

Chapter 4

Computational Analysis of Behavioural and Neural Data Through Bayesian Statistical Modelling

Raymond Wai Mun Chan and Fabrizio Gabbiani

Abstract Computational analysis of behavioural and neural data is nowadays an essential part of neuroethology, allowing an ever deeper understanding of how natural behaviour and neural activity are interrelated at the molecular, cellular and network level. The range of computational techniques applied in neuroethological research is currently so broad as to preclude an exhaustive survey in a succinct chapter. Here, we focus on a specific approach termed Bayesian statistical modelling that has proven to be a powerful method for relating neural activity to natural behavioural performance. As we illustrate in a specific example, this approach naturally dovetails with classic neural coding concepts such as population vector codes. It is also flexible enough to be applicable to a broad range of neuroethological questions.

Keywords Barn owl • Bayesian models • Interaural time difference • Maximum likelihood • Neural correlations • Population codes • Population vector • Sound localization

R.W.M. Chan
Department of Neuroscience, Baylor College of Medicine,
One Baylor Plaza, Houston, TX 77030, USA
e-mail: rchan@cns.bcm.edu

F. Gabbiani (✉)
Department of Neuroscience, Baylor College of Medicine,
One Baylor Plaza, Houston, TX 77030, USA

Department of Computational and Applied Mathematics,
Rice University, 6100 Main Street, Houston, TX 77005, USA
e-mail: gabbiani@bcm.edu

4.1 Introduction

During the past decades, neuroscience and neuroethology have experienced a dramatic increase in the availability of methods to analyse neural data. Yet, computational data analysis has long been an integral part of this area of research, as attested by many historical examples. Hodgkin and Huxley, for instance, used numerical integration of differential equations to study the propagation of action potentials in the giant squid axon. In another classical study, Katz and colleagues applied probability theory to derive the properties of quantal synaptic release at the neuromuscular junction. Recent progress in neural modelling has been in large part fueled by an exponential increase in computing power, the widespread availability of powerful numerical and simulation packages, such as MATLAB and NEURON, as well as the need to cope with increasingly complex neural data sets spanning multiple spatio-temporal scales. The interested reader will find a comprehensive treatment of modelling techniques in Gabbiani and Cox [2010], including many worked-out numerical and programming examples.

In this chapter, we focus on a specific topic that has attracted renewed attention and that is pertinent to neuroethology: Bayesian statistical modelling. The Bayesian framework allows the computational analysis of neural data in the context of the animal's environment using rigorous mathematical methods. In the following sections, we start with a brief introduction to Bayesian modelling before illustrating its use to analyze the neural coding of natural sounds in the barn owl. The figures of this chapter were generated using short MATLAB programs that are available online and will help the reader assimilate the material covered. The name of these programs is specified at the end of the figure legends using the notation: (name.m).

4.2 Bayesian Statistical Modelling

The Bayesian approach to statistics interprets probabilities as measures of belief instead of empirical frequencies for event occurrence (Doya et al. 2007, Hoff 2009). This framework centred on belief allows one to model decision making in a principled manner by (1) taking into account the sensory input experienced by an organism, (2) integrating previous information (e.g. memories or biases) and (3) deciding on an appropriate motor output, based on this information.

In the Bayesian framework, the state of the outside world may be conceived as a model indexed by a variable θ . In general, the variable θ will be multidimensional. The main task of the organism is to infer from sensory data, d , an estimate of the current state of the world, $\hat{\theta}$, so as to react with an appropriate motor output. The sensory data could for instance be the firing rate of sensory neurons activated in the current state of the world. Because the transduction of external stimuli into neural signals is noisy, due to both intrinsic and extraneous variability, and because the processing of neural signals is noisy as well, the sensory data d will usually be a random variable determined by θ and characterized by the conditional distribution

