Exercise 1.1

For $n = 2$, P(RB) $= \frac{r}{(r+b)}\frac{b}{(r+b+1)}$
For $n = 3$, P(RBB) $= \frac{r}{(r+b)}\frac{b}{(r+b+1)}\frac{b+1}{(r+b+2)}$
P(BRB) $= \frac{b}{(r+b)}\frac{r}{(r+b+1)}\frac{b+1}{(r+b+2)}$
It clearly seems that P(RBB)=P(BRB).

Now let it is true for $n = N$ and let $X_1, .., X_N$ denote the sample from the urn. Also assume that first M draws are red and the rest $N - M$ draws are blue.
Therefore, by induction,
P$(X_1, .., X_N) = \frac{r(r+1)...(r+M-1)b(b+1)...(b+N-M-1)}{(r+b)(r+b+1)...(r+b+N-1)}$
For $n = N + 1$, P$(X_1, .., X_N)$ = P$(X_1, .., X_N|X_{N+1}$ is red) + P$(X_1, .., X_N|X_{N+1}$ is blue)
Therefore, P$(X_1, .., X_{N+1}) = \frac{r(r+1)...(r+M-1)b(b+1)...(b+N-M-1)(r+M)}{(r+b)(r+b+1)...(r+b+N-1)(r+b+N)} + \frac{r(r+1)...(r+M-1)b(b+1)...(b+N-M-1)(b+N-M)}{(r+b)(r+b+1)...(r+b+N-1)(r+b+N)}$
= P$(X_1, .., X_N)$

It seems that the distribution does not depend on the order of the sequence. However, pmf at a single point is not identical. So the sequence is exchangeable but not iid.

Exercise 1.2

$X_1, ..., X_M$: Finite sequence.

For $N \leq M$, P$(\sum_{i=1}^N X_i = s|\sum_{i=1}^M X_i = t) = \frac{\binom{t}{s}\binom{M-t}{N-s}}{\binom{M}{N}}$.

Proof:
P$(\sum_{i=1}^N X_i = s|\sum_{i=1}^M X_i = t)$
$\rightarrow \frac{\mathrm{P}(\sum_{i=1}^N X_i=s, \sum_{i=1}^M X_i=t)}{\mathrm{P}(\sum_{i=1}^M X_i=t)}$

$\rightarrow \frac{\mathrm{P}(\sum_{i=1}^N X_i=s, \sum_{i=1}^M X_i=t-s)}{\mathrm{P}(\sum_{i=1}^M X_i=t)}$

$\rightarrow \frac{\mathrm{P}(\sum_{i=1}^N X_i=s)\mathrm{P}(\sum_{i=1}^M X_i=t-s)}{\mathrm{P}(\sum_{i=1}^M X_i=t)}$

Now it is given that X_i is binary and independent of each other, i.e., $\sum_{i=1}^M X_i \sim$ binomial distribution.
$\rightarrow \frac{\binom{N}{s}\binom{M-N}{t-s}}{\binom{M}{t}}$

If we manipulate the expression little bit, we will get $\frac{\binom{t}{s}\binom{M-t}{N-s}}{\binom{M}{N}}$.

Thus we can think of an urn containing M items, of which t are 1's and $M-t$ are 0's. We pick N items without replacement. Then we get the expression: $\frac{\binom{t}{s}\binom{M-t}{N-s}}{\binom{M}{N}}$ which is the probability of obtaining s 1's and $N-s$ 0's.

Exercise 1.3

As $M \to \infty$, $(M\theta)_s \to (M\theta)^s$ and similarly $(M(1-\theta))_{N-s} \to (M(1-\theta))^{N-s}$, $(M)_N \to (M)^N$.
$\therefore \quad \frac{(M\theta)_s(M(1-\theta))_{N-s}}{(M)_N} \to \theta^s(1-\theta)^{N-s}$

Exercise 1.4

$p(x|\lambda) = \exp(\log\left(\frac{\lambda^x e^{-\lambda}}{x!}\right))$
$\to \exp(x\log\lambda - \lambda - \log x!)$
$\to \frac{\exp(x\log\lambda - \lambda)}{x!}$

Therefore, $\eta = \log\lambda$, $T(x) = x$, $A(\eta) = \exp(\log\lambda) = \exp(\eta)$, $h(x) = \frac{1}{x!}$

$p(\underset{\sim}{x}|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}$
$\to (\prod_{i=1}^n \frac{1}{x_i!})\exp\{\sum_{i=1}^n x_i \log\lambda - n\lambda\}$

