\epsilon\sim N(0,\sigma^2)$

Maximum likelihood estimation of β:

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

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

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

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

$$= -\frac{n}{2}\ log(2\pi)\ -\ nlog(\sigma)\ -\frac{1}{2\sigma^2}\ [\sum_{i=1}^{n} {y_i}^2-2\beta\sum_{i=1}^{n} y_i x_i+\sum_{i=1}^{n}(\beta x_i)^2]$$

$$\frac{d[-2\beta\sum_{i=1}^{n} y_i x_i\ +\ \sum_{i=1}^{n}(\beta x_i)^2]}{d\beta}\ =\ 0$$

$$-2\sum_{i=1}^{n} y_i x_i\ +2\beta\sum_{i=1}^{n} x_i^2\ =\ 0\ \rightarrow \widehat{B_{ML}}\ =\ \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}\ =\ \frac{\mathrm{X}^T\mathrm{Y}}{\mathrm{X}^T\mathrm{X}}\ = \mathrm{X}^T\mathrm{Y}(\mathrm{X}^T\mathrm{X})^{-1}$$

Exercise 1.21:

Loss function = $\sum_{i=1}^{n}(y_i\ - x_i^T\beta)^2$

$$\frac{d(Loss\ function)}{d\beta}\ =\ 0\ \ \rightarrow\ \frac{d(\sum_{i=1}^{n} y_i {y_i}^T-2\beta\sum_{i=1}^{n} x_i^T y_i+\beta^2\sum_{i=1}^{n} x_i^T x_i}{d\beta}\ =\ 0$$

$$\beta(\mathrm{X}^T\mathrm{X})\ =\ \mathrm{X}^T\mathrm{Y} \rightarrow \widehat{\beta_{LS}} =\ \mathrm{X}^T\mathrm{Y}(\mathrm{X}^T\mathrm{X})^{-1}$$

**Loss function is the same as the joint and log of normal dist. That's why the results are similar.

Exercise 1.22:

Minimize $\sum_{i=1}^{n}(y_i - x_i^T\beta)^2\ s.t\ \sum_{j=1}^{n}\beta_j^2 \leq t$

Reformulate this constrained optimization using a Lagrange multiplier:

$$f(\beta)\ =\ \sum_{i=1}^{n}(y_i - x_i^T\beta)^2\ +\ \lambda\sum_{j=1}^{n}\beta_j^2\ =\ (\mathrm{Y}-\mathrm{X}\beta)^T(\mathrm{Y}-\mathrm{X}\beta)\ +\ \lambda(\beta^T\beta\ -\ t)$$

$$\frac{df(\beta)}{d\beta}\ =\ -2\mathrm{X}^T(\mathrm{Y}-\mathrm{X}\beta)\ +\ 2\lambda\beta\ =\ 0$$

$$\hat{\beta}_{ridge}\ =\ (\mathrm{X}^T\mathrm{X}\ +\lambda I_p)^{-1}\mathrm{X}^T\mathrm{Y}$$

**Benefit of Ridge: way of regularization for avoiding overfitting of the function.

Exercise 1.23:

$\widehat{\beta_{LS}} = (\mathrm{X}^T\mathrm{X})\mathrm{X}^T\mathrm{Y} = (\mathrm{X}^T\mathrm{X})\mathrm{X}^T(\mathrm{X}\beta + \epsilon) = \beta + (\mathrm{X}^T\mathrm{X})\mathrm{X}\epsilon$

$\mathrm{E}(\widehat{\beta_{LS}}) = E(\beta + (\mathrm{X}^T\mathrm{X})\mathrm{X}\epsilon|\mathrm{X}) = \beta + E((\mathrm{X}^T\mathrm{X})\mathrm{X}\epsilon|\mathrm{X}) = \beta \qquad * E(\epsilon) = 0$

$\mathrm{var}(\widehat{\beta_{LS}}|\mathrm{X}) = E[(\widehat{\beta_{LS}} - \beta)(\widehat{\beta_{LS}} - \beta)^T|\mathrm{X}] = E[(\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\epsilon\epsilon^T\mathrm{X}(\mathrm{X}^T\mathrm{X})^{-1}|\mathrm{X}] = \sigma^2(\mathrm{X}^T\mathrm{X})^{-1}$

Normal distribution with mean of β and variance of $\sigma^2(\mathrm{X}^T\mathrm{X})^{-1}$.

