Excercise 1

Yinan Zhu

1 Bayesian inference in simple conjugated families

(A)

$$p(\omega|\vec{x}) = \frac{p(\vec{x}|\omega)p(\omega)}{\int_\Omega p(\vec{x}|\alpha)p(\alpha)d\alpha} \tag{1}$$

For Bernoulli sampling model, the distribution of $\vec{x}$ given ω is

$$p(\vec{x}|\omega) = \prod_i p(x_i|\omega) = \omega^p(1-\omega)^q \tag{2}$$

p and q are the number of two outcomes observed among all samples respectively. Of course $p+q=N$.

$$\begin{aligned} p(\omega|\vec{x}) &= \frac{\omega^{p+a-1}(1-\omega)^{N-q+b-1}}{\int_\Omega \alpha^{p+a-1}(1-\alpha)^{N-p+b-1}d\alpha} \\ &= \frac{\Gamma(N+a+b)}{\Gamma(p+a)\Gamma(N-p+b)}\omega^{p+a-1}(1-\omega)^{N-p+b-1} \end{aligned} \tag{3}$$

(B) The PDF of the sum of two independent random variables is the convolution of their PDF.

$$\begin{aligned} p(Y_2=y) &= \int p(X_1=y-x)p(X_2=x)dx \\ &= \frac{e^{-y}}{\Gamma(a_1)\Gamma(a_2)}\int_{x<y} x^{a_2-1}(y-x)^{a_1-1}dx \\ &= \frac{y^{a_1+a_2-1}e^{-y}}{\Gamma(a_1)\Gamma(a_2)}\int_{x<y}(\frac{x}{y})^{a_2-1}(1-\frac{x}{y})^{a_1-1}d(\frac{x}{y}) \\ &= \frac{1}{\Gamma(a_1+a_2)}y^{a_1+a_2-1}e^{-y} \end{aligned} \tag{4}$$

Or directly use the property: the sum of independent gamma distributions $\sum Ga(a_i,b)$ is equivalent to $Ga(\sum a_i, b)$

The transformation can be written as

$$
\begin{aligned}
X_1 &= Y_1 Y_2 \\
X_2 &= Y_2 - Y_1 Y_2 \qquad (5)
\end{aligned}
$$

Jaccobian is

$$
J(X_1 X_2 | Y_1 Y_2) = \begin{vmatrix} \frac{\partial X_1}{\partial Y_1} & \frac{\partial X_1}{\partial Y_2} \\ \frac{\partial X_2}{\partial Y_1} & \frac{\partial X_2}{\partial Y_2} \end{vmatrix} = |Y_2| \qquad (6)
$$

$$
\begin{aligned}
p(Y_1 = y) &= \int p(X_1 = xy) p(X_2 = x - xy) |x| dx \\
&= \frac{y^{a_1 - 1}(1-y)^{a_2-1}}{\Gamma(a_1)\Gamma(a_2)} \int x^{a_1+a_2-1} e^{-x} dx \\
&= \frac{\Gamma(a_1+a_2) y^{a_1-1}(1-y)^{a_2-1}}{\Gamma(a_1)\Gamma(a_2)} \qquad (7)
\end{aligned}
$$

We can simlate $Beta(a_1, a_2)$ by making $X_1 = Ga(a_1, 1)$ and $X_2 = Ga(a_2, 1)$ and then compute $X_1/(X_1 + X_2)$ for each data point generated.