Therefore, $\eta = \log\lambda$, $T(x) = \sum_{i=1}^n x_i$, $A(\eta) = n\exp(\eta)$, $h(x) = \prod_{i=1}^n \frac{1}{x_i!}$

Exercise 1.5

$X_i \sim \text{Gamma}(\alpha, \beta)$
$p(\underset{\sim}{x}|\alpha,\beta) = \prod_{i=1}^n \frac{\beta^\alpha x_i^{\alpha-1} e^{-\beta x_i}}{\Gamma(\alpha)}$
$\to \frac{\beta^{n\alpha}\prod(x_i^{\alpha-1})e^{-\beta\sum x_i}}{(\Gamma(\alpha))^n}$
$\to \exp\{n\alpha\log\beta - n\log\Gamma(\alpha) + (\alpha-1)\sum\log x_i - \beta\sum x_i\}$

$$\text{Let } \underset{\sim}{\eta} = \begin{pmatrix}\eta_1 \\ \eta_2\end{pmatrix} = \begin{pmatrix}\alpha-1 \\ -\beta\end{pmatrix},$$

$$T(x) = \begin{pmatrix} \sum \log x_i \\ \sum x_i \end{pmatrix},$$
$$h(x) = 1,\ A(\underset{\sim}{\eta}) = n \log \Gamma(\alpha) - n\alpha \log \beta = n\{\log \Gamma(\eta_1 + 1) - (\eta_1 + 1)\log(-\eta_2)\}.$$

Exercise 1.6

$M_{T(X)} = \mathrm{E}(e^{s'T(x)}) = \int \exp\{s'T(x)\}h(X)\exp\{\eta' T(X) - A(\eta)\}dX$
$= \int h(X)\exp\{(s+\eta)'T(X) - A(\eta+s)\}dX \exp\{A(\eta+s) - A(\eta)\} = \exp\{A(\eta+s) - A(\eta)\}$

Exercise 1.7 MGF

$\mathrm{E}(e^{sX}) = \sum_{x=0}^{\infty} \frac{e^{sx}\lambda^x e^{-\lambda}}{x!}$

$\to e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^s)^x}{x!}$

$\to e^{-\lambda} e^{\lambda e^s}$

$\to e^{-\lambda(1-e^s)}$

$m_1 = \frac{d}{ds}\mathrm{E}(e^{sX})|_{s=0} = \frac{d}{ds} e^{-\lambda(1-e^s)}|_{s=0} = e^{-\lambda(1-e^s)}\lambda e^s|_{s=0} = \lambda = \mathrm{E}(X)$

$m_2 = \frac{d^2}{ds^2}\mathrm{E}(e^{sX})|_{s=0} = \frac{d}{ds} e^{-\lambda(1-e^s)}\lambda e^s|_{s=0} = e^{-\lambda(1-e^s)}\lambda e^s + e^{-\lambda(1-e^s)}\lambda^2 e^{2s}|_{s=0} = \lambda + \lambda^2 = \mathrm{E}(X^2)$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$

Thus by MGF, $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$.

Exercise 1.7 CGF

$\mathrm{C}_X(s) = \log \mathrm{E}(e^{sX})$

$\log(\mathrm{M}_X(s)) = -\lambda(1 - e^s)$

$\frac{d}{ds}\log(\mathrm{M}_X(s))|_{s=0} = \lambda e^s|_{s=0} = \lambda = \mathrm{E}(X)$

$\frac{d^2}{ds^2}\log(\mathrm{M}_X(s))|_{s=0} = \lambda e^s|_{s=0} = \lambda = \mathrm{Var}(X)$

Exercise 1.8

$X_1, ...X_N \overset{iid}{\sim} N(\mu, \sigma^2), \sigma^2$:known, $\mu \sim N(\mu_0, \sigma_0^2)$.

$f(\mu|X_1, .., X_N) \propto f(\underset{\sim}{x}|\mu, \sigma^2)f(\mu)$

$= \prod_{i=1}^N f(x_i|\mu, \sigma^2)\frac{1}{\sigma_0\sqrt{2\pi}}e^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}$

$$\propto \exp\left\{-\tfrac{1}{2}[\underbrace{(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})}_{\text{a}}\mu^2 - 2\underbrace{(\frac{\sum x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2})}_{\text{b}}\mu)] - \tfrac{\sum x_i^2}{2\sigma^2} - \tfrac{\mu_0^2}{2\sigma_0^2}\right\}$$

$\propto \exp\left\{-\frac{1}{2}a(\mu - \frac{b}{a})^2\right\}$

$\therefore \quad \mu|X_1, .., X_N \sim N(\mu_*, \sigma_*^2)$, where $\mu_* = \frac{b}{a} = \frac{\frac{\sum x_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}$ and $\sigma_*^2) = a^{-1} = (\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2})^{-1}$

