

Handout 2: Revision in mixture of probability distributions

Lecturer & author: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

Aim

To practice on probability calculations on compound distribution functions. To become familiar with distributions, Inverted Gamma, multivariate Normal, and multivariate Student T distributions.

References:

- DeGroot, M. H. (1970, or 2005). Optimal statistical decisions (Vol. 82). John Wiley & Sons.
 - Part one: Survey of probability theory. Chapters 1-5 ; However the treatment of the Normal and Student T distributions is different than ours.
- Raiffa, H., & Schlaifer, R. (1961). Applied statistical decision theory.
 - Chapters 8.2, 8.3 ; However the treatment of the Normal and Student T distributions is different than ours.

Web-applets

- https://georgios-stats-3.shinyapps.io/demo_multivariatenormaldistribution/
- https://github.com/georgios-stats/Shiny_applets/tree/master/demo_MultivariateNormalDistribution

1 Mixture probability distribution

Definition 1. Consider a random variable x distributed according to $F(x|z)$ with a pdf/pmf $f(x|z)$ labeled by an unknown parameter z . Consider that z is distributed according to distribution $\Pi(z)$ with pdf/pmf $\pi(z)$. Then:

1. This dependency can be represented as

$$\begin{aligned} x|z &\sim F(x|z) \\ z &\sim \Pi(z) \end{aligned}$$

where $F(x|z)$ is called conditional (or parametrized) distribution, $\Pi(z)$ is called mixing (or latent) distribution; and z is called mixing (or latent) variable.

2. Distribution $G(x)$ that results by integrating the conditional distribution $F(x|z)$ with respect to $\Pi(z)$ as

$$G(x) = \int F(x|z) d\Pi(z)$$

is called the mixture (or compound) distribution of x . The mixture (or compound) distribution $G(x)$ has PDF/PMF

$$g(x) = \int f(x|z) d\Pi(z) = \begin{cases} \int f(x|z) \pi(z) dz & , z \text{ cont.} \\ \sum_{\forall z} f(x|z) \pi(z) & , z \text{ discr.} \end{cases}$$

Definition 2. $G(x)$ is also called continuous mixture distribution when the mixing/latent variable z is continuous.

$G(x)$ is also called finite mixture when z is discrete.

Remark 3. Recall from Handout 1 that

$$\begin{aligned} E_G(x) &= E_{\Pi}(E_F(x|z)) \\ \text{Var}_G(x) &= E_{\Pi}(\text{Var}(x|z)) + \text{Var}_{\Pi}(E_F(x|z)) \end{aligned}$$

2 Inverted Gamma distribution $x|a, b \sim \text{IG}(a, b)$

Definition 4. The random variable $x \in (0, +\infty)$ follows an Inverted Gamma distribution $x \sim \text{IG}(a, b)$, if and only if $x = \frac{1}{y}$ follows a Gamma distribution, $y \sim \text{Ga}(a, b)$, with $a > 0$ and $b > 0$.

Example 5. Let $x \sim \text{IG}(a, b)$, then the PDF of x is

$$f_{\text{IG}(a,b)}(x) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\left(-\frac{b}{x}\right) 1_{(0,+\infty)}(x) \quad (1)$$

Solution. It is

$$f_{\text{IG}(a,b)}(x) = f_{\text{Ga}(a,b)}\left(\frac{1}{x}\right) \left| \frac{d}{dx} \left(\frac{1}{x}\right) \right| = \frac{b^a}{\Gamma(a)} \left(\frac{1}{x}\right)^{a-1} \exp\left(-\frac{b}{x}\right) 1_{(0,+\infty)}\left(\frac{1}{x}\right) \left| -\frac{1}{x^2} \right|$$

Example 6. Let a random variable $x \sim \text{IG}(a, b)$, then

$$E_{\text{IG}(a,b)}(x) = \frac{b}{a-1}; \quad a > 1 \quad \text{and} \quad \text{Var}_{\text{IG}(a,b)}(x) = \frac{b^2}{(a-1)^2(a-2)}; \quad a > 2$$

Solution. It is

$$E_{\text{IG}(a,b)}(x) = \int x f_{\text{IG}(a,b)}(x) dx = \int_{(0,+\infty)} x \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\left(-\frac{b}{x}\right) dx$$

