Bibliography


Appendix A

Probability distributions

A.1 The normal distribution

The p.d.f. of a \( k \)-dimensional multivariate normal distribution is specified as follows [27]:

\[
g(x|\mu, \Sigma) = (2\pi)^{-D/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}. \tag{A.1}
\]

A normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \) is usually written as \( \mathcal{N}(\mu, \Sigma) \). If we assume that the individual elements of the random variable \( x \) are independent (diagonal covariance matrix), we can write Equation A.1 as follows

\[
g(x|\mu, \Sigma) = (2\pi)^{-D/2} \left( \prod_{d=1}^D \sigma_d \right)^{-1} e^{-\frac{1}{2} \sum_{d=1}^D \sigma_d^{-1} (x_d-\mu_d)^2} \tag{A.2}
\]

and
Given a random sample of $k$-dimensional random vectors $(X_1, X_2, \ldots, X_n)$ from a multivariate normal distribution with zero mean and covariance matrix $\Sigma$, the random variable $V$ has a Wishart distribution [27] with $n$ degrees of freedom and parametric matrix $\Sigma$ when,

$$V = \sum_{i=1}^{n} X_i X_i^T. \tag{A.4}$$

For any $k \times k$, positive definite, symmetric matrix $v$, we have

$$g(v|n, \Sigma) = c|\Sigma|^{-n/2}|v|^{-(n-D-1)/2}e^{-\frac{1}{2}tr(\Sigma^{-1}v)}. \tag{A.5}$$

Here, $tr(\Sigma^{-1}v)$ is the trace of the matrix $\Sigma^{-1}v$. The value $c$ is a normalizing constant which ensures that the integral of $g(v|n, \Sigma)$ is equal to one.

If $\Sigma$ and $v$ are assumed to be diagonal, then Equation A.5 can be rewritten as follows:

$$g(v|n, \Sigma) = c\left(\prod_{d=1}^{D} \tau_d\right)^{n/2}\left(\prod_{d=1}^{D} v_d\right)^{-(n-D-1)/2}e^{-\frac{1}{2}\Sigma_d v_d \tau_d}. \tag{A.6}$$

where $\tau$ is the precision matrix for the Wishart distribution ($\tau = \Sigma^{-1}$) and

$$\log(g(v|n, \Sigma)) = \log(c) + \frac{n}{2} \cdot \log\left(\prod_{d=1}^{D} \tau_d\right) + \frac{n-D-1}{2} \log\left(\prod_{d=1}^{D} v_d\right) - \frac{1}{2} \sum_{d=1}^{D} v_d \tau_d. \tag{A.7}$$
A.3 Dirichlet distribution

Given the random vector \( X = (X_1, X_2, \ldots, X_k)^T \) with the following properties: For a point \( x = (x_1, x_2, \ldots, x_k)^T \) in \( \mathbb{R}^k \), \( x_i > 0; i = 1, \ldots, k \) and \( \sum_{i=1}^{k} x_i = 1 \), then the random vector \( X \) has a Dirichlet distribution [27]:

\[
g(x|\alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_k)} x_1^{\alpha_1-1}x_2^{\alpha_2-1}\cdots x_k^{\alpha_k-1}, \tag{A.8}
\]

where \( \Gamma(\alpha) \) is the gamma function and \( \alpha \) is the parametric vector of the distribution and

\[
\log(g(x|\alpha)) = \log(\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)) - \log(\Gamma(\alpha_1)) - \cdots - \log(\Gamma(\alpha_k))
+ (\alpha_1 - 1)\log(x_1) + \cdots + (\alpha_k - 1)\log(x_k). \tag{A.9}
\]

A.4 The gamma distribution

A random variable \( X \) has a gamma distribution [27] with parameters \( \alpha \) and \( \beta \) \((\alpha > 0, \beta > 0)\) if \( X \) has an absolutely continuous distribution whose p.d.f. is

\[
g(x|\alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.10}
\]

where \( \Gamma(\alpha) \) is the gamma function, which is defined as

\[
\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt. \tag{A.11}
\]

If a random variable \( X \) has a gamma distribution as given in (A.10), then
Appendix A

Probability distributions

\[ E(X) = \frac{\alpha}{\beta} \]  \hspace{1cm} (A.12)

\[ Var(X) = \frac{\alpha}{\beta^2}. \]  \hspace{1cm} (A.13)

If the mean \( \mu \) and variance \( \sigma^2 \) of a gamma distribution are known, the distribution parameters \( \alpha \) and \( \beta \) can easily be obtained as follows:

\[ \alpha = \frac{\mu^2}{\sigma^2} \]  \hspace{1cm} (A.14)

\[ \beta = \frac{\mu}{\sigma^2}. \]  \hspace{1cm} (A.15)

In this thesis, a gamma distribution with parameters \( \alpha \) and \( \beta \) has been referred to as \( \mathcal{G}(\alpha, \beta) \).

A.5 Conjugate families of distributions

If the prior distribution of \( \theta \) belongs to a conjugate family of distributions [27], then for any sample size \( n \) and any values of the observations in the sample, the posterior distribution of \( \theta \) must also belong to the same family. A family of distributions with this property is said to be closed under sampling.