Appendix C Probabilistic distribution functions

This appendix lists the probabilistic distribution functions (pdfs) used in the main discussions in this book. We basically use the definitions of these functions followed by Bernardo & Smith (2009). Each pdf section also provides values of mean, mode, variance, etc., which are used in the book, although it does not include a complete set of distribution values to avoid complicated descriptions in this appendix. The sections also provide some derivations of these values if these derivations are not complicated.

C.1 Discrete uniform distribution

When we have a discrete variable $a \in \{a_1, a_2, \cdots, a_n, \cdots, a_N\}$, the discrete uniform distribution is defined as:

- pdf:

$$\mathrm{Unif}(a) \triangleq \frac{1}{|\{a_n\}|} = \frac{1}{N}, \tag{C.1}$$

where N is the number of distinct elements.

C.2 Multinomial distribution

- pdf:

$$\mathrm{Mult}(\{x_j\}_{j=1}^J | \{\omega_j\}_{j=1}^J) \triangleq \frac{(\sum_{j=1}^J x_j)!}{\prod_{j=1}^J x_j!} \prod_{j=1}^J \omega_j^{x_j}, \tag{C.2}$$

where x_j is a nonnegative integer, and has the following constraint:

$$x_j \in \mathbb{Z}^+, \quad \sum_j^J x_j = N. \tag{C.3}$$

The parameter $\{\omega_1, \cdots, \omega_J\}$ has the following constraint:

$$\sum_j^J \omega_j = 1, \quad 0 \leq \omega_j \leq 1 \quad \forall j. \tag{C.4}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

- pdf ($x_j = 1$ and $x_{j'} = 0 \ \forall j' \neq j$):

$$\mathrm{Mult}(x_j|\{\omega_j\}_{j=1}^J) \triangleq \omega_j. \tag{C.5}$$

C.3 Beta distribution

Beta distribution is a special case of the Dirichlet distribution with two probabilistic variables x and $y = 1 - x$.

- pdf:

$$\mathrm{Beta}(x|\alpha, \beta) \triangleq C_{\mathrm{Beta}}(\alpha, \beta) x^{\alpha-1}(1-x)^{\beta-1}, \tag{C.6}$$

where

$$x \in [0, 1], \tag{C.7}$$

and

$$\alpha, \beta > 0. \tag{C.8}$$

- Normalization constant:

$$C_{\mathrm{Beta}}(\alpha, \beta) \triangleq \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{C.9}$$

where $\Gamma(\cdot)$ is a gamma function.
- Mean:

$$\mathbb{E}_{(x)}[x] = \frac{\alpha}{\alpha+\beta}. \tag{C.10}$$

This is derived as follows:

$$\begin{aligned}
\mathbb{E}_{(x)}[x] &= \int x\mathrm{Beta}(x|\alpha, \beta)dx \\
&= C_{\mathrm{Beta}}(\alpha, \beta) \int x^{\alpha}(1-x)^{\beta-1}dx \\
&= \frac{C_{\mathrm{Beta}}(\alpha, \beta)}{C_{\mathrm{Beta}}(\alpha+1, \beta)} \\
&= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}{\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\
&= \frac{\alpha}{\alpha+\beta},
\end{aligned} \tag{C.11}$$

where we use the following property of the gamma function:

$$\Gamma(x+1) = x\Gamma(x). \tag{C.12}$$

- Mode:

$$\frac{\alpha-1}{\alpha+\beta-2}, \quad \alpha, \beta > 1. \tag{C.13}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

It is derived as follows:

$$
\begin{aligned}
&\frac{d}{dx}\text{Beta}(x|\alpha,\beta) \\
&= C_{\text{Beta}}(\alpha,\beta)\left((\alpha-1)x^{\alpha-2}(1-x)^{\beta-1}-(\beta-1)x^{\alpha-1}(1-x)^{\beta-2}\right) \\
&= C_{\text{Beta}}(\alpha,\beta)x^{\alpha-2}(1-x)^{\beta-2}\left((\alpha-1)(1-x)-(\beta-1)x\right) \\
&= C_{\text{Beta}}(\alpha,\beta)x^{\alpha-2}(1-x)^{\beta-2}\left(\alpha-1-(\alpha+\beta-2)x\right). \qquad \text{(C.14)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\frac{d}{dx}\text{Beta}(x|\alpha,\beta) = 0 \\
&\Rightarrow x^{\text{mode}} = \frac{\alpha-1}{\alpha+\beta-2}.
\end{aligned} \qquad \text{(C.15)}
$$