Assume that $a > 1$. Then

$$\begin{aligned} E_{\text{IG}(a,b)}(x) &= \int_{(0,+\infty)} \frac{b^{a-1}}{\Gamma(a)} b \frac{\Gamma(a-1)}{\Gamma(a-1)} x^{-a+1-1} \exp\left(-\frac{b}{x}\right) dx = b \frac{\Gamma(a-1)}{\Gamma(a)} \int_{(0,+\infty)} \frac{b^{a-1}}{\Gamma(a-1)} x^{-a+1-1} \exp\left(-\frac{b}{x}\right) dx \\ &= b \frac{\Gamma(a-1)}{\Gamma(a)} \int_{(0,+\infty)} \frac{b^{a-1}}{\Gamma(a-1)} x^{-a+1-1} \exp\left(-\frac{b}{x}\right) dx = b \frac{\Gamma(a-1)}{(a-1)\Gamma(a-1)} \int f_{\text{IG}(a-1,b)}(x) dx = \frac{b}{a-1} \end{aligned}$$

Similarly

$$\begin{aligned} E_{\text{IG}(a,b)}(x^2) &= \int_{(0,+\infty)} x^2 \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\left(-\frac{b}{x}\right) dx = \dots = b \frac{\Gamma(a-1)}{(a-1)\Gamma(a-1)} \int x f_{\text{IG}(a-1,b)}(x) dx \\ &= \frac{b}{a-1} \frac{b}{a-2}; \quad a > 2 \end{aligned}$$

So

$$\text{Var}_{\text{IG}(a,b)}(x) = E_{\text{IG}(a,b)}(x^2) - (E_{\text{IG}(a,b)}(x))^2 = \frac{b^2}{(a-1)^2(a-2)}$$

3 Multivariate Normal distribution $x|\mu, \Sigma \sim \mathbf{N}_d(\mu, \Sigma)$

Definition 7. A d -dimensional random variable $x \in \mathbb{R}^d$ is said to have a multivariate Normal (Gaussian) distribution, if for every d -dimensional fixed vector $\alpha \in \mathbb{R}^d$, the random variable $\alpha^\top x$ has a univariate Normal (Gaussian) distribution.

Proposition 8. A random vector $x \in \mathbb{R}^d$ has a d -dimensional Normal distribution with mean $\mu = E(x)$ and covariance matrix $\Sigma = \text{Var}(x)$ if and only if random vector $x \in \mathbb{R}^d$ has a characteristic function

$$\varphi_x(t) = \exp\left(it^\top \mu - \frac{1}{2}t^\top \Sigma t\right) \quad (2)$$

Hence: the d -dimensional Normal distribution is uniquely defined by the mean and the covariance matrix.

Proof. (\implies) If x has a d -dimensional distribution then the characteristic function is $\varphi_x(t) = \varphi_{t^\top x}(1)$. Since x has a d -dimensional Normal distribution with mean $\mu = E(x)$ and covariance matrix $\Sigma = \text{Var}(x)$, $t^\top x$ has a Normal distribution with mean $E(t^\top x) = t^\top \mu$ and variance $\text{Var}(t^\top x) = t^\top \Sigma t$. Then (Handout 1, Example 31)

$$\varphi_x(t) = \varphi_{t^\top x}(1) = \exp\left(iE(t^\top x) - \frac{1}{2}\text{Var}(t^\top x)\right) = \exp\left(it^\top E(x) - \frac{1}{2}t^\top \text{Var}(x)t\right) = \exp\left(it^\top \mu - \frac{1}{2}t^\top \Sigma t\right)$$

(\impliedby) If random vector $x \in \mathbb{R}^d$ has a characteristic function $\varphi_x(t) = \exp(it^\top \mu - \frac{1}{2}t^\top \Sigma t)$, then for every d -dimensional fixed vector $\alpha \in \mathbb{R}^d$ the characteristic function of $\alpha^\top x$ is

$$\varphi_{\alpha^\top x}(t) = \varphi_x(t\alpha) = \exp\left(it\alpha^\top \mu - \frac{1}{2}t\alpha^\top \Sigma \alpha t\right) = \exp\left(it(\alpha^\top \mu) - \frac{1}{2}(\alpha^\top \Sigma \alpha)t^2\right)$$

which defines that $\alpha^\top x$ has a univariate Normal distribution with mean $\alpha^\top \mu$ and variance $\alpha^\top \Sigma \alpha$. \square

Notation 9. We denote the d -dimensional Normal distribution with mean μ and covariance matrix $\Sigma \geq 0$ as $N_d(\mu, \Sigma)$.

