

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

^aAuthor: 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

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

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

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

Solution. It is

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

Assume that $a > 1$. Then

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

Similarly

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

So

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

2 Multivariate Normal distribution¹ $x|\mu, \Sigma \sim \mathbf{N}_d(\mu, \Sigma)$

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 $\Sigma = \text{Var}(x)$ if and only if random vector $x \in \mathbb{R}^d$ has a characteristic function

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

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

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

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

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

¹Try 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 $\tilde{\alpha} = A^\top \alpha$. 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, $E(y) = E(c + Ax) = c + AE(x)$, and $\text{Var}(y) = \text{Var}(c + Ax) = A\text{Var}(x)A^\top$. \square

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

1. 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.
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. \square

Proposition 10. 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 11. [Marginalization & conditioning]² 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

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

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

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

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.

²It 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 $x_1 \sim N(A\mu, A\Sigma A^\top)$ 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 $[-\Sigma_{21}\Sigma_1^{-1}, I]$. Then $x_{2.1} \sim N(B\mu, B\Sigma B^\top)$ where

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

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 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 $dF(x_{2.1}|x_1) = dF(x_{2.1}) \in \text{Normal}$. From the above I observe that it is

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

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

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

Proposition 12. 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

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

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

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

95 Then

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

99 □
 100 **Fact 13.** [In exercises they will be given as a Hint.] Useful formulas about the PDF of the multivariate Normal
 101 distribution that we may use.

102 1. If $\Sigma_1 > 0$ and $\Sigma_2 > 0$ symmetric

$$103 \quad -\frac{1}{2} (x - \mu_1)^\top \Sigma_1^{-1} (x - \mu_1) - \frac{1}{2} (x - \mu_2)^\top \Sigma_2^{-1} (x - \mu_2) = -\frac{1}{2} (x - m)^\top V^{-1} (x - m) + C$$

104 where

$$105 \quad V^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}; \quad m = V (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2); \quad C = \frac{1}{2} m^\top V^{-1} m - \frac{1}{2} (\mu_1^\top \Sigma_1^{-1} \mu_1 + \mu_2^\top \Sigma_2^{-1} \mu_2)$$

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

$$107 \quad f_{N_d(\mu_1, \Sigma_1)}(x) f_{N_d(\mu_2, \Sigma_2)}(x) = f_{N_d(m, V)}(x) f_{N_d(\mu_2, \Sigma_1 + \Sigma_2)}(\mu_1)$$

108 where

$$109 \quad V^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}; \quad m = V (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2)$$

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

$$111 \quad -\frac{1}{2} \sum_{i=1}^n (x - \mu_i)^\top \Sigma_i^{-1} (x - \mu_i) = -\frac{1}{2} (x - m)^\top V^{-1} (x - m) + C \quad (6)$$

112 where

$$113 \quad V^{-1} = \sum_{i=1}^n \Sigma_i^{-1}; \quad m = V \left(\sum_{i=1}^n \Sigma_i^{-1} \mu_i \right); \quad C = \frac{1}{2} m^\top V^{-1} m - \frac{1}{2} \left(\sum_{i=1}^n \mu_i^\top \Sigma_i^{-1} \mu_i \right) \quad (7)$$

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

116 3 Multivariate Student's T distribution³ $x \sim \mathbf{T}_d(\mu, \Sigma, v)$

117 **Definition 14.** A d -dimensional random variable $x \in \mathbb{R}^d$ is said to have a multivariate Student's T distribution with
 118 location parameter μ , scale matrix Σ , and degrees of freedom v , and it is denoted as $x \sim \mathbf{T}_d(\mu, \Sigma, v)$, if and only if

$$119 \quad x = \mu + y \sqrt{v \xi}$$

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

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

121

122

123

124

125

126

127

128

129

131

132

138

139

3. It is

$$\text{Var}_{\text{IG}(\mu, \Sigma, \nu)}(x) = \text{E}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})} \left(\text{Var}_{\text{N}_d(\mu, \Sigma \xi \nu)}(x|\xi) \right) + \text{Var}_{\text{IG}(\frac{\nu}{2}, \frac{1}{2})} \left(\text{E}_{\text{N}_d(\mu, \Sigma \xi \nu)}(x|\xi) \right)$$

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

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

4 Practice

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