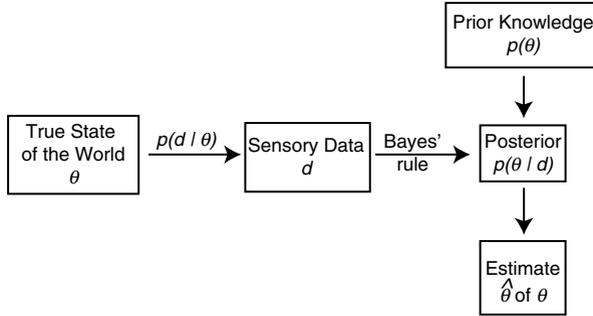


Fig. 4.1 Bayesian statistical models. In the Bayesian scheme, the observer gathers data from the outside world which is modelled based on parameters represented by θ . The data, d , is combined with prior knowledge on the parameters, θ , to infer a posterior probability distribution for the parameters using Bayes’ rule [Eq. (4.1)]. This in turn allows to determine an estimate for the model parameters

$p(d|\theta)$. The main function on which decisions are based is the posterior distribution $p(\theta|d)$ which gives the conditional probability of the model parameter θ given the observed data d . To calculate this posterior distribution, we use Bayes’ rule

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}, \quad (4.1)$$

where $p(d|\theta)$ is the conditional probability of the data given the model parameter, θ (Fig. 4.1).

In the field of statistics, the probability distribution $p(d|\theta)$ is also called a generative model since it maps outside stimuli into sensory neural responses (in our context). The probability distribution of the model parameter θ , $p(\theta)$, is the prior distribution that is related to properties of the outside world. The experimenter will often be able to manipulate these priors. The distribution of responses irrespective of the stimuli (or of the current state of the world) is called the marginal distribution, $p(d)$. This marginal distribution normalizes Eq. (4.1) such that the integral of the posterior distribution over θ is unity. When we fix the data d and let θ vary, the generative model $p(d|\theta)$ becomes what is known as the likelihood function in statistics. One conventional method of estimating θ consists in selecting the value, $\hat{\theta}$, that maximizes the likelihood given the data. This decision rule is called “maximum likelihood”. The analogous principle in the Bayesian case consists in selecting the maximum of the posterior distribution, or “maximum a posteriori” (MAP) estimate. Alternatively, another valid rule consists in computing the mean of the posterior distribution. The use of these decision rules will be illustrated in the following sections.

Often, knowing $p(d)$ is not necessary as we only need to know the dependence of the posterior on the model parameter, and $p(d)$ only acts as a normalizing constant. This is exploited by computing the product of the likelihood function and the prior distribution and ignoring the marginal distribution,

$$p(\theta | d) \propto p(d | \theta)p(\theta), \quad (4.2)$$

since $p(d)$ does not depend on θ . This last equation also makes clear that the Bayesian framework uses both the likelihood and the prior distribution of θ to arrive at an informed estimate $\hat{\theta}$. To render these general remarks more concrete, we turn to the example of sound localization in barn owls as recently described in Fischer and Peña [2011].

4.3 Sound Localization in Barn Owls

In the wild, owls use sound localization to detect and locate prey in the dead of night. Psychophysically, the time lag between a sound picked up in each ear but generated by a single source allows the owl to reconstruct the horizontal direction (or azimuth) to the source (Fig. 4.2a). A second and distinct cue, the interaural level difference allows the owl to reconstruct the elevation of the source but will not be considered further here (Konishi 2003). The time lag between sound arrival at both ears is called the interaural time difference (ITD), and is related to the azimuth direction of the sound source as shown in Fig. 4.2b. The horizontal axis shows the source direction centred on the owl's sagittal plane, while the vertical axis shows the corresponding ITD. This relationship is obtained by fitting the function

$$\text{ITD}(\theta) = A \sin(\omega\theta) \quad (4.3)$$

to head related transfer function data as a function of the source angle θ . Such a fit yields $A=260\mu\text{s}$ and $\omega=0.0143\text{rad}/^\circ$ as the fitted parameters (Fischer and Peña 2011). From this graph, one notices immediately that the inverse mapping from ITD to source direction is not always one to one. Hence, the owl must somehow pick one of the possible states of the world consistent with the observed ITD. Ethologically, we know that it does so by biasing its choice to the one straight ahead (Hausmann et al. 2009, Knudsen et al. 1979). This bias can be quantified using Bayesian statistics.

