POINT PROCESS ANALYSIS VIA GENERALIZED REGRESSION (LECTURE 4)

Rob Kass

Department of Statistics
Machine Learning Department
Center for the Neural Basis of Cognition
Carnegie Mellon University

Outline:

1. Generalized regression
2. Application to point processes---and spike trains

ordinary linear regression model

$$
\begin{aligned}
Y_i &= \mu_i + \epsilon_i \\
\mu_i &= \beta_0 + \beta_1 x_i
\end{aligned}
$$

$$\epsilon_i \sim N(0, \sigma^2).$$

generalized regression model

$$
\begin{aligned}
Y_i &\sim p(y_i | \theta_i) \\
\theta_i &= f(x_i)
\end{aligned}
$$

ordinary linear regression

$$\begin{aligned} Y_i &= \mu_i + \epsilon_i \\ \mu_i &= \beta_0 + \beta_1 x_i \end{aligned}$$

nonparametric regression

$$Y_i = f(x_i) + \epsilon_i$$

generalized linear model

$$f(y|\eta(\theta)) = h(y)\exp(\eta(\theta)T(y) - B(\theta))$$

$$g(\mu_i) = \beta_0 + \beta_1 x_i$$

generalized regression model

$$\begin{aligned} Y_i &\sim p(y_i|\theta_i) \\ \theta_i &= f(x_i) \end{aligned}$$

logistic regression model

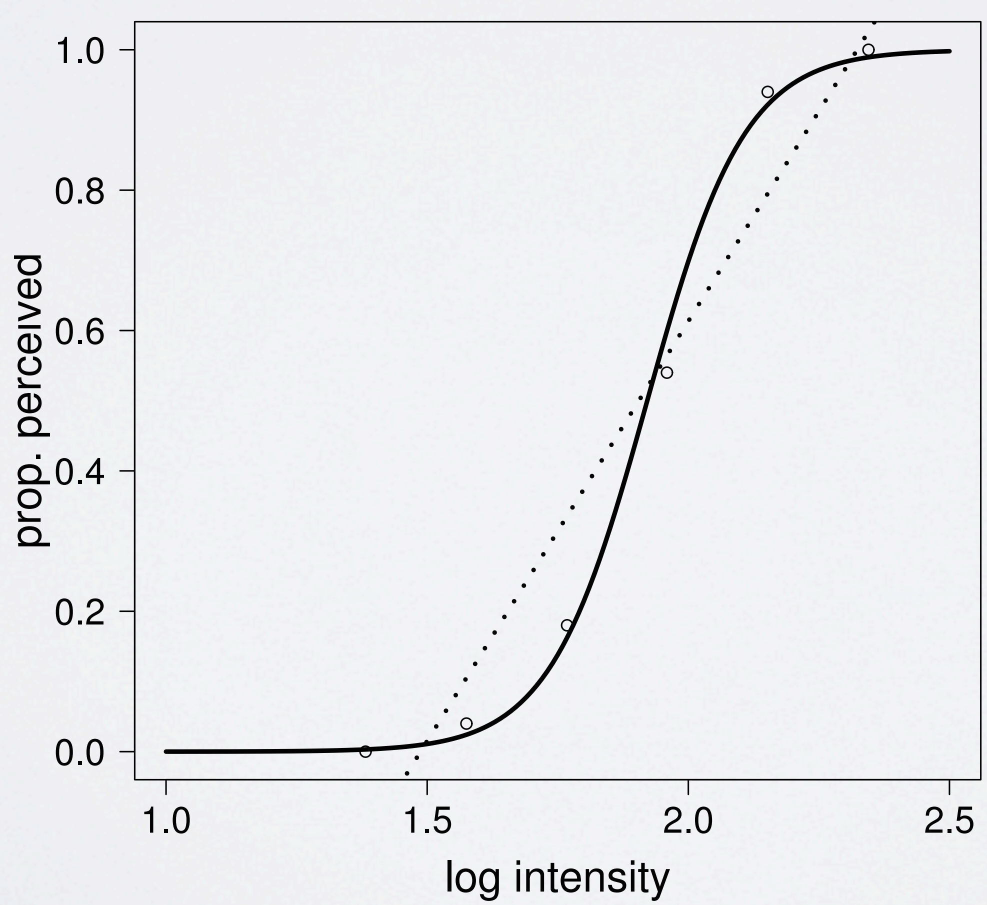
$$
\begin{aligned}
Y_i &\sim B(n_i, p_i) \\
p_i &= \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}.
\end{aligned}
$$

logistic regression model

$$\begin{aligned} Y_i &\sim B(n_i, p_i) \\ \log \frac{p_i}{1-p_i} &= \beta_0 + \beta_1 x_i. \end{aligned}$$

logistic regression model

$$
\begin{aligned}
Y_i &\sim B(n_i, p_i) \\
\log \frac{p_i}{1-p_i} &= \beta_0 + \beta_1 x_i.
\end{aligned}
$$



Poisson regression model

$$
\begin{aligned}
Y_i &\sim P(\lambda_i) \\
\lambda_i &= \exp(\beta_0 + \beta_1 x_i).
\end{aligned}
$$

$$\log \lambda_i = \beta_0 + \beta_1 x_i.$$

$$
\begin{aligned}
Y_i &\sim P(\lambda_i) \\
\lambda_i &= \exp(\beta_0 + \beta_1 x_i).
\end{aligned}
$$

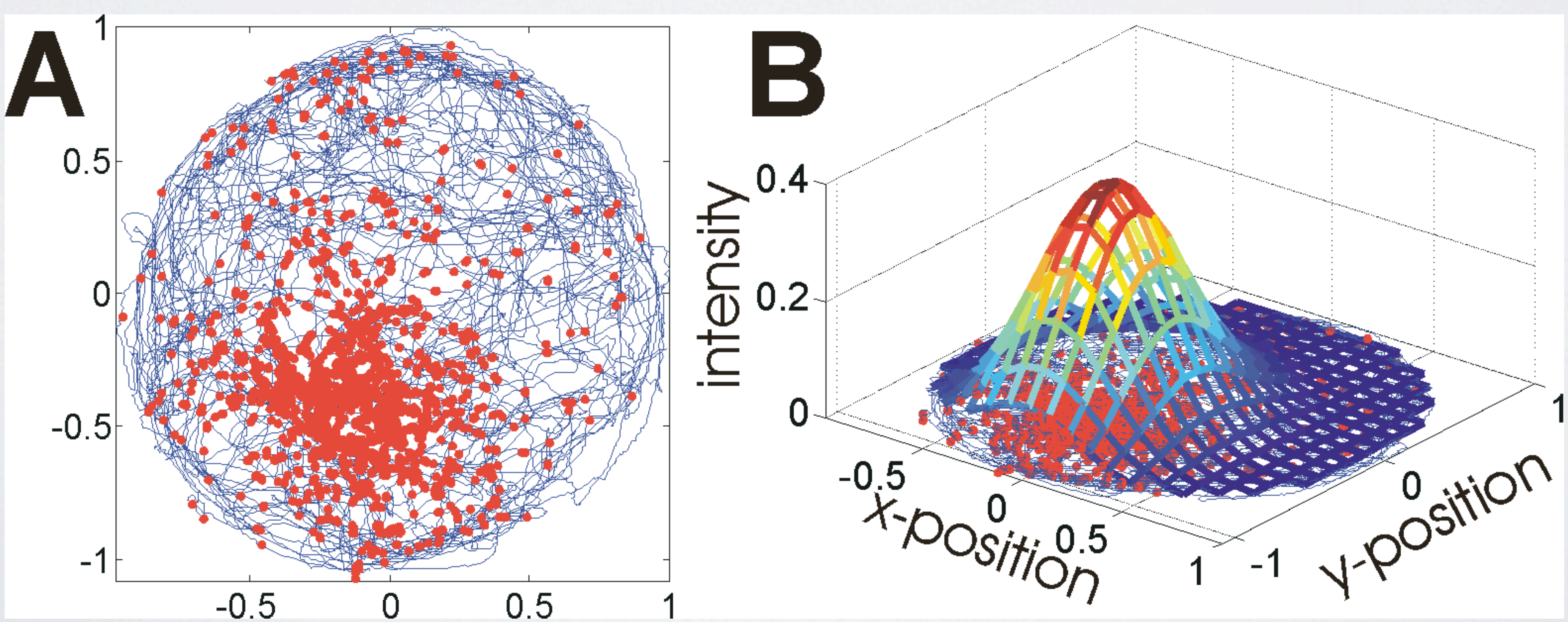
$$\log \lambda_i = \beta_0 + \beta_1 x_i.$$

Example 14.2.1 Directional sensitivity of an SEF neuron

```
left
9 6 9 9 6 6 8 5 7 9 4 8 8 3 6
up
2 0 6 4 4 0 0 0 5 2 1 0 3 0
right
4 8 2 2 4 0 3 4 1 1 0 3 4 0 2
down
1 5 1 2 0 4 4 4 4 3 6 1 1 1
```

14.4.2 Generalized linear models often involve multiple explanatory variables.

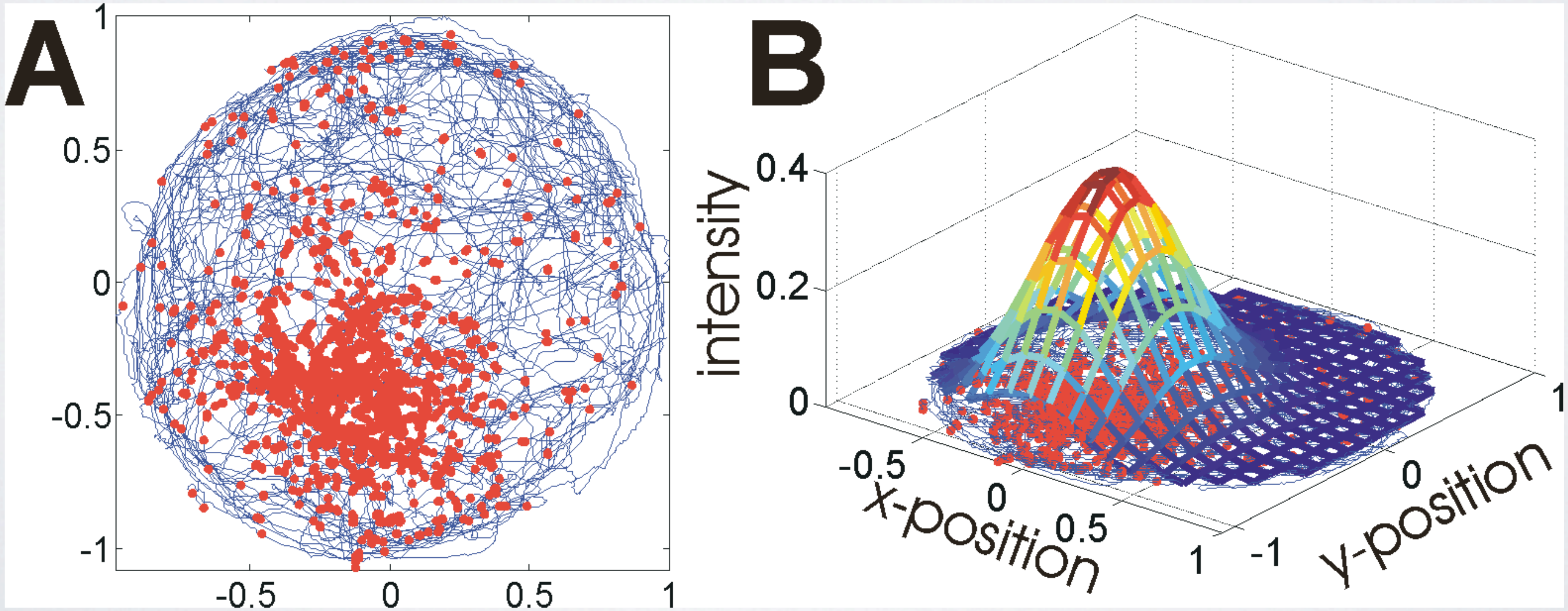
Example 14.4.1 A Poisson regression model for a hippocampal place cell Neurons in rodent hippocampus have spatially specific firing properties, whereby the spiking intensity is highest when the animal is at a specific location in an environment, and falls off as the animal moves further away from that point. Such receptive fields are called *place fields*, and neurons that have such firing properties are called *place cells*. Panel A of Figure 14.5 shows an example of the spiking activity of one such place cell, as a rat executes a free-foraging task in a circular environment. The rat's path through this environment is shown in blue, and the location of the animal at spike times is overlain in red. It is clear that the firing intensity is highest slightly to the southwest of the center of the environment, and decreases when the rat moves away from this point.



