Bayesian Statistics III/IV (MATH3341/4031) Michaelmas term, 2019

Handout 2: Probability calculations & Known distributions [a]

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

Aim

To practice on probability calculations. To become familiar with distributions, Inverted Gamma, multivariate Normal, and multivariate Student T distributions.
It is not required to memorize the formulas in Equations: 2, 3, 4, 6 7, 8, 9, and 10.

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

- Multivariate Normal and Student T distributions:
  `https://georgios-stats-3.shinyapps.io/demo_multivariatenormaldistribution/`
  `https://github.com/georgios-stats/Shiny_applets/tree/master/demo_MultivariateNormalDistribution`

[a] Author: Georgios P. Karagiannis.

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

Definition 1. 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 2. Let $x \sim \text{IG}(a,b)$, then the PDF of x is

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

Solution. It is

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

Example 3. Let a random variable $x \sim \text{IG}(a,b)$, then

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

Solution. It is

$$\mathrm{E}_{\mathrm{IG}(a,b)}(x) = \int x f_{\mathrm{IG}(a,b)}(x)\mathrm{d}x \quad = \int_{(0,+\infty)} x \frac{b^a}{\Gamma(a)} x^{-a-1} \exp(-\frac{b}{x})\mathrm{d}x$$

Assume that $a > 1$. Then

$$\mathrm{E}_{\mathrm{IG}(a,b)}(x) = \int_{(0,+\infty)} \frac{b^{a-1}}{\Gamma(a)} b \frac{\Gamma(a-1)}{\Gamma(a-1)} x^{-a+1-1} \exp(-\frac{b}{x})\mathrm{d}x \quad = b\frac{\Gamma(a-1)}{\Gamma(a)} \int_{(0,+\infty)} \frac{b^{a-1}}{\Gamma(a-1)} x^{-a+1-1} \exp(-\frac{b}{x})\mathrm{d}x$$

$$= b\frac{\Gamma(a-1)}{\Gamma(a)} \int_{(0,+\infty)} \frac{b^{a-1}}{\Gamma(a-1)} x^{-a+1-1} \exp(-\frac{b}{x})\mathrm{d}x = b\frac{\Gamma(a-1)}{(a-1)\Gamma(a-1)} \int f_{\mathrm{IG}(a-1,b)}(x)\mathrm{d}x \quad = \frac{b}{a-1}$$

Similarly

$$\mathrm{E}_{\mathrm{IG}(a,b)}(x^2) = \int_{(0,+\infty)} x^2 \frac{b^a}{\Gamma(a)} x^{-a-1} \exp(-\frac{b}{x})\mathrm{d}x = ... = b\frac{\Gamma(a-1)}{(a-1)\Gamma(a-1)} \int x f_{\mathrm{IG}(a-1,b)}(x)\mathrm{d}x$$

$$= \frac{b}{a-1}\frac{b}{a-2};\ a > 2$$

So

$$\mathrm{Var}_{\mathrm{IG}(a,b)}(x) = \mathrm{E}_{\mathrm{IG}(a,b)}(x^2) - \left(\mathrm{E}_{\mathrm{IG}(a,b)}(x)\right)^2 = \frac{b^2}{(a-1)^2(a-2)}$$

2 Multivariate Normal distribution[1] $x|\mu,\Sigma \sim \mathbf{N}_d(\mu,\Sigma)$

Definition 4. 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 5. *A random vector $x \in \mathbb{R}^d$ has a d-dimensional Normal distribution with mean $\mu = E(x)$ and covariance matrix $\Sigma = Var(x)$ if and only if random vector $x \in \mathbb{R}^d$ has a characteristic function*

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

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

Proof. ($\Longrightarrow$) 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 $\mu = \mathrm{E}(x)$ and covariance matrix $\Sigma = \mathrm{Var}(x)$, $t^\top x$ has a Normal distribution with mean $\mathrm{E}\left(t^\top x\right) = t^\top\mu$ and variance $\mathrm{Var}\left(t^\top x\right) = t^\top\Sigma t$. Then

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

($\Longleftarrow$) 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

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

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

Notation 6. We denote the d-dimensional Normal distribution with mean μ and covariance matrix $\Sigma \geq 0$ as $\mathrm{N}_d(\mu,\Sigma)$.

Notation 7. The d-dimensional standardized Normal distribution is $\mathrm{N}_d(0,I)$.