C.4 Dirichlet distribution

A Dirichlet distribution is a generalized beta distribution with J probabilistic variables.

- pdf:

$$
\text{Dir}(\{\omega_j\}_{j=1}^J|\{\phi_j\}_{j=1}^J) \triangleq C_{\text{Dir}}(\{\phi_j\}_{j=1}^J)\prod_j(\omega_j)^{\phi_j-1}, \qquad \text{(C.16)}
$$

  where

$$
\sum_j^J \omega_j = 1, \quad 0 \leq \omega_j \leq 1, \quad 0 < \phi_j. \qquad \text{(C.17)}
$$

- Normalization constant:

$$
C_{\text{Dir}}(\{\phi_j\}_{j=1}^J) \triangleq \frac{\Gamma\left(\sum_{j=1}^J \phi_j\right)}{\prod_{j=1}^J \Gamma(\phi_j)}. \qquad \text{(C.18)}
$$

- Mean:

$$
\mathbb{E}_{(\omega_j)}[\omega_j] = \frac{\phi_j}{\sum_{j'=1}^J \phi_{j'}}, \qquad \text{(C.19)}
$$

$$
\begin{aligned}
\mathbb{E}_{(\omega_j)}\left[\prod_j \omega_j^{\gamma_j}\right] &= \frac{C_{\text{Dir}}(\{\phi_j\}_{j=1}^J)}{C_{\text{Dir}}(\{\phi_j+\gamma_j\}_{j=1}^J)} \\
&= \frac{\Gamma\left(\sum_{j=1}^J \phi_j\right)}{\Gamma\left(\sum_{j=1}^J \phi_j+\gamma_j\right)}\frac{\prod_{j=1}^J \Gamma(\phi_j+\gamma_j)}{\prod_{j=1}^J \Gamma(\phi_j)},
\end{aligned} \qquad \text{(C.20)}
$$

$$
\mathbb{E}_{(\omega_j)}[\log\omega_j] = \exp\left(\Psi(\omega_j) - \Psi\left(\sum_{j'=1}^J \omega_{j'}\right)\right). \qquad \text{(C.21)}
$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

- Mode:

$$\frac{d}{d\omega_{j'}}\left(\log\left(\mathrm{Dir}(\{\omega_j\}_{j=1}^J|\{\phi_j\}_{j=1}^J)\right)+\lambda\left(1-\sum_{j=1}^J\omega_j\right)\right)$$
$$=\frac{\phi_j-1}{\omega_j}+\lambda=0,$$
$$\Rightarrow\omega_j^{\mathrm{mode}}\propto\phi_j-1$$
$$\Rightarrow\frac{\phi_j-1}{\sum_{j'=1}^J(\phi_{j'}-1)},\quad\phi_j>1. \tag{C.22}$$

C.5 Gaussian distribution

Note that this book uses Σ as a variance. The standard deviation σ is represented as $\sigma=\sqrt{\Sigma}$. We also list the pdf with precision scale parameter $r\triangleq 1/\Sigma$, which can also be used in this book.

