Output of a neuronal population code

Jianfeng Feng

Computational Neuroscience Laboratory, The Babraham Institute, Cambridge CB2 4AT, United Kingdom

Rita Cassia-Moura*

The Abdus Salam International Centre for Theoretical Physics, 34100 Trieste, Italy

In the population coding framework, we consider how the response distributions affect output distribution. A general theory for the output of neuronal population code is presented when the spike train is a renewal process. Under a given condition on the response distribution, the most probable value of the output distribution is the center of input-preferred values, whereas in the other cases the most *improbable* value of the output distribution is the center of input-preferred values or there are no most probable states. Depending on the exact form of the response distributions, the variance of the output distributions can either enlarge or reduce the tuning width of the tuning curves. [S1063-651X(99)09606-3]

PACS number(s): 87.19.La, 87.10.+e

How neurones unambiguously encode external inputs and then decode them is obviously one of the fundamental and challenging problems in theoretical neuroscience. In recent years there have been many research activities devoted to tackling the problem and a wide variety of possibly coding strategies have been put forward (see, for example, [1–3]). One of the straightforward generalizations of the classic notation of coding—the assumption of rate coding—is population coding, which has been closely examined both in experiments and in theory recently.

Experimentally it has been suggested that population coding might be employed in the generation of saccadic eye movements in primates [4], the direction of arm movements [5], visual orientation estimation [6–8], and sound localization [9]. In particular, in [10] the authors considered a neural network containing four input neurones (encoding neurones), ten output neurones (decoding neurones), and a few interneurones and successfully mimicked the biological data, provided that input and output neurones employed the population coding strategy.

Theoretically in the framework of population coding, a few issues such as the relationship between the maximum likelihood estimate and a network estimate [11], how the broadness of the tuning function and the correlation of afferent signals affect the outputs, etc. [12–15], have been discussed. In particular, in [11] the authors have pointed out that if the nervous system in fact managed to implement maximum likelihood, then its performance would be uniformly good for all stimuli.

The basic idea of the population code is as follows. A set of neurones fire at rates according to their tuning functions, which usually take the largest values at their preferred values. Turning functions are encoded in the spike trains of these neurones. Output distributions are then decoded in terms of the efferent spike trains and an action (for example,

the bend direction of the leech [10]) is taken which usually corresponds to the most probable value of the output distribution.

Population coding, as a postulated encode and decode strategy of the brain, is now widely accepted as a generalization of the rate coding assumption. However, in the literature all theoretical considerations are confined to the case in which the interspike interval distributions, i.e., the response distributions, are exponential, which is probably a good approximation of the true response distribution in some cases (however, see [15] for a general approach in terms of Fisher information). It is well known that neurones might fire with a considerably different response distribution [16–19], ranging from short-tailed to long-tailed distributions, and thus it is of vital importance for us to check population coding in the more general setting of response distribution. To this end we consider the population code with the response distribution as a renewal process. We find that under a given condition on the response distributions, the center of the preferred values is the most probable state of the output distribution, whereas the center is the most *improbable* state of the output distribution or the output distribution is independent of the preferred values. Furthermore, an exact parameter region is given in which the variance of the output distribution can enlarge the tuning width of the tuning curves, an issue which has been addressed and unresolved in the literature [14,15].

Response distributions. Suppose that at time zero an input is onset, and we record N cells' activities which are statistically independent. For a given cell, denote the efferent interspike intervals as T_i, $i=1,2,\ldots$, which are independent random variables with distribution density (the so-called one-parameter exponential family)

$$\lambda \exp(-\lambda g(t))g'(t), \quad t\geq 0, \tag{1}$$

where λ is the parameter. The distribution of T_i, $i=1,\ldots,N$, is termed as response distribution as well. For the concreteness of our calculation we are first going to consider three cases:

---

*Permanent address: Universidade de Pernambuco, ICB-DCF/ Biofisica, C.P. 7817, Recife, PE, Brazil.

$$g(t)=t, \quad g(t)=\ln(t+1), \quad g(t)=t^2,$$

discussed in more detail in the following examples (i)–(iii). We will also consider a slightly more general setting of the response distribution in examples (iv) taking the form

$$h(\lambda,t)\exp(-g(t,\lambda)), \quad t\geq 0, \quad (2)$$

where h,g are positive functions of t,λ. Hence in this paper we consider the case in which the spike train of each cell is a renewal process.

