

Supplemental Materials

Inferring the shape of data: A probabilistic framework for analysing experiments in the natural sciences

Korak Kumar Ray^{1,*}, Anjali R. Verma^{1,*}, Ruben L. Gonzalez, Jr.^{1,†}, and Colin D. Kinz-Thompson^{2,†}

¹*Department of Chemistry, Columbia University, New York, NY 10027 USA*

²*Department of Chemistry, Rutgers University-Newark, Newark, NJ 07102 USA*

*These authors contributed equally to this work.

†To whom correspondence may be addressed: Ruben L. Gonzalez, Jr, Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, Tel: (212) 854-1096, Email: rlg2118@columbia.edu; Colin D. Kinz-Thompson, Department of Chemistry, Rutgers University-Newark, 73 Warren St, Newark, NJ 07102, Tel: (973) 353-0671, Email: colin.kinzthompson@rutgers.edu.

1 Maximum entropy prior distributions

The maximum entropy approach relates informational constraints to specific probability distributions, so it is well suited to deriving the mathematical form of prior probability distributions.¹ Specifically, the maximum entropy prior for any given constraint is the probability distribution which has the maximum information entropy subject to those constraints. For an arbitrary continuous probability distribution $p(\theta)$, the entropy is

$$H = - \int p(\theta|I) \ln \left(\frac{p(\theta|I)}{m(\theta)} \right) d\theta,$$

where θ is a model parameter, I represents our background conditional information, and $m(\theta)$ is the invariant measure.¹ Given any constraint on θ , the above term for entropy may be maximised using the method of Lagrange multipliers to obtain the corresponding maximum entropy prior distribution. All probability distributions are subject to a normalisation constraint, such that

$$\int_{-\infty}^{\infty} p(\theta)d\theta = 1.$$

In this section, we derive several maximum entropy priors that are utilised in the subsequent sections and in the main text of the manuscript. It should be noted that these only form a subset of the possible maximum entropy priors, which can be generated for any given number of constraints.

1.1 Uniform prior

If the only prior information for a parameter θ is that $\theta_{\min} \leq \theta \leq \theta_{\max}$ and there is no reason to favour one value of θ within this range over another, then the invariant measure is $m(\theta) = 1$, and

the corresponding Lagrange function for the prior probability distribution can be written using λ as a Lagrange multiplier as

$$Q = - \int p(\theta|I) \ln(p(\theta|I))d\theta + \lambda \left(1 - \int p(\theta|I)d\theta \right),$$

which, when maximised, yields

$$\lambda = - \ln p(\theta|I).$$

Ensuring the prior is normalised then yields

$$p(\theta|I) = \frac{1}{\theta_{\max} - \theta_{\min}} \equiv \frac{1}{\Delta\theta} \quad \text{for } \theta_{\min} \leq \theta \leq \theta_{\max},$$

which is the continuous uniform distribution.

1.2 Log-uniform prior

A similar situation to that in Section 1.1 is when $\theta_{\min} \leq \theta \leq \theta_{\max}$ but there is no reason to favour one particular magnitude of θ over another (*i.e.*, $\ln \theta$). In this case, the invariant measure is $m(\theta) = \theta^{-1}$, and the corresponding maximum entropy prior is

$$p(\theta|I) = \frac{\theta^{-1}}{\ln \theta_{\max} - \ln \theta_{\min}} \equiv \frac{\theta^{-1}}{\Delta \ln \theta} \quad \text{for } \theta_{\min} \leq \theta \leq \theta_{\max}.$$

In this situation, the prior is uniform on a logarithmic scale, and is useful for scale parameters such as variance or precision.

