AstroStatistics and Cosmology Homework

Jacopo Tissino

2020-12-02

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 $\exp\left(-i\vec{k}\cdot\vec{x}\right)$:

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

Claim 1.1. *A multivariate normal distribution*

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

has a characteristic function equal to

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

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

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

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 $V = C^{-1}$:

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

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

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

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 $\mathrm{d}^n x$ to one in $\mathrm{d}^n(x+p)$ does not change the measure, since the Jacobian of a shift is the identity. Then, we have

$$
\begin{aligned}
\phi(\vec{k}) &= N \int \mathrm{d}^n(x+p) \exp\left(-① - ②\right) && (1.12)\\
&= N\sqrt{\frac{(2\pi)^n}{\det V}} \exp\left(-②\right) && (1.13)\\
&= \frac{1}{\underbrace{\sqrt{\det V \det C}}_{=1}} \exp\left(-②\right), && (1.14)
\end{aligned}
$$

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

Now, we only need to simplify ②:

$$
\begin{aligned}
② &= -\frac{1}{2}\left[-\vec{k}^{\top} V^{-1}\vec{k} - 2i\vec{\mu}^{\top} V V^{-1}\vec{k} + \vec{\mu}^{\top} V V^{-1} V\vec{\mu}\right] + \frac{1}{2}\vec{\mu}^{\top} V\vec{\mu} && (1.15)\\
&= \frac{1}{2}\vec{k}^{\top} C\vec{k} + i\vec{\mu}^{\top}\vec{k}, && (1.16)
\end{aligned}
$$

inserting which into the exponent yields the desired result. □

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

$$
\begin{aligned}
&\frac{1}{2}\left(\vec{x} + A^{-1}\vec{b}\right)^{\top} A\left(\vec{x} + A^{-1}\vec{b}\right) - \frac{1}{2}\vec{b}^{\top} A^{-1}\vec{b} = && (1.8)\\
&= \frac{1}{2}\left[\vec{x}^{\top} A\vec{x} + \vec{x}^{\top} A A^{-1}\vec{b} + \left(A^{-1}\vec{b}\right)^{\top} A\vec{x} + \left(A^{-1}\vec{b}\right)^{\top} A A^{-1}\vec{b} - \vec{b}^{\top} A^{-1}\vec{b}\right] && (1.9)\\
&= \frac{1}{2}\left[\vec{x}^{\top} A\vec{x} + \vec{x}^{\top}\vec{b} + \vec{b}^{\top}(A^{-1})^{\top} A\vec{x} + \vec{b}^{\top}(A^{-1})^{\top}\vec{b} - \vec{b}^{\top} A^{-1}\vec{b}\right] && (1.10)\\
&= \frac{1}{2}\vec{x}^{\top} A\vec{x} + \vec{b}^{\top}\vec{x}, && (1.11)
\end{aligned}
$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Exercise 5

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

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

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

$$\begin{aligned}
\mathbb{E}(x_\alpha) &= \left.\frac{\partial\phi(\vec{k})}{\partial(-ik_\alpha)}\right|_{\vec{k}=0} && (1.32)\\
&= \left.\frac{\partial}{\partial(-ik_\alpha)}\right|_{\vec{k}=0}\exp\left(-\frac{1}{2}\vec{k}^\top C\vec{k} - i\vec{k}\cdot\vec{\mu}\right) && (1.33)\\
&= \left[-i\sum_\beta k_\beta C_{\beta\alpha} + \mu_\alpha\right]\exp\left(-\frac{1}{2}\vec{k}^\top C\vec{k} - i\vec{k}\cdot\vec{\mu}\right)\Bigg|_{\vec{k}=0} && (1.34)\\
&= \mu_\alpha\,, && (1.35)
\end{aligned}$$

