Assignment 5
Theoretical Neuroscience
Liam Paninski, Quentin Huys
November 5, 2004

1. Linear algebra brush-up

A. Suppose we're given a bunch of data $\vec{x}_i, y_i$, and we want to fit the model

$$y_i = k_0 + \vec{k}_1 \cdot \vec{x}_i + \vec{x}_i^t K_2 \vec{x}_i + \epsilon n_i,$$

with k_0 a scalar constant, $\vec{k}_1$ a weight vector, K_2 a square weight matrix, and n_i is standard Gaussian noise (independent, mean zero, standard deviation of one).

i. Describe an ML solution for the parameters $\{k_0, \vec{k}_1, K_2, \epsilon\}$. Is the ML solution unique? What is the ML solution if we restrict K_2 to be symmetric?

ii. Assume we have a Gaussian prior on $\vec{k}_1$,

$$p_0(\vec{k}_1) = \mathcal{N}(\vec{\mu}, C).$$

What is the MAP estimate of the regression parameters? What is the conditional mean estimate? Now let the covariance matrix C be such that the elements of $\vec{k}_1$ are i.i.d. with

$$E(\vec{k}_1(i)^2) = \lambda_1^{-1}.$$

What are the MAP and conditional mean solutions in this special case? Are these solutions unique?

iii. (optional) Construct a covariance matrix C that models smooth (that is, not uncorrelated) samples $\vec{k}_1$. How does this affect the MAP solution?

iv. (optional) Look up the "James-Stein" estimator, either in any standard statistics textbook (Lehmann, Schervish, Wasserman, etc.) or by going back to the original work [James and Stein, 1960]. How does the James-Stein phenomenon relate to the MAP estimators you just constructed?

B. Define

$$\hat{k} = C^{-1}\vec{k}_{STA},$$

with C the stimulus covariance matrix, $E(\vec{x}^t\vec{x})$. (Assume $E(\vec{x}) = 0$.)

i. Prove that this is an unbiased estimate for $\vec{k}$, the linear component of the basic LNP model, whenever the stimulus distribution $p_0(\vec{x})$ is elliptically symmetric, using the geometric logic outlined in the lecture.

ii. (optional) What if the LNP model is modified to include a refractory term,

$$p(spike|\vec{x}, spike\ history) = f(\vec{k} \cdot \vec{x})r(\tau),$$

with τ denoting the time since the last spike. Describe $\vec{k}_{STA}$ in this case. Calculate the MLE for $r(.)$

2. Point processes

A. (No need to write your answers here, but think through the answers.) Prove or disprove that all of the following specify the same stochastic point process on some finite one-dimensional interval B:

i) The process for which the joint distribution of counts in an arbitrary collection of nonoverlapping sets $\{A_1, A_2, \dots\}$ is given by a product of Poisson distributions, where the rate parameter λ_i of each Poisson is given exactly by the total length of the corresponding set A_i.

ii) The process specified by choosing N according to a Poisson distribution with rate λ equal to the length of B, then throwing down N i.i.d. r.v.'s, uniformly distributed on B.

iii) The process specified by drawing i.i.d. exponential inter-point intervals x_i, and placing points at the points $\sum_{i<j} x_i$ in turn.

iv) The process specified as the limit of the Bernoulli process as the binwidth goes to zero. The Bernoulli process of binwidth b is defined by breaking the interval B into a collection of disjoint subintervals A_i, each of length b, then placing points in any given subinterval A_i by flipping an independent biased coin with $p(heads) = \min(1, b)$; if heads comes up on the i-th coin, then we place a point in the center of the subinterval A_i, and leave A_i empty otherwise.

Which of these definitions lends itself most easily to efficient sampling algorithms? Which of these definitions generalize to the case of an infinitely long observation interval B? To the multidimensional case? Give an example of a physical system that can be reasonably modeled as a two- or three-dimensional (e.g., a spatial or spatiotemporal) Poisson process.

B. Define the "hazard function" (aka the "intensity function") for a point process $N(t)$ as the conditional spike rate

$$h(t|N(s), s \le t_{i-1}) = \frac{p(N(t) = 1|N(s), s \le t_{i-1})}{1 - \int_{t_{i-1}}^{t} p(N(u) = 1|N(s), s \le t_{i-1}) du},$$

where t_{i-1} is the time of the last spike before the current time t and the probabilities on the right hand side are to be interpreted as the density of finding a spike at time t (or u).

i. Calculate the hazard function in the case of a renewal model whose interspike interval densities are given by a) a uniform density on the interval $[a, b], a, b > 0$; b) an exponential density with rate μ. What is the significance of the differences in $h(t)$ in these two cases?

ii. Characterize the class of point processes which are "memoryless," that is, whose hazard function $h(t|N(s), s \le t_{i-1})$ depends only on t, and not on the history of the process $N(s), s \le t_{i-1}$. Do we gain anything if we let $h(t)$ also depend on t_{i-1} (but not any of the history in the further past, $N(s), s < t_{i-1}$)?

3. Grab the data file c1p8.mat from http://people.brandeis.edu/∼abbott/book/exercises.html. This data is described in questions 8-9 of the neural coding exercises from the Theoretical Neuroscience book, online at http://people.brandeis.edu/∼abbott/book/exercises/c1/c1.pdf.

A. (optional) Compute the STA, correlation-corrected STA, regularized STA, the vector corresponding to the largest positive change in the spike-triggered covariance, and the vector corresponding to the largest negative change. Plot the nonlinearities corresponding to each of these vectors. Of the LNP models constructed by appending these linear filters and corresponding nonlinearities, which model assigns the highest likelihood to a set of held-out test data?

B. Construct a statistical test based on the time-rescaling idea: by rescaling time properly (where the proper time-rescaling depends both on your model of the spike train and on the history of the observed spike train), we can transform (basically) any point process into a Poisson process. Can you statistically distinguish the rescaled point processes corresponding to each of the models fit above from a Poisson process? Remember, the standard Poisson process has two key properties you might want to exploit: the isi's are 1) independent and 2) exponentially distributed with mean one. Which of the models you fit above pass the time-rescaling test? Of the models that fail the test (that is, whose rescaled isi's are statistically distinguishable from a Poisson process), discuss what you think the problems might be, and how to fix them.

C. Compute the MLE based on an exponential nonlinearity, with and without regularization, then with and without spike-history terms. Which estimator does the best job of predicting held-out data? Which models pass your time-rescaling test? Draw sample spike trains from the fit models; which bear the closest resemblance to the true data in terms of visual appearance, mean firing rate, and interspike interval properties? Discuss where each model fails and why; can you modify any of the models we have used here to correct any problems you see?

References

[James and Stein, 1960] James, W. and Stein, C. (1960). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:361–379.