We begin by asking what knowledge of the world the owl already has and what it wishes to know. In the sound localization problem, it knows approximately (see below) the ITD of the source, but wishes to know its associated direction θ . In Bayesian terms, we say that the owl wishes to infer the probability of θ given that it knows the ITD, or equivalently, the probability distribution of θ given the ITD, $p(\theta | \text{ITD})$. This is exactly the posterior distribution in Eq. (4.1) with ITD replacing d . In order to use Bayes' rule, we need a generative model and a prior distribution.

The sound reaching each ear may be corrupted by noise in the environment, like that caused by wind; in addition, the neural computation of ITD is noisy as well. Thus, we use

$$\text{ITD}(\theta) = A \sin(\omega\theta) + W, \quad (4.4)$$

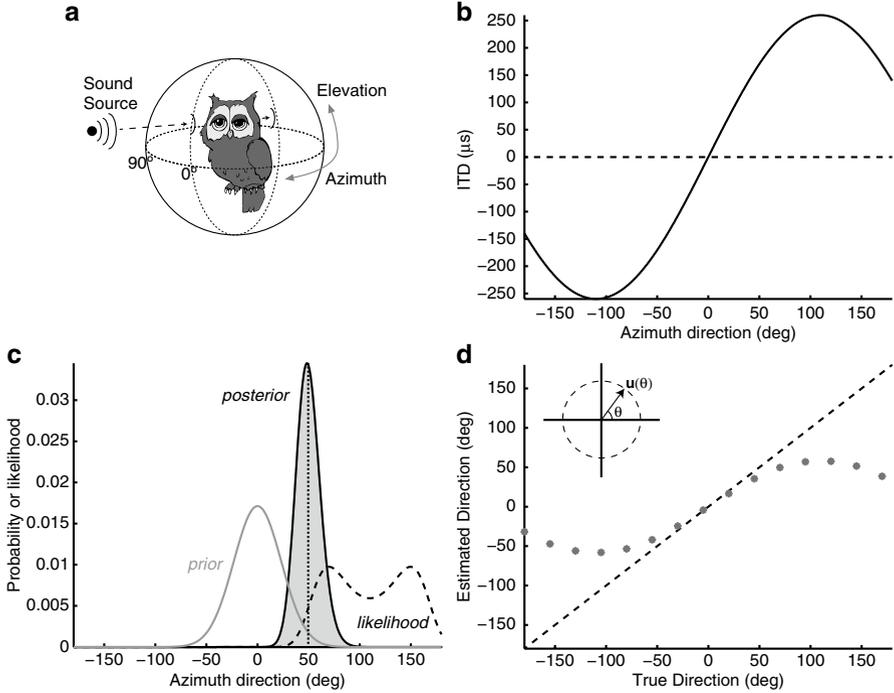


Fig. 4.2 Bayesian estimation of sound source. (a) Schematic illustration of the coordinate system used to describe sound direction, characterized by azimuth and elevation angles. A sound source with a horizontal direction or azimuth of 0° lies straight in front of the animal. A sound source with an azimuth different from 0° will arrive at a different time at the two ears. (b) Model of ITD as a function of azimuth fitted from experimental data. The *dashed horizontal line* indicates zero ITD. (c) The likelihood function or generative model for the θ , $p(\text{ITD}|\theta)$, is illustrated as the *dashed black line* (ITD=218.9 ms). The *grey line* is the prior distribution expected by the owl, while the *black line* and *grey area curve* is the posterior probability distribution. The *dotted black line* represents the owl's sound source direction estimate ($\theta = 49.7^\circ$). (d) Estimated azimuth as a function of true (presented) azimuth based on the model described in the main text. The *dotted line* denotes model performance while the *dashed line* denotes ideal performance. Note the bias towards central positions exhibited by the model. The *upper left inset* illustrates the relation between the azimuthal angle θ and its associated unit vector $\mathbf{u}(\theta)$ on the unit circle (*dashed*) (`bayesian.m`)

where W is a Gaussian random variable with zero mean and standard deviation $\sigma_g = 41.2 \mu\text{s}$. This gives the generative model

$$p(\text{ITD} | \theta) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(\text{ITD} - A \sin(\omega\theta))^2}{2\sigma_g^2}}. \quad (4.5)$$

The corresponding likelihood function is illustrated as the dashed curve of Fig. 4.2c for a specific value of the ITD. Notice that it is bimodal with two identical peaks. Hence the owl cannot simply select from this model a single most likely θ , according to the usual “maximum likelihood” principle.