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

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

The covariance matrix can be computed by linearity as

$$\widetilde{C}_{\alpha\beta} = \mathbb{E}\Big[\big(x_\alpha - \mathbb{E}(x_\alpha)\big)\big(x_\beta - \mathbb{E}(x_\beta)\big)\Big] = \mathbb{E}\big[x_\alpha x_\beta\big] - \mu_\alpha\mu_\beta\,, \tag{1.37}$$

the first term of which reads as follows:

$$\begin{aligned}
\mathbb{E}[x_\alpha x_\beta] &= \left.\frac{\partial^2\phi(\vec{k})}{\partial(-ik_\beta)\partial(-ik_\alpha)}\right|_{\vec{k}=0} && (1.38)\\
&= \left.\frac{\partial}{\partial(-ik_\beta)}\right|_{\vec{k}=0}\left[-i\sum_\beta k_\beta C_{\beta\alpha} + \mu_\alpha\right]\exp\left(-\frac{1}{2}\vec{k}^\top C\vec{k} - i\vec{k}\cdot\vec{\mu}\right) && (1.39)\\
&= C_{\alpha\beta} + \mu_\alpha\mu_\beta\,, && (1.40)
\end{aligned}$$

therefore, as expected, $\widetilde{C}_{\alpha\beta}$ 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)

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

which means that the characteristic function is in the form

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

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. □

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:

$$\mu_\alpha = t_{i\alpha}A_i\,, \tag{1.43}$$

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

$$\mathscr{L}(d_\alpha|A_i) \propto \exp\left(-\frac{1}{2}(d_\alpha - A_i t_{i\alpha})C^{-1}_{\alpha\beta}\left(d_\beta - A_j t_{j\beta}\right)\right). \tag{1.44}$$

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

$$\chi^2 = (d_\alpha - A_i t_{i\alpha})C^{-1}_{\alpha\beta}\left(d_\beta - A_j t_{j\beta}\right). \tag{1.45}$$

We want to minimize this as the amplitudes vary: therefore, we set the derivative with respect to A_k to zero,[3]

$$\frac{\partial\chi^2}{\partial A_k} = -2t_{k\alpha}C^{-1}_{\alpha\beta}\left(d_\beta - A_j t_{j\beta}\right) = 0\,, \tag{1.47}$$

which means that

$$t_{k\alpha}C^{-1}_{\alpha\beta}d_\beta = (t_{k\alpha}C^{-1}_{\alpha\beta}t_{j\beta})A_j\,, \tag{1.48}$$

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

[3] The fact that the stationary point we will find is indeed a minimum can be checked by looking at the second derivative of χ^2:

$$\frac{\partial^2\chi^2}{\partial A_k\partial A_m} = 2t_{k\alpha}C^{-1}_{\alpha\beta}t_{m\beta}\,, \tag{1.46}$$

and recalling that the inverse of the covariance matrix is positive definite.

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

$$(t|C^{-1}|d)_k = (t|C^{-1}|t)_{kj} A_j \tag{1.49}$$

$$\left[(t|C^{-1}|t)^{-1}\right]_{mk} (t|C^{-1}|d)_k = \underbrace{\left[(t|C^{-1}|t)^{-1}\right]_{mk} (t|C^{-1}|t)_{kj}}_{=\delta_{mj}} A_j = A_m \tag{1.50}$$

$$A_m = \left[(t|C^{-1}|t)^{-1}\right]_{mk} (t|C^{-1}|d)_k \,, \tag{1.51}$$

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 $\mu(\Theta, A) = A\overline{x}(\Theta)$, where $\overline{x}$ is a generic function of Θ, while A is our scale parameter.[4] Our likelihood then reads

$$\mathscr{L}(x|\Theta, A) = \underbrace{\frac{1}{(2\pi)^{N/2}\sqrt{\det C}}}_{B_1} \exp\left(-\frac{1}{2}(x - A\overline{x}(\Theta))^\top C^{-1}(x - A\overline{x}(\Theta))\right) . \tag{1.52}$$

