

AstroStatistics and Cosmology Homework

Jacopo Tissino

2020-11-22

1 November exercises

Exercise 4

After being given a probability distribution $\mathbb{P}(x)$, we define the *characteristic function* ϕ as its Fourier transform, which can also be expressed as the expectation value of $\exp(-i\vec{k} \cdot \vec{x})$:

$$\phi(\vec{k}) = \int d^n x \exp(-i\vec{k} \cdot \vec{x}) \mathbb{P}(x) = \mathbb{E} \left[\exp(-i\vec{k} \cdot \vec{x}) \right]. \quad (1.1)$$

Claim 1.1. *A multivariate normal distribution*

$$\mathcal{N}(\vec{x}|\vec{\mu}, C) = \frac{1}{(2\pi)^{n/2} \sqrt{\det C}} \exp\left(-\frac{1}{2} \vec{y}^\top C^{-1} \vec{y}\right) \Big|_{\vec{y}=\vec{x}-\vec{\mu}}, \quad (1.2)$$

has a characteristic function equal to

$$\phi(\vec{k}) = \exp\left(-i\vec{\mu} \cdot \vec{k} - \frac{1}{2} \vec{k}^\top C \vec{k}\right). \quad (1.3)$$

Proof: completing the square. The integral we need to compute is given, absorbing the normalization into a factor N , by

$$\phi(\vec{k}) = N \int d^n x \exp\left(-i\vec{k} \cdot \vec{x} - \frac{1}{2} \vec{y}^\top C^{-1} \vec{y}\right) \Big|_{\vec{y}=\vec{x}-\vec{\mu}}. \quad (1.4)$$

The only integrals we really know how to do are Gaussian ones, so we want to rewrite the argument of the exponential so that it is a quadratic form. The manipulation goes as follows, considering the opposite of the argument the exponential in order to have less minus signs and defining the symmetric matrix $V = C^{-1}$:

$$i\vec{k} \cdot \vec{x} + \frac{1}{2} \vec{y}^\top V \vec{y} = i\vec{k} \cdot \vec{x} + \frac{1}{2} \vec{x}^\top V \vec{x} - \vec{x}^\top V \vec{\mu} + \frac{1}{2} \vec{\mu}^\top V \vec{\mu} \quad (1.5)$$

$$= \frac{1}{2} \vec{x}^\top V \vec{x} + \vec{x}^\top (i\vec{k} - V\vec{\mu}) + \frac{1}{2} \vec{\mu}^\top V \vec{\mu} \quad (1.6)$$

$$\begin{aligned}
&= \underbrace{\frac{1}{2} \left(\vec{x} + V^{-1}(i\vec{k} - V\vec{\mu}) \right)^\top V \left(\vec{x} + V^{-1}(i\vec{k} - V\vec{\mu}) \right)}_{\textcircled{1}} + \\
&\quad \underbrace{-\frac{1}{2} \left(i\vec{k} - V\vec{\mu} \right)^\top V^{-1} \left(i\vec{k} - V\vec{\mu} \right) + \frac{1}{2} \vec{\mu}^\top V \vec{\mu}}_{\textcircled{2}}, \tag{1.7}
\end{aligned}$$

which we can now integrate, since it is now a quadratic form in terms of a shifted variable, $\vec{x} + \vec{p}$, where the constant (with respect to \vec{x}) vector \vec{p} is given by $V^{-1}(i\vec{k} - V\vec{\mu})$.¹

Now, shifting the integral from one in $d^n x$ to one in $d^n(x + p)$ does not change the measure, since the Jacobian of a shift is the identity. Then, we have

$$\phi(\vec{k}) = N \int d^n(x + p) \exp\left(-\textcircled{1} - \textcircled{2}\right) \tag{1.12}$$

$$= N \sqrt{\frac{(2\pi)^n}{\det V}} \exp\left(-\textcircled{2}\right) \tag{1.13}$$

$$= \underbrace{\frac{1}{\sqrt{\det V \det C}}}_{=1} \exp\left(-\textcircled{2}\right), \tag{1.14}$$

since the determinant of the inverse is the inverse of the determinant.