1.3 Gaussian prior

For a situation where the expectation value and the variance of the parameter θ are known to be μ and τ^{-1} , respectively, we have the constraints that

$$\begin{aligned} \mathbb{E}[\theta] &= \int_{-\infty}^{\infty} \theta p(\theta|I) d\theta \equiv \mu, \text{ and} \\ \mathbb{E}[\theta^2 - \mathbb{E}[\theta]^2] &= \mathbb{E}[\theta^2] - \mu^2 = \int_{-\infty}^{\infty} \theta^2 p(\theta|I) d\theta - \mu^2 \equiv \frac{1}{\tau}. \end{aligned}$$

To derive the maximum entropy prior for this situation, when the invariant measure $m(\theta) = 1$, we write the Lagrange function with Lagrange multipliers α , β , and γ as

$$\begin{aligned} Q = - \int_{-\infty}^{\infty} p(\theta|I) \ln(p(\theta|I))d\theta + \alpha \left(1 - \int_{-\infty}^{\infty} p(\theta|I)d\theta \right) + \beta \left(\mu - \int_{-\infty}^{\infty} \theta p(\theta|I)d\theta \right) \\ + \gamma \left((\mu^2 + \tau^{-1}) - \int_{-\infty}^{\infty} \theta^2 p(\theta|I)d\theta \right). \end{aligned}$$

When maximised, this yields

$$p(\theta|I) = \frac{\tau}{\sqrt{2\pi}} e^{-\frac{\tau}{2}(\theta-\mu)^2},$$

which is a Normal or Gaussian distribution. Thus, the use of a Gaussian prior represents previous information on the mean and the variance of the parameter.

1.4 Gamma prior

For a situation where the expectations of the parameter and the logarithm of the parameter are known to be μ and L , respectively, we have the constraints that

$$\begin{aligned}\mathbb{E}[\theta] &= \int \theta p(\theta|I) d\theta = \mu, \\ \mathbb{E}[\ln \theta] &= \int \ln \theta p(\theta|I) d\theta = L,\end{aligned}$$

and, given the logarithmic function, that $\theta > 0$. To derive the maximum entropy prior for this situation, we write the Lagrange function with Lagrange multipliers α , β , and γ as

$$\begin{aligned}Q &= - \int_0^\infty p(\theta|I) \ln(p(\theta|I)) d\theta + \alpha \left(1 - \int_0^\infty p(\theta|I) d\theta \right) + \beta \left(\mu - \int_0^\infty \theta p(\theta|I) d\theta \right) \\ &\quad + \gamma \left(L - \int_0^\infty \ln \theta p(\theta|I) d\theta \right).\end{aligned}$$

Upon maximisation, this yields the prior

$$p(\theta|I) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta},$$

where $\Gamma(\alpha)$ denotes the gamma function. This probability distribution is known as the gamma distribution.

1.5 A joint prior for modeling shapes

In the background information used in this work to calculate the shape of data, we have assumed that the three parameters in the main text of the manuscript (m , b , and τ) are independent of each other. This assumption yields the joint prior probability distribution

$$p(m, b, \tau|M_0) = p(m|M_0)p(b|M_0)p(\tau|M_0),$$

where the conditional M_0 represents our background knowledge for the inference problem. For m , we make assumptions that lead us to use either a uniform, Gaussian, or gamma maximum entropy prior, which are

$$\begin{aligned}p(m|M_0) &= \frac{1}{\Delta m}, \\ p(m|M_0) &= \frac{\Lambda_m}{\sqrt{2\pi}} e^{-\frac{\Lambda_m}{2}(m-\mu_m)^2}, \quad \text{or} \\ p(m|M_0) &= \frac{\beta_m^{\alpha_m}}{\Gamma(\alpha_m)} m^{\alpha_m-1} e^{-\beta_m m}.\end{aligned}$$