Tuning function. We assume that the tuning function of the ith cell, $i=1,\ldots,N$, takes the following form:

$$\lambda_i(x)=F\exp\left(-\frac{(x-x_i)^2}{\sigma^2}\right)+b, \quad (3)$$

where x_i is the preferred value of the ith cell, σ is the tuning width, F is the mean peak firing rate, and b is the basal firing rate. When $x=x_i$, the tuning function λ_i takes the largest value, i.e., the cell fires with its greatest activity. In this paper we restrict ourselves to the case of $\sigma\ll 1$ and refer the reader to our full paper for the general case [20].

Through the spike trains of N cells, or equivalently the response distributions, the information [the tuning function $\lambda_i(x)$, $i=1,\ldots,N$] is encoded. Now we discuss how the information is read out (decoded).

Decoding I. We first consider the case with the response distribution defined by Eq. (1). For simplicity of notation we take $b=0$ for exponential and Gaussian distribution and $b=1$ for Pareto distribution.

Example (i). Exponential distribution, $g(t)=t$. This is the most widely used case and has been extensively discussed in the literature. The actual behavior of neurones, of course, is never exactly distributed as an exponential function. However, in many cases it serves as a good approximation. Its mean and variance are

$$\mu_i(x)=\frac{1}{\lambda_i(x)}, \quad \sigma_i^2(x)=\frac{1}{\lambda_i^2(x)} \quad (4)$$

with a coefficient of variation (CV) being equal to 1.

Example (ii). Pareto distribution, $g(t)=\ln(t+1)$. In this case the distributions have long tails and we know that they are ubiquitous [16,17]. Its mean and variance are

$$\mu_i(x)=\frac{1}{\lambda_i(x)-1}, \quad \sigma_i^2(x)=\frac{\lambda_i(x)}{[\lambda_i(x)-1][\lambda_i(x)-2]}.$$

Hence we require that $\lambda_i(x)>2$ for the existence of $\mu_i(x)$, $\sigma_i(x)$ and its CV is greater than 1.

Example (iii). Positive Gaussian distribution, $g(t)=t^2$. This is one of the cases discussed in [19,21] with mean, variance

$$\mu_i(x)=\frac{1}{\lambda_i(x)}, \quad \sigma_i^2(x)=\frac{1}{2\lambda_i^2(x)},$$

and $CV<1$.

These simple response distributions of neuronal activities are most commonly discussed in the literature. Furthermore, from example (i)–(iii) we see that it is reasonable to think of $\lambda_i(x)$ as the firing rate.

Suppose during time window $[0,T]$ the ith cell, $i=1,\ldots,N$, emits r_i spikes. The information, i.e., the tuning function $\lambda_i(x)$, is thus encoded through r_i, $i=1,\ldots,N$. In terms of data r_i, $i=1,\ldots,N$, the neurone is able to read out the information, namely to recover the function $\lambda_i(x)$ in a certain way.

Let $N_i(t)$ be the number of emitted spikes of the ith cell during time interval $[0,t]$. More exactly, we have

$$P[N_i(t)=r_i|x]=P(T_1+\cdots+T_{r_i}>t)$$
$$-P(T_1+\cdots+T_{r_{i-1}}>t).$$

In general, it is difficult to obtain an analytical formula for the distribution above except for the Poisson case. However, we have the following asymptotic result [22].

Lemma 1. When t is large, we have $N_i(t)\sim N(\bar{\mu}_i(t,x),\bar{\sigma}_i^2(t,x))$, where

$$\bar{\mu}_i(t,x)=\frac{t}{\mu_i(x)}, \quad \bar{\sigma}_i^2(t,x)=\frac{t\sigma_i^2(x)}{\mu_i^3(x)}$$

and

$$\mu_i(x)=\int_0^\infty th(x,t)dt, \quad (5)$$

$$\sigma_i^2(x)=\int_0^\infty t^2h(x,t)dt-\mu_i^2(x), \quad (6)$$

where $h(x,t)$ is the distribution density of T_1.