Now, we only need to simplify $\textcircled{2}$:

$$\textcircled{2} = -\frac{1}{2} \left[-\vec{k}^\top V^{-1} \vec{k} - 2i\vec{\mu}^\top V V^{-1} \vec{k} + \vec{\mu}^\top V V^{-1} V \vec{\mu} \right] + \frac{1}{2} \vec{\mu}^\top V \vec{\mu} \tag{1.15}$$

$$= \frac{1}{2} \vec{k}^\top C \vec{k} + i\vec{\mu}^\top \vec{k}, \tag{1.16}$$

inserting which into the exponent yields the desired result. \square

Proof: by diagonalization. We now follow a different approach: the covariance matrix C is symmetric, so we will always be able to find an orthogonal matrix O (satisfying $O^\top = O^{-1}$) such that $C = O^\top D O$, where D is diagonal. We will then also have $V = C^{-1} = O^\top D^{-1} O$. Let us denote the eigenvalues of D as λ_i , and the eigenvalues of D^{-1} as $d_i = \lambda_i^{-1}$.

Defining $\vec{z} = O\vec{x}$, $\vec{m} = O\vec{\mu}$, $\vec{u} = O\vec{k}$ the negative of the argument of the integral becomes:

$$i\vec{k} \cdot \vec{x} + \frac{1}{2} (\vec{x} - \vec{\mu})^\top C^{-1} (\vec{x} - \vec{\mu}) = i\vec{u} \cdot \vec{z} + \frac{1}{2} (\vec{z} - \vec{m})^\top D^{-1} (\vec{z} - \vec{m}) \tag{1.17}$$

¹ In the last step we applied the matrix square completion formula: for a symmetric matrix A and vectors \vec{x} , \vec{b} we have

$$\frac{1}{2} (\vec{x} + A^{-1} \vec{b})^\top A (\vec{x} + A^{-1} \vec{b}) - \frac{1}{2} \vec{b}^\top A^{-1} \vec{b} = \tag{1.8}$$

$$= \frac{1}{2} \left[\vec{x}^\top A \vec{x} + \vec{x}^\top A A^{-1} \vec{b} + (A^{-1} \vec{b})^\top A \vec{x} + (A^{-1} \vec{b})^\top A A^{-1} \vec{b} - \vec{b}^\top A^{-1} \vec{b} \right] \tag{1.9}$$

$$= \frac{1}{2} \left[\vec{x}^\top A \vec{x} + \vec{x}^\top \vec{b} + \vec{b}^\top (A^{-1})^\top A \vec{x} + \vec{b}^\top (A^{-1})^\top \vec{b} - \vec{b}^\top A^{-1} \vec{b} \right] \tag{1.10}$$

$$= \frac{1}{2} \vec{x}^\top A \vec{x} + \vec{b}^\top \vec{x}, \tag{1.11}$$

which we used with $\vec{b} = i\vec{k} - V\vec{\mu}$.

$$= i\vec{u} \cdot \vec{z} + \frac{1}{2} \sum_i d_i (z_i - m_i)^2 \quad (1.18)$$

$$= \sum_i \left[iu_i z_i + \frac{d_i}{2} (z_i^2 + m_i^2 - 2m_i z_i) \right] \quad (1.19)$$

$$= \sum_i \left[z_i^2 \frac{d_i}{2} + z_i (iu_i - m_i d_i) + \frac{d_i}{2} m_i^2 \right]. \quad (1.20)$$

With this, and since by $\det O = 1$ we have $d^n z = d^n x$, we can decompose our Gaussian integral into a product of Gaussian integrals:

$$\phi(\vec{k}) = N \int d^n x \exp \left(-i\vec{k} \cdot \vec{x} - \frac{1}{2} (\vec{x} - \vec{\mu})^\top C^{-1} (\vec{x} - \vec{\mu}) \right) \quad (1.21)$$

$$= N \int d^n z \exp \left(- \sum_i \left[z_i^2 \frac{d_i}{2} + z_i (iu_i - m_i d_i) + \frac{d_i}{2} m_i^2 \right] \right) \quad (1.22)$$

