GENERALIZED LINEAR MODEL

Linear model

Linear model

Parameters

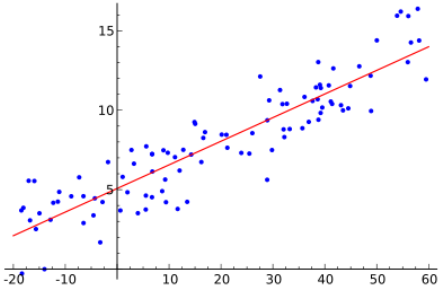
Measurement noise

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N\left(0,\sigma^2\right)$$

Linear predictor

Dependent variable

Explanatory variables



Probabilistic representation: A normal distribution

$$f\left(y \mid \mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{\sigma^2}},$$

where the mean is modeled as $\mu = X\beta.$

Linear predictor

Generalized linear model

What if the measurement noise follows non-Gaussian distribution?

GLM-analysis extends the regression analysis when the residual is given by an exponential family of distributions.

The exponential family of distributions include many popular distributions: normal, log-normal, exponential, gamma, chi-squared, beta, Dirichlet, Bernoulli, categorical, Poisson, inverse Gaussian, Wishart, Inverse Wishart, von Mises, etc.

Exponential family of distributions

$$f\left(y \mid \overset{\substack{\text{Parameter(s) of the family}\\ \Downarrow}}{\theta}\right) = h(y)\exp\left[\underset{\substack{\Uparrow\\ \text{Natural parameter}}}{\eta}(\theta)\overset{\substack{\text{(Sufficient statistics)}\\ \Downarrow}}{T}(y) - \underset{\substack{\Uparrow\\ \text{log-partition function}}}{A}(\theta)\right]$$

By re-parameterizing the distribution by η , an exponential family is simplified to a ‘natural form’ (canonical form) as

$$f\left(y \mid \eta\right) = h(y)\exp\left[\eta T(y) - A(\eta)\right]$$

Relation between a natural parameter and log-partition function:

$$\frac{\partial A(\eta)}{\partial \eta} = E_Y\left[T(Y)\right] \qquad \frac{\partial^2 A(\eta)}{\partial \eta^2} = E_Y\left[T(Y) - E_Y T(Y)\right]^2 = \mathrm{var}\left[T(Y)\right]$$

Examples

Bernoulli probability mass function.

$$f(y;\mu) = \mu^{y}(1-\mu)^{1-y} = \exp\left[y\log\frac{\mu}{1-\mu} + \log(1-\mu) \right] \quad y = \{0,1\}$$

Poisson probability mass function.

$$f(y;\mu) = \frac{\mu^{y}}{y!}e^{-\mu} = \frac{1}{y!}\exp[y\log\mu - \mu] \quad y = \{0,1,2,\cdots\}$$

Generalized linear model

Modeling the mean by a linear predictor.

$$f(y;\theta) = h(y)\exp\left[\eta(\theta)T(y) - A(\theta)\right]$$

Mean $\mu \equiv E_Y[Y]$ ↔ **Linear predictor** $X\beta$

Mean is bounded $0 < \mu < 1$ (Bernoulli)

$0 < \mu$ (Poisson)

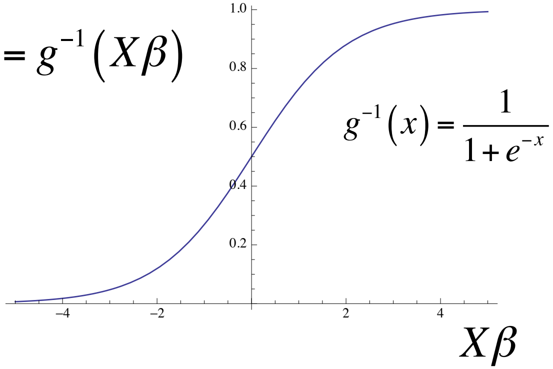
→ $\xcancel{\mu = X\beta}$

A link function relates the mean and a linear predictor:

$$\mu = g^{-1}(X\beta) \qquad X\beta = g(\mu)$$

so that μ is bounded.

$\mu = g^{-1}(X\beta)$



A link function can be selected arbitrary. However, there is a function that arises ‘naturally’, depending on the type of distribution.