14.4.2 Generalized linear models often involve multiple explanatory variables.

$$Y_t \sim P(\lambda_t)$$

$$\lambda_t = \exp\left\{\alpha - \frac{1}{2}\begin{pmatrix} x(t) - \mu_x & y(t) & - \mu_y \end{pmatrix}\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1}\begin{pmatrix} x(t) - \mu_x \\ y(t) - \mu_y \end{pmatrix}\right\}.$$



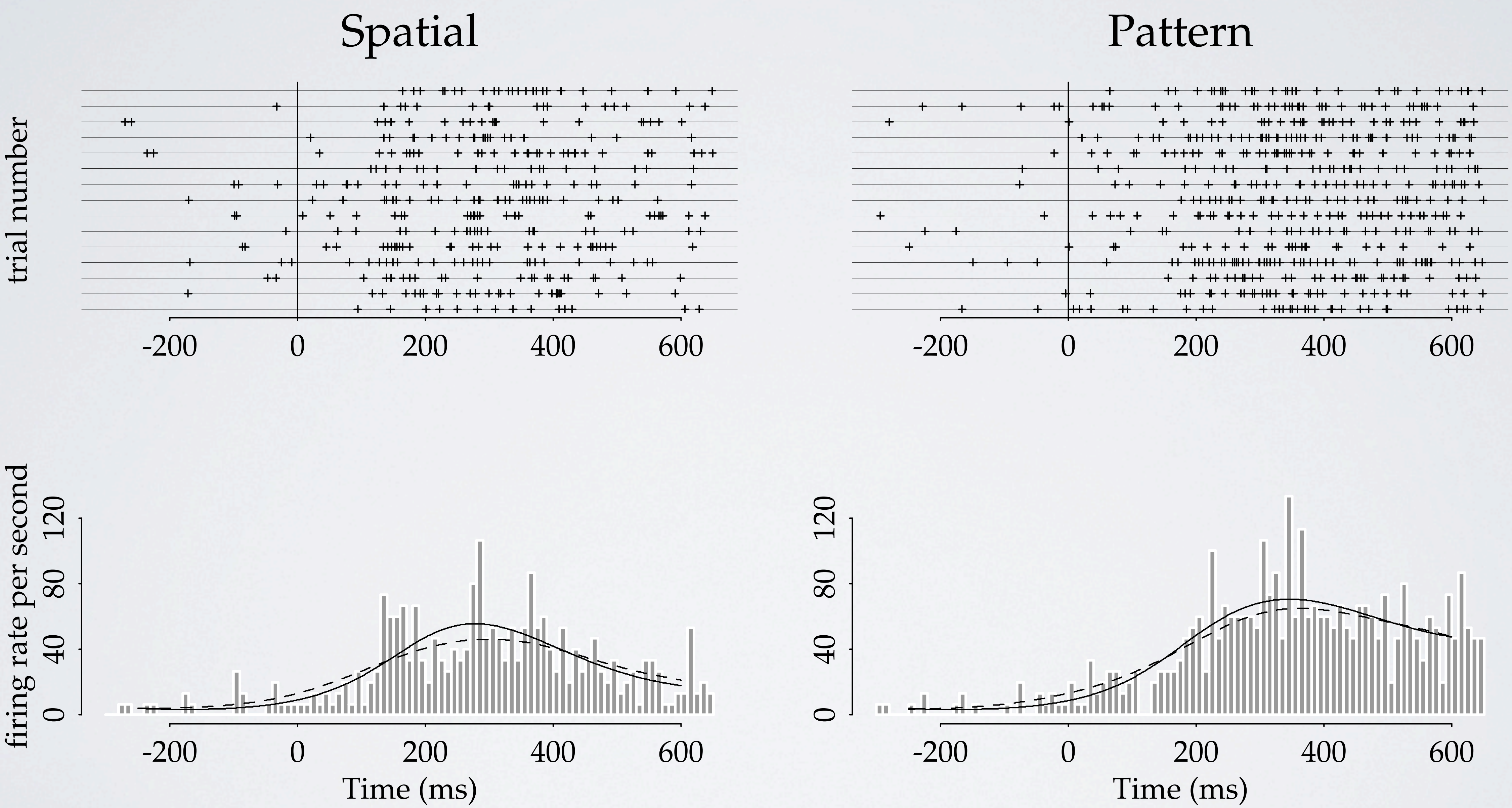
nonparametric regression model

$$Y_i = f(x_i) + \epsilon_i$$

generalized (nonparametric) regression model

$$\begin{aligned} Y_i &\sim p(y_i|\theta_i) \\ \theta_i &= f(x_i) \end{aligned}$$

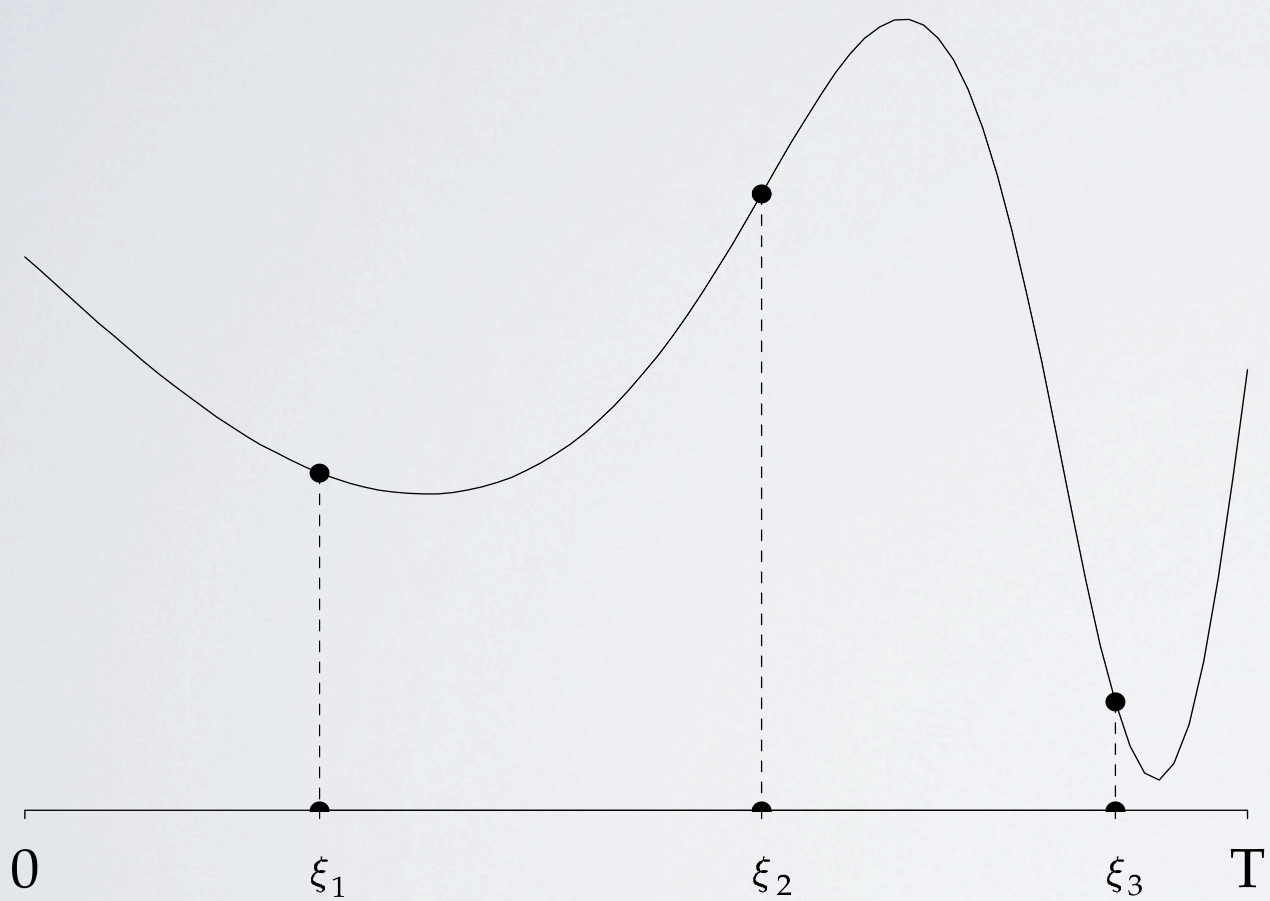
Two ideas: local fitting and basis functions



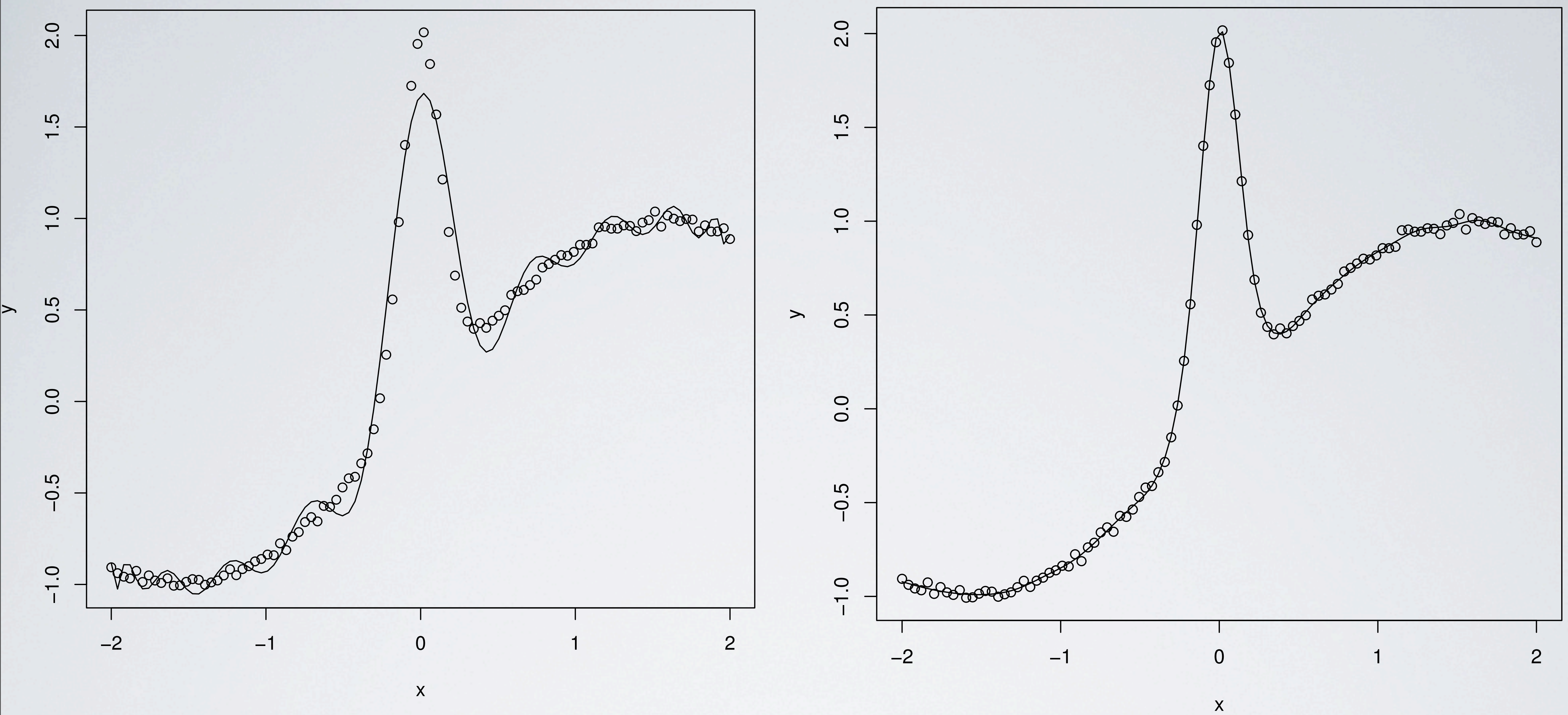
16.2.1 Splines may be used to represent complicated functions.