For b , we have only made assumptions that lead us to use a uniform maximum entropy prior, which is

$$p(b|M_0) = \frac{1}{\Delta b}.$$

For τ , we have made the assumption that we do not know the magnitude of the noise and thus use a log-uniform maximum entropy prior, which is

$$p(\tau|M_0) = \frac{\tau^{-1}}{\Delta \ln \tau}.$$

2 Evidence expressions for calculating the shape of data

2.1 Definitions

In the derivations presented in this section, we use the following terms and definitions:

$$\begin{aligned}
 N & : \text{Number of data points} \\
 M & \equiv \frac{N}{2} - 1 \\
 \langle \mathbf{x} \rangle & : \text{Average of } \mathbf{x} \left(\equiv \frac{1}{N} \sum_{i=1}^N x_i \right) \\
 V_x & \equiv \langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2 \\
 V_y & \equiv \langle \mathbf{y}^2 \rangle - \langle \mathbf{y} \rangle^2 \\
 C & \equiv \langle \mathbf{x}\mathbf{y} \rangle - \langle \mathbf{x} \rangle \langle \mathbf{y} \rangle \\
 r & \equiv \frac{C}{\sqrt{V_x V_y}} \\
 \text{erf}(\alpha) & : \text{Error function} \\
 B(\alpha, \beta) & : \text{Beta function} \\
 I_\nu(\alpha, \beta) & : \text{Regularised incomplete Beta function} \\
 {}_2F_1 \left(\begin{matrix} \alpha, \beta \\ \gamma \end{matrix}; z \right) & : \text{Gaussian hypergeometric function}
 \end{aligned}$$

2.2 Gaussian likelihood

The evidence functions in the following sub-sections are all computed assuming in M_0 that the likelihood function is a Gaussian distribution (*i.e.*, white noise). Specifically, this likelihood function for the data, \mathbf{y} , with the template, \mathbf{x} , is

$$p(\mathbf{y}|\mathbf{x}, m, b, \tau, M_0) = \prod_{i=1}^N \frac{\tau}{\sqrt{2\pi}} e^{-\frac{\tau}{2}(y_i - mx_i - b)^2}.$$

The integrals are computed using standard tabulated forms.^{2,3}

2.2.1 Uniform priors for \mathbf{m} and \mathbf{b} , log-uniform prior for τ ; $\mathbf{m} \in \mathbb{R}, \mathbf{b} \in \mathbb{R}, \tau > 0$

Intermediate integration steps are shown for this section. All other cases are computed using the same order of integration (*i.e.*, first b , then m , then τ).

$$\begin{aligned}
 p(\mathbf{y}|\mathbf{x}, m, \tau, M_0) &= \int_{-\infty}^{\infty} db p(\mathbf{y}|\mathbf{x}, m, b, \tau, M_0) p(b|M_0) \\
 &= \int_{-\infty}^{\infty} db \frac{\tau^{N/2}}{(2\pi)^{N/2} \Delta b} e^{-\sum_{i=1}^N \frac{(y_i - mx_i - b)^2 \tau}{2}} \\
 &= \frac{\tau^{(N-1)/2}}{(2\pi)^{(N-1)/2} \Delta b \sqrt{N}} e^{-\frac{N\tau}{2}(V_x + m^2 V_y - 2mC)}
 \end{aligned}$$

$$\begin{aligned}
 p(\mathbf{y}|\mathbf{x}, \tau, M_0) &= \int_{-\infty}^{\infty} dm p(\mathbf{y}|\mathbf{x}, m, \tau, M_0)p(m|M_0) \\
 &= \int_{-\infty}^{\infty} dm \frac{\tau^{(N-1)/2}}{(2\pi)^{(N-1)/2} \Delta b \Delta m \sqrt{N}} e^{-\frac{N\tau}{2}(V_x + m^2 V_y - 2mC)} \\
 &= \frac{\tau^M}{N(2\pi)^M \Delta b \Delta m \sqrt{V_x}} e^{-\frac{N\tau}{2} V_y (1-r^2)}
 \end{aligned}$$

$$\begin{aligned}
 P(\mathbf{y}|\mathbf{x}, M_0) &= \int_0^{\infty} d\tau p(\mathbf{y}|\mathbf{x}, \tau, M_0)p(\tau|M_0) \\
 &= \int_0^{\infty} d\tau \frac{\tau^{M-1}}{N(2\pi)^M \Delta b \Delta m \Delta \ln \tau \sqrt{V_x}} e^{-\frac{N\tau}{2} V_y (1-r^2)} \\
 &= \frac{\Gamma(M) N^{-\frac{N}{2}} V_x^{-\frac{1}{2}}}{\Delta b \Delta m \Delta \ln \tau} [\pi V_y (1-r^2)]^{-M}.
 \end{aligned}$$