Canonical link function: Bernoulli

Poisson probability mass function.

$$f(y;\mu) = \mu^{y}(1-\mu)^{1-y} = \exp\left[y\log\frac{\mu}{1-\mu} + \log(1-\mu) \right] \quad y = \{0,1\}$$

Linear predictor: $\eta = X\beta$

Model the mean using a link function: $\mu = g^{-1}(\eta)$

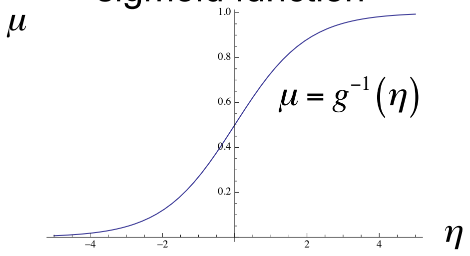
Canonical link function: (logit function)

If we use $\mu = g^{-1}(\eta) = \dfrac{1}{1+e^{-\eta}},$

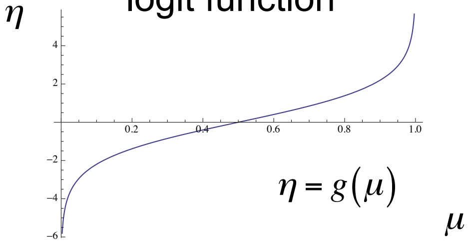
namely $\eta = g(\mu) = \log\dfrac{\mu}{1-\mu},$

then the distribution is written in a natural form as

$$f(y;\eta) = \exp\left[y\eta - A'(\eta) \right]$$



sigmoid function



logit function

Canonical link function: Poisson

Bernoulli probability mass function.

$$f(y;\mu)=\frac{\mu^y}{y!}e^{-\mu}=\frac{1}{y!}\exp\left[y\log\mu-\mu\right] \qquad y=\{0,1,2,\cdots\}$$

Linear predictor: $\eta = X\beta$

Model the mean using a link function: $\mu = g^{-1}(\eta)$

Canonical link function: (log function)

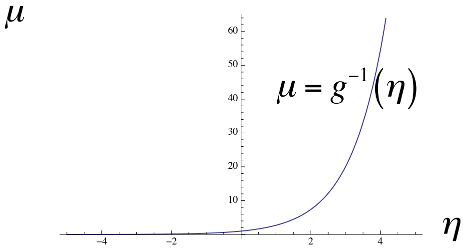
If we use $\mu = g^{-1}(\eta) = e^{\eta}$,

namely $\eta = g(\mu) = \log\mu$,

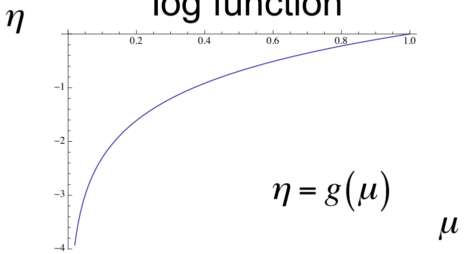
then the distribution is written in a natural form as