- pdf (with variance Σ):

$$\mathcal{N}(x|\mu,\Sigma)\triangleq C_{\mathcal{N}}(\Sigma)\exp\left(-\frac{(x-\mu)^2}{2\Sigma}\right), \tag{C.23}$$

where

$$x\in\mathbb{R} \tag{C.24}$$

and

$$\mu\in\mathbb{R},\Sigma>0. \tag{C.25}$$

- Normalization constant (with variance Σ):

$$C_{\mathcal{N}}(\Sigma)\triangleq(2\pi)^{-\frac{1}{2}}(\Sigma)^{-\frac{1}{2}}. \tag{C.26}$$

- pdf (with precision r):

$$\mathcal{N}(x|\mu,r^{-1})\triangleq C_{\mathcal{N}}(r^{-1})\exp\left(-\frac{r(x-\mu)^2}{2}\right), \tag{C.27}$$

where

$$r>0. \tag{C.28}$$

- Normalization constant (with precision r):

$$C_{\mathcal{N}}(r^{-1})\triangleq(2\pi)^{-\frac{1}{2}}(r)^{\frac{1}{2}}. \tag{C.29}$$

- Mean:

$$\mathbb{E}_{(x)}[x]=\mu. \tag{C.30}$$

- Variance:

$$\mathbb{E}_{(x)}[x^2] = \Sigma = r^{-1}. \tag{C.31}$$

- Mode:

$$\mu. \tag{C.32}$$

C.6 Multivariate Gaussian distribution

- pdf (with covariance matrix $\mathbf{\Sigma}$):

$$\begin{aligned}\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) &\triangleq C_{\mathcal{N}}(\mathbf{\Sigma}) \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \\ &= C_{\mathcal{N}}(\mathbf{\Sigma}) \exp\left(-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}}\right]\right),\end{aligned} \tag{C.33}$$

where

$$\mathbf{x} \in \mathbb{R}^D, \tag{C.34}$$

and

$$\boldsymbol{\mu} \in \mathbb{R}^D, \quad \mathbf{\Sigma} \in \mathbb{R}^{D \times D}. \tag{C.35}$$

$\mathbf{\Sigma}$ is positive definite.

- Normalization constant (with covariance matrix $\mathbf{\Sigma}$):

$$C_{\mathcal{N}}(\mathbf{\Sigma}) \triangleq (2\pi)^{-\frac{D}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}}. \tag{C.36}$$

- pdf (with precision matrix $\mathbf{R}$):

$$\begin{aligned}\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{R}^{-1}) &\triangleq C_{\mathcal{N}}(\mathbf{R}^{-1}) \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{R} (\mathbf{x} - \boldsymbol{\mu})\right) \\ &= C_{\mathcal{N}}(\mathbf{R}^{-1}) \exp\left(-\frac{1}{2}\mathrm{tr}\left[\mathbf{R}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}}\right]\right).\end{aligned} \tag{C.37}$$

- Normalization constant (with precision matrix$\mathbf{R}$):

$$C_{\mathcal{N}}(\mathbf{R}^{-1}) \triangleq (2\pi)^{-\frac{D}{2}} |\mathbf{R}|^{\frac{1}{2}}. \tag{C.38}$$

- Mean:

$$\mathbb{E}_{(\mathbf{x})}[\mathbf{x}] = \boldsymbol{\mu}. \tag{C.39}$$

- Variance:

$$\mathbb{E}_{(\mathbf{x})}[\mathbf{x}\mathbf{x}^{\mathsf{T}}] = \mathbf{\Sigma} = \mathbf{R}^{-1}. \tag{C.40}$$

- Mode:

$$\boldsymbol{\mu}. \tag{C.41}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

C.7 Multivariate Gaussian distribution (diagonal covariance matrix)

This is a special case of the multivariate Gaussian distribution. The diagonal covariance matrix reduces the number of parameters and makes the analytical treatment simple (it is simply represented as a product of scalar Gaussian distributions).

