

Statistical analysis of neural data: Addenda to Brown's point process notes

Liam Paninski
Department of Statistics and Center for Theoretical Neuroscience
Columbia University
<http://www.stat.columbia.edu/~liam>

January 31, 2007

4: Two applications of the point process likelihood formula

Two-alternative forced choice (2AFC) discrimination. Let's assume we have some model that assigns some conditional intensity function $\lambda_{\vec{x}}(t)$ to any given stimulus \vec{x} . We may use this model to perform discrimination tasks: given the observed spike train data $D = \{t_i\}$, which of the two stimuli \vec{x}_1 or \vec{x}_2 are more likely to have produced this spike train? (The detection task is a special case: here we would like to know whether \vec{x}_0 or $\vec{x} = 0$ produced the observed response.)

According to Neyman-Pearson, we should simply compare loglikelihoods:

$$\begin{aligned}\log \frac{p(D|\vec{x}_1)}{p(D|\vec{x}_2)} &= \sum_i \log \lambda_{\vec{x}_1}(t_i) - \int_0^T \lambda_{\vec{x}_1}(t) dt - \left(\sum_i \log \lambda_{\vec{x}_2}(t_i) - \int_0^T \lambda_{\vec{x}_2}(t) dt \right) \\ &= \sum_i \log \frac{\lambda_{\vec{x}_1}(t_i)}{\lambda_{\vec{x}_2}(t_i)} + \int_0^T [\lambda_{\vec{x}_2}(t) - \lambda_{\vec{x}_1}(t)] dt.\end{aligned}$$

This formula has a simple interpretation: we will prefer hypothesis \vec{x}_1 if the difference $\log \lambda_{\vec{x}_1}(t) - \log \lambda_{\vec{x}_2}(t)$ is large at the times t_i when we observed a spike, and when the difference $\lambda_{\vec{x}_1}(t) - \lambda_{\vec{x}_2}(t)$ is large at all other times.

In the special case of homogeneous Poisson processes, $\lambda_{\vec{x}_i}(t) \equiv \lambda_i$, this formula reduces further:

$$\log \frac{p(D|\vec{x}_1)}{p(D|\vec{x}_2)} = N_T \log \frac{\lambda_1}{\lambda_2} + T(\lambda_2 - \lambda_1),$$

with N_T denoting the total number of spikes observed on $[0, T]$. In other words, the total spike count N_T is sufficient here to perform discrimination optimally. Another way to state this is that N_T is a minimal sufficient statistic for performing inference on the the rate λ of a homogeneous Poisson process: all timing information may be ignored in this case. (Of course in general we do need to keep track of spike timing to perform inference optimally, as is clear from the more general formula above.)

Change-point detection. In many cases we are interested in detecting the time of a change in a neuron's firing rate, from some baseline rate λ_0 to λ_1 (Akman and Raftery, 1986; Loader, 1991; Herberts and Jensen, 2004). We may estimate the change-point time (the time t_0 at which the rate changes from λ_0 to λ_1) by maximum likelihood:

$$\hat{t}_{MLE} = \arg \max_t L(t),$$

with the likelihood of a change at time t defined as

$$L(t) = N_t \log \lambda_0 + (N_T - N_t) \log \lambda_1 - t\lambda_0 - (T - t)\lambda_1 = t(\lambda_1 - \lambda_0) + N_t \log \frac{\lambda_0}{\lambda_1} + \text{const.},$$

with N_t defined as the cumulative number of spikes up to time t . (For simplicity here, assume that the rates λ_0 and λ_1 are known.) Thus this likelihood has a very simple form: a linear drift interrupted by jumps of size $\log \frac{\lambda_0}{\lambda_1}$ at spike times. We may estimate the change point t_0 by optimizing $L(t)$,

$$\hat{t}_{MLE} = \arg \max_t L(t).$$

It is clear that we will not be able to estimate t_0 exactly unless the firing rate goes to infinity (the information in the spike train is roughly proportional to the total number of spikes observed, as we will discuss in more detail shortly). The consistency and asymptotic distribution of \hat{t}_{MLE} is fairly easy to derive in the limit $\lambda_0, \lambda_1 \rightarrow \infty, \lambda_0/\lambda_1 = c$; for consistency, we need only note that N_t may be approximated by its mean

$$E(N_t) = \begin{cases} \lambda_0 t & 0 \leq t < t_0 \\ \lambda_0 t_0 + \lambda_1 (t - t_0) & t_0 \leq t \leq T \end{cases}$$

in the large firing rate limit; therefore, we may approximate

$$L(t) \approx \text{const.} + t(\lambda_1 - \lambda_0) + \begin{cases} \lambda_0 \log \frac{\lambda_0}{\lambda_1} t & 0 \leq t < t_0 \\ [\lambda_0 t_0 + \lambda_1 (t - t_0)] \log \frac{\lambda_0}{\lambda_1} & t_0 \leq t \leq T \end{cases}.$$

Proving consistency is now (by standard M-estimator theory) just a matter of proving that $L(t)$ is asymptotically increasing up to t_0 and decreasing after t_0 (van der Vaart, 1998). This may be proven directly¹ or by the more standard approach of writing this expected log-likelihood ratio as a negative Kullback-Leibler (KL) divergence and using the non-negativity of the KL divergence (Cover and Thomas, 1991).

The asymptotic distribution of \hat{t}_{MLE} may be characterized as the maximizer of a Brownian motion with two-sided linear drift; generically, both the drift and scale of the Brownian change discontinuously at t_0 (van der Vaart, 1998).