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 $\overline{x}$ and defining $V = C^{-1}$ we find

$$P(\Theta|x) = B_1 \int \exp\left(-\frac{1}{2}(x - A\overline{x})^\top V (x - A\overline{x})\right) \mathrm{d}A \tag{1.53}$$

$$= B_1 \int \exp\left(-\frac{1}{2}\left(x^\top V x - 2A\overline{x}^\top V x + A^2 \overline{x}^\top V \overline{x}\right)\right) \mathrm{d}A \,. \tag{1.54}$$

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

$$P(\Theta|x) = \underbrace{B_1 \exp\left(-\frac{1}{2}x^\top V x\right)}_{B_2} \exp\left(\frac{(\overline{x}^\top V x)^2}{(\overline{x}^\top V \overline{x})}\right) \sqrt{\frac{2\pi}{\overline{x}^\top V \overline{x}}} \tag{1.55}$$

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

$$= B_2 \sqrt{\frac{2\pi}{\overline{x}^\top V \overline{x}}} \exp\left(\frac{\overline{x}^\top \Omega \overline{x}}{\overline{x}^\top V \overline{x}}\right), \tag{1.56}$$

where we defined the bilinear form $\Omega = V x x^\top V^\top$.[5]

An application of posterior marginalization in this fashion

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

$$V = \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix}. \tag{1.57}$$

Also, suppose that the observed data parameter is

$$x = A_x \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}. \tag{1.58}$$

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

$$B_2 = B_1 \exp\left(-\frac{1}{2} A_x^2 \left(\frac{\cos^2\varphi}{\sigma_x^2} + \frac{\sin^2\varphi}{\sigma_y^2}\right)\right); \tag{1.59}$$

while the bilinear form Ω is

$$\Omega = A_x^2 \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix} \begin{bmatrix} \cos^2\varphi & \cos\varphi\sin\varphi \\ \cos\varphi\sin\varphi & \sin^2\varphi \end{bmatrix} \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix} \tag{1.60}$$

$$= A_x^2 \begin{bmatrix} \cos^2\varphi/\sigma_x^4 & \cos\varphi\sin\varphi/\sigma_x^2\sigma_y^2 \\ \cos\varphi\sin\varphi/\sigma_x^2\sigma_y^2 & \sin^2\varphi/\sigma_y^4 \end{bmatrix}. \tag{1.61}$$

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

$$P(\Theta|x) = B_1\sqrt{2\pi}\left(\frac{\cos^2\Theta}{\sigma_x^2} + \frac{\sin^2\Theta}{\sigma_y^2}\right)^{-1/2} \exp\left(A_x^2 F(\Theta, \varphi)\right), \tag{1.62}$$

where $F(\Theta, \varphi)$ is some function whose specific form does not really matter.[6]

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] With explicit indices, $\Omega_{im} = V_{ij} x_j x_k V_{km}$.

[6] For completeness, here is the full expression:

$$\begin{aligned} F(\Theta, \varphi) = &-\frac{1}{2}\left(\frac{\cos^2\varphi}{\sigma_x^2} + \frac{\sin^2\varphi}{\sigma_y^2}\right) + \\ &+ \left(\frac{\cos^2\Theta}{\sigma_x^2} + \frac{\sin^2\Theta}{\sigma_y^2}\right)^{-1}\left[\frac{\cos^2\Theta\cos^2\varphi}{\sigma_x^4} + 2\frac{\cos\Theta\sin\Theta\cos\varphi\sin\varphi}{\sigma_x^2\sigma_y^2} + \frac{\sin^2\Theta\sin^2\varphi}{\sigma_y^4}\right]. \end{aligned} \tag{1.63}$$