Notation 10. The d -dimensional standardized Normal distribution is $N_d(0, I)$.

Proposition 11. Let random variable $x \sim N_d(\mu, \Sigma)$, fixed vector $c \in \mathbb{R}^q$ and fixed matrix $A \in \mathbb{R}^q \times \mathbb{R}^d$. The random vector $y = c + Ax$ has distribution $y \sim N_q(c + A\mu, A\Sigma A^\top)$.

Proof. First I show that y is Normally distributed. Let $\alpha \in \mathbb{R}^q$ any fixed vector. Then $\alpha^\top y = \tilde{\alpha}^\top x + \alpha^\top c$ where $\tilde{\alpha} = A^\top \alpha$. Because x is multivariate Normal, then $\tilde{\alpha}^\top x$ is univariate Normal (by Definition 7), then $\alpha^\top y$ is univariate Normal. So y is q -variate Normal. Also, $E(y) = E(c + Ax) = c + AE(x)$, and $\text{Var}(y) = \text{Var}(c + Ax) = A\text{Var}(x)A^\top$. \square

Proposition 12. Let a d -dimensional random vector $x \sim N_{(any)}(\mu, \Sigma)$.

1. Let $y = Ax$ and $z = Bx$, where $A \in \mathbb{R}^{q \times d}$ and $B \in \mathbb{R}^{k \times d}$: The vectors $y = Ax$ and $z = Bx$ are independent if and only if $A\Sigma B^\top = 0$.
2. Let $x = (x_1, \dots, x_d)^\top$: The x_1, \dots, x_d are mutually independent if and only if the corresponding off diagonal parts of the Σ are zero.

Proof. In both cases, the CF (2) factorizes as $\varphi_x(t) = \prod_j \varphi_{x_j}(t_j)$ only when the corresponding of diagonal parts of Σ are zero. \square

Proposition 13. Any sub-vector of a vector with multivariate Normal distribution has a multivariate Normal distribution.

Proof. Let $x \sim N_d(\mu, \Sigma)$. Any sub-vector y of x can be expressed as $y = 0 + Px$, where $P \in \mathbb{R}^{q \times d}$ is a suitable projection matrix. Then $y \sim N_d(P\mu, P\Sigma P^\top)$. \square

Proposition 14. [Marginalization & conditioning] Let $x \sim N_d(\mu, \Sigma)$. Consider partition such that

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix},$$

where $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$ Then:

1. For the marginal, it is $x_1 \sim N_{d_1}(\mu_1, \Sigma_1)$.

2. For $x_{2.1} = x_2 - \Sigma_{21}\Sigma_1^{-1}x_1$, with $\Sigma_1 > 0$, it is $x_{2.1} \sim N_{d_2}(\mu_{2.1}, \Sigma_{2.1})$ where

$$\mu_{2.1} = \mu_2 - \Sigma_{21}\Sigma_1^{-1}\mu_1 \text{ and } \Sigma_{2.1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top \quad (3)$$

3. Random variables x_1 and $x_{2.1}$ are independent.

4. For the conditional, if $\Sigma_1 > 0$, it is

$$x_2|x_1 \sim N_{d_2}(\mu_{2|1}, \Sigma_{2|1})$$

where

$$\mu_{2|1} = \mu_2 - \Sigma_{21}\Sigma_1^{-1}(x_1 - \mu_1) \text{ and } \Sigma_{2|1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top \quad (4)$$

Hint: If that was a Homework it will be given as a hint to use, in (1.): $x_1 = Ax$ with $A = [I, 0]$, and in (2.): $x_{2.1} = Bx$ with $[-\Sigma_{21}\Sigma_1^{-1}, I]$.

Solution.