- pdf (with covariance matrix $\mathbf{\Sigma} = \text{diag}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = [\Sigma_1, \cdots, \Sigma_D]^\mathsf{T}$):

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) \triangleq \prod_{d=1}^{D} \mathcal{N}(x_d|\mu_d, \Sigma_d)$$
$$= \prod_{d=1}^{D} C_{\mathcal{N}}(\Sigma_d) \exp\left(-\frac{1}{2}\frac{(x_d - \mu_d)^2}{\Sigma_d}\right), \tag{C.42}$$

where

$$\mathbf{x} \in \mathbb{R}^D, \tag{C.43}$$

and

$$\boldsymbol{\mu} \in \mathbb{R}^D, \quad \boldsymbol{\sigma} \in \mathbb{R}^D_{>0}. \tag{C.44}$$

- Normalization constant (with variance Σ_d):

$$C_{\mathcal{N}}(\Sigma_d) \triangleq (2\pi)^{-\frac{1}{2}} \Sigma_d^{-\frac{1}{2}}. \tag{C.45}$$

- pdf (with precision matrix $\mathbf{R} = \text{diag}(\mathbf{r})$, where $\mathbf{r} = [r_1, \cdots, r_D]^\mathsf{T}$):

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{R}^{-1}) \triangleq \prod_{d=1}^{D} \mathcal{N}(x_d|\mu_d, r_d^{-1})$$
$$= \prod_{d=1}^{D} C_{\mathcal{N}}(r_d^{-1}) \exp\left(-\frac{r_d}{2}(x_d - \mu_d)^2\right). \tag{C.46}$$

- Normalization constant (with precision r_d):

$$C_{\mathcal{N}}(r_d^{-1}) \triangleq (2\pi)^{-\frac{1}{2}} r_d^{\frac{1}{2}}. \tag{C.47}$$

- Mean:

$$\mathbb{E}_{(\mathbf{x})}[\mathbf{x}] = \boldsymbol{\mu}. \tag{C.48}$$

- Variance:

$$\mathbb{E}_{(x_d)}[x_d^2] = \Sigma_d = r_d^{-1}. \tag{C.49}$$

- Mode:

$$\boldsymbol{\mu}. \tag{C.50}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

C.8 Spherical Gaussian distribution

This is another special case of the multivariate Gaussian distribution where the diagonal covariance matrix is represented with the identity matrix with a single scaling parameter.

- pdf (with covariance matrix $\boldsymbol{\Sigma} = \Sigma \mathbf{I}_D$):

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq \prod_{d=1}^{D} \mathcal{N}(x_d|\mu_d, \Sigma)$$
$$= (C_{\mathcal{N}}(\Sigma))^D \prod_{d=1}^{D} \exp\left(-\frac{1}{2}\frac{(x_d - \mu_d)^2}{\Sigma}\right), \quad \text{(C.51)}$$

where

$$\mathbf{x} \in \mathbb{R}^D, \quad \text{(C.52)}$$

and

$$\boldsymbol{\mu} \in \mathbb{R}^D, \quad \Sigma \in \mathbb{R}_{>0}. \quad \text{(C.53)}$$

- Normalization constant (with variance Σ):

$$C_{\mathcal{N}}(\Sigma) \triangleq (2\pi)^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}}. \quad \text{(C.54)}$$

- pdf (with precision matrix $\mathbf{R} = r^{-1}\mathbf{I}_D$):

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{R}^{-1}) \triangleq \prod_{d=1}^{D} \mathcal{N}(x_d|\mu_d, r^{-1})$$
$$= \left(C_{\mathcal{N}}(r^{-1})\right)^D \prod_{d=1}^{D} \exp\left(-\frac{r}{2}(x_d - \mu_d)^2\right). \quad \text{(C.55)}$$

- Normalization constant (with precision r):

$$C_{\mathcal{N}}(r^{-1}) \triangleq (2\pi)^{-\frac{1}{2}} r^{\frac{1}{2}}. \quad \text{(C.56)}$$

- Mean:

$$\mathbb{E}_{(\mathbf{x})}[\mathbf{x}] = \boldsymbol{\mu}. \quad \text{(C.57)}$$

- Variance:

$$\mathbb{E}_{(x_d)}[x_d^2] = \Sigma = r^{-1}. \quad \text{(C.58)}$$

- Mode:

$$\boldsymbol{\mu}. \quad \text{(C.59)}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