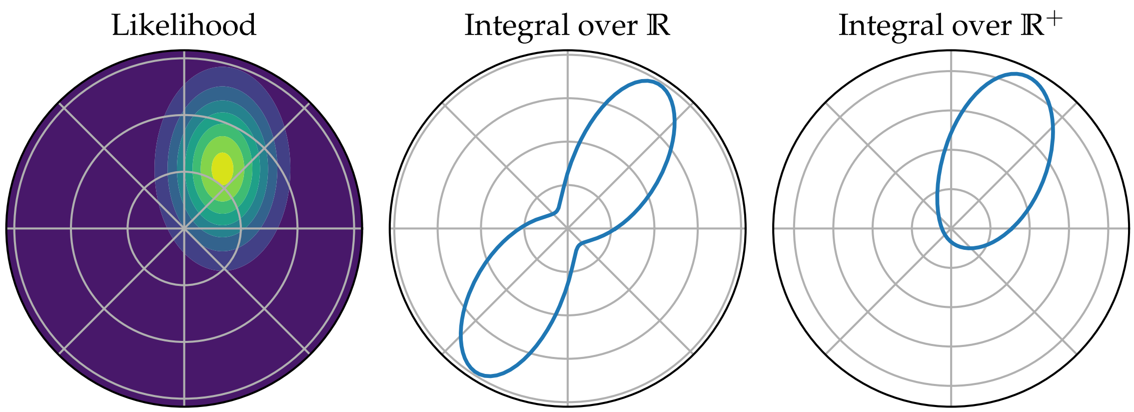

Figure 1: Marginalization: the left plot shows the full likelihood in terms of A and Θ; the middle plot shows the result of marginalization as shown in the previous calculation (the posterior as a function of Θ); the right plot shows the result of the more physically meaningful marginalization over $A \in (0, +\infty)$ 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 $\varphi = 1\,\mathrm{rad}$.

Likelihood marginalization

So far we have considered the posterior $P(\Theta|x)$, the marginalized posterior, a function of the parameter(s) Θ; however we may also be interested in the marginalized likelihood $\mathscr{L}(x|\Theta)$, whose expression is the same as the one we found for $P(\Theta|x)$; further, we do not even need to assume a form for the prior on Θ in order to arrive at that expression. Let us write it in a way which makes the dependence on x more explicit:

$$\mathscr{L}(x|\Theta) = \underbrace{B_1 \sqrt{\frac{2\pi}{\overline{x}^\top V \overline{x}}}}_{B_3} \exp\left(-\frac{1}{2} x^\top V x + \frac{(\overline{x}^\top V x)^2}{\overline{x}^\top V \overline{x}} \right) , \tag{1.64}$$

which can be simplified by making use of the fact that the best-fit template amplitude we found in the last exercise (equation (1.51)) can be applied here, with the single template $t = \overline{x}$, the single amplitude A, the data $d = x$, and the inverse covariance matrix $C^{-1} = V$: the fitting value for A is

$$\hat{A} = \frac{\overline{x}^\top V x}{\overline{x}^\top V \overline{x}}; \tag{1.65}$$

therefore the likelihood is

$$\mathscr{L}(x|\Theta) = B_3 \exp\left(-\frac{1}{2} x^\top V x + \hat{A}\overline{x}^\top V x\right). \tag{1.66}$$

This can be rewritten in the canonical MVN form by making use of the matrix square completion formula (1.8), with $A = -V$ and $\vec{b}^\top = \hat{A}\overline{x}^\top V$:

$$\begin{aligned} -\frac{1}{2} x^\top V x + \hat{A}\overline{x}^\top V x &= -\frac{1}{2}\left(x - V^{-1}\hat{A}(\overline{x}^\top V)^\top\right)^\top V \left(x - V^{-1}\hat{A}(\overline{x}^\top V)^\top\right) \\ &\quad + \frac{1}{2}\hat{A}^2(\overline{x}^\top V)V^{-1}(\overline{x}^\top V)^\top \end{aligned} \tag{1.67}$$

$$= -\frac{1}{2}\left(x - \hat{A}\overline{x}\right)^\top V \left(x - \hat{A}\overline{x}\right) + \frac{1}{2}\hat{A}^2 \overline{x}^\top V \overline{x}. \tag{1.68}$$

Therefore, the marginalized likelihood reads

$$\mathscr{L}(x|\Theta) = B_3 \exp\left(\frac{1}{2}\hat{A}^2\overline{x}^\top V \overline{x}\right) \exp\left(-\frac{1}{2}\left(x - \hat{A}\overline{x}\right)^\top V \left(x - \hat{A}\overline{x}\right)\right). \tag{1.69}$$

We must be careful with this expression: it looks like a multivariate normal in x, however $\hat{A}$ is definitely *not* independent of x, as it is in fact a linear function of it.