1. It is $x_1 = Ax$ with $A = [I, 0]$. Then $x_1 \sim N(A\mu, A\Sigma A^\top)$ where

$$A\mu = [I, 0] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu_1; \quad A\Sigma A^\top = [I, 0] \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix} \begin{bmatrix} I \\ 0 \end{bmatrix} = \Sigma_1$$

2. It is $x_{2.1} = Bx$ with $B = [-\Sigma_{21}\Sigma_1^{-1}, I]$. Then $x_{2.1} \sim N(B\mu, B\Sigma B^\top)$ where

$$\begin{aligned} B\mu &= [-\Sigma_{21}\Sigma_1^{-1}, I] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = -\Sigma_{21}\Sigma_1^{-1}\mu_1 + \mu_2; \\ B\Sigma B^\top &= [-\Sigma_{21}\Sigma_1^{-1}, I] \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix} \begin{bmatrix} -\Sigma_1^{-1}\Sigma_{21}^\top \\ I \end{bmatrix} = [0, -\Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top + \Sigma_2] \begin{bmatrix} -\Sigma_{21}\Sigma_1^{-1} \\ I \end{bmatrix} \\ &= -\Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top + \Sigma_2 \end{aligned}$$

3. x_1 and $x_{2.1}$ are independent, because (i.) x_1 and x_2 are Normally distributed and (ii.) for $x_1 = Ax$ with $A = [I, 0]$ and $x_{2.1} = Bx$ with $[-\Sigma_{21}\Sigma_1^{-1}, 0]$ are

$$\begin{aligned} \text{Cov}(x_1, x_{2.1}) &= \text{Cov}(Ax, Bx) = A\Sigma B^\top = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix} \begin{bmatrix} -\Sigma_1^{-1}\Sigma_{21}^\top \\ I \end{bmatrix} = \\ &= \begin{bmatrix} \Sigma_1, & \Sigma_{21}^\top \end{bmatrix} \begin{bmatrix} -\Sigma_1^{-1}\Sigma_{21}^\top \\ I \end{bmatrix} = -\Sigma_{21}^\top + \Sigma_{21}^\top = 0 \end{aligned}$$

4. From part ??, I observe that it is

$$x_{2.1} = x_2 - \Sigma_{21}\Sigma_1^{-1}x_1 \iff x_2 = x_{2.1} + \Sigma_{21}\Sigma_1^{-1}x_1. \quad (5)$$

In (5), if I condition x_2 on a given value for x_1 , the term $\Sigma_{21}\Sigma_1^{-1}x_1$ is a constant, namely I have $x_2|x_1 = x_{2.1}|x_1 + \text{const}$ where variation comes from $x_{2.1}|x_1$. From part ??, $x_{2.1}$ is independent on x_1 , and hence $F(x_{2.1}|x_1) = F(x_{2.1})$. Furthermore, $x_{2.1}|x_1 \in \text{Normal}$ because $x_{2.1} \in \text{Normal}$. Consequently, $x_2|x_1 \in \text{Normal}$ as a linear transformation of a Normal variate. Now, about the moments

$$\begin{aligned} E(x_2|x_1) &= E(x_{2.1} + \Sigma_{21}\Sigma_1^{-1}x_1|x_1) = E(x_{2.1}|x_1) + E(\Sigma_{21}\Sigma_1^{-1}x_1|x_1) = [\mu_2 - \Sigma_{21}\Sigma_1^{-1}\mu_1] + [\Sigma_{21}\Sigma_1^{-1}x_1] \\ \text{Var}(x_2|x_1) &= \text{Var}(x_{2.1} + \Sigma_{21}\Sigma_1^{-1}x_1|x_1) = \text{Var}(x_{2.1}|x_1) = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top \end{aligned}$$

Proposition 15. The density function of the d -dimensional Normal distribution with mean μ and covariance matrix Σ , when Σ is symmetric positive definite matrix ($\Sigma > 0$), exists and it is equal to

$$f(x) = (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \quad (6)$$

Proof. Let $x \sim N(\mu, \Sigma)$. Because $\Sigma > 0$, we use Cholesky decomposition to define L such that $\Sigma = LL^\top$. Let $z = L^{-1}(x - \mu)$. It is $E(z) = 0$, $\text{Var}(z) = I$, $z \sim N_d(0, I)$, and hence z_1, \dots, z_d are mutually independent So

$$f_z(z) = \prod_{i=1}^d (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}z_i^2\right) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}z^\top z\right)$$