(C)

$$
p(\theta|\vec{x}) = \frac{p(\vec{x}|\theta)p(\theta)}{\int_\Omega p(\vec{x}|\alpha)p(\alpha)d\alpha} \qquad (8)
$$

$$
p(\vec{x}|\theta) = \prod_i p(x_i|\theta) = \text{Constant} \times e^{-\frac{\sum_i (x_i-\theta)^2}{2\sigma^2}} \qquad (9)
$$

$$
\begin{aligned}
p(\omega|\vec{x}) &= \frac{e^{-\frac{\sum_i (x_i-\theta)^2}{2\sigma^2} - \frac{(\theta-m)^2}{2v^2}}}{\int_\Omega e^{-\frac{\sum_i (x_i-\alpha)^2}{2\sigma^2} - \frac{(\alpha-m)^2}{2v^2}} d\alpha} \\
&= \frac{1}{\sqrt{2V^2\pi}} e^{-\frac{(\omega-M)^2}{2V^2}} \qquad (10) \\
V^2 &= \frac{\sigma^2 v^2}{\sigma^2 + N v^2} \qquad (11) \\
M &= \frac{v^2 \sum_i x_i + \sigma^2 m}{\sigma^2 + N v^2} \qquad (12)
\end{aligned}
$$

The resut is another Gaussian distribution

(D)

$$
p(\omega|\vec{x}) = \frac{p(\vec{x}|\omega)p(\omega)}{\int_\Omega p(\vec{x}|\alpha)p(\alpha)d\alpha}
$$

$$
\begin{aligned}
&= \frac{e^{-(\frac{\sum_i (x_i-\theta)^2}{2}+b)\omega}\omega^{a+\frac{N}{2}-1}}{\int e^{-(\frac{\sum_i (x_i-\theta)^2}{2}+b)\alpha}\alpha^{a+\frac{N}{2}-1}d\alpha} \\
&= \frac{(\frac{\sum_i (x_i-\theta)^2}{2}+b)^{a+\frac{N}{2}}}{\Gamma(a+\frac{N}{2})} e^{-(\frac{\sum_i (x_i-\theta)^2}{2}+b)\omega}\omega^{a+\frac{N}{2}-1} \qquad (13)
\end{aligned}
$$

Transform back to the distribution of $\frac{1}{\omega}$, using the fact that $\sigma^2 = \frac{1}{\omega}$ is strictly positive and monotonous, it can be shown that their pdf has relation

$$
p(\sigma^2) = \frac{1}{\sigma^4} p(\omega)|_{\omega=\frac{1}{\sigma^2}} \qquad (14)
$$

so we have

$$
p(\sigma^2|\vec{x}) = \frac{(\frac{\sum_i (x_i-\theta)^2}{2}+b)^{a+\frac{N}{2}}}{\Gamma(a+\frac{N}{2})} \frac{e^{-(\frac{\sum_i (x_i-\theta)^2}{2}+b)\frac{1}{\sigma^2}}}{\sigma^{2a+N+2}} \qquad (15)
$$

(E)

$$
\begin{aligned}
p(\omega|\vec{x}) &= \frac{p(\vec{x}|\omega)p(\omega)}{\int_\Omega p(\vec{x}|\alpha)p(\alpha)d\alpha} \\
&= \frac{e^{-\sum_i \frac{(x_i-\theta)^2}{2\sigma_i^2} - \frac{(\theta-m)^2}{2v^2}}}{\int_\Omega e^{-\sum_i \frac{(x_i-\alpha)^2}{2\sigma_i^2} - \frac{(\alpha-m)^2}{2v^2}} d\alpha} \\
&= \frac{1}{\sqrt{2V^2\pi}} e^{-\frac{(\omega-M)^2}{2V^2}} \qquad (16) \\
\frac{1}{V^2} &= \frac{1}{v^2} + \sum_i \frac{1}{\sigma_i^2} \qquad (17) \\
M &= \frac{\sum_i \frac{x_i}{\sigma_i^2} + \frac{m}{v^2}}{\frac{1}{v^2} + \sum_i \frac{1}{\sigma_i^2}} \qquad (18)
\end{aligned}
$$