It is interesting to see the degree of accuracy of the approximation of Lemma 1 within a reasonable time period, say 200 msec. In [23] the author has carried out some numerical simulations for the response distributions of the integrate and fire model, showing that the approximation is strikingly good. Of course, this approximation *strongly* depends on the cell's firing rate. If it fires at a rate of 10 Hz, then within a time window of 200 msec only two spikes are recorded and the approximation will be certainly not true.

We do not yet have an exact idea on how neurones decode their input signals, however we know that the maximum likelihood gives us an optimal estimate [11] where a comparison between the performance of the maximum likelihood and a network has also been carried out. More generally, from a Baysian point of view the posterior function is obtained in the following way:

$$\hat{p}(x)\propto\prod_{i=1}^N p[N_i(T)=r_i x]p(x)$$

for a prior distribution $p(x)$. We fix the prior distribution in the following discussion by assuming $p(x)=$const, i.e., the noninformative prior, which gives the same results as the maximum likelihood.

Theorem 1. For the response distributions in example (i)–(iii) the posterior distribution is

$$\hat{P}(x) \propto \exp\left(-\frac{(x-\hat{\mu})^2}{\hat{\sigma}^2}\right)$$

with $\hat{\mu} = (\Sigma_{i=1}^N x_i)/N$ and $\hat{\sigma}^2 = \sigma^2/N$ (exponential distribution and positive Gaussian distribution), $\hat{\sigma}^2 = \sigma^2/(2N)$ (Pareto distribution, provided that $F \gg 1$).

Proof. In terms of Lemma 1, we see that the posterior distribution is given by

$$\ln \hat{P}(x) \propto -\sum_{i=1}^N \ln[2\pi\bar{\sigma}_i(T,x)^2]^{1/2} - \sum_{i=1}^N \frac{[r_i - \bar{\mu}_i(T,x)]^2}{2\bar{\sigma}_i(T,x)^2}.$$

From the strong law of large numbers (Lemma 1) we know that

$$\sum_{i=1}^N \frac{[r_i - \bar{\mu}_i(T,x)]^2}{2\bar{\sigma}_i(T,x)^2} \to \frac{N}{2}.$$

Therefore, we have

$$\ln \hat{P}(x) \propto [\tfrac{1}{2}\ln(2\pi) - 1]N - \ln T$$
$$-2\sum_{i=1}^N \ln \sigma_i(x) + 3\sum_{i=1}^N \ln \mu_i(x). \quad (7)$$

Now we discuss different cases separately.

(i) Exponential distribution. From Eq. (7) we see that

$$\ln \hat{P}(x) \propto [\tfrac{1}{2}\ln(2\pi) - 1]N - \ln T + \ln F$$
$$-\frac{N}{\sigma^2}(x-\hat{\mu})^2 + \frac{\Sigma_{i=1}^N(x_i - \hat{\mu})^2}{\sigma^2}.$$

(ii) Pareto distribution,

$$\ln \hat{P}(x) \propto [\tfrac{1}{2}\ln(2\pi) - 1]N - \ln T + \ln F$$
$$-\frac{2N}{\sigma^2}(x-\hat{\mu})^2 + \frac{2\Sigma_{i=1}^N(x_i - \hat{\mu})^2}{\sigma^2}.$$

(iii) Positive Gaussian distribution. The proof is similar to that of the exponential distribution.