C.9 Matrix variate Gaussian distribution

- pdf:

$$\begin{aligned} p(\mathbf{X}) &= \mathcal{N}(\mathbf{X}|\mathbf{M}, \boldsymbol{\Phi}, \boldsymbol{\Omega}) \\ &\triangleq C_{\mathcal{N}}(\boldsymbol{\Phi}, \boldsymbol{\Omega}) \exp\left(-\frac{1}{2}\mathrm{tr}\left[(\mathbf{X}-\mathbf{M})^{\intercal}\boldsymbol{\Phi}^{-1}(\mathbf{X}-\mathbf{M})\boldsymbol{\Omega}^{-1}\right]\right), \end{aligned} \tag{C.60}$$

  where

$$\mathbf{X} \in \mathbb{R}^{D \times D'}, \tag{C.61}$$

  and

$$\mathbf{M} \in \mathbb{R}^{D \times D'}, \quad \boldsymbol{\Phi} \in \mathbb{R}^{D \times D}, \quad \boldsymbol{\Omega} \in \mathbb{R}^{D' \times D'}. \tag{C.62}$$

  $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ are correlation matrices and are positive definite. When $D' = 1$, it becomes the multivariate Gaussian distribution.
- Normalization constant:

$$C_{\mathcal{N}}(\boldsymbol{\Phi}, \boldsymbol{\Omega}) \triangleq (2\pi)^{-\frac{DD'}{2}} |\boldsymbol{\Omega}|^{-\frac{D}{2}} |\boldsymbol{\Phi}|^{-\frac{D'}{2}}. \tag{C.63}$$

- Mean:

$$\mathbb{E}_{(\mathbf{X})}[\mathbf{X}] = \mathbf{M}. \tag{C.64}$$

- Variance:

$$\mathbb{E}_{(\mathbf{X})}[\mathbf{X}\mathbf{X}^{\intercal}] = \boldsymbol{\Phi}, \tag{C.65}$$

$$\mathbb{E}_{(\mathbf{X})}[\mathbf{X}^{\intercal}\mathbf{X}] = \boldsymbol{\Omega}. \tag{C.66}$$

- Mode:

$$\mathbf{M}. \tag{C.67}$$

C.10 Laplace distribution

- pdf:

$$\begin{aligned} \mathrm{Lap}(x|\mu, \beta) &\triangleq C_{\mathrm{Lap}}(\beta) \exp\left(-\frac{|x-\mu|}{\beta}\right) \\ &= C_{\mathrm{Lap}}(\beta) \begin{cases} \exp\left(-\frac{x-\mu}{\beta}\right) & \text{if } x \geq u, \\ \exp\left(-\frac{\mu-x}{\beta}\right) & \text{if } x < u, \end{cases} \end{aligned} \tag{C.68}$$

  where

$$x \in \mathbb{R}, \tag{C.69}$$

  and

$$\mu \in \mathbb{R}, \quad \beta > 0. \tag{C.70}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

- Normalization constant:

$$C_{\mathrm{Lap}}(\beta) \triangleq \frac{1}{2\beta}. \tag{C.71}$$

- Mean:

$$\mathbb{E}_{(x)}[x] = \mu. \tag{C.72}$$

- Mode:

$$\mu. \tag{C.73}$$

Note that the probabilistic distribution function is not continuous at μ, and it is not differentiable.

C.11 Gamma distribution

A gamma distribution is used as a prior/posterior distribution of precision parameter r of a Gaussian distribution.

- pdf:

$$\mathrm{Gam}(y|\alpha, \beta) \triangleq C_{\mathrm{Gam}}(\alpha, \beta) y^{\alpha-1} \exp\left(-\beta y\right), \tag{C.74}$$

  where

$$y > 0, \tag{C.75}$$

  and

$$\alpha, \beta > 0. \tag{C.76}$$

- Normalization constant:

$$C_{\mathrm{Gam}}(\alpha, \beta) \triangleq \frac{\beta^{\alpha}}{\Gamma(\alpha)}. \tag{C.77}$$