$$= N \prod_i \int dz_i \exp \left(-z_i^2 \frac{d_i}{2} - z_i (iu_i - m_i d_i) - \frac{d_i}{2} m_i^2 \right) \quad (1.23)$$

$$= N \prod_i \sqrt{\frac{2\pi}{d_i}} \exp \left(\frac{(iu_i - m_i d_i)^2}{2d_i} - \frac{d_i m_i^2}{2} \right) \quad (1.24)$$

$$= \frac{1}{\sqrt{\det C \det V}} \prod_i \exp \left(\frac{-u_i^2 + m_i^2 d_i^2 - 2iu_i m_i d_i}{2d_i} - \frac{d_i m_i^2}{2} \right) \quad (1.25)$$

$$= \exp \left(\sum_i \left[-\frac{u_i^2}{2d_i} - iu_i m_i \right] \right) \quad (1.26)$$

$$= \exp \left(-\frac{1}{2} \vec{u}^\top C \vec{u} - i\vec{u} \cdot \vec{m} \right) \quad (1.27)$$

$$= \exp \left(-\frac{1}{2} \vec{k}^\top C \vec{k} - i\vec{k} \cdot \vec{\mu} \right), \quad (1.28)$$

where we have used the expression for the single-variable Gaussian integral:

$$\int dz \exp(-az^2 + bz + c) = \sqrt{\frac{\pi}{a}} \exp \left(\frac{b^2}{4a} + c \right), \quad (1.29)$$

which comes from the one-variable completion of the square:

$$-az^2 + bz + c = -a \left(z - \frac{b}{2a} \right)^2 + \frac{b^2}{4a} + c. \quad (1.30)$$

Also, we used the fact that orthogonal transformation do not change fully-contracted objects, such as scalar products or bilinear forms. \square

Exercise 5

We can calculate the moments of a distribution through its characteristic function:

$$\mathbb{E}[x_\alpha^{n_\alpha} \dots x_\beta^{n_\beta}] = \frac{\partial^{n_\alpha \dots n_\beta} \phi(\vec{k})}{\partial(-ik_\alpha)^{n_\alpha} \dots \partial(-ik_\beta)^{n_\beta}} \Big|_{\vec{k}=0}. \quad (1.31)$$

In the multivariate Gaussian case we can then calculate the mean (component by component) as

$$\mathbb{E}(x_\alpha) = \frac{\partial \phi(\vec{k})}{\partial(-ik_\alpha)} \Big|_{\vec{k}=0} \quad (1.32)$$

$$= \frac{\partial}{\partial(-ik_\alpha)} \Big|_{\vec{k}=0} \exp\left(-\frac{1}{2} \vec{k}^\top C \vec{k} - i \vec{k} \cdot \vec{\mu}\right) \quad (1.33)$$

$$= \left[-i \sum_{\beta} k_\beta C_{\beta\alpha} + \mu_\alpha \right] \exp\left(-\frac{1}{2} \vec{k}^\top C \vec{k} - i \vec{k} \cdot \vec{\mu}\right) \Big|_{\vec{k}=0} \quad (1.34)$$

$$= \mu_\alpha, \quad (1.35)$$

where we used the fact that the differentiation of a symmetric bilinear form is as follows:

$$\frac{\partial}{\partial k_\alpha} \left(\sum_{\beta\gamma} k_\beta k_\gamma C_{\beta\gamma} \right) = 2 \sum_{\beta\gamma} \delta_{\beta\alpha} k_\gamma C_{\beta\gamma} = 2 \sum_{\gamma} k_\gamma C_{\alpha\gamma}. \quad (1.36)$$