A bias for one peak over the other has to be introduced to make a unique choice. This bias is derived from the prior distribution for the model parameter, θ , which we model as a Gaussian distribution of the form

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\theta^2/(2\sigma_p^2)}, \quad (4.6)$$

where $\sigma_p=23.3^\circ$ is the standard deviation. In this case, the mean is zero which reflects the owl's strong bias for sound sources at the front, while tending to ignore possible sources from the sides. The prior distribution is shown in Fig. 4.2c as the grey curve. Note that θ is a valid angle on the unit circle only when $\theta \in (-180^\circ, 180^\circ]$, and we make the approximation that the probability mass is negligible outside these bounds. This prior is consistent with the known interactions of barn owls and their potential preys (Edut and Eilam 2004).

Applying Eqs. (4.5) and (4.6) to Eq. (4.2), we can compute the shape of the posterior distribution, shown as the grey area curve in Fig. 4.2c. With the posterior distribution, the owl can ask how probable the various source directions θ are given a measured ITD and use this information to make a behaviourally relevant choice. Making this choice involves using a decision rule to reduce a distribution over source directions $p(\theta|\text{ITD})$ to a single estimated source direction $\hat{\theta}$. One decision rule that is consistent with the behavioural data is to take an average of unit vectors weighted by their posterior probability of the form

$$\hat{\theta}(\text{ITD}) = \int \mathbf{u}(\theta)p(\theta|\text{ITD})d\theta, \quad (4.7)$$

where $\mathbf{u}(\theta)$ is the two-dimensional unit vector for each angle (Fig. 4.2d, inset) and the integral is taken over the unit circle. This is also referred to as the circular mean. The result of this estimation procedure for a number of azimuth directions is given by the black points in Fig. 4.2d. As a reference, the black dashed line represents the perfect estimation. Note that the algorithm exhibits a bias towards central positions, that is, it tends to underestimate the true azimuth direction when the source is positioned at eccentric positions. This bias has been shown to exist in the barn owl by means of behavioural experiments. The good agreement between Eq. (4.7) and experimental data suggests that this equation may be implemented neurally, a topic we address in the following section.

4.4 Neural Encoding and Population Vector Decoding

We next ask how sound source localization in the barn owl is implemented by a population of neurons. One approach consists in building a model of the encoding process and then decoding the resulting neural activity using a population vector (PV). The population vector decoding method was pioneered more than 20 years