- Mean:

$$\mathbb{E}_{(y)}[y] = \frac{\alpha}{\beta}. \tag{C.78}$$

- Variance:

$$\mathbb{E}_{(y)}[y^2] = \frac{\alpha}{\beta^2}. \tag{C.79}$$

- Mode:

$$\begin{aligned} \frac{d}{dy}\mathrm{Gam}(y|\alpha, \beta) = C_{\mathrm{Gam}}(\alpha, \beta)\,(\alpha - 1 - \beta y)\, y^{\alpha-2} \exp(-\beta y) = 0 \\ \Rightarrow \frac{\alpha - 1}{\beta}, \quad \alpha > 1. \end{aligned} \tag{C.80}$$

The shape of the gamma distribution is not symmetric, and the mode and mean values are different.

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

To make the notation consistent with the Wishart distribution in Appendix C.14, we also use the following definition for the gamma distribution, with $\alpha \to \frac{\phi}{2}$ and $\beta \to \frac{r^0}{2}$ in the original gamma distribution defined in Eq. (C.74):

- pdf (with $\frac{1}{2}$ factor):

$$\begin{aligned}\mathrm{Gam}_2(y|\phi, r^0) &\triangleq \mathrm{Gam}\left(y \middle| \frac{\phi}{2}, \frac{r^0}{2}\right) \\ &= C_{\mathrm{Gam}_2}\left(\phi, r^0\right) y^{\frac{\phi}{2}-1} \exp\left(-\frac{r^0 y}{2}\right).\end{aligned} \tag{C.81}$$

- Normalization constant (with $\frac{1}{2}$ factor):

$$C_{\mathrm{Gam}_2}(\phi, r^0) \triangleq \frac{\left(\frac{r^0}{2}\right)^{\frac{\phi}{2}}}{\Gamma\left(\frac{\phi}{2}\right)}. \tag{C.82}$$

- Mean (with $\frac{1}{2}$ factor):

$$\mathbb{E}_{(y)}[y] = \frac{\frac{\phi}{2}}{\frac{r^0}{2}} = \frac{\phi}{r^0}. \tag{C.83}$$

- Variance (with $\frac{1}{2}$ factor):

$$\mathbb{E}_{(y)}[y^2] = \frac{\frac{\phi}{2}}{\left(\frac{r^0}{2}\right)^2} = \frac{2\phi}{\left(r^0\right)^2}. \tag{C.84}$$

- Mode (with $\frac{1}{2}$ factor):

$$\frac{\frac{\phi}{2} - 1}{\frac{r^0}{2}} = \frac{\phi - 2}{r^0}, \quad \phi > 2. \tag{C.85}$$

It is shown in Appendix C.14 that this gamma distribution ($\mathrm{Gam}_2(\cdot)$) with $\frac{1}{2}$ factor is equivalent to the Wishart distribution with the number of dimensions 1 ($D = 1$).

C.12 Inverse gamma distribution

An inverse gamma distribution is used as a prior/posterior distribution of variance parameter Σ. It can be obtained by simply replacing y in a gamma distribution by $\frac{1}{y}$ with $f(y) = g(y') \left|\frac{dy'}{dy}\right|$.

- pdf:

$$\begin{aligned}\mathrm{IGam}(y|\alpha, \beta) &= \mathrm{Gam}\left(\frac{1}{y} \middle| \alpha, \beta\right) \left|\frac{d(y^{-1})}{dy}\right| \\ &= C_{\mathrm{Gam}}(\alpha, \beta) \left(\frac{1}{y}\right)^{\alpha-1} \exp\left(-\frac{\beta}{y}\right) y^{-2} \\ &\triangleq C_{\mathrm{IGam}}(\alpha, \beta) y^{-\alpha-1} \exp\left(-\frac{\beta}{y}\right),\end{aligned} \tag{C.86}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

where

$$y > 0, \tag{C.87}$$

and

$$\alpha, \beta > 0. \tag{C.88}$$

- Normalization constant:

$$C_{\mathrm{IGam}}(\alpha, \beta) = C_{\mathrm{Gam}}(\alpha, \beta) \triangleq \frac{\beta^\alpha}{\Gamma(\alpha)}. \tag{C.89}$$