The covariance matrix can be computed by linearity as

$$\tilde{C}_{\alpha\beta} = \mathbb{E} \left[(x_\alpha - \mathbb{E}(x_\alpha)) (x_\beta - \mathbb{E}(x_\beta)) \right] = \mathbb{E}[x_\alpha x_\beta] - \mu_\alpha \mu_\beta, \quad (1.37)$$

the first term of which reads as follows:

$$\mathbb{E}[x_\alpha x_\beta] = \frac{\partial^2 \phi(\vec{k})}{\partial(-ik_\beta) \partial(-ik_\alpha)} \Big|_{\vec{k}=0} \quad (1.38)$$

$$= \frac{\partial}{\partial(-ik_\beta)} \Big|_{\vec{k}=0} \left[-i \sum_{\beta} k_\beta C_{\beta\alpha} + \mu_\alpha \right] \exp\left(-\frac{1}{2} \vec{k}^\top C \vec{k} - i \vec{k} \cdot \vec{\mu}\right) \quad (1.39)$$

$$= C_{\alpha\beta} + \mu_\alpha \mu_\beta, \quad (1.40)$$

therefore, as expected, $\tilde{C}_{\alpha\beta}$ is indeed $C_{\alpha\beta}$.

Exercise 6

Claim 1.2. *The characteristic function of a multivariate Gaussian is, up to normalization, a multivariate Gaussian.*

Proof. The characteristic function is the exponential of (minus)

$$\frac{1}{2}\vec{k}^\top C\vec{k} + i\vec{k} \cdot \vec{\mu} = \frac{1}{2}(\vec{k} + iC^{-1}\vec{\mu})^\top C(\vec{k} + iC^{-1}\vec{\mu}) + \frac{1}{2}\vec{\mu}^\top C^{-1}\vec{\mu}, \quad (1.41)$$

which means that the characteristic function is in the form

$$\phi(\vec{k}) = \text{const} \times \exp\left(-\frac{1}{2}(\vec{k} - \vec{m})^\top C(\vec{k} - \vec{m})\right), \quad (1.42)$$

a multivariate normal with mean $\vec{m} = -iC^{-1}\vec{\mu}$ and covariance matrix C^{-1} , the inverse of the covariance matrix of the corresponding MVN. \square

Exercise 8

For clarity, we denote with Greek indices those ranging from 1 to N , the size of the vector of data; and with Latin indices those ranging from 1 to M , the number of templates.

We are assuming that the data have a Gaussian distribution with a covariance matrix C , and we are modelling their mean μ_α as a sum of templates $t_{i\alpha}$ with coefficients A_i :

$$\mu_\alpha = t_{i\alpha}A_i, \quad (1.43)$$

where the Einstein summation convention has been used. Therefore, the likelihood is proportional to

$$\mathcal{L}(d_\alpha|A_i) \propto \exp\left(-\frac{1}{2}(d_\alpha - A_it_{i\alpha})C_{\alpha\beta}^{-1}(d_\beta - A_jt_{j\beta})\right). \quad (1.44)$$

The normalization only depends on the covariance matrix $C_{\alpha\beta}$, which we assume is fixed. Therefore, maximizing the likelihood² is equivalent to minimizing the χ^2 , which reads

$$\chi^2 = (d_\alpha - A_it_{i\alpha})C_{\alpha\beta}^{-1}(d_\beta - A_jt_{j\beta}). \quad (1.45)$$

We want to minimize this as the amplitudes vary: therefore, we set the derivative with respect to A_k to zero,

$$\frac{\partial \chi^2}{\partial A_k} = -2t_{k\alpha}C_{\alpha\beta}^{-1}(d_\beta - A_jt_{j\beta}) = 0, \quad (1.46)$$

which means that

$$t_{k\alpha}C_{\alpha\beta}^{-1}d_\beta = (t_{k\alpha}C_{\alpha\beta}^{-1}t_{j\beta})A_j, \quad (1.47)$$

a linear system of M equations (indexed by k) in the M variables A_j . If we denote the evaluations of bilinear forms in the data (N -dimensional) space with brackets, as $a_\alpha C_{\alpha\beta} b_\beta \stackrel{\text{def}}{=} (a|C|b)$, this reads

$$(t|C^{-1}|d)_k = (t|C^{-1}|t)_{kj}A_j \quad (1.48)$$

² Which is equivalent to maximizing the posterior if we are using a flat prior.

$$\left[(t|C^{-1}|t)^{-1} \right]_{mk} (t|C^{-1}|d)_k = \underbrace{\left[(t|C^{-1}|t)^{-1} \right]_{mk} (t|C^{-1}|t)_{kj}}_{=\delta_{mj}} A_j = A_m \quad (1.49)$$

$$A_m = \left[(t|C^{-1}|t)^{-1} \right]_{mk} (t|C^{-1}|d)_k, \quad (1.50)$$

where the inverse of $(t|C^{-1}|t)$ is to be computed in the M -dimensional vector space.