2.2.2 Uniform priors for \mathbf{m} and \mathbf{b} , log-uniform prior for τ ; $\mathbf{m} > \mathbf{0}$, $\mathbf{b} \in \mathbb{R}$, $\tau > \mathbf{0}$

This integral is used in the main text of the manuscript and in Section 3. From above,

$$p(\mathbf{y}|\mathbf{x}, m, \tau, M_0) = \frac{\tau^{(N-1)/2}}{(2\pi)^{(N-1)/2} \Delta b \sqrt{N}} e^{-\frac{N\tau}{2}(V_x + m^2 V_y - 2mC)}$$

Further,

$$\begin{aligned}
 p(\mathbf{y}|\mathbf{x}, \tau) &= \int_0^{\infty} dm p(\mathbf{y}|\mathbf{x}, m, \tau, M_0)p(m|M_0) \\
 &= \frac{\tau^{(N-1)/2}}{(2\pi)^M \Delta b \Delta m N \sqrt{V_x}} e^{-\frac{N\tau}{2} V_y (1-r^2)} \times \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{C \sqrt{N\tau}}{\sqrt{2V_x}} \right) \right]
 \end{aligned}$$

$$\begin{aligned}
 P(\mathbf{y}|\mathbf{x}, M_0) &= \int_0^{\infty} d\tau p(\mathbf{y}|\mathbf{x}, \tau, M_0)p(\tau|M_0) \\
 &= \frac{\Gamma(M) N^{-\frac{N}{2}} V_x^{-\frac{1}{2}}}{\Delta m \Delta b \Delta \ln \tau} [\pi V_y (1-r^2)]^{-M} \times \\
 &\quad \left[\frac{1}{2} + r(1-r^2)^{-1/2} B \left(\frac{1}{2}, M \right) {}_2F_1 \left(\frac{1}{2}, M + \frac{1}{2}; \frac{3}{2}; \frac{-r^2}{1-r^2} \right) \right] \\
 &= \frac{\Gamma(M) N^{-\frac{N}{2}} V_x^{-\frac{1}{2}}}{2 \Delta m \Delta b \Delta \ln \tau} [\pi V_y (1-r^2)]^{-M} \left[1 + \frac{r}{|r|} I_{r^2} \left(\frac{1}{2}, M \right) \right].
 \end{aligned}$$

2.2.3 Uniform priors for \mathbf{m} and \mathbf{b} , log-uniform for τ ; $\mathbf{m} > \mathbf{0}$, $\mathbf{b} > \mathbf{0}$, $\tau > \mathbf{0}$

While this condition is the most ‘realistic’ model for some applications, a closed form for this integral, to the best of our knowledge, unfortunately does not exist. Using different priors did not

yield a closed form either. If required, this integral may be numerically calculated (*e.g.*, using a triple quadrature method), however, this is much more computationally expensive than using an analytical form.

$$P(\mathbf{y}|\mathbf{x}, M_0) = \int_0^\infty \int_0^\infty \int_0^\infty d\tau \, dm \, db \frac{\tau^{-1}}{\Delta b \Delta m \Delta \ln \tau} \left(\frac{\tau}{2\pi}\right)^{\frac{N}{2}} e^{-\frac{\tau}{2} \sum_{i=1}^N (y_i - mx_i - b)^2}.$$

2.2.4 Uniform prior for \mathbf{b} , log-uniform prior for τ ; $\mathbf{m} = \mathbf{0}, \mathbf{b} \in \mathbb{R}, \tau > \mathbf{0}$

This evidence function is for the case where there is no template, \mathbf{x} , since the condition $m = 0$ removes any dependence on \mathbf{x} . Thus, this evidence represents a ‘flat’ shape composed only of background and noise. This integral is used in the main text of the manuscript and in Section 3.