Then

$$\begin{aligned} f_x(x) &= f_z(z) \left| \frac{dz}{dx} \right| = f_z\left(L^{-1}(x - \mu)\right) \left| \det\left(\frac{d}{dx} L^{-1}(x - \mu)\right) \right| \\ &= (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}(x - \mu)^\top (L^{-1})^\top L^{-1}(x - \mu)\right) \det(L^{-1}) \\ &= (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \det(\Sigma)^{-\frac{1}{2}} \end{aligned}$$

□

4 Multivariate Student's T distribution $x \sim T_d(\mu, \Sigma, v)$

Definition 16. A d -dimensional random variable $x \in \mathbb{R}^d$ is said to have a multivariate Student's T distribution with location parameter μ , scale matrix Σ , and degrees of freedom v , and it is denoted as $x \sim T_d(\mu, \Sigma, v)$, if and only if

$$x = \mu + y\sqrt{v\xi}$$

where $y \sim N_d(0, \Sigma)$ and $\xi \sim \text{IG}(\frac{v}{2}, \frac{1}{2})$ are independent random variables.

Example 17. Definition 16 can be equivalently represented in terms of a mixture probability distribution (Section 1) as the compound (or mixture) distribution

$$x \sim T_d(\mu, \Sigma, v)$$

of

$$\begin{aligned} x|\xi &\sim N_d(\mu, \Sigma v\xi) \\ \xi &\sim \text{IG}\left(\frac{v}{2}, \frac{1}{2}\right) \end{aligned}$$

Solution. Straightforward derivation from Definition 16. We have that $y \sim N_d(0, \Sigma)$, $x = \mu + y\sqrt{v\xi}$, and consider the rest Σ , μ , v , and ξ as constant/known. So $x|\mu, \Sigma, v, \xi$ follows a Normal distribution with mean $E(x|\xi) = \mu$ and covariance matrix $\text{Var}(x|\xi) = \Sigma v\xi$. Hence, we can just write down $x|\xi \sim N_d(\mu, \Sigma v\xi)$ by suppress the uninteresting constants Σ , μ , and v from the conditioning.

Proposition 18. If $x \sim T_d(\mu, \Sigma, v)$ and $\Sigma > 0$ then

1. The expected value is

$$E_{T_d(\mu, \Sigma, v)}(X) = \mu$$

2. The covariance matrix is

$$\text{Var}_{T_d(\mu, \Sigma, \nu)}(X) = \begin{cases} \frac{\nu}{\nu-2} \Sigma & , \text{ if } \nu > 2 \\ \text{undefined} & , \text{ else} \end{cases} \quad (7)$$

Proof. Given Definition 16, $x \sim T_d(\mu, \Sigma, \nu)$ results as the marginal distribution of (x, ξ) where $x|\xi \sim N_d(\mu, \Sigma\xi v)$ and $\xi \sim \text{IG}(\frac{\nu}{2}, \frac{1}{2})$.

1. It is

$$\mathbb{E}_{T_d(\mu, \Sigma, \nu)}(x) = \mathbb{E}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\mathbb{E}_{N_d(\mu, \Sigma\xi v)}(x|\xi)) = \mathbb{E}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\mu) = \mu \quad (8)$$

2. It is

$$\text{Var}_{T_d(\mu, \Sigma, \nu)}(x) = \mathbb{E}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\text{Var}_{N_d(\mu, \Sigma\xi v)}(x|\xi)) + \text{Var}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\mathbb{E}_{N_d(\mu, \Sigma\xi v)}(x|\xi)) \quad (9)$$

$$= \mathbb{E}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\Sigma\xi v) + \text{Var}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\mu) \stackrel{=0}{=} \Sigma v \mathbb{E}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\xi) + 0$$

$$= \begin{cases} \Sigma v \frac{\frac{1}{2}}{\frac{\nu}{2}-1} & , \text{ if } \frac{\nu}{2} > 1 \\ \text{undefined} & , \text{ else} \end{cases}$$

□

Example 19. If $x \sim T_d(\mu, \Sigma, \nu)$ and $\Sigma > 0$ the PDF of x is

$$f_X(x|\mu, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2}) \nu^{\frac{d}{2}} \pi^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \left(1 + \frac{1}{\nu} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right)^{-\frac{\nu+d}{2}} \quad (10)$$