- Mean:

$$\mathbb{E}_{(y)}[y] = \frac{\beta}{\alpha - 1}, \quad \alpha > 1. \tag{C.90}$$

- Mode:

$$\frac{\beta}{\alpha + 1}. \tag{C.91}$$

C.13 Gaussian–gamma distribution

We provide a Gaussian–gamma distribution, which is used as a conjugate prior distribution of a Gaussian distribution with mean μ and precision r. This is also known as the normal-gamma distribution. Note that we use the gamma distribution defined in Eq. (C.81) instead of the original definition of the gamma distribution in Eq. (C.74).

- pdf:

$$\begin{aligned} &\mathcal{N}\mathrm{Gam}(\mu, r|\mu^0, \phi^\mu, r^0, \phi^r) \\ &\quad \triangleq \mathcal{N}(\mu|\mu^0, (r\phi^\mu)^{-1})\mathrm{Gam}_2\left(r \middle| \phi^r, r^0\right) \\ &\quad \triangleq C_{\mathcal{N}\mathrm{Gam}}(\phi^\mu, r^0, \phi^r) r^{\frac{\phi^r - 1}{2}} \exp\left(-\frac{r^0 r}{2} - \frac{\phi^\mu r(\mu - \mu^0)^2}{2}\right). \end{aligned} \tag{C.92}$$

Note that the power of r is $\frac{\phi^r - 1}{2}$, which is different from the original gamma distribution definition in Section (C.11), because it also considers an $r^{\frac{1}{2}}$ factor in the Gaussian distribution.

- Normalization constant:

$$C_{\mathcal{N}\mathrm{Gam}}(r, \alpha, \beta) \triangleq \frac{\sqrt{\phi^\mu}\left(\frac{r^0}{2}\right)^{\frac{\phi^r}{2}}}{\sqrt{2\pi}\,\Gamma\left(\frac{\phi^r}{2}\right)}. \tag{C.93}$$

- Mean:

$$\mathbb{E}_{(\mu, r)}[\{\mu, r\}] = \left\{\mu^0, \frac{\phi^r}{r^0}\right\}. \tag{C.94}$$

https://doi.org/10.1017/CBO9781107295360.012 Published online by Cambridge University Press

- Mode:

$$\left\{\mu^0, \frac{\phi^r - 2}{r^0}\right\}, \quad \phi^r > 2. \tag{C.95}$$

The means and modes of mean and precision parameters are the same as those of the Gaussian and gamma distributions, respectively.

C.14 Wishart distribution

- pdf:

$$\mathcal{W}(\mathbf{Y}|\mathbf{R}^0, \phi) \triangleq C_{\mathcal{W}}(\mathbf{R}^0, \phi)|\mathbf{Y}|^{\frac{\phi-D-1}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[\mathbf{R}^0\mathbf{Y}\right]\right), \tag{C.96}$$

  where

$$\mathbf{Y} \in \mathbb{R}^{D\times D}, \tag{C.97}$$

  and

$$\mathbf{R}^0 \in \mathbb{R}^{D\times D}, \quad \phi > D - 1. \tag{C.98}$$

  $\mathbf{Y}$ and $\mathbf{R}^0$ are positive definite.
- Normalization constant:

$$C_{\mathcal{W}}(\mathbf{R}^0, \phi) \triangleq \frac{|\mathbf{R}^0|^{\frac{\phi}{2}}}{(2)^{\frac{D\phi}{2}}\Gamma_D\left(\frac{\phi}{2}\right)}, \tag{C.99}$$

  where $\Gamma_D(\cdot)$ is a multivariate Gamma function.
- Mean:

$$\mathbb{E}_{(\mathbf{Y})}[\mathbf{Y}] = \phi(\mathbf{R}^0)^{-1}. \tag{C.100}$$

- Mode:

$$(\phi - D - 1)\,(\mathbf{R}^0)^{-1}, \quad \phi \geq D + 1. \tag{C.101}$$

When $D \to 1$, the Wishart distribution is equivalent to the gamma distribution defined in Eq. (C.81).