16.2.2 Splines may be fit to data using linear models.

$$
\begin{aligned}
f(x) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \\
&+ \beta_4 (x-\xi_1)_+^3 + \beta_5 (x-\xi_2)_+^3 + \cdots + \beta_{p+3} (x-\xi_p)_+^3
\end{aligned}
$$



here p=3 for 3 knots



$$(\xi_1, \xi_2, \ldots, \xi_7) = (-1.8, -.4, -.2, 0, .2, .4, 1.8).$$

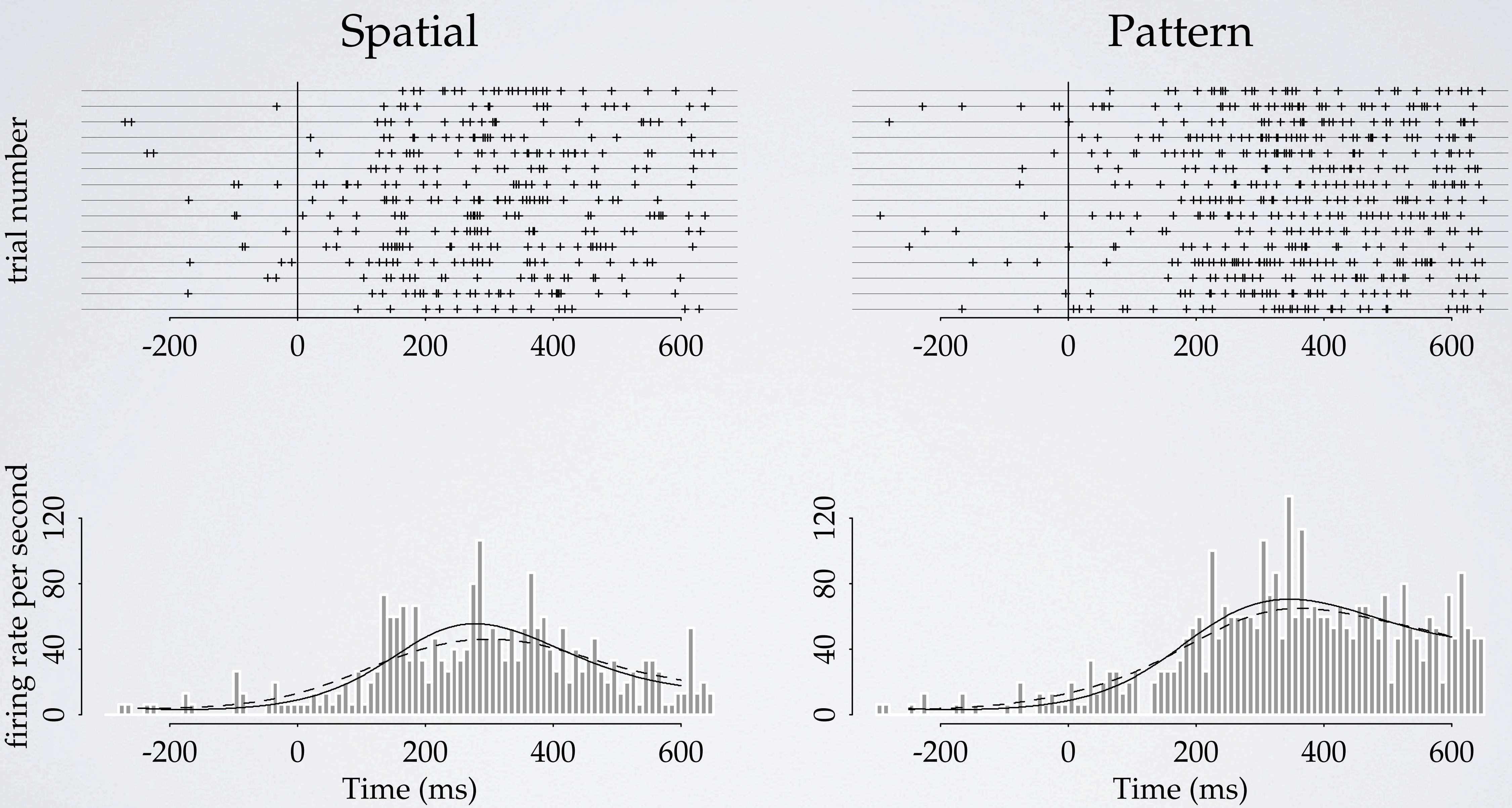
Figure 16.2: *Data simulated from function* $f(x) = \sin(x) + 2\exp(-30x^2)$ *together with twentieth-order polynomial fit (shown as line). Note that the polynomial is over-fitting (under-smoothing) in the relatively smooth regions of* $f(x)$*, and under-fitting (over-smoothing) in the peak. (In the data shown*

previous example: SEF PSTH

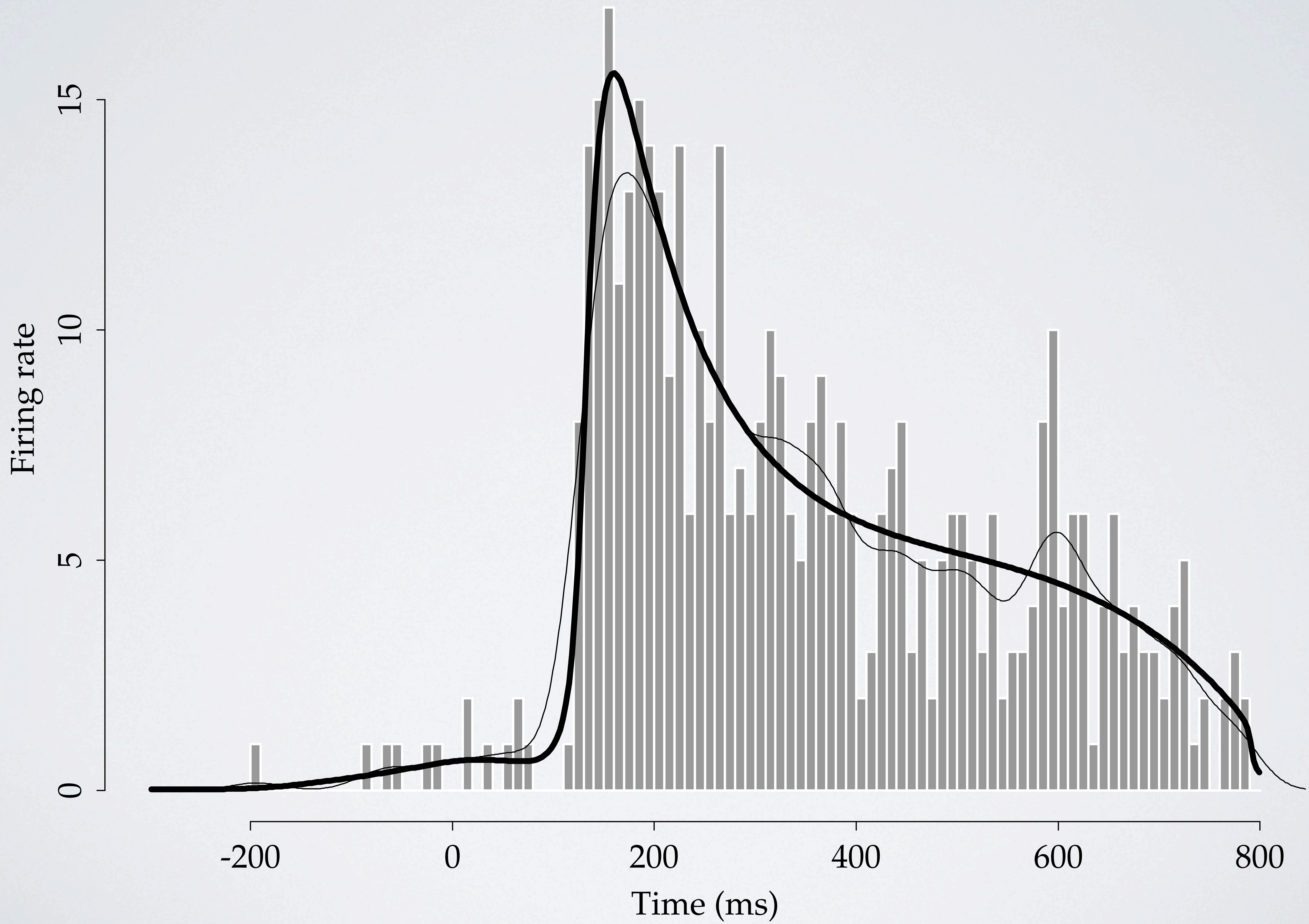
$$\begin{aligned} Y_i &\sim P(\lambda_i) \\ \log \lambda_i &= f(x_i) \end{aligned}$$

with $f(x)$ being a regression spline having knots at $-200, 200$. The fitted values $\hat{f}(x_i)$ were obtained using generalized linear model software and x_{max} was the value of x_i at which maximum among the $\hat{f}(x_i)$ values occurred.

Two ideas: local fitting and basis functions



BARS



$$
\begin{aligned}
f(x) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \\
&+ \beta_4 (x - \xi_1)_+^3 + \beta_5 (x - \xi_2)_+^3 + \cdots + \beta_{p+3} (x - \xi_p)_+^3
\end{aligned}
$$

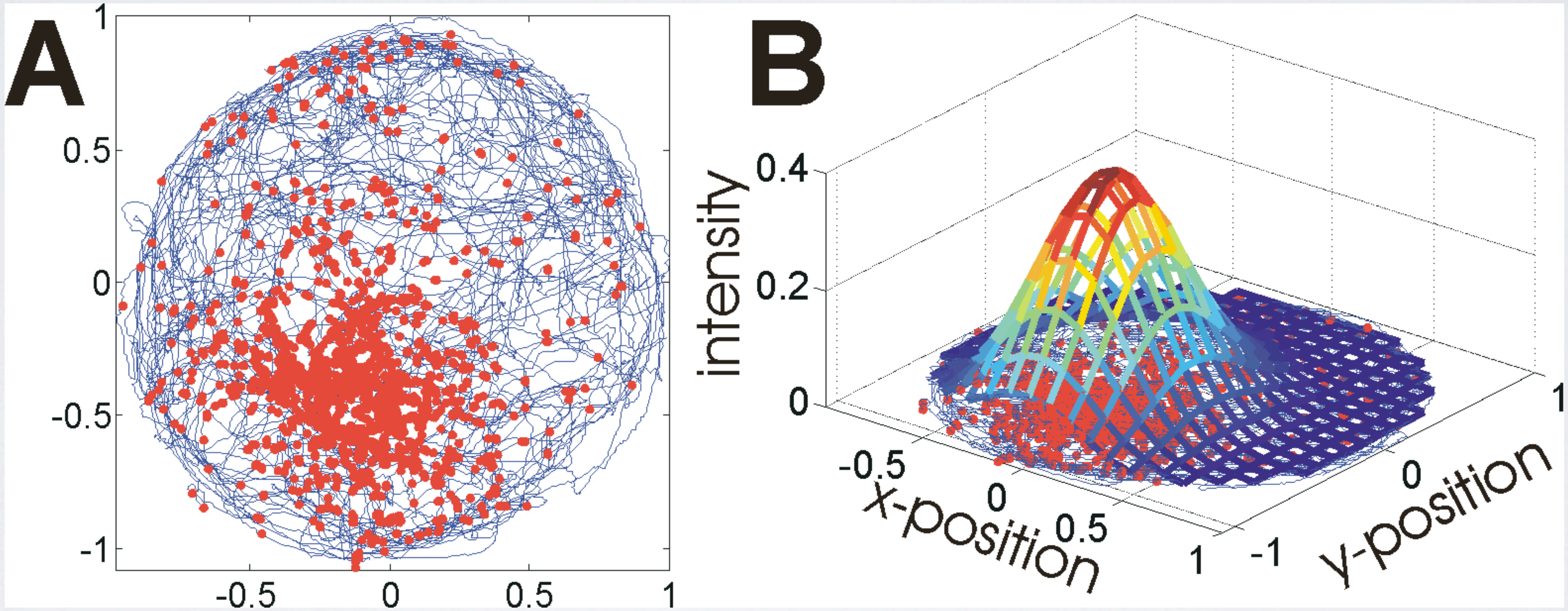
basis functions

Outline:

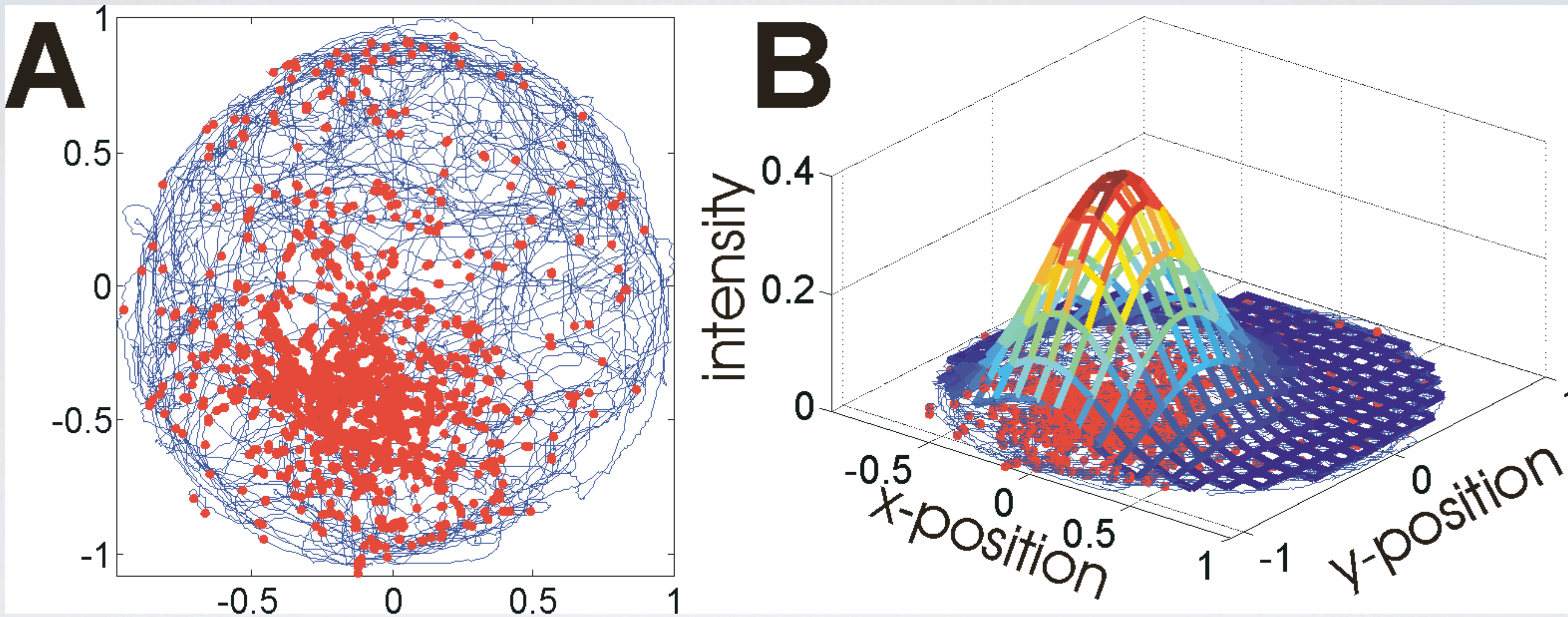
1. Generalized regression
2. Application to point processes---and spike trains

$$\log FR = \text{stimulus effects}$$

$$\log \lambda_t = \alpha - \frac{1}{2} \begin{pmatrix} x(t) - \mu_x & y(t) & - \mu_y \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1} \begin{pmatrix} x(t) - \mu_x \\ y(t) - \mu_y \end{pmatrix}$$



$$\log \lambda_t = \alpha - \frac{1}{2} \begin{pmatrix} x(t) - \mu_x & y(t) - \mu_y \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1} \begin{pmatrix} x(t) - \mu_x \\ y(t) - \mu_y \end{pmatrix}$$



previously: time bins (could be large)

here: instantaneous (small time bins)

(theoretically infinitesimal)

$$\log FR = \text{stimulus effects}$$

$\Rightarrow$ Poisson process
(no memory)

$$\log FR = \text{stimulus effects}$$

$$\Rightarrow$$ Poisson process (no memory)

$$\log FR = \text{stimulus effects} + \text{history effects}$$

$$\Rightarrow$$ non-Poisson process

homogeneous Poisson:

$$P(\text{event in } (t, t+dt]) = \lambda dt.$$

inhomogeneous Poisson:

$$P(\text{event in } (t, t+dt]) = \lambda(t) dt.$$

general:

$$P(\text{event in } (t, t+dt] | H_t) = \lambda(t|H_t) dt,$$

$$(\text{for Poisson: } \lambda(t|H_t) = \lambda(t))$$

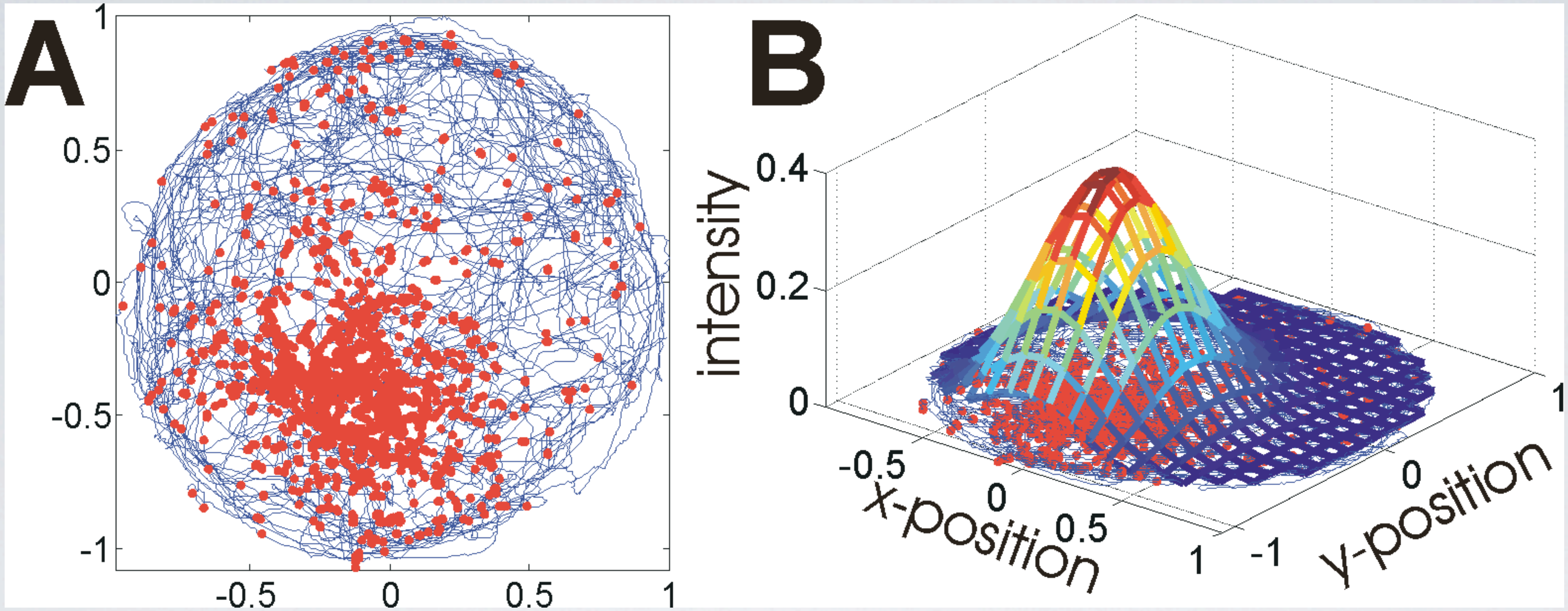
Theorem The event time sequence $S_1, S_2, \ldots, S_{N(T)}$ from a Poisson process with intensity function $\lambda(t)$ on an interval $(0, T]$ has joint pdf

$$f_{S_1,\ldots,S_{N(T)}}(s_1, \ldots, s_n) = \exp\left\{-\int_0^T \lambda(t)dt\right\} \prod_{i=1}^{n} \lambda(s_i). \qquad (19.6)$$

Theorem Let $\lambda(t)$ be a continuous function on $[0, T]$, set $\lambda_i = \lambda(t_i)$ for subinterval midpoints t_i, and let $p_i = (\Delta t)\lambda_i$. Then as $\Delta t \to 0$ we have

$$\frac{1}{(\Delta t)^n} \prod_{i=1}^{n} p_i^{x_i}(1-p_i)^{1-x_i} \to e^{-\int_0^T \lambda(t)dt} \prod_{i=1}^{n} \lambda(s_i). \qquad (19.8)$$

Therefore: can implement point process statistical modeling using binary generalized regression



binary data (small time bins)