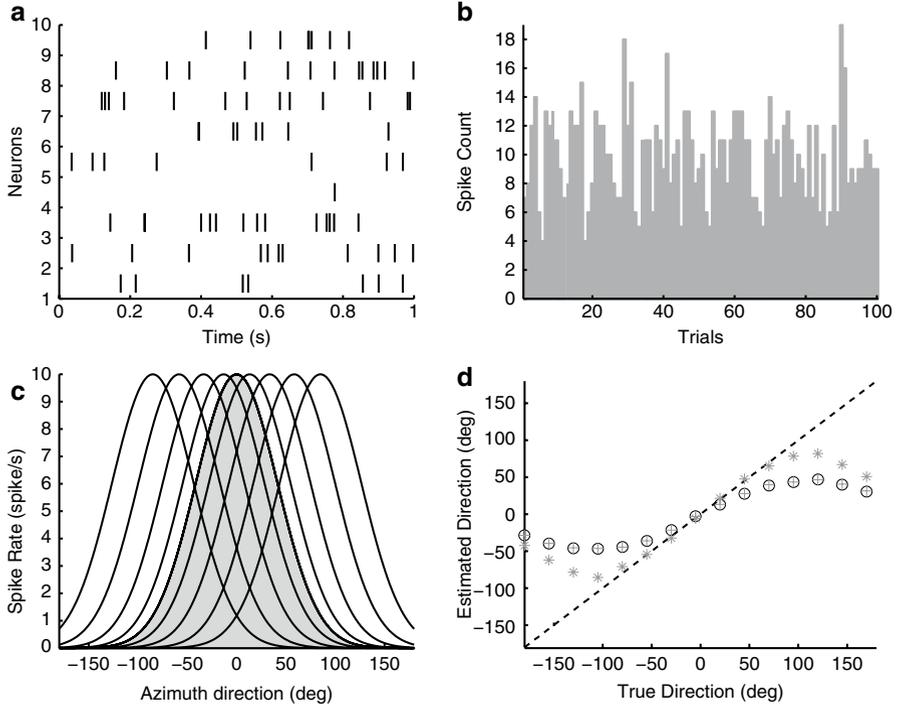


Fig. 4.3 Population vector (PV) estimate of sound source. **(a)** Raster plots of ten Poisson neurons' spike trains with a spike rate of 10 spike/s over a 1 s trial. **(b)** Histogram of a single Poisson spiking neuron's spike count across 100 trials (1 s long; 10 spike/s). **(c)** *Tuning curves* of nine neurons from the neural population encoding model. The *grey tuning curve* has a preferred direction of 0° for the sound source direction, with a peak firing rate of 10 spike/s. The preferred direction of the population is normally distributed around 0° [see Eq. (4.10)]. **(d)** Estimated azimuth as a function of true (presented) azimuth based on the PV and the probabilistic population code (PPC, see Sect. 4.5). The *grey crossed curve* represents the PV estimate while the *grey starred curve* represents the estimate obtained from the PPC (see Sect. 4.5). The *dashed line* denotes ideal performance. The *black circled curve* represents the estimate obtained from the PV when the neuron's firing rates are correlated ($\rho=0.5$, Sect. 4.6). Note that the bias towards central positions exhibited by the PV estimate is similar to that obtained from the PPC Bayesian Model (`nnetwork.m`)

ago in the superior colliculus and motor cortex to decode eye and arm movements from experimentally determined neuronal firing rates. Figure 4.3a shows the activity of ten Poisson neurons with spike rates of 10 spikes/s in a single trial (1 s long). For such Poisson neurons, the distribution of the spike count, k , in each 1 s bin has the form,

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (4.8)$$

where $\lambda = 10$ is the mean, as well as the variance (Ross 2007). The histogram of 100 trials from a single Poisson neuron with a rate of 10 spikes/s is shown in Fig. 4.3b. To test if such a spike histogram from an unknown distribution can be approximated by a Poisson spiking neuron, one can as a first step check if the ratio of the spike count variance and its mean is close to one. This ratio is called the Fano factor (Gabbiani and Cox 2010).

Sensory neurons change their spiking rate based on the specific external stimuli presented to an animal. This change in spike rate can be quantified using a tuning curve as shown by the grey area curve in Fig. 4.3c. The horizontal axis is the parameter of the stimulus, in our example the sound source direction, while the vertical axis is the average firing rate of the neuron responding to that stimulus parameter.

For our Poisson neuron, this would correspond to a 0° sound source and coincides with the peak firing rate of the neuron. We model the tuning curve with the form

$$r_n(\text{ITD}) = r_{\max} e^{-(\text{ITD} - A \sin(\omega \theta_n))^2 / (2\sigma_g^2)}, \quad n = 1, \dots, N, \quad (4.9)$$

where r_{\max} is the peak firing rate and θ_n is the n th neuron's preferred direction. For our Poisson neuron, this would be 10 spikes/s and 0° respectively. The parameters A , ω and σ_g are the same as in Eq. (4.5); hence, Eq. (4.9) is proportional to Eq. (4.5). We shall assume that the population of neurons responsible for sound localization is homogeneous except that neurons have varying preferred directions, θ_n , as shown by the black curves in Fig. 4.3c.