A clearer way to see that this is indeed still a MVN is to come back to the original expression (1.64), and to write it as

$$\mathscr{L}(x|\Theta) = B_3 \exp\left(-\frac{1}{2} x^\top \left(V - 2\frac{V\overline{x}\overline{x}^\top V}{\overline{x}^\top V \overline{x}}\right) x\right), \tag{1.70}$$

thus showing that the likelihood is a *zero-mean* MVN with covariance given by

$$\left[V - 2\frac{V\overline{x}\overline{x}^\top V}{\overline{x}^\top V \overline{x}}\right]^{-1}. \tag{1.71}$$

2 December exercises

Exercise 10

We have a time series of N data points, $D = \{d_i\}$, corresponding to the times t_i, which are separated by the constant spacing Δ.

We model them as

$$d_i = \underbrace{B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)}_{f(t_i)} + n_i\,, \tag{2.1}$$

where $f(t)$ the signal we want to characterize, which depends on the unknown amplitudes B_1 and B_2 and the unknown frequency ω; while n_i is the noise: each n_i is i.i.d. as a zero-mean Gaussian with known variance σ^2.

The full likelihood

The likelihood of a single datum of index i attaining the value d_i is given[7] by

$$\mathscr{L}(d_i|\omega, B_1, B_2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(d_i - f(t_i)\right)^2\right). \tag{2.2}$$

Now, since the noise at each point is independent, the full likelihood is the product of the likelihoods of each datum:

$$\begin{aligned}
\mathscr{L}(D|\omega, B_1, B_2) &= \frac{1}{(\sqrt{2\pi}\sigma)^N} \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}\left(d_i - f(t_i)\right)^2\right) & (2.3)\\
&= \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(d_i - f(t_i)\right)^2\right) & (2.4)\\
&= \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{N} \left(d_i - B_1\cos(\omega t_i) - B_2 \sin(\omega t_i)\right)^2}_{Q}\right). & (2.5)
\end{aligned}$$

Let us manipulate the sum in the exponent, which we denote as Q:

$$\begin{aligned}
Q &= \sum_i d_i^2 - 2\sum_i d_i\left(B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)\right) + \sum_i \left(B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)\right)^2 & (2.6)\\
&= N\overline{d}^2 - 2B_1 \underbrace{\sum_i d_i \cos(\omega t_i)}_{R_1(\omega)} - 2B_2 \underbrace{\sum_i d_i \sin(\omega t_i)}_{R_2(\omega)} + \\
&\quad + B_1^2 \underbrace{\sum_i \cos^2(\omega t_i)}_{c} + B_2^2 \underbrace{\sum_i \sin^2(\omega t_i)}_{s} + 2B_1B_2 \sum_i \cos(\omega t_i)\sin(\omega t_i) & (2.7)\\
&= N\overline{d}^2 - 2B_1R_1(\omega) - 2B_2R_2(\omega) + B_1^2 c + B_2 s + B_1B_2 \underbrace{\sum_i \sin(2\omega t_i)}_{h}\,. & (2.8)
\end{aligned}$$

[7] Omitting the dependence on previous information for simplicity.

Large pulsation limit

Typically, in the limit $\omega \gg \Delta^{-1}$ we expect to have $c \approx s \approx N/2$ and $h \approx 0$, since if this the case then after each Δ of time many periods will have passed, therefore each term in the sum c will be equivalent to $\cos^2(x)$ for x uniformly distributed between 0 and 2π, therefore the sum will converge to the N times the mean value of the argument, which is $1/2$ for both $\cos^2$ and $\sin^2$, and 0 for sin.