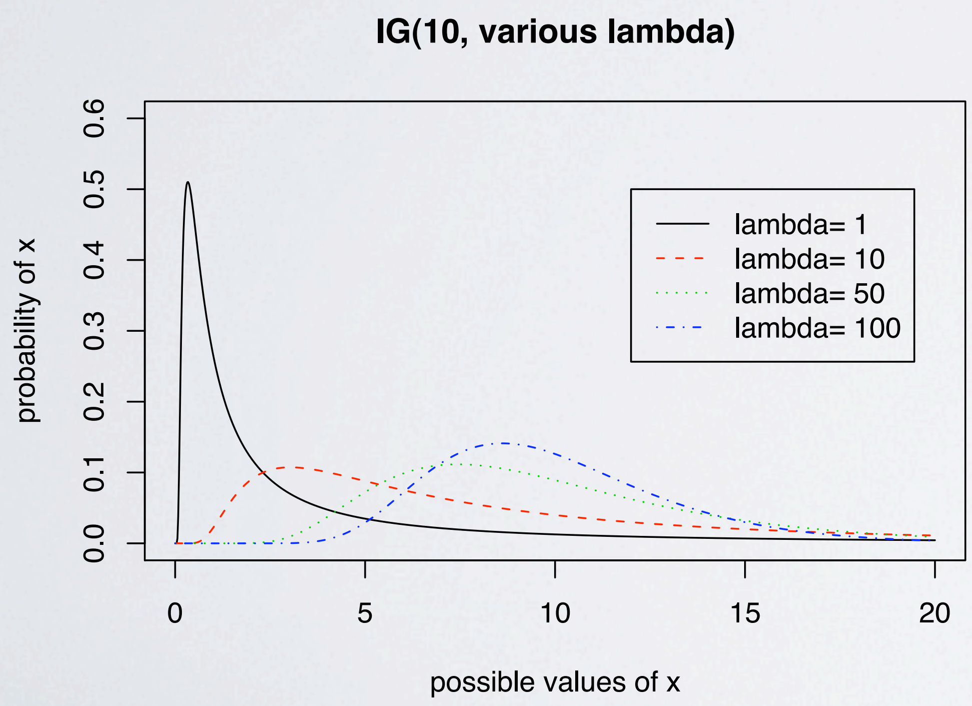
fit
using Poisson regression

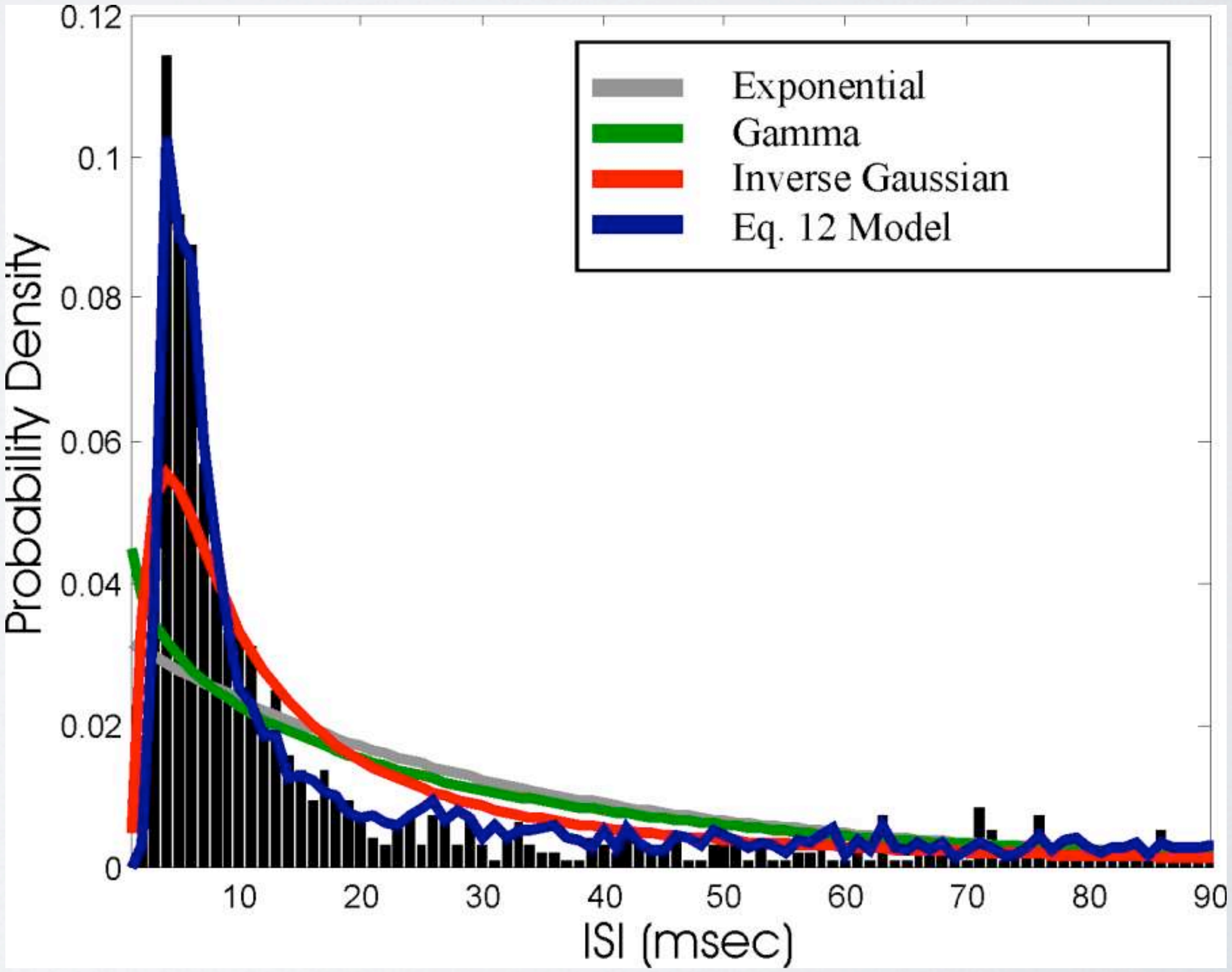
The ability to go back and forth between continuous time, where neuroscience and statistical theory reside, and discrete time, where measurements are made and data are analyzed, is central to statistical modeling of spike trains.

Non-Poisson Point Processes

19.3.1 Renewal processes have i.i.d. inter-event waiting times.

regular --> irregular --> bursty





Non-Poisson Point Processes

19.3.1 Renewal processes have i.i.d. inter-event waiting times.

Result The superposition of independent renewal processes having waiting times with continuous pdfs and finite means is, approximately, a Poisson process.

$$H_t = (s_1, s_2, \ldots, s_n)$$ observed spike times

$$H_t = (S_1, S_2, \ldots, S_{N(t-)})$$ random spike times

conditional intensity:

$$\lambda(t|H_t) = \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t]} = 1|H_t)}{\Delta t},$$

$$P(\text{event in } (t, t+dt]|H_t) = \lambda(t|H_t)dt,$$

Theorem The event time sequence $S_1, S_2, \ldots, S_{N(T)}$ from a Poisson process with intensity function $\lambda(t)$ on an interval $(0, T]$ has joint pdf

$$f_{S_1,\ldots,S_{N(T)}}(s_1, \ldots, s_n) = \exp\left\{-\int_0^T \lambda(t)dt\right\}\prod_{i=1}^{n}\lambda(s_i). \tag{19.6}$$

Theorem The event time sequence $S_1, S_2, \ldots, S_{N(T)}$ of an orderly point process on an interval $(0, T]$ has joint pdf

$$f_{S_1,\ldots,S_{N(T)}}(s_1, \ldots, s_n) = \exp\left\{-\int_0^T \lambda(t|H_t)dt\right\}\prod_{i=1}^{n}\lambda(s_i|H_{s_i}) \tag{19.17}$$

where $\lambda(t|H_t)$ is the conditional intensity function of the process.

For small Δt, the probability of an event in an interval $(t, t + \Delta t]$

$$P(\text{ event in } (t, t + \Delta t]|H_t) \approx \lambda(t|H_t)\Delta t \tag{19.19}$$

and the probability of no event is

$$P(\text{no event in } (t, t + \Delta t]|H_t) \approx 1 - \lambda(t|H_t)\Delta t. \tag{19.20}$$

If we consider the discrete approximation, analogous to the Poisson process case, we may define $p_i = \int \lambda(t|H_t)dt$ where the integral is over the ith time bin. We again get Bernoulli random variables Y_i with $P(Y_i = 1) = p_i$ but now these Y_i random variables are *dependent*, e.g., we may have $P(Y_i = 1|Y_{i-1} = 1) \neq p_i$. This is somewhat more complicated than the Poisson case, but it remains relatively easy to formulate history-dependent models for these Bernoulli trials. We give examples in Section 19.3.4.

---> together with pdf result implies we can again use Poisson regression

now define the unconditional or *marginal intensity function* as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t]} = 1)}{\Delta t}.$$

According to the law of total probability (page **??**), for a pair of random variables Y and X and an event A we have $P(X \in A) = E_Y(P(X \in A|Y))$. Letting H_t play the role of Y and $\Delta N_{(t,t+\Delta t]} = 1$ the role of $X \in A$, we get, similarly,

$$P(\Delta N_{(t,t+\Delta t]} = 1) = E_{H_t}\left(P(\Delta N_{(t,t+\Delta t]} = 1|H_t)\right)$$

(no memory)

$$\lambda(t) = E_{H_t}(\lambda(t|H_t)).$$

estimated by PSTH

If we consider spike trains to be point processses, within trials the instantaneous firing rate is $\lambda(t|H_t)$ and we have

$$P(\text{spike in } (t, t+dt]|H_t) = \lambda(t|H_t)dt,$$

while the across-trial average firing rate is $\lambda(t)$ and we have

$$P(\text{spike in } (t, t+dt]) = \lambda(t)dt.$$

Example 19.1.1 (continued) Let us take time bins to have width $\Delta t =$ 1 ms and write $\lambda_k = \lambda(t_k|H_{t_k})$, where t_k is the midpoint of the kth time bin. Defining

$$\lambda_k = \exp\{\alpha_0 + \sum_{j=1}^{120} \alpha_j \Delta N_{k-j}\}, \qquad (19.24)$$

we get a model with 120 history-related covariates, each indicating whether or not a spike was fired in a 1 millisecond interval at a different time lag.

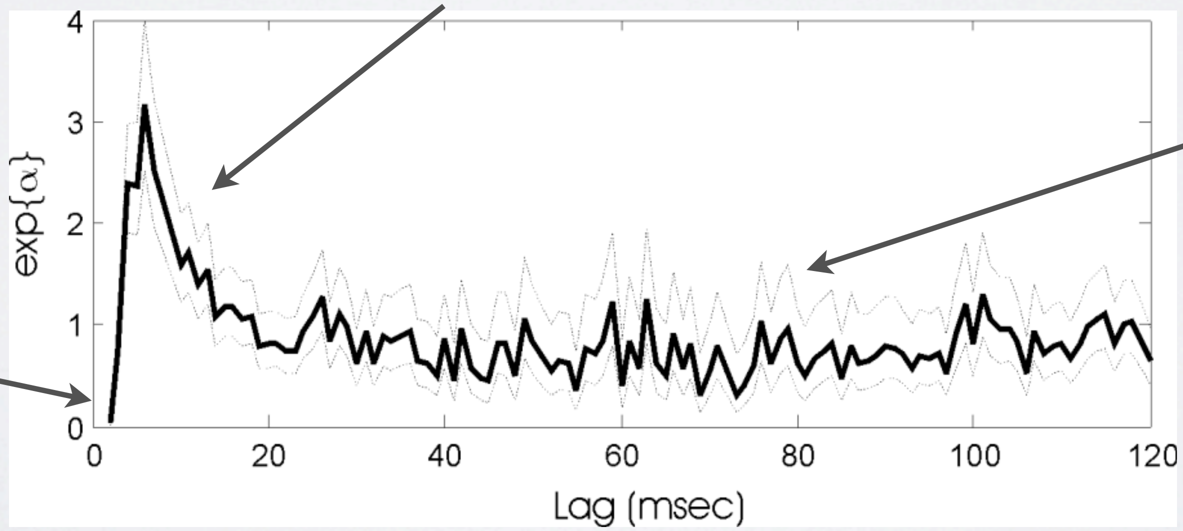
The parameter α_0 provides the log background firing rate in the absence of prior spiking activity within the past 120 milliseconds. Using Poisson regression with ML estimation (as in Chapter **??**) we obtained $\hat{\alpha}_0 = 3.8$ so that, if there were no spikes is the previous 120 milliseconds, the conditional intensity would become $\lambda_k = \exp(\hat{\alpha}_0) = 45$ spikes per second, corresponding to an average ISI of 22 ms. The MLEs $\hat{\alpha}_i$ obtained from the data are plotted in Figure 19.5, in the form $\exp\{\hat{\alpha}_i\}$. The values related to 0-2 ms after a