$$P(\mathbf{y}|\mathbf{x}_{null}, M_0) = \frac{\Gamma(M + \frac{1}{2})N^{-\frac{N}{2}}}{\Delta b \Delta \ln \tau} (\pi V_y)^{-(M+\frac{1}{2})}.$$

2.2.5 Uniform prior for \mathbf{b} , log-uniform prior for τ ; $\mathbf{m} = \mathbf{1}, \mathbf{b} \in \mathbb{R}, \tau > \mathbf{0}$

$$P(\mathbf{y}|\mathbf{x}, M_0) = \frac{\Gamma(M + \frac{1}{2})N^{-\frac{N}{2}}}{\Delta b \Delta \ln \tau} [\pi (V_y + V_x - 2C)]^{-(M+\frac{1}{2})}.$$

2.2.6 Uniform prior for \mathbf{m} , log-uniform prior for τ ; $\mathbf{m} \in \mathbb{R}, \mathbf{b} = \mathbf{0}, \tau > \mathbf{0}$

$$P(\mathbf{y}|\mathbf{x}, M_0) = \frac{\Gamma(M + \frac{1}{2})N^{-\frac{N}{2}} \langle x^2 \rangle^{-\frac{1}{2}}}{\Delta m \Delta \ln \tau} \left[\pi \left(\langle y^2 \rangle - \frac{\langle xy \rangle^2}{\langle x^2 \rangle} \right) \right]^{-(M+\frac{1}{2})}.$$

2.2.7 Log-uniform prior for τ ; $\mathbf{m} = \mathbf{1}, \mathbf{b} = \mathbf{0}, \tau > \mathbf{0}$

$$P(\mathbf{y}|\mathbf{x}, M_0) = \frac{\Gamma(\frac{N}{2})}{\Delta \ln \tau} [\pi N (\langle y^2 \rangle + \langle x^2 \rangle - 2 \langle xy \rangle)]^{-\frac{N}{2}}.$$

2.2.8 Gaussian prior for \mathbf{m} , uniform for \mathbf{b} , log-uniform for τ ; $\mathbf{m} \in \mathbb{R}, \mathbf{b} \in \mathbb{R}, \tau > \mathbf{0}$

To the best of our knowledge, the closed form of this integral only exists under the assumption $NV_x \gg \Lambda_m$, in which case, the approximate integral is given below. While computationally more expensive, it is recommended that the value of the integral is computed numerically instead.

$$\begin{aligned} P(\mathbf{y}|\mathbf{x}, M_0) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty d\tau \, dm \, db \frac{\tau^{\frac{N}{2}-1} \Lambda_m e^{-\frac{\Lambda_m}{2}(m-\mu_m)^2}}{(\sqrt{2\pi})^{\frac{N+1}{2}} \Delta b \Delta \ln \tau} e^{-\frac{\tau}{2} \sum_{i=1}^N (y_i - mx_i - b)^2} \\ &\approx \frac{\Gamma(M)N^{-\frac{N}{2}}}{\Delta b \Delta \ln \tau} \left(\frac{\Lambda_m}{2\pi V_x} \right)^{\frac{1}{2}} e^{-\frac{\Lambda_m}{2}(\mu_m - \frac{C}{V_x})^2} [\pi V_y (1 - r^2)]^{-M}. \end{aligned}$$

2.2.9 Gaussian prior for \mathbf{m} , log-uniform prior for τ ; $\mathbf{m} \in \mathbb{R}, \mathbf{b} = \mathbf{0}, \tau > \mathbf{0}$

To the best of our knowledge, the closed form of this integral only exists under the assumption $N \langle x^2 \rangle \gg \Lambda_m$, in which case, the approximate integral is given below. While computationally more

expensive, it is recommended that the value of the integral is computed numerically instead.