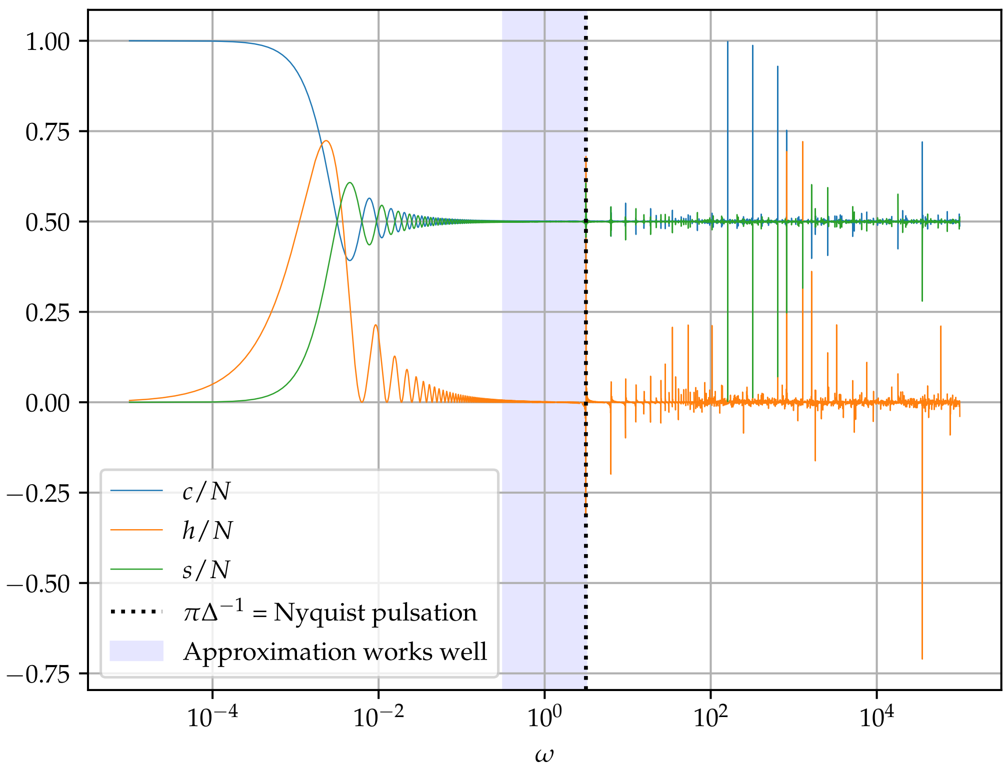

Figure 2: Values of c, s and h for different orders of magnitude of ω.

However, as we can see in figure 2, the three functions do not really *converge* to those values, and stating something like "$\lim_{\omega\to\infty} c = N/2$" would be incorrect mathematically. This is due to the presence of *resonance*: if the ratio $\omega\Delta$ is a rational multiple of π, especially with a small denominator, there will be a bias in the points sampled, resulting in values which may range all the way from 0 to N for c and s, and from $-N$ to N for h. This should not really be an issue in realistic cases, as the set of points for which happens has measure zero.

Really, working in the $\omega \gg \Delta^{-1}$ regime is not wise, since we will necessarily have aliasing in the measured signal, as we are trying to measure a signal well above the Nyquist frequency of our sampler.

Fortunately, there is a regime in the region $\omega \sim \Delta^{-1}$ where the approximation we are discussing works well, and there are no aliasing issues. Let us then assume that we are working in that region, and set $c = s = N/2$ and $h = 0$.