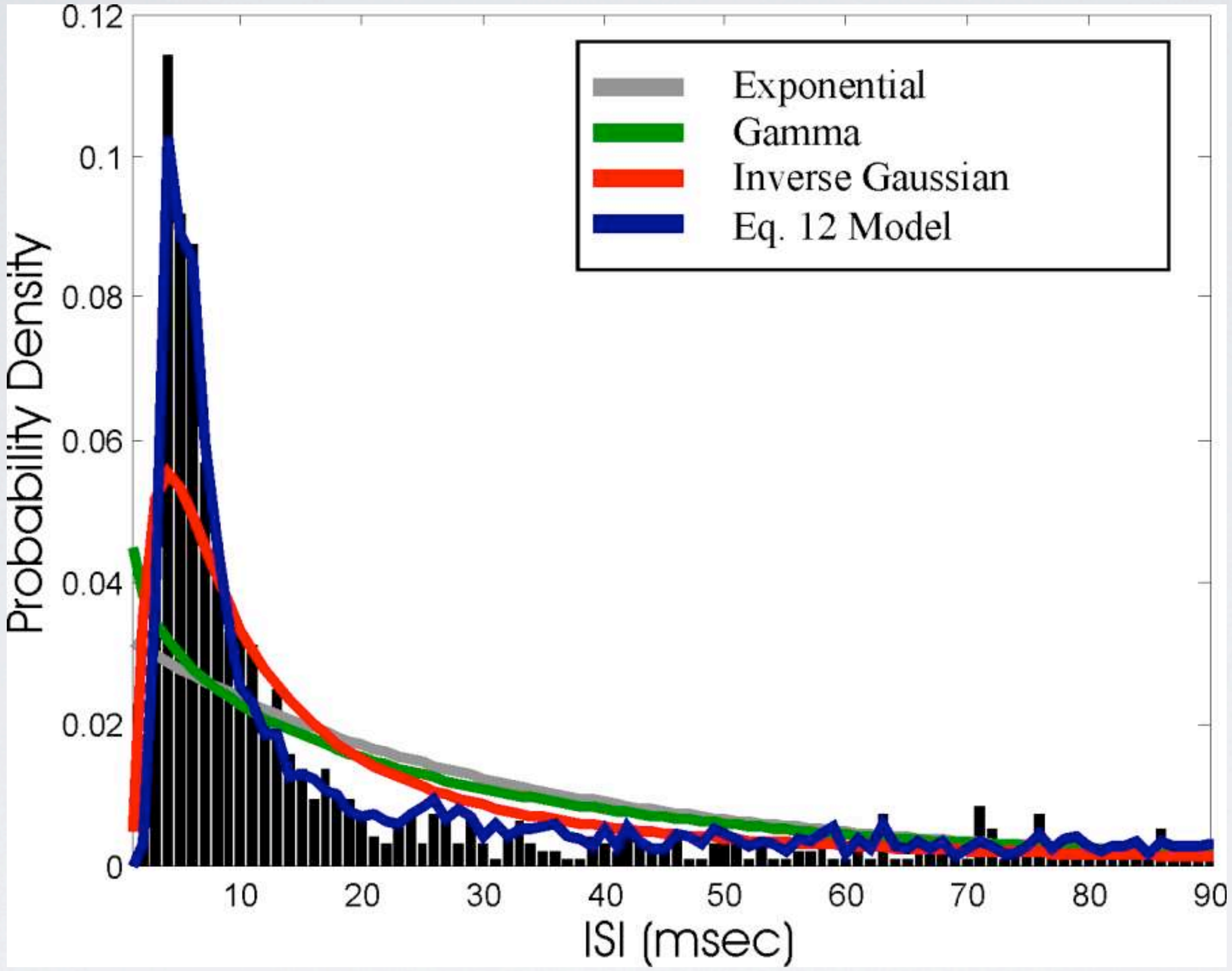
bursting

noise

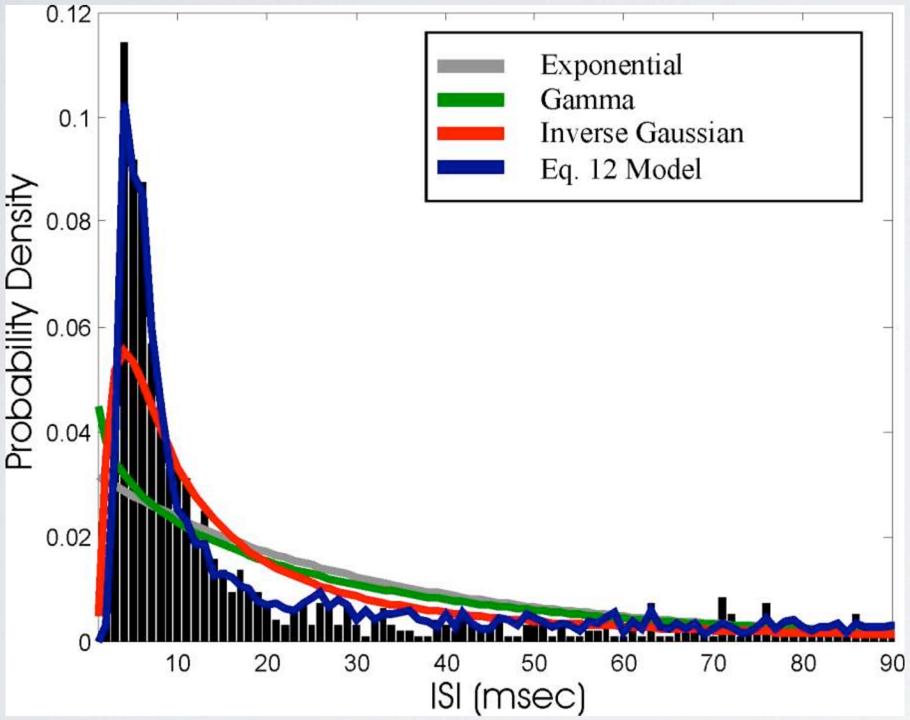
$\hat{\alpha}_j$ not significantly different than 0

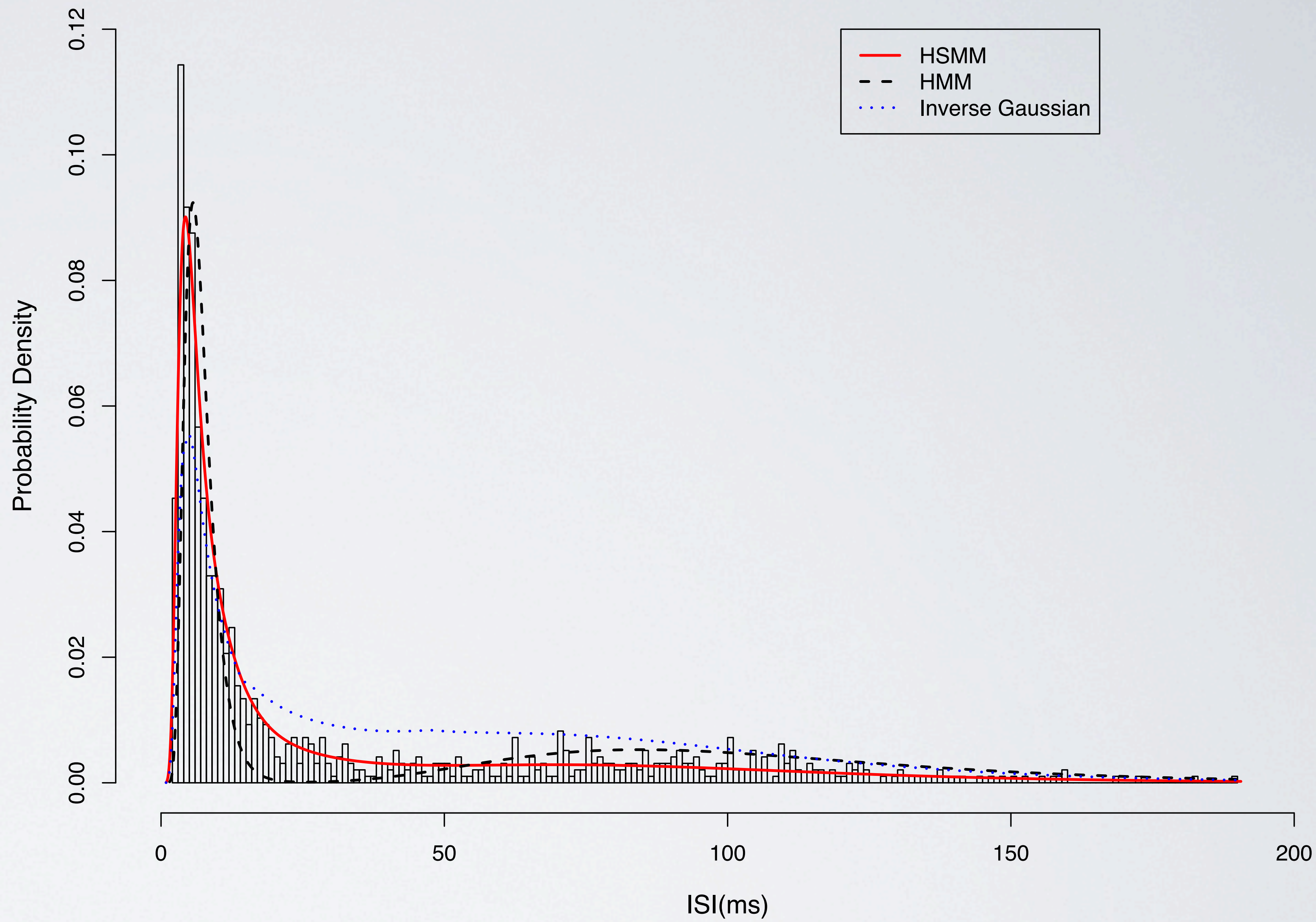
refractory period





Exponential
Gamma
Inverse Gaussian
Eq. 12 Model
Probability Density
ISI (msec)
0.12
0.1
0.08
0.06
0.04
0.02
0
10
20
30
40
50
60
70
80
90





Tokdar, Xi, Kelly, Kass (2010, *JCNS*)

event occurred. To get an inhomogeneous Poisson process, we retain the memorylessness but introduce a time-varying conditional intensity. A simple idea is to take a renewal process and, similarly, introduce a time-varying factor. For a renewal process, the probability of an event at time t depends on the timing of the most recent previous event $s_*(t)$, but not on any events prior to $s_*(t)$. If we allow the conditional intensity intensity to depend on both time t and the time of the previous event $s_*(t)$ we obtain a form

$$\lambda(t|H_t) = g(t, s_*(t)) \tag{19.25}$$

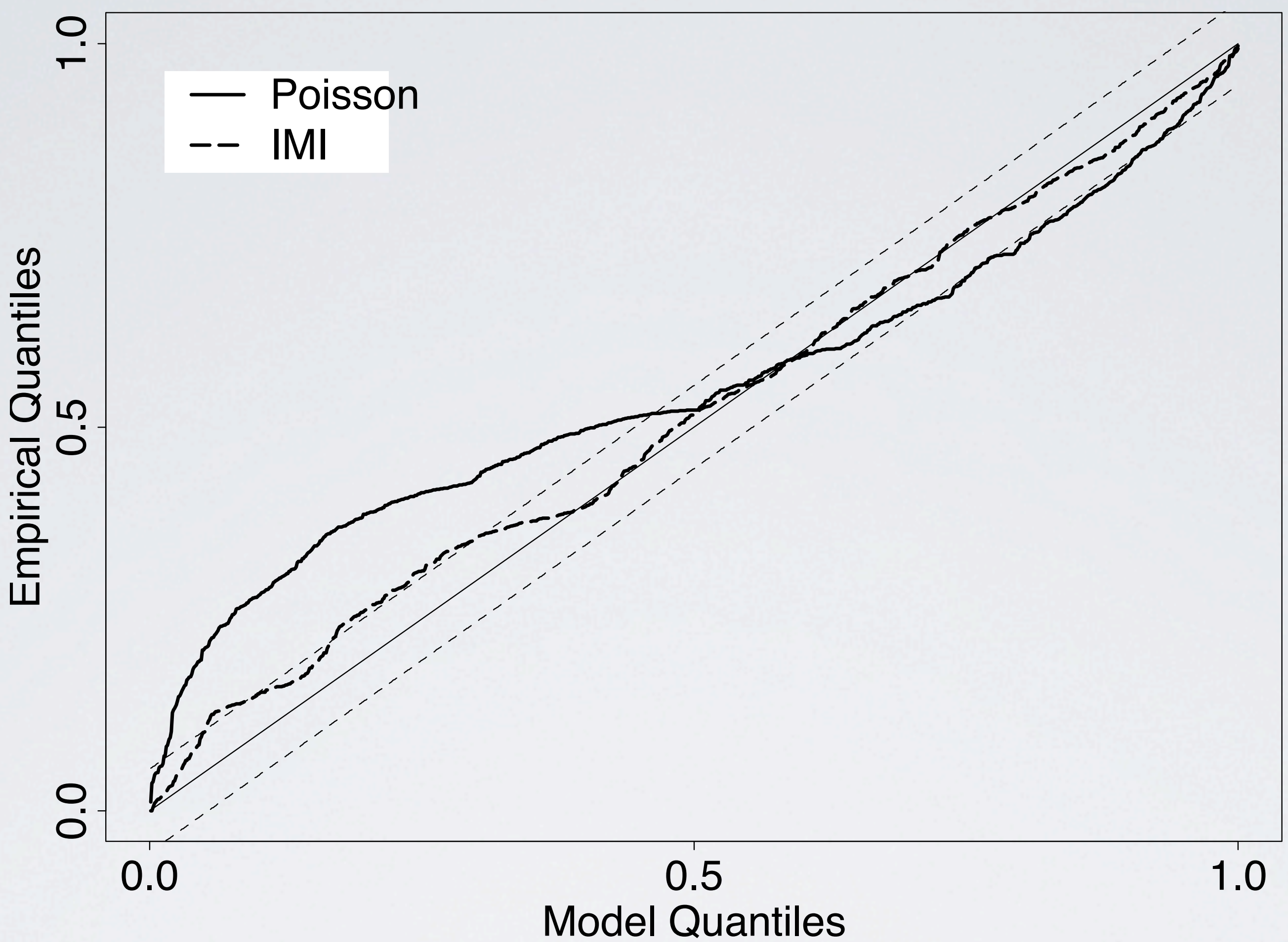
where $g(x, y)$ is a function to be specified. Models of this type are sometimes called *Markovian* or *Inhomogeneous Markov Interval* (IMI) models.[6] In an