Let us now analyze the implication of Theorem 1. For a set of tuning functions with their preferred values x_i, $i = 1, \ldots, N$, we obtain an output function (the posterior distribution) with its most probable value at the center of x_i, $i = 1, \ldots, N$. Here we want to emphasize that Theorem 1 is obtained for those cells whose activities fulfill the central limit theorem (Lemma 1). As we have pointed out earlier, within a given short time window the central limit theorem is in general true for only a subset of all cells, and therefore Theorem 1 is valid for *this subset* of cells. Hence the safest way to apply Theorem 1 is that we pick up the cells which fire most intensively and in this case the results of Theorem 1 are approximately in agreement with results in the literature [11]. There are examples which tell us that if we do use data of bursting cells, the estimate accuracy is improved [24]. On the other hand, Theorem 1 also reveals another interesting phenomenon: if the time window of recording is long enough, the most probable state of posterior distribution is independent of inputs. Therefore, to find an appropriate time window is subtle in the application of the maximum likelihood estimate, a topic which has not been fully considered in the literature.

Decoding II. It seems that the preceding section gives us the desired results: the center of the input preferred values is the most probable state of the output distribution, no matter what the response distributions are. However, as we can see from the proof of Theorem 1, the posterior distribution critically depends on the difference between $3\ln\mu_i(x)$ and $2\ln\sigma_i(x)$. This fact suggests consideration of the following example.

Example (iv). Suppose that the response distribution is

$$\lambda^{(k+1)\lambda^k} \frac{t^{\lambda^k-1}\exp(-\lambda^{k+1}t)}{\Gamma(\lambda^k)}, \quad (8)$$

where $k > 0$ and $\Gamma()$ is the Gamma function. The distribution defined by Eq. (8) is simply the Gamma distribution with parameters $(\lambda^{k+1}, \lambda^k)$. We know that its mean and variance are

$$\mu_i(x) = \frac{1}{\lambda_i(x)}, \quad \sigma_i^2(x) = \frac{1}{\lambda_i^{k+2}(x)}. \quad (9)$$

Again we see that the firing rate is $\lambda_i(x)$. In fact from the fitting of the histogram of efferent spike trains, we know that the Gamma distribution is more suitable as the response distributions than exponential ($k=0$) and Pareto distribution. Typically, the response distribution equals zero at time zero (if we ignore the refractory periods), similar to the Gamma distribution. See, for example, [19] for an excellent fit of biological data in terms of the Gamma distribution. Mimicking the proof of Theorem 1, we have [Theorem 2(i) is a generalization of Theorem 1] the following.

Theorem 2. For a response distribution with the properties $\sigma_i^2(x) \sim \lambda_i^{-(k+2)}(x)$ and $\mu_i(x) \sim \lambda_i^{-1}(x)$, we have the following situations.

(i) If and only if $k < 1$,

$$\hat{P}(x) \propto \exp\left(-\frac{(x-\hat{\mu})^2}{\hat{\sigma}^2}\right)$$

with $\hat{\mu} = (\Sigma_{i=1}^N x_i)/N$ and $\hat{\sigma}^2 = \sigma^2/[N(1-k)]$. The most probable state is the center of the input preferred values.

(ii) If $k=1$, the posterior distribution is independent of x and therefore there are no most probable states.

(iii) If $k > 1$,

$$\hat{P}(x) \propto \exp\left(\frac{(x-\hat{\mu})^2}{\hat{\sigma}^2}\right)$$

with $\hat{\mu} = (\Sigma_{i=1}^N x_i)/N$ and $\hat{\sigma}^2 = \sigma^2/[N(k-1)]$. The most improbable state of the posterior distribution is the center of the input preferred values.

Theorem 2 reveals an interesting critical phenomenon: $k=1$ is the critical case with the property that the output distribution is uniformly distributed over all possible states. Therefore, in this case we are not able to get a maximum likelihood estimate. $k < 1$ is the case discussed in the literature [11] and the maximum likelihood estimate holds true.

The worst case is $k>1$, where the center of the input preferred values is the most improbable state of the output distribution.

Another interesting issue arising from Theorem 2(i) is whether the tuning width of the tuning function is enlarged or reduced in the output distribution. If it is enlarged, the output will have less accuracy than the input tuning function [9], otherwise the output will be more concentrated on its most probable states. In the literature there is some discussion on this issue [9,14]. Here we find a full spectrum of behaviors: the variance of the output distribution can increase, decrease, or remain the same, compared with the tuning width of the tuning function and depending on the parameter k. More exactly we have the following.