Marginalization

With these simplifications, the likelihood looks like

$$\mathscr{L}(D|\omega, B_1, B_2) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{Q}{2\sigma^2}\right) \tag{2.9}$$

$$Q = N\overline{d}^2 - 2B_1R_1(\omega) - 2B_2R_2(\omega) + B_1^2\frac{N}{2} + B_2\frac{N}{2} \tag{2.10}$$

$$= N\left(\overline{d}^2 + \frac{B_1^2 + B_2^2}{2}\right) - 2B_1R_1(\omega) - 2B_2R_2(\omega)\,. \tag{2.11}$$

The posterior is proportional to the likelihood, since we are assuming the priors on ω and B_i are uniform. We wish to marginalize it over the parameters $B_i \in \mathbb{R}$, for $i = 1, 2$. This amounts to solving the integral

$$P(\omega|D) \propto \int_{\mathbb{R}^2} \mathrm{d}B_1\,\mathrm{d}B_2\, P(\omega, B_1, B_2|D) \tag{2.12}$$

$$\propto \int_{\mathbb{R}^2} \mathrm{d}B_1\,\mathrm{d}B_2 \exp\left(-\frac{N}{2\sigma^2}\left(\underbrace{\overline{d}^2}_{\text{constant}} + \frac{B_1^2 + B_2^2}{2} - 2B_1R_1(\omega) - 2B_2R_2(\omega)\right)\right) \tag{2.13}$$

$$\propto \int_{\mathbb{R}^2} \mathrm{d}B_1\,\mathrm{d}B_2 \exp\left(-\frac{1}{2\sigma^2}\left(\sum_i \frac{NB_i^2}{2} - 2B_iR_i\right)\right) \tag{2.14}$$

$$\propto \prod_i \int_{\mathbb{R}} \mathrm{d}B_i \exp\left(-\frac{NB_i^2}{4\sigma^2} + \frac{B_iR_i}{\sigma^2}\right) \tag{2.15}$$

$$\propto \prod_i \sqrt{\frac{\pi}{N/(4\sigma^2)}} \exp\left(\frac{R_i^2}{\sigma^4}\frac{1}{4}\frac{4\sigma^2}{N}\right) \tag{2.16}$$

$$\propto N^{-1} \prod_i \exp\left(\frac{R_i^2}{\sigma^2 N}\right) \tag{2.17}$$

$$\propto N^{-1} \exp\left(\frac{R_1^2(\omega) + R_2^2(\omega)}{\sigma^2 N}\right). \tag{2.18}$$

In the last step we have used the usual expression for a univariate Gaussian integral (1.29).

Since the exponential is monotonic and we are keeping σ and N constant, the Maximum A-Posteriori (MAP) estimate is given by the maximum of $R_1^2(\omega) + R_2^2(\omega)$.

The periodogram

The periodogram C is defined as

$$C(\omega) = \frac{2}{N}\left|\sum_{k=1}^{N} d_k \exp(-i\omega t_k)\right|^2, \tag{2.19}$$

and while this definition could be applied for an arbitrary set of times t_k, we will only consider it for times $t_k = k\Delta + t_0$ for some t_0: a discrete-time Fourier transform.

We can rewrite the periodogram as

$$\begin{aligned} C(\omega) &= \frac{2}{N}\left|\sum_{k=1}^{N} d_k\big(\cos(\omega t_k) - i\sin(\omega t_k)\big)\right|^2 && (2.20)\\ &= \frac{2}{N}\left[\left(\sum_{k=1}^{N} d_k\cos(\omega t_k)\right)^2 + \left(\sum_{k=1}^{N} d_k\sin(\omega t_k)\right)^2\right] && (2.21)\\ &= \frac{2}{N}\Big[R_1^2(\omega) + R_2^2(\omega)\Big]. && (2.22) \end{aligned}$$

Therefore, the value of ω which maximizes $C(\omega)$ is the same which maximizes $R_1^2(\omega) + R_2^2(\omega)$, which is the MAP estimate.

Least-squares fitting

Least-squares fitting the sinusoid with the same model means we minimize Q.