(integrate-and-fire is special case)

important special case:

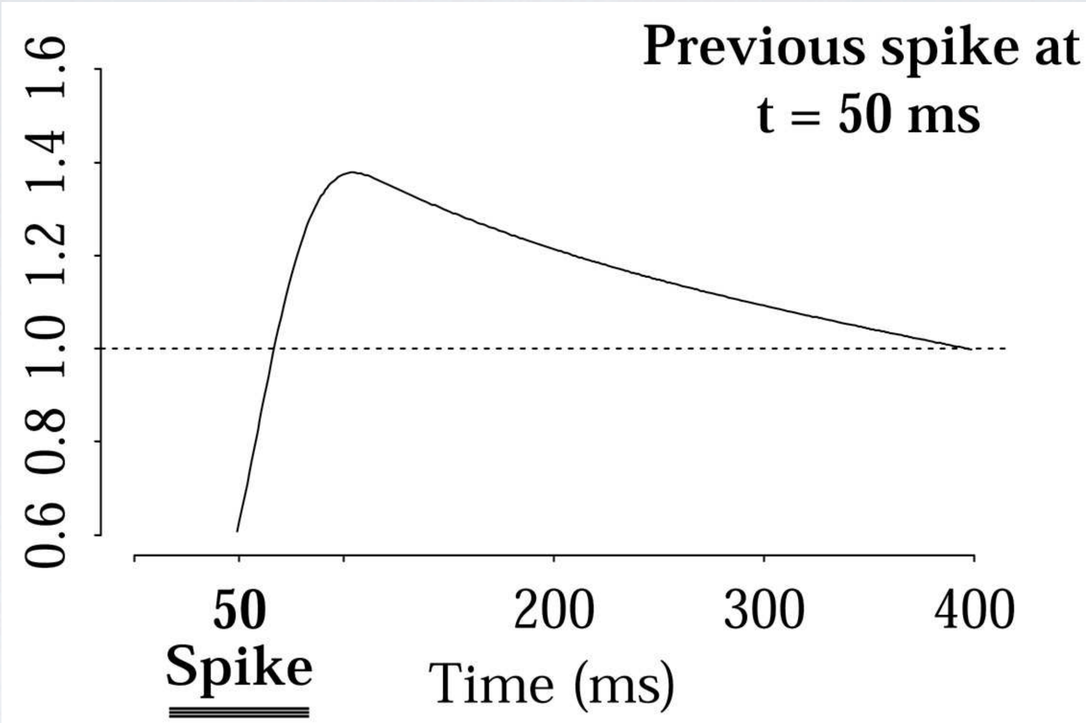
$$\lambda(t|H_t) = g_0(t)g_1(t - s_*(t)).$$

$$\log \lambda(t|H_t) = \log g_0(t) + \log g_1(t - s_*(t))$$



Poisson
IMI
Empirical Quantiles
Model Quantiles
0.0
0.5
1.0
0.0
0.5
1.0

$$\log \lambda(t|H_t) = \log g_0(t) + \log g_1(t - s_*(t))$$

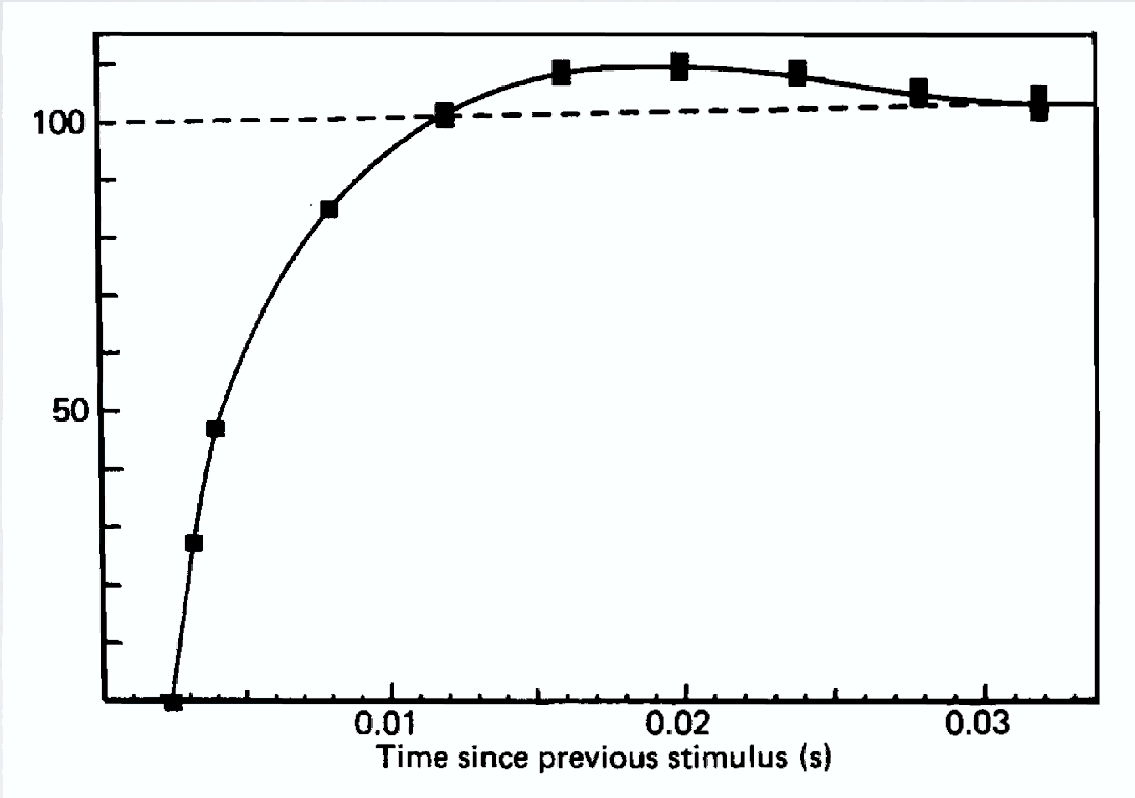


second term---history effects

"post-spike filter"

(Kass and Ventura, 2001)

Adrian and Lucas (1912)



STEPS IN ANALYSIS:

1. Formulate Poisson nonparametric regression model including choice of basis functions (often called "generalized linear model" --- in statistics "generalized additive model")
2. Fit via ML
3. Assess fit via P-P plots (using time rescaling)
4. Inference: interpret relevant estimates with CIs (via propagation of uncertainty or bootstrap), or decoding accuracy

Time Rescaling Theorem. Suppose we have a point process with conditional intensity function $\lambda(t|H_t)$ and with occurrence times $0 < S_1 < S_2, ..., < S_{N(T)} \leq T$. Let $Z_1 = \int_0^{S_1} \lambda(t|H_t)dt$, and

$$Z_i = \int_{S_{i-1}}^{S_i} \lambda(t|H_t)dt$$

for $j = 2, ..., N(T)$. Then $Z_1, \ldots, Z_{N(T)}$ are i.i.d. $Exp(1)$ random variables.

Proof: Omitted. $\square$

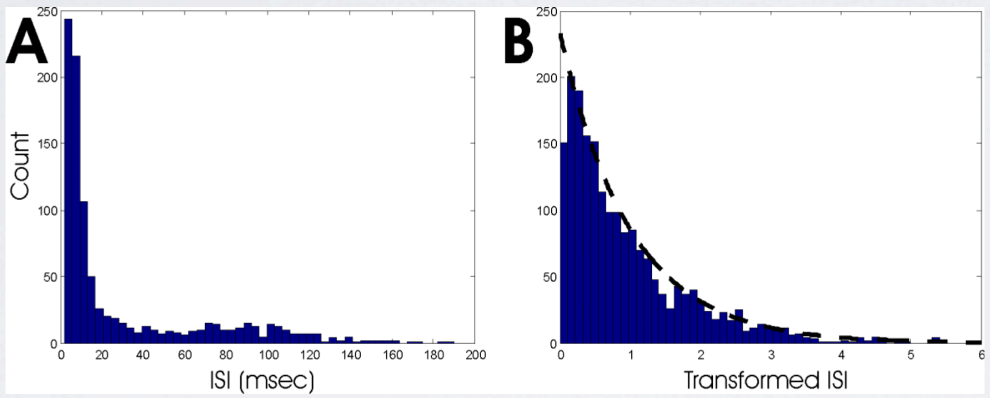


Figure 19.9: *Histograms of (A) ISIs and (B) time-rescaled ISIs for the retinal ganglion cell spike train. Dashed line in panel B is the* $Exp(1)$ *pdf.*

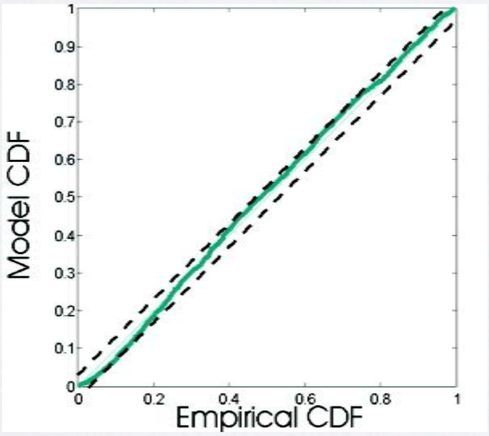


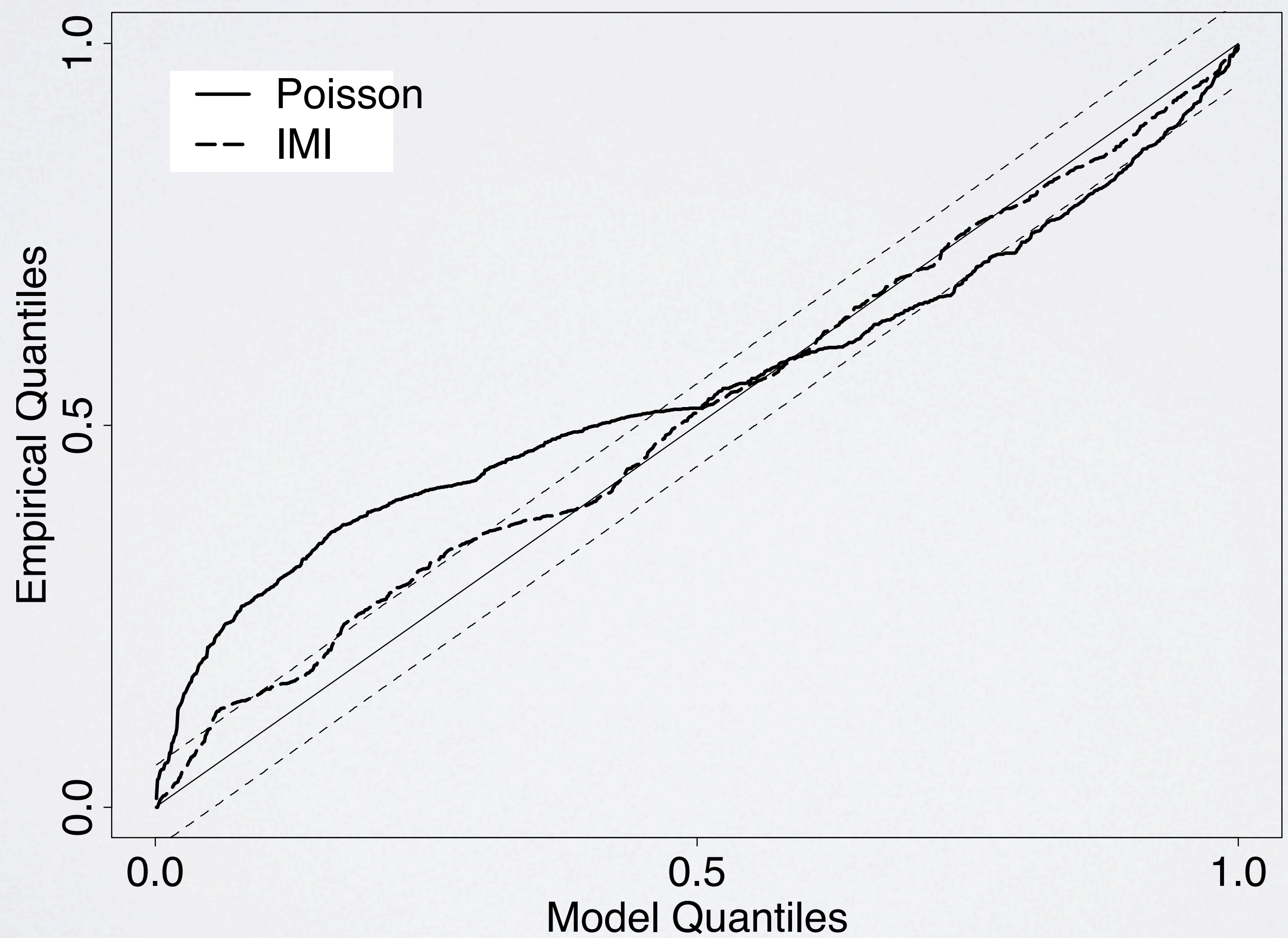
Figure 19.10: *P-P plot for the distribution of rescaled intervals shown in Figure 19.9.*