Solution. Given Definition 16, $x \sim T_d(\mu, \Sigma, \nu)$ results as the marginal distribution of (x, ξ) where $x|\xi \sim N_d(\mu, \Sigma\xi v)$ and $\xi \sim \text{IG}(\frac{\nu}{2}, \frac{1}{2})$. So it is

$$\begin{aligned} f_x(x) &= \int f_{x|\xi}(x|\xi) f_\xi(\xi) d\xi = \int f_{N_d(\mu, \Sigma v \xi)}(x|\xi) f_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})}(\xi) d\xi \\ &= \int \underbrace{\left(\frac{1}{2\pi} \right)^{\frac{d}{2}} \frac{1}{\sqrt{\det(\Sigma v \xi)}} \exp \left(-\frac{1}{2} (x - \mu)^\top \frac{\Sigma^{-1}}{v \xi} (x - \mu) \right)}_{=N_d(x|\mu, \Sigma v \xi)} \underbrace{\frac{1}{\Gamma(\frac{\nu}{2})} \xi^{-\frac{\nu}{2}-1} \exp \left(-\frac{1}{\xi} \frac{1}{2} \right) 1_{(0, \infty)}(\xi)}_{=IG(\xi|\frac{\nu}{2}, \frac{1}{2})} d\xi \\ &= \left(\frac{1}{2\pi} \right)^{\frac{d}{2}} \frac{1}{\sqrt{\det(\Sigma v)}} \frac{1}{\Gamma(\frac{\nu}{2})} \int \underbrace{\xi^{-\frac{\nu}{2}-\frac{d}{2}-1} \exp \left(-\frac{1}{\xi} \left[\frac{1}{2v} (x - \mu)^\top \Sigma^{-1} (x - \mu) + \frac{1}{2} \right] \right)}_{=\Gamma(\frac{\nu}{2} + \frac{d}{2}) \left[\frac{1}{2v} (x - \mu)^\top \Sigma^{-1} (x - \mu) + \frac{1}{2} \right]^{-\left(\frac{\nu}{2} + \frac{d}{2}\right)}} d\xi \\ &= \left(\frac{1}{2\pi} \right)^{\frac{d}{2}} \frac{1}{\sqrt{\det(\Sigma v)}} \frac{1}{\Gamma(\frac{\nu}{2})} \Gamma \left(\frac{\nu}{2} + \frac{d}{2} \right) \left[\frac{1}{2v} (x - \mu)^\top \Sigma^{-1} (x - \mu) + \frac{1}{2} \right]^{-\left(\frac{\nu}{2} + \frac{d}{2}\right)} \\ &= \left(\frac{1}{2\pi} \right)^{\frac{d}{2}} \frac{1}{\sqrt{\det(\Sigma v)}} \frac{1}{\Gamma(\frac{\nu}{2})} \Gamma \left(\frac{\nu}{2} + \frac{d}{2} \right) \left(\frac{1}{2} \right)^{-\frac{(\nu+d)}{2}} \left[\frac{1}{v} (x - \mu)^\top \Sigma^{-1} (x - \mu) + 1 \right]^{-\frac{\nu+d}{2}} \\ &= \left(\frac{1}{\pi} \right)^{\frac{d}{2}} \frac{1}{\sqrt{\det(\Sigma)}} \left(\frac{1}{v} \right)^{\frac{d}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} \Gamma \left(\frac{\nu+d}{2} \right) \left[\frac{1}{v} (x - \mu)^\top \Sigma^{-1} (x - \mu) + 1 \right]^{-\frac{\nu+d}{2}} \end{aligned} \quad (11)$$

158 where the integral in (11) was calculated by recognizing the IG density from (1).

159 **5 Practice**

160 **Question 20.** *For practice try the Exercises 18, 19, and, 21, from the Exercise Sheet.*

Supplementary material

A The useful formulas regarding Normal PDF

The following formulas will be given as hints in the exercises and there is no need to be memorized. I present them for your information –no need to memorize.