Exercise 1.24:

$\hat{\beta}_{Ridge} = (\mathrm{X}^T\mathrm{X} + \lambda I_p)^{-1}\mathrm{X}^T y = (\mathrm{X}^T\mathrm{X} + \lambda I_p)^{-1}\mathrm{X}^T\mathrm{X}\,\hat{\beta}_{ls} = w_\lambda\hat{\beta}_{ls}$

$\mathrm{E}(\hat{\beta}_{Ridge}) = E(w_\lambda\hat{\beta}_{ls}|\mathrm{X}) = w_\lambda\, E(\hat{\beta}_{ls}) = (\mathrm{X}^T\mathrm{X} + \lambda I_p)^{-1}\mathrm{X}^T\mathrm{X}\,\beta$

$\mathrm{var}(\hat{\beta}_{ls}) = \sigma^2(\mathrm{X}^T\mathrm{X})^{-1}$

$\mathrm{var}(\hat{\beta}_{Ridge}) = \sigma^2 w_\lambda(\mathrm{X}^T\mathrm{X})^{-1}{w_\lambda}^T$

Normal distribution with mean of $(\mathrm{X}^T\mathrm{X} + \lambda I_p)^{-1}\mathrm{X}^T\mathrm{X}\,\beta$ and variance of $\sigma^2 w_\lambda(\mathrm{X}^T\mathrm{X})^{-1}{w_\lambda}^T$.

Exercise 1.25:

We need to have an unbiased estimator of σ^2.

$\hat{y} = My$

$\mathrm{M} = \mathrm{x}\,(x^T x)^{-1}x^T$ $\qquad$ * M is the projection Matrix.

$\mathrm{SSE} = (\mathrm{y} - \hat{y})^T(y - \hat{y}) = (y - My)^T(y - My) = y^T(I - M)^T(I - M)y = y^T(I - M)y$

$\mathrm{E(SSE)} = \mathrm{E}\,((\hat{y} + \epsilon)^T(I - M)(\hat{y} + \epsilon)) = E[\epsilon^T\epsilon(I - M)] + E[\hat{y}^T(I - M)\hat{y}] = \mathrm{tr}(\sigma^2(I - M)) + E(y^T(I - M)y) = \sigma^2\, tr(I - M) + \hat{\beta}^T x(I - M)x^T\hat{\beta} = \sigma^2\, tr(I - M) = \sigma^2\,(n - p)$

**tr: the trace of an n-by-n square matrix A is defined to be the sum of the elements on the main diagonal.

So, $\frac{SSE}{n-p}$ *is the unbiased estimator of* σ^2.

Which means that $\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-p}$ results in an unbiased estimator of σ^2.

In this question p is number of predictors = 3 and n is the number of observations = 102.

For the estimator of the standard error we have:

$$\sqrt{\text{diag}(\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-p}\ (x^T x)^{-1}}$$

Code is on Github.

According to the result, the estimation of standard error is almost similar to that one obtained from lm.

Exercise 1.26:

$f(\theta) = \sum_i \theta_i$

$$\text{var } f(\theta) = \text{var }(\sum_{i=1}^{p} \theta_i) = \sum_{i=1}^{p}\sum_{j=1}^{p} cov(\theta_i, \theta_j) = \sum_{i=1}^{p} var(\theta_i) + 2 \sum_{1 \le i \le j \le p} cov(\theta_i, \theta_j)$$

Exercise 1.27:

$f(\theta) \neq \sum_i \theta_i$

$\mu = \hat{\theta}$ *our estimator of* θ

Taylor Expansion of $f(\theta)$:

$$f(\theta) = f(\hat{\theta}) + \frac{\partial f(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}\ (\theta - \hat{\theta}) + \ldots = \underbrace{f(\hat{\theta}) + \frac{\partial f(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}}_{\boxed{\text{Constant Part}}}\ \theta - \frac{\partial f(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}\hat{\theta} + \ldots$$

$$\text{var}(f(\theta)) = (\frac{\partial f(\theta)}{\partial \theta}|_{\theta=\hat{\theta}})^T cov(\theta) \frac{\partial f(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}$$

**

$$f'(\theta)|_{\theta=\hat{\theta}}\ (\theta - \hat{\theta}) = \sum \frac{\partial f}{\partial \theta_i}(\theta_i - \hat{\theta}_i)$$

$\rightarrow$ *variance of this is the variance of a constant part times the vector of* $(\theta - \hat{\theta})$.