Time Rescaling Theorem. Suppose we have a point process with conditional intensity function $\lambda(t|H_t)$ and with occurrence times $0 < S_1 < S_2, ..., < S_{N(T)} \leq T$. Let $Z_1 = \int_0^{S_1} \lambda(t|H_t)dt$, and

$$Z_i = \int_{S_{i-1}}^{S_i} \lambda(t|H_t)dt$$

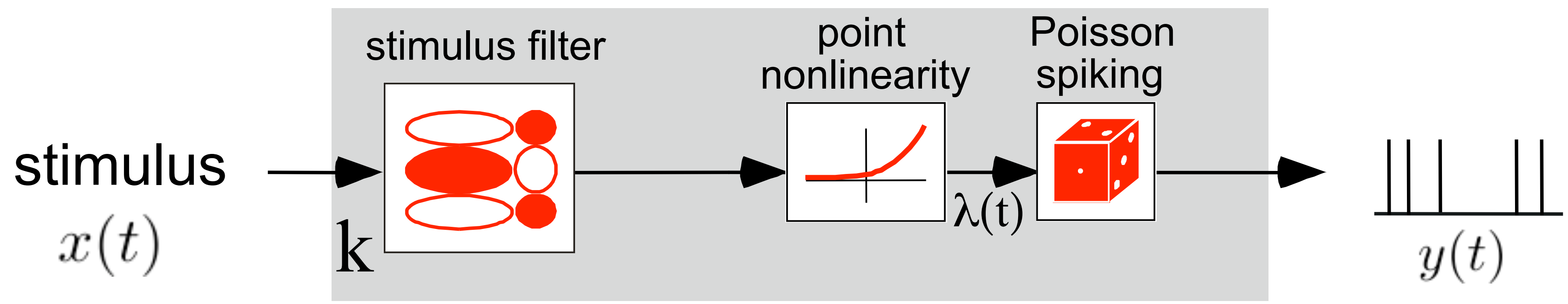
for $j = 2, ..., N(T)$. Then $Z_1, \ldots, Z_{N(T)}$ are i.i.d. $Exp(1)$ random variables.

Proof: Omitted. □



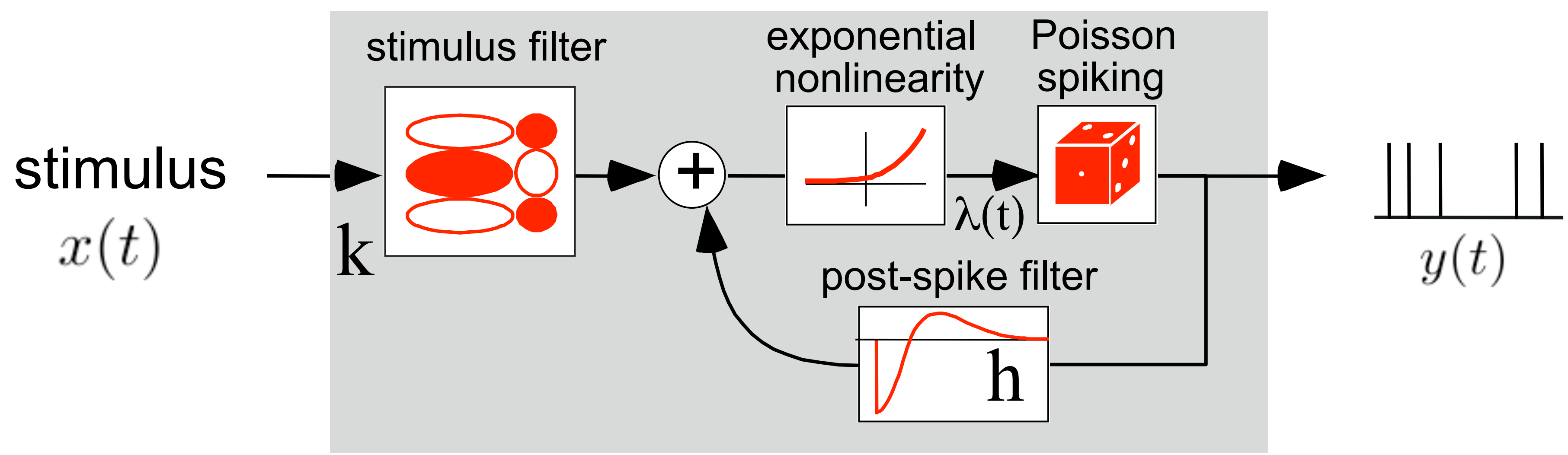
from Jonathan Pillow

Linear-Nonlinear-Poisson (GLM)



GLM with history-dependence

(Truccolo et al 04)



conditional intensity
(spike rate)

$$\lambda(t) = f(k \cdot x(t) \ + \ h \cdot y(t))$$
$$= e^{k \cdot x(t)} \ \cdot \ e^{h \cdot y(t)}$$

$$
\begin{aligned}
\log \lambda(t|H_t) &= \log g_0(t) + \log g_1(t - s_*(t)) \\
&+ \log g_2(t - s_{2*}(t)) + \log g_3(t - s_{3*}(t)) \\
&+ \cdots
\end{aligned}
$$

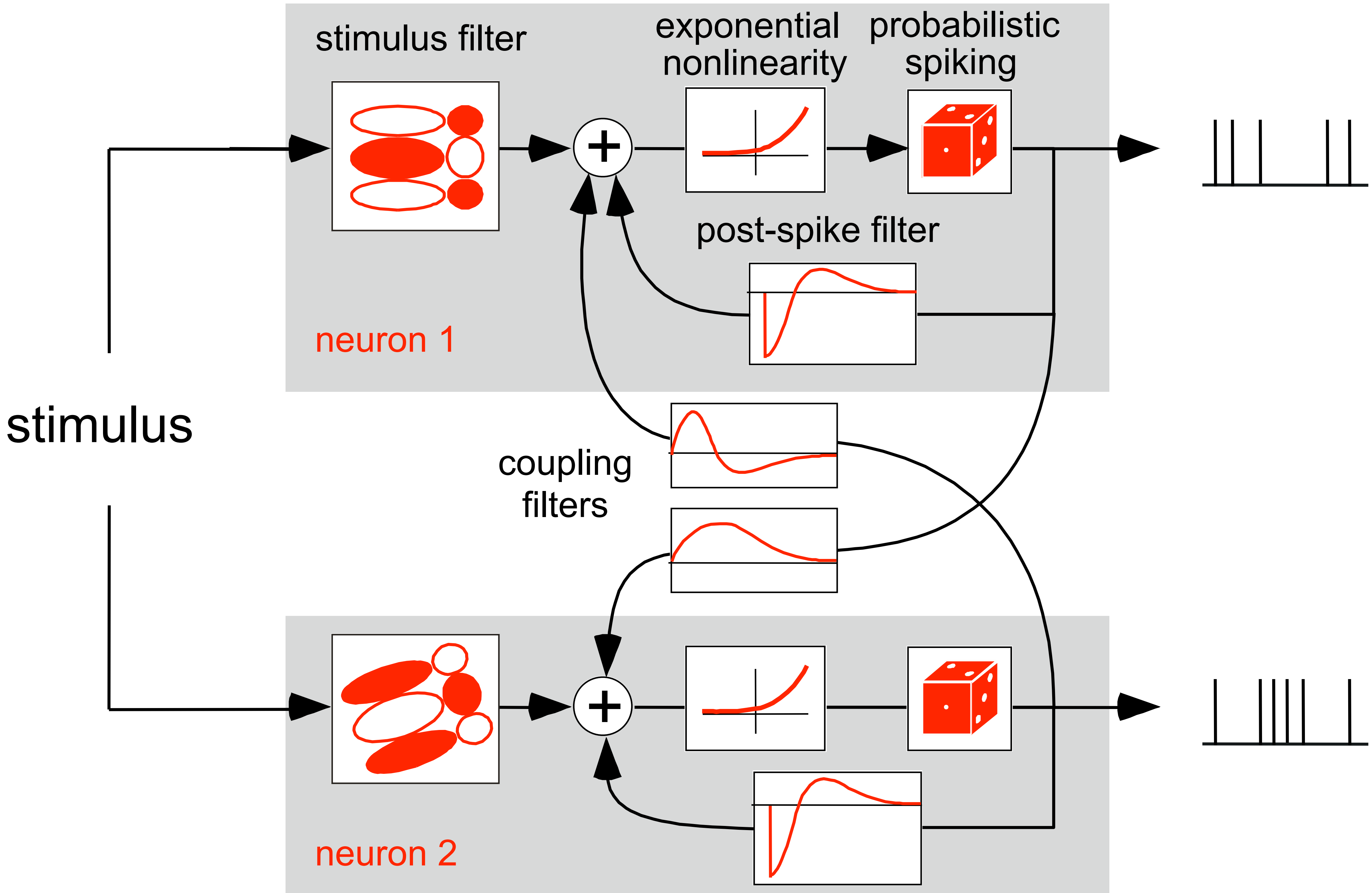
Multiple spike train analysis is analogous

$$
\begin{aligned}
\log \lambda(t|H_t) &= \log g_0(t) + \log g_1(t - s_*(t)) \\
&+ \log g_2(t - s_{2*}(t)) + \log g_3(t - s_{3*}(t)) \\
&+ \log h_1(t - u_*(t)) + \log h_2(t - u_{2*}(t)) \\
&+ \log h_3(t - u_{3*}(t)).
\end{aligned}
$$

coupling terms

$$\log FR = \text{stimulus effects} + \text{history effects} + \text{coupling effects}.$$

multi-neuron GLM



$$\log FR = \text{stimulus effects} + \text{history effects} + \text{coupling effects}.$$

Example 19.1.2 Spatiotemporal correlations in visual signalling To better understand the role of correlation among retinal ganglion cells, Pillow *et al.* (2008, *Nature*) examined 27 simultaneously-recorded neurons from an isolated monkey retina during stimulation by binary white noise. The authors used a model having the form of (19.2). They concluded, first, that spike times appear more precise when the spiking behavior of coupled neighboring neurons is taken into account and, second, that in predicting (decoding) the stimulus from the spike trains inclusion of the coupling term improved prediction by 20% compared with a method that ignored coupling and instead assumed independence among the neurons. $\square$