Exercise 1.9

$f(\omega|X_1, .., X_N) \propto f(\underset{\sim}{x}|\mu, \omega)f(\omega)$

$= (\frac{\omega}{2\pi})^{\frac{n}{2}} \exp\left\{-\frac{\omega}{2}\sum(x_i-\mu)^2\right\}\frac{\beta^\alpha}{\Gamma\alpha}\omega^{\alpha-1}\exp\{-\beta\omega\}$

$\propto \omega^{\frac{n}{2}+\alpha-1}\exp\left\{-(\beta+\frac{1}{2}\sum(x_i-\mu)^2)\omega\right\}$

$\therefore \quad \omega|X_1, .., X_N \sim \text{Gamma}(\frac{n}{2}+\alpha, \quad \beta + \frac{1}{2}\sum(x_i-\mu)^2)$

Exercise 1.10

$f(x) = \int_0^\infty f(x|\omega)f(\omega)d\omega$

$= \int(\frac{\omega}{2\pi})^{1/2}e^{-\frac{x^2}{2\sigma^2}}\frac{\beta^\alpha}{\Gamma\alpha}\omega^{\alpha-1}e^{-\beta\omega}d\omega$

$$= \tfrac{1}{2}\tfrac{\beta^\alpha}{\Gamma\alpha}\tfrac{\Gamma(\frac{1}{2}+\alpha)}{(\beta+\frac{x^2}{2})^{\frac{1}{2}+\alpha}}\underbrace{\int\frac{(\beta+\frac{x^2}{2})^{\frac{1}{2}+\alpha}\omega^{\frac{1}{2}+\alpha-1}\exp\left\{-(\beta+\frac{x^2}{2})\omega\right\}}{\Gamma(\frac{1}{2}+\alpha)}d\omega}_{1}$$

Exercise 1.11

(a) $\text{E}[(X-\mu)(X-\mu)^T] = \text{E}(XX^T - 2X\mu^T + \mu\mu^T)$

$= \text{E}(XX^T) - 2\underbrace{\text{E}(X)}_{\mu}\mu^T + \mu\mu^T$

$= \text{E}(XX^T) - \mu\mu^T$

(b) $\text{Cov}(X) = \Sigma$
$\text{Cov}(AX + b) = A\text{Cov}(X)A^T = A\Sigma A^T$

Exercise 1.12

(a) $X \sim N(m, v^2)$
$M_X(t) = \text{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2v^2}(x-m)^2} dx$
$= \int \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2}[\frac{x^2}{v^2} - 2(\frac{m}{v^2} + t)x]} e^{-\frac{m^2}{2v^2}} dx$

After some adjustment we get:
$$= e^{mt + \frac{1}{2}v^2t^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2v^2}(x-(m+tv^2))^2} dx}_{1}$$

$\therefore \quad M_X(t) = e^{mt + \frac{1}{2}v^2t^2}$

(b) $Y_1, ...Y_N \overset{iid}{\sim} N(0, 1)$
$f(\underset{\sim}{y}) = \frac{1}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\underset{\sim}{y}^T \underset{\sim}{y}\right\}$
$M_Y(t) = \text{E}(e^{t'y}) = \text{E}(e^{\sum t_i y_i}) = \prod \text{E}(e^{t_i y_i}) = e^{\frac{1}{2}\sum t_i^2} = e^{\frac{1}{2}t't}$

Exercise 1.13

$X \sim N_p(\mu, \Sigma)$, then a linear combination $a'X$ is distributed as $N_1(a'\mu, a'\Sigma a)$. MGF of multivariate normal distribution is: $\exp\left\{t'\mu + \frac{1}{2}t'\Sigma t\right\}$.
Assuming that $Z = a'X$.
MGF of Z: $M_Z(t) = \exp\left\{t'a'\mu + \frac{1}{2}t'a'\Sigma at\right\}$
Let at=s, then $M_Z(t) = \exp\left\{s'\mu + \frac{1}{2}s'\Sigma s\right\}$
Therefore, the above expression of MGF conforms with the MGF of a univariate normal distribution $\rightarrow a'X$ is normally distributed.

Exercise 1.14

$X = DZ + \mu$

$M_X(t) = \text{E}(e^{t'X}) = \text{E}(e^{t'(DZ+\mu)}) = e^{t'\mu}\text{E}(e^{t'DZ})$.