(F) Compute the distribution of σ^2 from that of $\frac{1}{\sigma^2}$

$$
p(\sigma^2 = \omega^2) = \frac{1}{\omega^4} p(\frac{1}{\sigma^2} = \frac{1}{\omega^2}) = \frac{b^a \omega^{-2a-2} e^{-\frac{b}{\omega^2}}}{\Gamma(a)} \qquad (19)
$$

$$
\begin{aligned}
p(x) &= \int p(x|\sigma^2=\omega^2)p(\sigma^2=\omega^2)d\omega^2 \\
&= \frac{b^a}{\sqrt{2\pi}\Gamma(a)} \int_0^\infty \omega^{-2a-3} e^{-\frac{b}{\omega^2} - \frac{x^2}{2\omega^2}} d\omega^2 \qquad (20)
\end{aligned}
$$

doing transformation $d\omega^2 = -\mu^{-2} d\mu$ where $\mu = \frac{1}{\omega^2}$. Also use Gamma integral:

$$\begin{aligned} p(x) &= -\frac{b^a}{\sqrt{2\pi}\Gamma(a)} \int_\infty^0 \mu^{a-\frac{1}{2}} e^{-(b+\frac{x^2}{2})\mu} d\mu \\ &= \frac{\Gamma(a+\frac{1}{2})}{\sqrt{2b\pi}\Gamma(a)} (1+\frac{x^2}{2b})^{-\frac{1}{2}-a} \end{aligned} \tag{21}$$

This is Student's t-distribution.

2 The multivariate normal distribution

(A) The expectation value is linear. If x and y are random variables and a and b are constants:

$$E(ax+by) = aE(x) + bE(y) \tag{22}$$

This can be generaized for matrix coefficients and random variables

$$(E(AX))_{ij} = E(\sum_k A_{ik}X_{kj}) = \sum_k A_{ik}E(X_{kj}) = (AE(X))_{ij} \tag{23}$$

$$E((\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T) = E(\vec{x}\vec{x}^T) - \vec{\mu}E(\vec{x}^T) - E(\vec{x})\vec{\mu}^T + \vec{\mu}\vec{\mu}^T = E(\vec{x}\vec{x}^T) - \vec{\mu}\vec{\mu}^T \tag{24}$$

$$\begin{aligned} Cov(A\vec{x}) &= E(A\vec{x}\vec{x}^T A^T) - AE(\vec{x})E(\vec{x})^T A^T \\ &= AE(\vec{x}\vec{x}^T)A^T - AE(\vec{x})E(\vec{x})^T A^T = ACov(\vec{x})A^T \end{aligned} \tag{25}$$

Adding b will not change the result as it is cancelled in $\vec{x}-\vec{\mu}$, so we have

$$Cov(A\vec{x}+b) = ACov(\vec{x})A^T \tag{26}$$

(B)

$$p(\vec{z}) = \prod_i p(z_i) = \frac{1}{(2\pi)^{\frac{dim(\vec{z})}{2}}} e^{-\frac{|\vec{z}|^2}{2}} \tag{27}$$

$$M(\vec{x},\vec{t}) = \prod_i M(x_i,t_i)) = e^{\frac{|\vec{t}|^2}{2}} \tag{28}$$

(C) If $\vec{x} \sim N(\mu,\Sigma)$, since $\vec{a}^T\vec{x} \sim N(\vec{a}^T\vec{\mu}, \vec{a}^T\Sigma\vec{a})$, its moment generating function evaluated at $t=1$ is

$$E(e^{\vec{a}^T\vec{x}}) = e^{\vec{a}^T\vec{\mu}+\frac{\vec{a}^T\Sigma\vec{a}}{2}} \tag{29}$$

this is the moment generating function of $\vec{x}$ evaluated at $\vec{t}=\vec{a}$. It is the same as the form we want.

If x has the proposed moment generating function, we can evaluate it at $\vec{t} = \vec{a}$ for any nonzero $\vec{a}$. It will be a moment generating function of $\vec{a}^T \vec{x}$ at $t = 1$. Easy to see it is of the form of a moment generating function from Gaussian distribution.