The N neurons in the population have preferred direction θ_n sampled from the distribution

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\theta^2 / (2\sigma_p^2)}, \quad (4.10)$$

which is exactly the same as the prior in Eq. (4.6). Using the neuronal population vector of the form

$$\hat{\theta}(\theta) = \frac{1}{N} \sum_{n=1}^N \mathbf{u}(\theta_n) k_n \quad (4.11)$$

to decode the estimated sound source direction, we get results shown by the grey crossed curve in Fig. 4.3d. In Eq. (4.11), θ is the true sound source, k_n is the firing rate from a single trial of neuron n with tuning function given in Eq. (4.9) and $N=400$ is the number of neurons.

Notice that the curve looks strikingly similar to Fig. 4.2d. This is no coincidence, as our neural implementation can be shown to converge to the Bayesian estimate as the number of neurons $N \rightarrow \infty$ (Fischer and Peña 2011). Note also that in Eq. (4.9), $A \sin(\omega \theta)$ is the mean ITD, according to Eq. (4.4). Thus, a simple averaging

mechanism is able to account for the behavioural data, based on the firing rate of a neuronal population tuned to ITD in a similar manner as barn owl neurons.

4.5 Probabilistic Population Codes

The PV is not the only method for decoding sensory responses from a population of neurons. An alternative scheme is based on a probabilistic population code (PPC; Ma et al. 2006). The PPC assumes that neuronal populations encode probability distributions through their joint firing rate tuning curves. As a result, the entire tuning curve of the neuronal population and not just the preferred direction is used in the decoding process. Let $\mathbf{k} = (k_1, k_2, \dots, k_N)$ represent the response in a single trial of N neurons to a fixed sound source direction θ . The posterior distribution has the form

$$p(\theta | \mathbf{k}) \propto p(\mathbf{k} | \theta)p(\theta), \quad (4.12)$$

where $p(\theta)$ is the same as in Eq. (4.6) and $p(\mathbf{k} | \theta)$ is a distribution which models the probability of a neuronal response given the stimulus.

If we assume that the neurons representing $p(\mathbf{k} | \theta)$ are independent and Poisson, then the probabilistic population code for the distribution $p(\mathbf{k} | \theta)$ has the form

$$p(\mathbf{k} | \theta) = \prod_{n=1}^N \frac{r_n(\theta)^{k_n}}{k_n!} e^{-r_n(\theta)}, \quad (4.13)$$

where k_n is the response of neuron n and $r_n(\theta)$ is its tuning function. We model the tuning functions similarly as in Sect. 4.4, $r_n(\theta) = r_n(\text{ITD}(\theta))$ using Eqs. (4.9) and (4.3), but with a uniform distribution of preferred directions over the unit circle instead of being normally distributed. Note that the right-hand side of Eq. (4.13) is formed by taking products of Eq. (4.8) with k_i replacing k and $r_i(\theta)$ replacing λ , since we assume independent Poisson neurons.

Based on this probabilistic population code, an alternative decision rule to averaging over unit vectors is to pick the azimuth that maximizes the posterior probability. This rule is called the maximum a posteriori probability (MAP) rule and has the form

$$\hat{\theta}(\mathbf{k}) = \arg_{\theta} \max p(\theta | \mathbf{k}).$$

The result of this estimation method based on the probabilistic population code is given in Fig. 4.3d as the grey starred curve. It is significantly different from the population vector result and does not match the behavioural data very well. On the other hand, a probabilistic population code has been successfully used to explain the sensory integration of visual and vestibular cues in neurons of the monkey visual cortex using a slightly different decoding mechanism (Fetsch et al. 2011).