Let $D't = s$

$= e^{t'\mu}\mathrm{E}(e^{s'Z}) = e^{t'\mu + \frac{1}{2}t'DD't}.$

$\therefore \quad \Sigma = DD'$

First, we generate samples from standard normal distribution and then we transform the variables using linear combinations.

Exercise 1.15

$X \sim N(\mu, \Sigma), \quad Z = D^{-1}(X - \mu).$

$|J| = |DD'|^{-1/2} = |\Sigma|^{-1/2}, \quad \Sigma = DD'$

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

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

Exercise 1.16

$X_1 = (I \quad 0)X =$ Linear combination of X $\rightarrow$ normally distributed.

$\mathrm{E}(X_1) = (I \quad 0)\mu = \mu_1, \quad \mathrm{Cov}(X_1) = \Sigma_{11}, \quad X_1 \sim N(\mu_1, \Sigma_{11})$

Exercise 1.17

I found a solution to this answer online. The link is given below:
Inverse of a partitioned symmetric matrix

Exercise 1.18

Define $Z = X_1 + AX_2, \quad A = -\Sigma_{12}\Sigma_{22}^{-1}$, Z is a linear combination of normally distributed random vectors, so the distribution of Z is normal.
$\mathrm{Cov}(Z, X_2) = \mathrm{Cov}(X_1 + AX_2, X_2) = \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(AX_2, X_2) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0$
Correlation 0 within normally distributed random variables indicates that Z and X_2 are independent.

$\mathrm{E}(X_1|X_2) = \mathrm{E}(Z - AX_2|X_2) = \mathrm{E}(Z) - AX_2 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \Sigma_{12}\Sigma_{22}^{-1}x_2$
$\mathrm{Cov}(X_1|X_2) = \mathrm{Var}(X_1) + A\mathrm{Var}(X_2)A' + 2\mathrm{Cov}(X_1, X_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Exercise 1.19

$\mathrm{E}(y_1|x_i) = x_i'\beta, \quad \mathrm{Cov}(x_i, \epsilon_i) = 0$

$\mathrm{Cov}(x_i, \epsilon_i) = 0 \to \mathrm{E}(x_i\epsilon_i) = \mathrm{E}(x_i(y_1 - x_i'\beta)) = 0 = g(\beta)$ (say).

Sample moment: $g(\hat{\beta}) = \frac{1}{n}\sum x_i(y_1 - x_i'\beta) = 0$
Solving the above equation we get: MM estimator $\hat{\beta}_{MM} = (\sum x_i x_i')^{-1}\sum x_i y_i$.

Exercise 1.20

$f(\underset{\sim}{y}|\underset{\sim}{x}) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\beta)'(\underset{\sim}{y} - X\beta)\right\}$
Maximizing the above function will give the same estimate as least square.

Exercise 1.21

Maximizing log-likelihood is equivalent to minimizing $(\underset{\sim}{y} - X\beta)'(\underset{\sim}{y} - X\beta)$.

Exercise 1.22

$l(\beta) = (y - X\beta)'(y - X\beta) + \lambda(\beta'\beta - t)$

$\nabla l(\beta) = -2X'(y - X\beta) + 2\lambda\beta = 0 \quad \to \hat{\beta}_{ridge} = (X'X + \lambda I_p)^{-1}X'y$

λ: tuning parameter, inclusion of λ makes problem non-singular even if $X'X$ is not invertible.
λ: controls the size of the coefficients, controls the amount of regularization.
$\lambda \to 0$, we obtain the least square solution.
$\lambda \to \infty$, $\hat{\beta}_{ridge} = 0$ (Intercept-only model).

Exercise 1.23

$\hat{\beta}_{LS} = (X'X)^{-1}X'y$
$y|X \sim N(X\beta, \sigma^2 I)$, $\hat{\beta}_{LS}$ is a linear combination of normally distributed variables $\to$ normally distributed.

$\mathrm{E}(\hat{\beta}_{LS}) = \beta$ and $\mathrm{Cov}(\hat{\beta}_{LS}) = \sigma^2(X'X)^{-1}$.
Let β is $p \times 1$ vector, then $\hat{\beta}_{LS} \sim N_p(\beta, \sigma^2(X'X)^{-1})$.

Exercise 1.24

$\hat{\beta}_{ridge} = (X'X + \lambda I_p)^{-1}X'y = (X'X + \lambda I_p)^{-1}(X'X)(X'X)^{-1}X'y = \underbrace{(X'X + \lambda I_p)^{-1}(X'X)}_{\mathrm{W}}\hat{\beta}_{LS}$

$\hat{\beta}_{ridge} = W\hat{\beta}_{LS}$: Linear combination of normally distributed random vectors $\rightarrow$ normally distributed.