(D)

$$\begin{aligned} E(e^{t^T Lz + t^T \mu}) &= e^{t^T \mu} E(e^{(L^T t)^T z}) \\ &= e^{t^T \mu + \frac{(L^T t)^T (L^T t)}{2}} \\ &= e^{t^T \mu + \frac{t^T (LL^T) t}{2}} \end{aligned} \tag{30}$$

From (A) LL^T is the covariance matrix of $Lz + \mu$. From (C) this proves $Lz + \mu$ is multivariate normal.

(E) Because covariance matrix is symmetric and positive semi-definite, it can be written as $\Sigma = L^T DL$, where D is diagonal with every entry nonegative and and L is orthogonal.

Define $y = D^{-\frac{1}{2}} L(x - \mu)$. Using similar derivation as in last question.

$$\begin{aligned} E(e^{t^T y}) &= E(e^{t^T D^{-\frac{1}{2}} Lx - t^T D^{-\frac{1}{2}} \mu}) \\ &= E(e^{t^T D^{-\frac{1}{2}} Lx}) \\ &= e^{\frac{t^T (D^{-1} L \Sigma L^T) t}{2}} = e^{\frac{t^T (DD^{-1}) t}{2}} \end{aligned} \tag{31}$$

(D^{-1} is not the inverse of D).y is a collection of independent standard normal distribution and zeros. $x = L^T D^{\frac{1}{2}} y + \mu$ is the desired transformation. To simulate a multivariate Gaussian. Diagonalize its covariance matrix and simulate univariate Gaussian whose mean is zero and standard deviation being diagonal entries of D. Then perform the transformation.

(F) There exists an array of independent normal distribution with mean 0 which can be transformed to x after an affine transformation. They have PDF:

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{dim(x)}{2}} \prod_i \sigma_i} e^{-\sum_i \frac{z_i^2}{2\sigma_i^2}} \tag{32}$$

Let L be the rotation in the affine transformation. From previous question we have $\Sigma = LDL^T$ where $L = diag(\sigma_1^2, \sigma_2^2, \cdots)$ and $\det(\Sigma) = \prod_i \sigma_i$. Set $y = Lz$, we have

$$f(\vec{y}) = |L| f(L\vec{z}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{y^T \Sigma^{-1} y}{2}} \tag{33}$$

This is a quadratic form of $y = x - \mu$.

(G) $y_1 = Ax_1 \sim N(A\mu_1, A\Sigma_1 A^T)$, $y_2 = Bx_2 \sim N(B\mu_2, B\Sigma_2 B^T)$. Obviously $E(\vec{x}) = E(\vec{y_1}) + E(\vec{y_2}) = A\vec{\mu_1} + B\vec{\mu_2}$

For covariance matrix, since they are dependent, $E(\vec{y_1}\vec{y_2}^T) = E(\vec{y_1})E(\vec{y_2})$

$$\begin{aligned} Cov(\vec{x}) &= E((\vec{y_1}+\vec{y_2})(\vec{y_1}+\vec{y_2})^T) \\ &= Cov(\vec{y_1}) + Cov(\vec{y_2}) + E(\vec{y_2})E(\vec{y_1})^T + E(\vec{y_1})E(\vec{y_2})^T \\ &= A\Sigma_1 A^T + B\Sigma_2 B^T + B\vec{\mu_2}\vec{\mu_1}^T A^T + A\vec{\mu_1}\vec{\mu_2}^T B^T \end{aligned} \tag{34}$$

So $x \sim N(A\vec{\mu_1} + B\vec{\mu_2}, A\Sigma_1 A^T + B\Sigma_2 B^T + B\vec{\mu_2}\vec{\mu_1}^T A^T + A\vec{\mu_1}\vec{\mu_2}^T B^T)$