4.6 Correlated Tuning Curves

In actual neural networks, the trial by trial firing rates of neurons may be correlated with each other (Averbeck et al. 2006, Ecker et al. 2010). To model these correlations, we assume that our N neurons have the same tuning curves and distributions of preferred directions as in Eqs. (4.9) and (4.10). In addition, we assume that their firing rates on a single trial are drawn from the multinormal probability distribution of the form

$$p(\mathbf{k} \mid \text{ITD}) = \frac{1}{(2\pi)^{1/(2N)} |\Sigma_g|^{1/2}} e^{-\frac{1}{2}(\mathbf{k}-\mathbf{r}(\text{ITD}))' \Sigma_g^{-1} (\mathbf{k}-\mathbf{r}(\text{ITD}))} \quad (4.14)$$

(Anderson 2003). In this equation, $\mathbf{r}(\text{ITD}) = (r_1(\text{ITD}), \dots, r_N(\text{ITD}))'$ is the mean firing rate of the N neurons given the ITD, and \mathbf{v}' denotes the transpose of vector \mathbf{v} . The covariance matrix is represented by Σ_g and its determinant by $|\Sigma_g|$. If the covariance matrix is given by $\Sigma_g = (\Sigma_{ij})$, with

$$\Sigma_{ij} = \sqrt{r_i(\text{ITD})r_j(\text{ITD})}\delta_{ij}, \quad i, j = 1, \dots, N,$$

and $\delta_{ij}=1$ when $i=j$, while $\delta_{ij}=0$ when $i \neq j$, we get an uncorrelated multinormal distribution. Because each neuron's firing rate variance is proportional to its mean firing rate, this formulation is close to that of Sect. 4.4 using Poisson neurons (mean equal to variance). If the covariance matrix elements have the form

$$\Sigma_{ij} = \sqrt{r_i(\text{ITD})r_j(\text{ITD})}(\delta_{ij} + \rho(1 - \delta_{ij})), \quad (4.15)$$

where $\rho \in [0, 1)$ is the correlation coefficient, we have introduced correlations of magnitude ρ into all pairs of neurons' firing rates.

When applying the PV, we see no difference in sound source estimates from independent neurons. This is illustrated in Fig. 4.3d by the black circled curve, which is exactly overlapping with the grey crossed curve obtained from independent neurons, in spite of sizable correlations between single neurons' firing rates ($\rho=0.5$). Thus, neuronal correlations do not affect the results exposed in the previous sections. Intuitively, this may be understood from the fact that the direction of the PV will not be changed by correlated noise, if the noise scales uniformly with the mean firing rate of the neurons, as implemented by Eq. (4.15). In conclusion, Bayesian statistical modelling is a computational analysis technique that can provide insight in the coding of sensory information from a neuroethological perspective, as illustrated in this chapter.

References

- Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn Wiley, Hoboken
- Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation *Nat Rev Neurosci* 7(5):358–366
- Doya K, Ishii I, Pouget A, Rao R (eds) (2007) Bayesian brain: probabilistic approaches to neural coding The MIT Press, Cambridge
- Ecker AS, Berens P, Keliris Ga, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits *Science* 327(5965):584–587
- Edut S, Eilam D (2004) Protean behavior under barn-owl attack: voles alternate between freezing and fleeing and spiny mice flee in alternating patterns *Behav Brain Res* 155(2):207–216
- Fetsch CR, Pouget A, DeAngelis GGC, Angelaki DDE (2011) Neural correlates of reliability-based cue weighting during multisensory integration *Nat Neurosci* 15(1):1–18
- Fischer BJ, Peña JL (2011) Owl's behavior and neural representation predicted by Bayesian inference *Nat Neurosci* 14(8):1061–1066
- Gabbiani F, Cox SJ (2010) Mathematics for neuroscientists Academic Press, San Diego
- Hausmann L, Von Campenhausen M, Endler F, Singheiser M, Wagner H (2009) Improvements of sound localization abilities by the facial ruff of the barn owl (*Tyto alba*) as demonstrated by virtual ruff removal *PLoS One* 4(11):e7721
- Hoff PD (2009) A first course in Bayesian statistical methods Springer texts in statistics Springer, Dordrecht
- Knudsen EI, Blasdel GG, Konishi M (1979) Sound localization by the barn owl (*Tyto alba*) measured with the search coil technique *J Comp Physiol* 133(1):1–11
- Konishi M (2003) Coding of auditory space *Annu Rev Neurosci* 26:31–55
- Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes *Nat Neurosci* 9(11):1432–1438
- Ross SM (2007) Introduction to probability models, 10th edn Academic Press, San Diego