$$\begin{aligned}
 P(\mathbf{y}|\mathbf{x}, M_0) &= \int_0^\infty \int_{-\infty}^\infty d\tau \, dm \frac{\tau^{\frac{N}{2}-1} \Lambda_m e^{-\frac{\Lambda_m}{2}(m-\mu_m)^2}}{(\sqrt{2\pi})^{\frac{N+1}{2}} \Delta b \Delta \ln \tau} e^{-\frac{\tau}{2} \sum_{i=1}^N (y_i - mx_i - b)^2} \\
 &\approx \frac{\Gamma(M + \frac{1}{2}) N^{-\frac{N}{2}}}{\Delta \ln \tau} \left(\frac{\Lambda_m}{2\pi \langle x^2 \rangle} \right)^{\frac{1}{2}} e^{-\frac{\Lambda_m}{2} \left(\mu_m - \frac{\langle xy \rangle}{\langle x^2 \rangle} \right)^2} \left[\pi \left(\langle y^2 \rangle - \frac{\langle xy \rangle^2}{\langle x^2 \rangle} \right) \right]^{-(M + \frac{1}{2})}.
 \end{aligned}$$

2.2.10 Gamma prior for \mathbf{m} , uniform for \mathbf{b} , log-uniform for τ ; $\mathbf{m} > \mathbf{0}$, $\mathbf{b} \in \mathbb{R}$, $\tau > \mathbf{0}$

To the best of our knowledge, the closed form of this integral only exists under the assumption $NV_x \gg \beta_m$, in which case, the approximate integral is given below. While computationally more expensive, it is recommended that the value of the integral is computed numerically instead.

$$\begin{aligned}
 P(\mathbf{y}|\mathbf{x}, M_0) &= \int_0^\infty \int_0^\infty \int_{-\infty}^\infty d\tau \, dm \, db \frac{\tau^{\frac{N}{2}-1} \beta_m^{\alpha_m} m^{\alpha_m-1} e^{-\beta_m m}}{(\sqrt{2\pi})^{\frac{N}{2}} \Gamma(\alpha_m) \Delta b \Delta \ln \tau} e^{-\frac{\tau}{2} \sum_{i=1}^N (y_i - mx_i - b)^2} \\
 &\approx \frac{\beta_m^{\alpha_m} V_x^{-\frac{\alpha_m}{2}} V_y^{-\frac{N-\alpha_m-1}{2}} N^{-\frac{N}{2}}}{(\pi)^{\frac{N-1}{2}} \Gamma(\alpha_m) \Delta b \Delta \ln \tau} \left\{ \frac{1}{2} \Gamma\left(\frac{N - (\alpha_m + 1)}{2}\right) \Gamma\left(\frac{\alpha_m}{2}\right) {}_2F_1\left(\frac{\alpha_m}{2}, \frac{N - (\alpha_m + 1)}{2}; \frac{1}{2}; r^2\right) \right. \\
 &\quad \left. + r \Gamma\left(\frac{N - \alpha_m}{2}\right) \Gamma\left(\frac{\alpha_m + 1}{2}\right) {}_2F_1\left(\frac{\alpha_m + 1}{2}, \frac{N - \alpha_m}{2}; \frac{3}{2}; r^2\right) \right\}.
 \end{aligned}$$

2.3 Non-Gaussian Likelihoods

In Section 2.2, we report evidence functions corresponding to models, M_0 , where the likelihood is a Gaussian (*i.e.*, for models with white noise). In other situations, the evidence for a model with a different likelihood (*e.g.*, M'_0) may be of interest. In particular, in the case of Poisson noise, the likelihood function would be

$$p(\mathbf{y}|\mathbf{x}, m, b, M'_0) = \prod_{i=1}^N \frac{(mx_i + b)^{y_i} e^{-(mx_i + b)}}{\Gamma(y_i)}.$$

As in Section 2.2.3, the analytical form of the corresponding evidence function proved difficult to obtain by integrating the joint probability $P(\mathbf{y}, m, b|\mathbf{x}, M'_0)$. If required, it can be numerically integrated (*e.g.*, using a double quadrature method).

