

Excercise 1

Yinan Zhu

1 Bayesian inference in simple conjugated families

(A)

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

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

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

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

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

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

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

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

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

Jacobian is

$$J(X_1 X_2 | Y_1 Y_2) = \left| \begin{array}{cc} \frac{\partial X_1}{\partial Y_1} & \frac{\partial X_1}{\partial Y_2} \\ \frac{\partial X_2}{\partial Y_1} & \frac{\partial X_2}{\partial Y_2} \end{array} \right| = |Y_2| \quad (6)$$

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

We can simulate $Beta(a_1, a_2)$ 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)

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

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

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

$$V^2 = \frac{\sigma^2 v^2}{\sigma^2 + N v^2} \quad (11)$$

$$M = \frac{v^2 \sum_i x_i + \sigma^2 m}{\sigma^2 + N v^2} \quad (12)$$

The result is another Gaussian distribution

(D)

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

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

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

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

so we have

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

(E)

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

$$\frac{1}{V^2} = \frac{1}{v^2} + \sum_i \frac{1}{\sigma_i^2} \quad (17)$$

$$M = \frac{\sum_i \frac{x_i}{\sigma_i^2} + \frac{m}{v^2}}{\frac{1}{v^2} + \sum_i \frac{1}{\sigma_i^2}} \quad (18)$$

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

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

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

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

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

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:

$$E(ax + by) = aE(x) + bE(y) \quad (22)$$

This can be generalized for matrix coefficients and random variables

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

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

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

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

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

(B)

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

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

(C)

(D)

$$\begin{aligned} E(e^{t^T Lz + t^T \mu}) &= e^{t^T \mu} E(e^{(L^T t)^T z}) \\ &= e^{t^T \mu + \frac{(L^T t)^T (L^T t)}{2}} \\ &= e^{t^T \mu + \frac{t^T (LL^T) t}{2}} \end{aligned} \quad (29)$$

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 $\Sigma = L^T D L$, where D is diagonal with every entry nonnegative and L is orthogonal.

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

$$E(e^{t^T y}) = e^{\frac{t^T (D^{-1} L \sigma L^T) t}{2}} = e^{\frac{|t|^2}{2}} \quad (30)$$

(D^{-1} is not the inverse of D). y is a collection of independent standard normal distribution. Let $y' = (y, 0, \dots, 0)$, the number of zeros same to the number of zero diagonal elements. $x = L^T D^{\frac{1}{2}} (y' + \mu)$ is the desired transformation.