Exercise 9

Our model for the mean value is in the form $\mu(\Theta, A) = A\bar{x}(\Theta)$, where \bar{x} is a generic function of Θ , while A is our scale parameter.³ Our likelihood then reads

$$\mathcal{L}(x|\Theta, A) = \underbrace{\frac{1}{(2\pi)^{N/2} \sqrt{\det C}}}_{B_1} \exp\left(-\frac{1}{2}(x - A\bar{x}(\Theta))^\top C^{-1}(x - A\bar{x}(\Theta))\right). \quad (1.51)$$

If the priors for both A and Θ are flat, this corresponds to the joint posterior $P(\Theta, A|x)$. We want to marginalize over A , which amounts to integrating over it: dropping the dependence on Θ of \bar{x} and defining $V = C^{-1}$ we find

$$P(\Theta|x) = B_1 \int \exp\left(-\frac{1}{2}(x - A\bar{x})^\top V(x - A\bar{x})\right) dA \quad (1.52)$$

$$= B_1 \int \exp\left(-\frac{1}{2}\left(x^\top Vx - 2A\bar{x}^\top Vx + A^2\bar{x}^\top V\bar{x}\right)\right) dA. \quad (1.53)$$

Used the symmetry of V .

The amplitude being negative makes little sense in a typical physical context, however the Gaussian integral can be done analytically only over the whole of \mathbb{R} .

In order to get analytical results, here we will marginalize by integrating over negative amplitudes as well ($A \in \mathbb{R}$); the last figure (1) will show how only integrating over positive amplitudes only would have looked (by numerical calculation) in a simple case. In general if one wishes to perform the integral over $A \in (0, +\infty)$ the tabulated values of the error function may be used.

Applying the formula for the single-variable Gaussian integral (1.29) (the bilinear forms are all evaluated to yield scalars, we are only integrating over the scalar A !) we then get

$$P(\Theta|x) = \underbrace{B_1 \exp\left(-\frac{1}{2}x^\top Vx\right)}_{B_2} \exp\left(\frac{(\bar{x}^\top Vx)^2}{(\bar{x}^\top V\bar{x})}\right) \sqrt{\frac{2\pi}{\bar{x}^\top V\bar{x}}} \quad (1.54)$$

$$= B_2 \sqrt{\frac{2\pi}{\bar{x}^\top V\bar{x}}} \exp\left(\frac{\bar{x}^\top \Omega \bar{x}}{\bar{x}^\top V\bar{x}}\right), \quad (1.55)$$

where we defined the bilinear form $\Omega = Vxx^\top V^\top$.⁴

³ This is not specified in the problem, but it seems natural to think that $|\bar{x}(\Theta)|$ is a constant for varying Θ .

⁴ With explicit indices, $\Omega_{im} = V_{ij}x_jx_kV_{km}$.

Let us consider a simple example of this as a sanity check: suppose that x is two-dimensional, and $\bar{x}(\Theta) = (\cos \Theta, \sin \Theta)^\top$; further, suppose that V is diagonal, so that

$$V = \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix}. \quad (1.56)$$

Also, suppose that the observed data parameter is

$$x = A_x \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix}. \quad (1.57)$$

Then, the multiplicative constant in front of the marginalized posterior reads

$$B_2 = B_1 \exp \left(-\frac{1}{2} A_x^2 \left(\frac{\cos^2 \varphi}{\sigma_x^2} + \frac{\sin^2 \varphi}{\sigma_y^2} \right) \right); \quad (1.58)$$

while the bilinear form Ω is

$$\Omega = A_x^2 \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix} \begin{bmatrix} \cos^2 \varphi & \cos \varphi \sin \varphi \\ \cos \varphi \sin \varphi & \sin^2 \varphi \end{bmatrix} \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix} \quad (1.59)$$

$$= A_x^2 \begin{bmatrix} \cos^2 \varphi / \sigma_x^4 & \cos \varphi \sin \varphi / \sigma_x^2 \sigma_y^2 \\ \cos \varphi \sin \varphi / \sigma_x^2 \sigma_y^2 & \sin^2 \varphi / \sigma_y^4 \end{bmatrix}. \quad (1.60)$$