(i) If $k<1-1/N$, the tuning width is reduced.

(ii) If $k=1-1/N$, the tuning width and the variance remain the same.

(iii) If $1>k>1-1/N$, the variance is enlarged.

Recently, attention has been focused on how a certain quantity of afferent signals changes after neuronal transformation. In [21] the authors have observed that the output variance in time domain is smaller than the input variance in time domain. We have pointed out that whether the output variance in time domain is smaller or larger than the input variance in time domain depends on response distributions [16,17] and other factors [18]. Here again the ratio between output variance and input variance (tuning width) depends on the response distribution and the number of active neurones. However, we want to emphasize that when the number of active neurones is large, the parameter region of k in which the output variance is larger than that of input is small, i.e., $1-1/N<k<1$.

Discussion. We found that the output distribution, in the framework of population coding, is sensitive to the concrete form of the response distribution under the condition that it is a renewal process. The most probable state of the output distribution is the center of the input preferred states provided that $k<1$. If $k>1$, the center of the input preferred states is the most *improbable* state of the output distribution. The output distribution can be uniformly distributed over all possible states if $k=1$, which is the critical case.

There are a few issues needed to be clarified here.

(i) General cases without the condition $\sigma \ll 1$ are considered in [20]. Although we confine ourselves to neuronal population code, which is the motivation of our study, the conclusions in the present paper are true in the general context of statistical inference.

(ii) Large t. Our conclusions in fact are obtained under the assumption of large time behaviors (Lemma 1). However, as we pointed out before, it characterizes the essential behaviors of the decoding and encoding in the population coding framework [20].

---

[1] W. Gerstner, A. K. Kreiter, H. Markram, and A. V. M. Herz, Proc. Natl. Acad. Sci. USA **94**, 12 740 (1997).

[2] J. J. Hopfield, Nature (London) **376**, 33 (1995).

[3] F. Rieke, D. Warland, R. de Ruyter van Steveninch, and W. Bialek, *Spikes: Exploring The Neural Code* (MIT Press, Cambridge, MA, 1997).

[4] C. Lee, W. H. Rohrer, and D. L. Sparks, Nature (London) **332**, 357 (1988).

[5] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, Science **233**, 1416 (1986).

[6] C. D. Gilbert and T. N. Wiesel, Vision Res. **30**, 1689 (1990).

[7] R. Vogels, Biol. Cybern. **64**, 25 (1990).

[8] E. Zohary, Biol. Cybern. **66**, 262 (1992).

[9] D. C. Fitzpatrick, R. Batra, R. S. Terrence, and S. Kuwada, Nature (London) **388**, 871 (1997).

[10] J. E. Lewis and W. B. Kristan, Jr., Nature (London) **391**, 76 (1998).

[11] H. S. Seung and H. Sompolinsky, Proc. Natl. Acad. Sci. USA **90**, 10 749 (1993).

[12] L. Abbott and P. Dayan, Neural Comput. **11**, 91 (1998).

[13] A. Pouget, K. C. Zhang, S. Deneve, and P. E. Latham, Neural Comput. **10**, 373 (1998).

[14] R. S. Zemel, P. Dayan, and A. Pouget, Neural Comput. **10**, 403 (1998).

[15] K. C. Zhang and T. Sejnowski, Neural Comput. **11**, 75 (1998).

[16] J. Feng, Phys. Rev. Lett. **79**, 4505 (1997).

[17] J. Feng and D. Brown, J. Phys. A **31**, 1239 (1998).

[18] J. Feng and D. Brown, Biol. Cybern. **78**, 369 (1998).

[19] G. L. Gerstein and B. Mandelbrot, Biophys. J. **4**, 41 (1964).

[20] J. Feng and D. Brown (unpublished).

[21] P. Marsalek, C. Koch, and J. Maunsell, Proc. Natl. Acad. Sci. USA **94**, 735 (1997).

[22] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1966).

[23] M. Stemmler, Network **7**, 687 (1996).

[24] J. E. Lisman, Trends Neurosci. **20**, 38 (1997).