[1] Try the applet: `https://georgios-stats-3.shinyapps.io/demo_multivariatenormaldistribution/`

Proposition 8. *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 $\tilde{\alpha} = A^\top b$. Because x is multivariate Normal, then $\tilde{\alpha}^\top x$ is univariate Normal (by Definition 4), then $\alpha^\top y$ is univariate Normal. So y is q-variate Normal. Also, $\mathrm{E}(y) = \mathrm{E}(c + Ax) = c + A\mathrm{E}(x)$, and $\mathrm{Var}(y) = \mathrm{Var}(c + Ax) = A\mathrm{Var}(x)A^\top$. □

Proposition 9. *Let a* d*-dimensional random vector* $x \sim N_{(any)}(\mu, \Sigma)$*.*

1. *Let* $x = (x_1, ..., x_d)^\top$*: The* $x_1, ..., x_d$ *are mutually independent if and only if the corresponding off diagonal parts of the* Σ *are zero.*

2. *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* $A\Sigma B^\top = 0$*.*

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. □

Proposition 10. *Any sub-vector of a vector with multivariate Normal distribution has a multivariate Normal distribution.*

Proof. Let $x \sim \mathrm{N}_d(\mu, \Sigma)$. 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 $y \sim \mathrm{N}_d(P\mu, P\Sigma P^\top)$. □

Proposition 11. *[Marginalization & conditioning]* [2] *Let* $x \sim \mathrm{N}_d(\mu, \Sigma)$. Consider partition such that

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

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

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

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

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

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

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

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

where

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

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.

[2] It is good (although not required) to memorize the formulas in (1) and (4) as they are important in Statistics.

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

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

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

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

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

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

$$= \begin{bmatrix} \Sigma_1, & \Sigma_{21}^\top \end{bmatrix} \begin{bmatrix} -\Sigma_1^{-1}\Sigma_{21}^\top \\ I \end{bmatrix} = -\Sigma_{21}^\top + \Sigma_{21}^\top = 0$$

4. From the above, $x_{2.1}$ is independent on x_1; hence the conditional distribution of $x_2|x_1$ (x_2 given x_1 is known) is the same as the marginal distribution of $x_{2.1}$ aka Normal. Namely $\mathrm{d}F(x_{2.1}|x_1) = \mathrm{d}F(x_{2.1}) \in$ Normal. From the above I observe that it is

$$x_{2.1} = x_2 - \Sigma_{21}\Sigma_1^{-1}x_1 \Longleftrightarrow x_2 = x_{2.1} + \Sigma_{21}\Sigma_1^{-1}x_1;$$

hence 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 $x_2|x_1 = x_{2.1} + \text{const.}$, which implies that the conditional distribution of $x_2|x_1$ is Normal. Now, about the moments

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

Proposition 12. *The density function of the d-dimensional Normal distribution with mean μ and covariance matrix Σ, when $\underline{\Sigma \text{ is symmetric positive definite matrix } (\Sigma > 0)}$, exists and it is equal to*

$$f(x) = (2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) \tag{5}$$

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

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

□

Fact 13. *[In exercises they will be given as a Hint.] Useful formulas about the PDF of the multivariate Normal distribution that we may use.*

1. *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

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

2. *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

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

3. *If $\Sigma_i > 0$ symmetric for $i = 1, ..., n$*

$$-\frac{1}{2}\sum_{i=1}^{n}(x-\mu_i)\Sigma_i^{-1}(x-\mu_i)^\top = -\frac{1}{2}(x-m)V^{-1}(x-m)^\top + C \tag{6}$$

where

$$V^{-1} = \sum_{i=1}^{n}\Sigma_i^{-1}; \quad m = V\left(\sum_{i=1}^{n}\Sigma_i^{-1}\mu_i\right); \quad C = \frac{1}{2}mV^{-1}m^\top - \frac{1}{2}\left(\sum_{i=1}^{n}\mu_i\Sigma_i^{-1}\mu_i^\top\right) \tag{7}$$

Proof. (1.) is derived by $\pm$ing terms and doing matrix calculations. (2.) is derived by exponentiation and completing the associated constants. (3.) is shown by induction from the (1.). □

3 Multivariate Student's T distribution[3] $x \sim \mathbf{T}_d(\mu, \Sigma, v)$

Definition 14. 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 $x \sim \mathrm{T}_d(\mu, \Sigma, v)$, if and only if

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

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

[3] Try the applet: `https://georgios-stats-3.shinyapps.io/demo_multivariatenormaldistribution/`

Example 15. If $x \sim \mathrm{T}_d(\mu, \Sigma, v)$ and $\Sigma > 0$ then

1. The PDF of x is
$$f_X(x|\mu, \Sigma, v) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\nu^{\frac{d}{2}}\pi^{\frac{d}{2}}\det(\Sigma)^{\frac{1}{2}}}\left(1 + \frac{1}{v}(t-\mu)^{\mathrm{T}}\Sigma^{-1}(t-\mu)\right)^{-\frac{\nu+d}{2}} \tag{8}$$

2. The expected value is
$$\mathrm{E}_{\mathrm{T}_d(\mu,\Sigma,\nu)}(X) = \mu \tag{9}$$

3. The covariance matrix is
$$\mathrm{Var}_{\mathrm{T}_d(\mu,\Sigma,\nu)}(X) = \begin{cases} \frac{\nu}{\nu-2}\Sigma & \text{, if } \nu > 2 \\ \\ \text{undefined} & \text{, else} \end{cases} \tag{10}$$

Hint: Use that: $x = \mu + y\sqrt{v\xi}$ where $y \sim \mathrm{N}_d(0, \Sigma)$ and $\xi \sim \mathrm{IG}(\frac{v}{2}, \frac{1}{2})$ independent.