$\therefore \quad \mathrm{E}(\hat{\beta}_{ridge}) = \mathrm{E}(W\hat{\beta}_{LS}) = W\beta$
$\mathrm{Var}(\hat{\beta}_{ridge}) = W\mathrm{Var}(\hat{\beta}_{LS})W' = \sigma^2 W(X'X)^{-1}W'$

Exercise 1.25

Please refer to this link for code: **Solution of prob 1.25**

When σ^2 is unknown, we can estimate by $\frac{y'(I-M)y}{n-p}$ where p is the rank of $X'X$ ($n = 102$ and $p = 4$). Then we can use the result of Exercise 1.23 which shows that $\mathrm{Cov}(\hat{\beta}_{LS}) = \sigma^2(X'X)^{-1}$.

Residual standard error obtained from lm function is 2575 on 98 degrees of freedom which is exactly same as the standard error ($\sqrt{\sigma^2}$) obtained from the second method $2574.709 \approx 2575$.
Standard errors of the coefficients from two methods:

	LM	2nd method
Intercept	254.9	254.9342
education	511.9	511.9442
women	271.4	271.4444
prestige	514.6	514.5795

Exercise 1.26

$f(\theta) = \sum_{i=1}\theta_i$.
Standard Error: $\sqrt{\mathrm{Var}(\sum\theta_i)} = \sqrt{\sum\mathrm{Var}(\theta_i) + 2\sum_{i<j}\mathrm{Cov}(\theta_i, \theta_j)}$.

Exercise 1.27

1st Method:
X is a r.v. with the central moments $\mu_i = \mathrm{E}(X-\mu)^i$

Consider a Taylor series expansion of $f(x)$ around $\mathrm{E}(X) = \mu$:
$f(X) = f(\mu) + \frac{X-\mu}{1!} f'(\mu) + \frac{(X-\mu)^2}{2!} f''(\mu) + \frac{(X-\mu)^3}{3!} f'''(\mu) + ...$

$\mathrm{E}(f(X)) = \mathrm{E}(f(\mu)) + \frac{\mathrm{E}(X-\mu)}{1!} f'(\mu) + \frac{\mathrm{E}((X-\mu)^2)}{2!} f''(\mu) + ...$

$\mathrm{E}(f(X)) = \mathrm{E}(f(\mu)) + 0 + \frac{\sigma^2}{2!} f''(\mu) + \frac{\mu_3}{3!} f'''(\mu) + ...$

$f(X) - \mathrm{E}(f(X)) = (X-\mu) f'(\mu) + \frac{(X-\mu)^2}{2} f''(\mu) - \frac{\sigma^2}{2} f''(\mu) + ...$

$\mathrm{E}[(f(X) - \mathrm{E}[f(X)])^2] = \mathrm{E}[(X-\mu)^2 (f'(\mu))^2] + \mathrm{E}[\frac{(X-\mu)^4}{4} (f''(\mu))^2] + \frac{\sigma^4}{4} (f''(\mu))^2 + \mathrm{E}[2\frac{1}{2}(X-\mu)^3 f'(\mu) f''(\mu)] - \mathrm{E}[2\frac{1}{2}(X-\mu) f'(\mu) \sigma^2 f''(\mu)] - \mathrm{E}[2\frac{1}{2}\frac{(X-\mu)^2}{2} f''(\mu) \sigma^2 f''(\mu)]$.

$\mathrm{E}[(f(X) - \mathrm{E}[f(X)])^2] = \sigma^2 (f'(\mu))^2 + \frac{\mu_4 - \sigma^4}{4} (f''(\mu))^2 + \mu_3 f'(\mu) f''(\mu) + ... = \mathrm{Var}[f(X)]$
SD of $f(X) = \sqrt{\mathrm{Var}[f(X)]}$

2nd Method:
Using Taylor expansion on $\theta = \hat{\theta}$,
$f(\theta) = f(\hat{\theta}) + f'(\hat{\theta})(\theta - \hat{\theta}) + O((\theta - \hat{\theta})^2)$
$\mathrm{Var}(f(\theta)) = \mathrm{Var}(f(\hat{\theta})) + \mathrm{Var}(f'(\hat{\theta})(\theta - \hat{\theta})) = \mathrm{Var}(f'(\hat{\theta})(\theta - \hat{\theta})) = f'(\hat{\theta}) \mathrm{Var}(\theta - \hat{\theta})(f'(\hat{\theta}))' = f'(\hat{\theta}) \sum [f'(\hat{\theta})]^T$