3 Simplifications for model selection

The posterior probability for the k^{th} template, \mathbf{x}_k , given the dataset, \mathbf{y} , is

$$\begin{aligned}
 P(\mathbf{x}_k|\mathbf{y}, M_0) &= \frac{P(\mathbf{y}|\mathbf{x}_k, M_0) P(\mathbf{x}_k|M_0)}{\sum_{i=1}^K P(\mathbf{y}|\mathbf{x}_i, M_0) P(\mathbf{x}_i|M_0)} \\
 &= \frac{1}{\sum_{i=1}^K \frac{P(\mathbf{y}|\mathbf{x}_i, M_0) P(\mathbf{x}_i|M_0)}{P(\mathbf{y}|\mathbf{x}_k, M_0) P(\mathbf{x}_k|M_0)}} \\
 &\equiv \frac{1}{\sum_{i=1}^K R_{i,k} S_{i,k}},
 \end{aligned}$$

where $R_{i,k}$ denotes the ratio of the evidences for \mathbf{x}_i over \mathbf{x}_k , and $S_{i,k}$ denotes the ratio of the model priors for \mathbf{x}_i over \mathbf{x}_k . The former ratio, $R_{i,k}$, is known as a Bayes factor. In the following sections, we provide some simplified formulas for Bayes factors in which many terms have been canceled. We also note that it is possible that a significant amount of the approximation error in the approximate evidence functions given in Sections 2.2.8 - 2.2.10 cancel in such ratios.

3.1 Identifying a shape from the background (\mathbf{x}_k vs. \mathbf{x}_{null})

Consider calculating the shape of \mathbf{y} , where at least one model is given by the template \mathbf{x}_k , and the likelihood and prior assumptions used in Section 2.2.2, and where another of the models is the absence of a template (*i.e.*, a flat background) as given by \mathbf{x}_{null} in Section 2.2.4. In this situation, calculating the probability of \mathbf{x}_k involves the Bayes factor

$$R_{\mathbf{x}_{null}, \mathbf{x}_k} = \frac{2\Delta m}{B(\frac{1}{2}, M)} \left(\frac{V_{\mathbf{x}_k}}{V_{\mathbf{y}}} \right)^{\frac{1}{2}} \frac{(1 - r_k^2)^M}{\left(1 + \frac{r_k}{|r_k|} I_{r_k^2}(\frac{1}{2}, M) \right)},$$

where subscript k denotes terms related to \mathbf{x}_k .

3.2 Selecting between two templates (\mathbf{x}_k vs. \mathbf{x}_i)

Consider calculating the shape of \mathbf{y} , when at least two of the models use templates, \mathbf{x}_k and \mathbf{x}_i , and both use the likelihood and prior assumptions from Section 2.2.2. In this situation, calculating the probability of \mathbf{x}_k involves the Bayes factor

$$R_{\mathbf{x}_i, \mathbf{x}_k} = \left(\frac{V_{\mathbf{x}_k}}{V_{\mathbf{x}_i}} \right)^{1/2} \left(\frac{1 - r_k^2}{1 - r_i^2} \right)^M \frac{\left(1 + \frac{r_i}{|r_i|} I_{r_i^2}(\frac{1}{2}, M) \right)}{\left(1 + \frac{r_k}{|r_k|} I_{r_k^2}(\frac{1}{2}, M) \right)},$$

where the subscripts denote the specified template.

References

- ¹E. T. Jaynes, *Probability theory: the logic of science* (Cambridge University Press, Cambridge, UK; New York, NY, 2003).
- ²I. S. Gradshteyn, I. M. Ryzhik, D. Zwillinger, and V. Moll, *Table of integrals, series, and products; 8th ed.* (Academic Press, Amsterdam, 2014).
- ³E. W. Ng and M. Geller, “A table of integrals of the error functions”, *Journal of Research of the National Bureau of Standards, Section B: Mathematical Sciences* **73B**, 1–20 (1969).