Then, when we evaluate the marginalized posterior we will find something in the form

$$P(\Theta|x) = B_1 \sqrt{2\pi} \left(\frac{\cos^2 \Theta}{\sigma_x^2} + \frac{\sin^2 \Theta}{\sigma_y^2} \right)^{-1/2} \exp \left(A_x^2 F(\Theta, \varphi) \right), \quad (1.61)$$

where $F(\Theta, \varphi)$ is some function whose specific form does not really matter.⁵

The amplitude of the observed data vector, A_x , appears in a rather simple way, as a multiplicative prefactor in the exponent: it can affect the shape of the distribution, but not its mean. Specifically, we can see that scaling A_x is equivalent to scaling σ_x and σ_y simultaneously in the opposite direction — this is rather intuitive, since the angular size of the distribution as seen from the origin is smaller if it is further away.

⁵ For completeness, here is the full expression:

$$\begin{aligned} F(\Theta, \varphi) = & -\frac{1}{2} \left(\frac{\cos^2 \varphi}{\sigma_x^2} + \frac{\sin^2 \varphi}{\sigma_y^2} \right) + \\ & + \left(\frac{\cos^2 \Theta}{\sigma_x^2} + \frac{\sin^2 \Theta}{\sigma_y^2} \right)^{-1} \left[\frac{\cos^2 \Theta \cos^2 \varphi}{\sigma_x^4} + 2 \frac{\cos \Theta \sin \Theta \cos \varphi \sin \varphi}{\sigma_x^2 \sigma_y^2} + \frac{\sin^2 \Theta \sin^2 \varphi}{\sigma_y^4} \right]. \end{aligned} \quad (1.62)$$

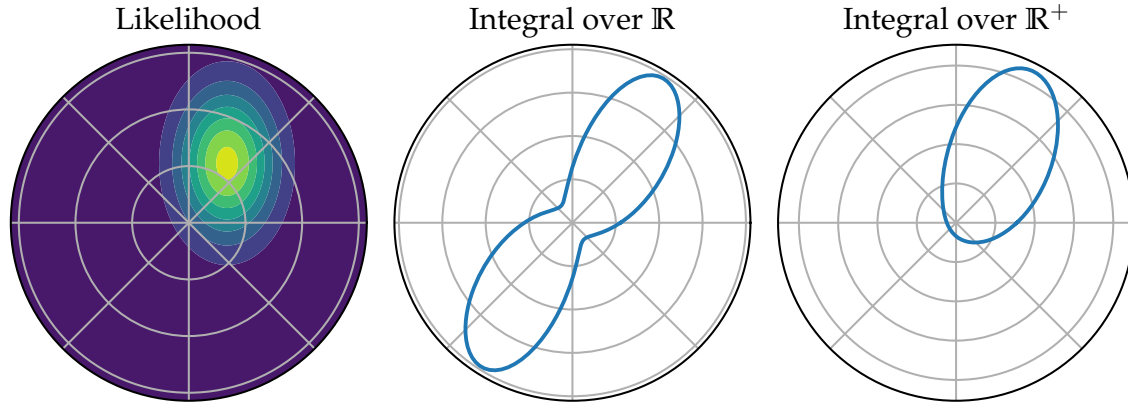


Figure 1: Marginalization: the left plot shows the full likelihood in terms of A and Θ ; the middle plot shows the result of marginalization as shown in the previous calculation (the posterior as a function of Θ); the right plot shows the result of the more physically meaningful marginalization over $A \in (0, +\infty)$ only. Here the likelihood is a diagonal Gaussian with $\sigma_x = 1.2$ and $\sigma_y = 1.8$, centered in $A_x = 2.5$ and $\varphi = 1$ rad.