Fact 21. If $\Sigma_1 > 0$ and $\Sigma_2 > 0$ symmetric

$$-\frac{1}{2}(x - \mu_1) \Sigma_1^{-1} (x - \mu_1)^\top - \frac{1}{2}(x - \mu_2) \Sigma_2^{-1} (x - \mu_2)^\top = -\frac{1}{2}(x - m) V^{-1} (x - m)^\top + C$$

where

$$V^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}; \quad m = V (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2); \quad C = \frac{1}{2} m^\top V^{-1} m - \frac{1}{2} (\mu_1^\top \Sigma_1^{-1} \mu_1 + \mu_2^\top \Sigma_2^{-1} \mu_2)$$

Proof. It is derived by \pm ing terms and doing matrix calculations. □

Fact 22. If $f_{N_d(\mu, \Sigma)}(x)$ denotes the PDF of $N_d(\mu, \Sigma)$, then

$$f_{N_d(\mu_1, \Sigma_1)}(x) f_{N_d(\mu_2, \Sigma_2)}(x) = f_{N_d(m, V)}(x) f_{N_d(\mu_2, \Sigma_1 + \Sigma_2)}(\mu_1)$$

where

$$V^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}; \quad m = V (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2)$$

Proof. It is derived by exponentiation □

Fact 23. If $\Sigma_i > 0$ symmetric for $i = 1, \dots, n$

$$-\frac{1}{2} \sum_{i=1}^n (x - \mu_i) \Sigma_i^{-1} (x - \mu_i)^\top = -\frac{1}{2} (x - m) V^{-1} (x - m)^\top + C \quad (12)$$

where

$$V^{-1} = \sum_{i=1}^n \Sigma_i^{-1}; \quad m = V \left(\sum_{i=1}^n \Sigma_i^{-1} \mu_i \right); \quad C = \frac{1}{2} m^\top V^{-1} m - \frac{1}{2} \left(\sum_{i=1}^n \mu_i^\top \Sigma_i^{-1} \mu_i \right) \quad (13)$$

Proof. It is shown by induction from the (1.). □

B Examples of mixture distributions

The following distributions below are given as examples of mixture distributions –no need to memorize.

B.1 Fisher-Snedecor distribution $x \sim F(d_1, d_2)$

Definition 24. A random variable $x \in \mathbb{R}$ is said to have a Fisher-Snedecor distribution with degrees of freedom n, m , and it is denoted as $x \sim F(n, m)$, if and only if

$$x = \frac{y_1/d_1}{y_2/d_2}$$

where $y_1 \sim G(d_1/2, 1/2)$ and $y_2 \sim G(d_2/2, 1/2)$ are independent random variables.

Example 25. The distribution function $x \sim L(\mu, b)$ is such that

$$dF(x) = \left[xB\left(\frac{d_1}{2}, \frac{d_2}{2}\right) \right]^{-1} \sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}} dx$$

with mean $E(x) = d_2/(d_2 - 1)$.

Solution. See Statistical Concepts II (Exercises) which is proved by transformation of random variables and marginalization.

Example 26. The distribution function $x \sim F(d_1, d_2)$ can be re-presented as a mixture distribution as

$$\begin{cases} x|v \sim \text{Ga}(d_1/2, zd_1/2) \\ z \sim \text{Ga}(d_2/2, d_2/2) \end{cases}$$

Solution. Show that $f(x) = \int f(x|z)f(z)dz$; see Exercise 19 in the Exercise sheet.

B.2 Laplace distribution $x \sim L(\mu, b)$

Definition 27. A random variable $x \in \mathbb{R}$ is said to have a Laplace (or Double exponential) distribution with location parameter μ , and scale parameter b , and it is denoted as $x \sim L_d(\mu, b)$, if and only if

$$x = \mu + y\sqrt{2b}$$

where $y \sim N_d(0, b^2)$ and $v \sim G(1, 1)$ are independent random variables.

Example 28. The distribution function $x \sim L(\mu, b)$ can be re-presented as a mixture distribution as

$$\begin{cases} x|v \sim N(\mu, 2vb^2) \\ v \sim \text{Ga}(1, 1) \equiv \text{Ex}(1) \end{cases} \quad (14)$$

Show that the distribution function $x \sim L(\mu, b)$ is such that

$$dF(x) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) dx$$

with mean $E(x) = \mu$ and variance $\text{Var}(x) = 2b^2$.

Solution. By transformation of random variables you show (14). To show that $f(x) = \int f(x|v)f(v)dv$. Perform change of variable to compute the integral. See C. Robert's solution: <https://stats.stackexchange.com/questions/175458/show-that-a-scale-mixtures-of-normals-is-a-power-exponential>