(A) $\vec{x}_1 = A\vec{x}$ where

$$A = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \tag{35}$$

$$\begin{aligned} f(\vec{x}_1) &= \frac{1}{\sqrt{\det(2\pi A\Sigma A^T)}} e^{-\frac{(\vec{x}_1-\vec{\mu}_1)^T (A\Sigma A^T)^{-1}(\vec{x}_1-\vec{\mu}_1)}{2}} \\ &= \frac{1}{\sqrt{\det(2\pi\Sigma_{11})}} e^{-\frac{(\vec{x}_1-\vec{\mu}_1)^T \Sigma_{11}^{-1}(\vec{x}_1-\vec{\mu}_1)}{2}} \end{aligned} \tag{36}$$

(B)

$$\Omega_{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} \tag{37}$$

$$\Omega_{12} = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1} \tag{38}$$

$$\Omega_{22} = (\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1} \tag{39}$$

(C)

$$\begin{aligned} f(\vec{x}_1|\vec{x}_2) &= f(\vec{x}_1, \vec{x}_2)/f(\vec{x}_2) \\ &\propto e^{-\frac{(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu}) - (\vec{x}_2-\vec{\mu}_2)^T\Sigma_{22}^{-1}(\vec{x}_2-\vec{\mu})}{2}} \end{aligned} \tag{40}$$

Compute the arguments of exponentials

$$\begin{aligned} & (\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu}) - (\vec{x}_2-\vec{\mu}_2)^T\Sigma_{22}^{-1}(\vec{x}_2-\vec{\mu}) \\ = & (\vec{x}_1^T - \vec{\mu}_1^T, \vec{x}_2^T - \vec{\mu}_2^T) \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{pmatrix} \begin{pmatrix} \vec{x}_1 - \vec{\mu}_1 \\ \vec{x}_2 - \vec{\mu}_2 \end{pmatrix} \\ & -(\vec{x}_2-\vec{\mu}_2)^T\Sigma_{22}^{-1}(\vec{x}_2-\vec{\mu}) \\ = & ((\vec{x}_1-\vec{\mu}_1)^T\Omega_{11} + (\vec{x}_2-\vec{\mu}_2)^T\Omega_{12}^T, (\vec{x}_1-\vec{\mu}_1)^T\Omega_{12} + (\vec{x}_2-\vec{\mu}_2)^T\Omega_{22}) \\ & \cdot \begin{pmatrix} \vec{x}_1 - \vec{\mu}_1 \\ \vec{x}_2 - \vec{\mu}_2 \end{pmatrix} - (\vec{x}_2-\vec{\mu}_2)^T\Sigma_{22}^{-1}(\vec{x}_2-\vec{\mu}) \\ = & (\vec{x}_1-\vec{\mu}_1)^T\Omega_{11}(\vec{x}_1-\vec{\mu}_1) + 2(\vec{x}_2-\vec{\mu}_2)^T\Omega_{12}^T(\vec{x}_1-\vec{\mu}_1) \\ & +(\vec{x}_2-\vec{\mu}_2)^T(\Omega_{22} - \Sigma_{22}^{-1})(\vec{x}_2-\vec{\mu}_2) \\ = & ((\vec{x}_1-\vec{\mu}_1)^T + (\vec{x}_2-\vec{\mu}_2)^T\Omega_{12}^T\Omega_{11}^{-1})\Omega_{11}(\vec{x}_1-\vec{\mu}_1+\Omega_{11}^{-1}\Omega_{12}(\vec{x}_2-\vec{\mu}_2)) \\ & +g(\vec{x_2}) \end{aligned} \tag{41}$$

So the functin $f(\vec{x_1}|\vec{x_2})$ peaks at values that makes $\vec{x}_1 - \vec{\mu}_1 + \Omega_{11}^{-1}\Omega_{12}(\vec{x}_2 - \vec{\mu}_2)$ vanish. This gives $\vec{x}_1$ as a linear function of $\vec{x}_2$

3 Multiple regression: three classical principles for inference

1.