Solution. Given Definition 14, $x \sim \mathrm{T}_d(\mu, \Sigma, v)$ results as the marginal distribution of (x, ξ) where $x|\xi \sim \mathrm{N}_d(\mu, \Sigma\xi v)$ and $\xi \sim \mathrm{IG}(\frac{v}{2}, \frac{1}{2})$.

1. So it is
$$\begin{aligned}
f_x(x) &= \int f_{x|\xi}(x|\xi)f_\xi(\xi)\mathrm{d}\xi = \int f_{\mathrm{N}_d(\mu,\Sigma v\xi)}(x|\xi)f_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}(\xi)\mathrm{d}\xi \\
&= \int \underbrace{\left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{1}{\sqrt{\det(\Sigma v\xi)}}\exp\left(-\frac{1}{2}(x-\mu)^\top\frac{\Sigma^{-1}}{v\xi}(x-\mu)\right)}_{=\mathrm{N}_d(x|\mu,\Sigma v\xi)}\underbrace{\frac{\frac{1}{2}^{\frac{v}{2}}}{\Gamma(\frac{v}{2})}\xi^{-\frac{v}{2}-1}\exp\left(-\frac{1}{\xi}\frac{1}{2}\right)1_{(0,\infty)}(\xi)}_{=\mathrm{IG}(\xi|\frac{v}{2},\frac{1}{2})}\mathrm{d}\xi \\
&= \left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{1}{\sqrt{\det(\Sigma v)}}\frac{\frac{1}{2}^{\frac{v}{2}}}{\Gamma(\frac{v}{2})}\underbrace{\int \xi^{-\frac{v}{2}-\frac{d}{2}-1}\exp\left(-\frac{1}{\xi}\left[\frac{1}{2v}(x-\mu)^\top\Sigma^{-1}(x-\mu)+\frac{1}{2}\right]\right)\mathrm{d}\xi}_{=\Gamma\left(\frac{v}{2}+\frac{d}{2}\right)\left[\frac{1}{2v}(x-\mu)^\top\Sigma^{-1}(x-\mu)+\frac{1}{2}\right]^{-\left(\frac{v}{2}+\frac{d}{2}\right)}} \qquad (11) \\
&= \left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{1}{\sqrt{\det(\Sigma v)}}\frac{\frac{1}{2}^{\frac{v}{2}}}{\Gamma(\frac{v}{2})}\Gamma\left(\frac{v}{2}+\frac{d}{2}\right)\left[\frac{1}{2v}(x-\mu)^\top\Sigma^{-1}(x-\mu)+\frac{1}{2}\right]^{-\left(\frac{v}{2}+\frac{d}{2}\right)} \\
&= \left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\frac{1}{\sqrt{\det(\Sigma v)}}\frac{\frac{1}{2}^{\frac{v}{2}}}{\Gamma(\frac{v}{2})}\Gamma\left(\frac{v}{2}+\frac{d}{2}\right)\left(\frac{1}{2}\right)^{-\frac{(v+d)}{2}}\left[\frac{1}{v}(x-\mu)^\top\Sigma^{-1}(x-\mu)+1\right]^{-\frac{v+d}{2}} \\
&= \left(\frac{1}{\pi}\right)^{\frac{d}{2}}\frac{1}{\sqrt{\det(\Sigma)}}\left(\frac{1}{v}\right)^{\frac{d}{2}}\frac{1}{\Gamma(\frac{v}{2})}\Gamma\left(\frac{v+d}{2}\right)\left[\frac{1}{v}(x-\mu)^\top\Sigma^{-1}(x-\mu)+1\right]^{-\frac{v+d}{2}}
\end{aligned}$$
where the integral in (11) was calculated by recognizing the IG density from (1).

2. It is
$$\mathrm{E}_{\mathrm{t}_d(\mu,\Sigma,\nu)}(x) = \mathrm{E}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}\left(\mathrm{E}_{\mathrm{N}_d(\mu,\Sigma\xi v)}(x|\xi)\right) = \mathrm{E}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}(\mu) = \mu$$

3. It is

$$
\begin{aligned}
\mathrm{Var}_{\mathrm{t}_d(\mu,\Sigma,\nu)}(x) =& \mathrm{E}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}\left(\mathrm{Var}_{\mathrm{N}_d(\mu,\Sigma\xi v)}(x|\xi)\right) + \mathrm{Var}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}\left(\mathrm{E}_{\mathrm{N}_d(\mu,\Sigma\xi v)}(x|\xi)\right) \\
=& \mathrm{E}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}(\Sigma\xi v) + \cancelto{0}{\mathrm{Var}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}(\mu)} = \Sigma v \mathrm{E}_{\mathrm{IG}(\frac{v}{2},\frac{1}{2})}(\xi) + 0 \\
=& \begin{cases} \Sigma v \frac{\frac{1}{2}}{\frac{v}{2}-1} & \text{, if } \frac{v}{2} > 1 \\ \text{undefined} & \text{, else} \end{cases}
\end{aligned}
$$

4 Practice

Question 16. *For practice try the Exercises 13, 14, and, 16, from the Exercise Sheet.*