¹Taking derivatives, we need to prove that

$$(\lambda_1 - \lambda_0) + \lambda_0 \log \frac{\lambda_0}{\lambda_1} > 0$$

and

$$(\lambda_1 - \lambda_0) + \lambda_1 \log \frac{\lambda_0}{\lambda_1} < 0$$

for any positive $\lambda_0 \neq \lambda_1$. To see this, simply write

$$(\lambda_1 - \lambda_0) + \lambda_0 \log \frac{\lambda_0}{\lambda_1} = \lambda_0 (1/c - 1 + \log c)$$

and

$$(\lambda_1 - \lambda_0) + \lambda_1 \log \frac{\lambda_0}{\lambda_1} = \lambda_1 (1 - c + \log c)$$

for $c = \lambda_0/\lambda_1$; the desired inequalities follow noting the positivity of λ_0 and λ_1 and then using the inequalities

$$1 - 1/c \leq \log c \leq c - 1, \quad 0 < c < \infty.$$

5.2: Renewal processes

A key equation for a spike train corresponding to a renewal process is the conditional firing rate given the occurrence of a spike at time $t = 0$:

$$\lambda_t|(n(t) = 0) = \sum_{i>0} p *^{i-1} p(t) = \left(\sum_{i>0} (p^\wedge)^i \right)^\vee (t) = \left(\frac{p^\wedge}{1 - p^\wedge} \right)^\vee (t),$$

where $*^i$ denotes the i -fold convolution and \wedge and \vee denote the Fourier and inverse Fourier transforms, respectively. The renewal theorem (Karlin and Taylor, 1981) states that

$$\lim_{t \rightarrow \infty} \lambda_t = 1/E(T)$$

under weak conditions on $p(t)$; i.e., the asymptotic firing rate is given by the inverse of the mean interspike interval. (Note that in general, by Jensen's inequality,

$$1/E(T) \leq E(1/T),$$

since the function $1/x$ is convex for $x > 0$.)

5: Spatial Poisson processes

While we will spend most of our time discussing one-dimensional temporal point processes, it is worth noting that the concept of a Poisson process extends readily to the multidimensional case (Snyder and Miller, 1991; Moeller and Waagepetersen, 2004). The same definition applies: a Poisson process is a nonnegative discrete stochastic process for which the counts in any collection of disjoint sets C_1, C_2, \dots are independent Poisson random variables, with rate parameter $\lambda_C = \int_{\vec{x} \in C} \lambda(\vec{x}) d\vec{x}$, for some rate function $\lambda(\vec{x})$. We obtain the same likelihood formula in the continuous-space limit:

$$\sum_{\vec{x}_j} \log \lambda(\vec{x}_j) - \int \lambda(\vec{x}) d\vec{x},$$

where \vec{x}_j are the points where samples were observed. We will discuss this further in the context of models which are useful for optimal image smoothing.

Point processes in one dimension (e.g. time) are simpler because sampling and likelihood calculations may be done recursively for many natural models of the conditional intensity function $\lambda(\cdot)$. In general it may be difficult or awkward to construct analogous recursions in the case that the state space has dimension larger than one; see (Moeller and Waagepetersen, 2004) for further details.

7: Time-rescaling: network simulation example

To be concrete, let's imagine a population of model neurons with spike rate given by the generalized linear model form

$$\lambda_i(t) = f \left(I_i(t) + \sum_{t_{i',j} < t} h_{i',i}(t - t_{i',j}) \right),$$

where $I_i(t)$ is some input signal to neuron i and $h_{i',i}$ a post-spike effect from the i' -th observed neuron; these terms are summed over all past spike times $\{t_{i',j}\}$. The $h_{i,i}$ terms correspond to the i -th cell's own past activity, and may encode e.g. refractory or burstiness effects; the $h_{i',i}$ terms from the other cells in the population correspond to interneuronal interaction effects, and may be excitatory and/or inhibitory.

To sample from this model we may adapt the approach taken in Brown's eq (7.6):

Initialization (assuming no spikes have been observed in the interval $(-\infty, 0)$):

Set the clocks $u_i = 0$ for all cells i .

Set the summed currents $m_i(t) = 0$ for $t \in [0, T]$ and all cells i .

Draw the random times $\tau_i = e_i/dt, e_i \sim \exp(1)$ for all cells i .

Set the time index $t = 0$.

While $t \leq T$:

Advance the clocks $u_i = u_i + f[I_i(t) + h_i(t)]$ for all cells i .

If $u_{i'} \geq \tau_{i'}$ for any cell i' :

— record that neuron i' spiked at time t ;

— update the currents $m_i(s) = m_i(s) + h_{i',i}(s - t)$ for $s \in [t, t + t_0]$ for all cells i which are connected to cell i' ;

— draw a new (independent) random time $\tau_{i'} = e_{i'}/dt, e_{i'} \sim \exp(1)$;

— reset the clock $u_{i'} = 0$.

Advance the time index $t = t + dt$.

Note that only one call to the random number generator is required per spike, making the method quite computationally efficient.

References

- Akman, V. and Raftery, A. (1986). Asymptotic inference for a change-point poisson process. *Annals of Statistics*, 14:1583–1590.
- Cover, T. and Thomas, J. (1991). *Elements of information theory*. Wiley, New York.
- Herberts, T. and Jensen, U. (2004). Optimal detection of a change point in a poisson process for different observation schemes. *Scandinavian Journal of Statistics*, 31:347–366.
- Karlin, S. and Taylor, H. (1981). *A First Course in Stochastic Processes*. Academic Press, New York.
- Loader, C. (1991). Inference for a hazard rate change point. *Biometrika*, 78:749–757.
- Moeller, J. and Waagepetersen, R. (2004). *Statistical inference and simulation for spatial point processes*. Chapman Hall.
- Snyder, D. and Miller, M. (1991). *Random Point Processes in Time and Space*. Springer-Verlag.
- van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.