C.15 Gaussian–Wishart distribution

- pdf:

$$\begin{aligned}
&\mathcal{NW}(\boldsymbol{\mu}, \mathbf{R}|\boldsymbol{\mu}^0, \phi^\mu, \mathbf{R}^0, \phi^{\mathbf{R}}) \\
&\quad \triangleq \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}^0, (\phi^\mu\mathbf{R})^{-1})\mathcal{W}(\mathbf{R}|\mathbf{R}^0, \phi^{\mathbf{R}}) \\
&\quad \triangleq C_{\mathcal{NW}}(\phi^\mu, \mathbf{R}^0, \phi^{\mathbf{R}})|\mathbf{R}|^{\frac{\phi^{\mathbf{R}}-D}{2}} \\
&\qquad \times \exp\left(-\frac{1}{2}\text{tr}\left[\mathbf{R}^0\mathbf{R}\right] - \frac{\phi^\mu}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}^0)^\intercal\mathbf{R}(\boldsymbol{\mu} - \boldsymbol{\mu}^0)\right).
\end{aligned} \tag{C.102}$$

Note that when $D \to 1$, Gaussian–Wishart distribution $\mathcal{NW}(\boldsymbol{\mu}, \mathbf{R}|\boldsymbol{\mu}^0, \phi^\mu, \mathbf{R}^0, \phi^{\mathbf{R}})$ approaches Gaussian–gamma distribution $\mathcal{N}\mathrm{Gam}(\mu, r|\mu^0, \phi^\mu, r^0, \phi^r)$ in Appendix C.13.

- Normalization constant:

$$C_{\mathcal{NW}}(\phi^\mu, \mathbf{R}^0, \phi^R) \triangleq (2\pi)^{\frac{D}{2}} (\phi^\mu)^{-\frac{D}{2}} \frac{|\mathbf{R}^0|^{\frac{\phi^R}{2}}}{(2)^{\frac{D\phi^R}{2}} \Gamma_D\left(\frac{\phi^R}{2}\right)}. \tag{C.103}$$

- Mean:

$$\mathbb{E}_{(\boldsymbol{\mu},\mathbf{R})}[\{\boldsymbol{\mu}, \mathbf{R}\}] = \left\{\boldsymbol{\mu}^0, \phi^R(\mathbf{R}^0)^{-1}\right\}. \tag{C.104}$$

- Mode:

$$\left\{\boldsymbol{\mu}^0, \left(\phi^R - D - 1\right)(\mathbf{R}^0)^{-1}\right\}, \quad \phi^R \geq D + 1. \tag{C.105}$$

Again, when $D \to 1$, these modes are equivalent to the modes of the Gaussian–gamma distribution in Appendix C.13.

C.16 Student's *t*-distribution

- pdf:

$$\mathrm{St}(x|\mu, \lambda, \kappa) \triangleq C_{\mathrm{St}} \left(1 + \frac{1}{\kappa\lambda}(x - \mu)^2\right)^{-\frac{\kappa+1}{2}}, \tag{C.106}$$

where

$$x \in \mathbb{R}, \tag{C.107}$$

and

$$\mu \in \mathbb{R}, \quad \kappa > 0, \quad \lambda > 0. \tag{C.108}$$

- Normalization constant:

$$C_{\mathrm{St}}(\kappa, \lambda) \triangleq \frac{\Gamma\left(\frac{\kappa+1}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)\Gamma\left(\frac{1}{2}\right)} \left(\frac{1}{\kappa\lambda}\right)^{\frac{1}{2}}, \tag{C.109}$$

- Mean:

$$\mathbb{E}_{(x)}[x] = \mu, \tag{C.110}$$

- Mode:

$$\mu. \tag{C.111}$$

Parameter κ is called the degrees of freedom, and if κ is large, the distribution approaches the Gaussian distribution. Note that Student's t-distribution is not included in the exponential family (Section 2.1.3).