$$f(y;\eta)=\frac{1}{y!}\exp\left[y\eta - e^{\eta}\right]$$

Exp. function



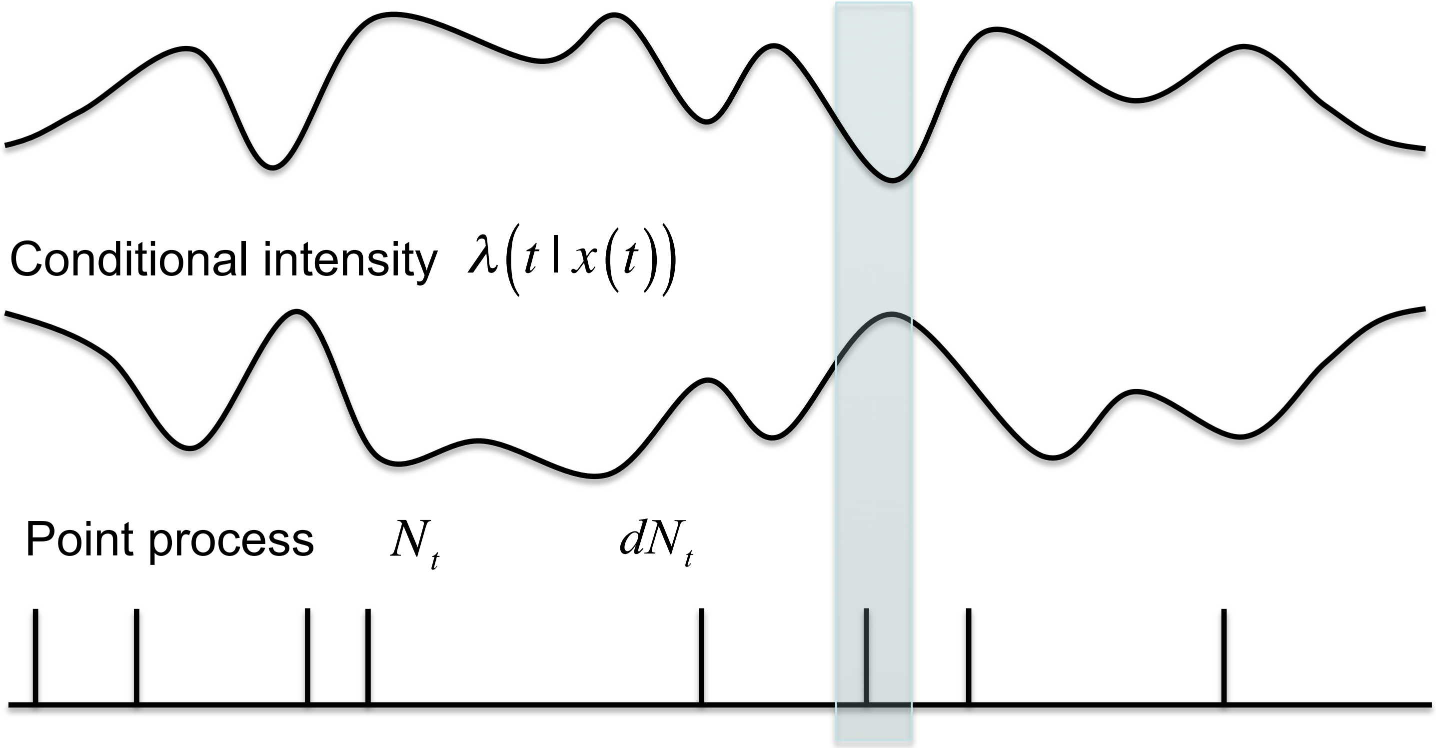
log function



Relating a stimulus input/motor output with neural activity

POINT PROCESS-GLM

Relating stimulus and neural activity



Stimulus signal $x(t)$

Conditional intensity $\lambda\left(t \mid x(t)\right)$

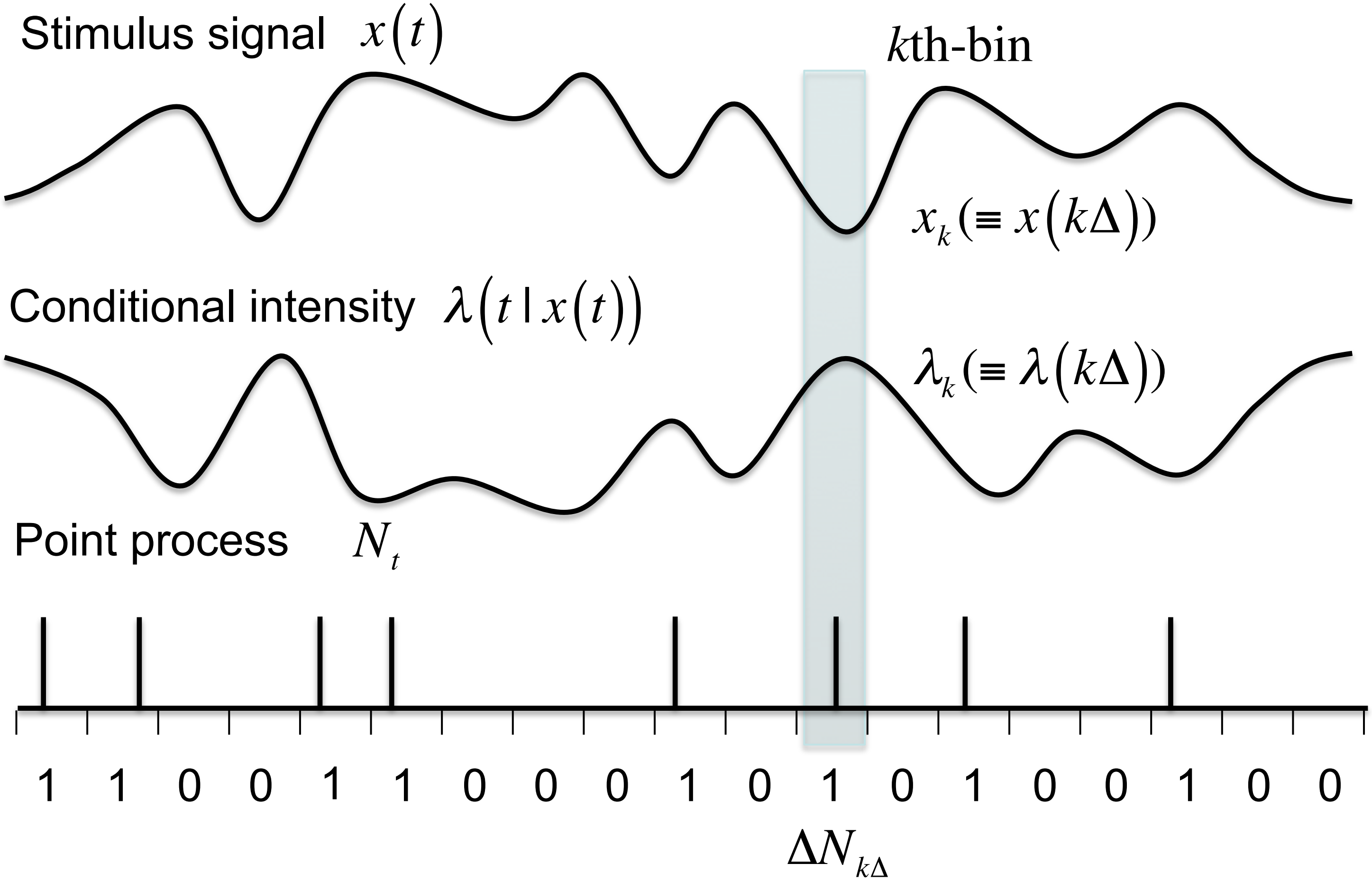
Point process N_t dN_t

t $t+\Delta$

$$P\left(\text{a spike in } [t,t+\Delta) \mid x(t)\right) = \lambda\left(t \mid x(t)\right)\Delta + o(\Delta)$$

Stimulus signal (Covariate)

Discretizing time into bins



Likelihood (Joint probability mass function)

$$P\left(\Delta N_{\Delta}, \Delta N_{2\Delta}, \cdots, \Delta N_{N\Delta}\right) = \prod_{k=1}^{N} P\left(\Delta N_{k\Delta} \mid \Delta N_{\Delta}, \cdots, \Delta N_{(k-1)\Delta}\right)$$

Conditional probability mass function

Bernoulli-GLM

Conditional Bernoulli probability:

$$P\left(\Delta N_{k\Delta} \mid \Delta N_{\Delta}, \cdots, \Delta N_{(k-1)\Delta}\right) = \left(\lambda_k \Delta\right)^{\Delta N_{k\Delta}} \left(1 - \lambda_k \Delta\right)^{1-\Delta N_{k\Delta}} \qquad \lambda_k \equiv \lambda\left(k\Delta \mid x(k\Delta)\right)$$

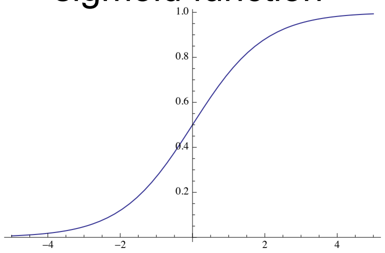
$$= \exp\left[\Delta N_{k\Delta} \log \frac{\lambda_k \Delta}{1 - \lambda_k \Delta} + \log\left\{1 - \lambda_k \Delta\right\}\right]$$

Mean of the Bernoulli probability at k-th bin: $\mu = E\left(\Delta N_{k\Delta}\right) = \lambda_k \Delta$

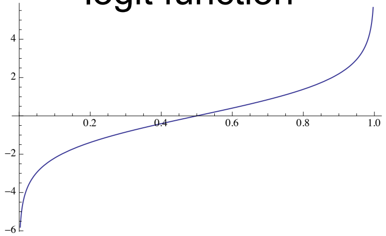
Modeling the mean response via a canonical link function:

$$\lambda_k \Delta = \frac{1}{1 + e^{-(\beta_0 + \beta x_k)}} \qquad \log \frac{\lambda_k \Delta}{1 - \lambda_k \Delta} = \beta_0 + \beta x_k$$

sigmoid function



logit function



misprint!

Poisson-GLM

Approximation by Poisson distribution:

$$P\left(\Delta N_{k\Delta} \mid \Delta N_{\Delta}, \cdots, \Delta N_{(k-1)\Delta}\right) = \left(\frac{\lambda_k \Delta}{1-\lambda_k \Delta}\right)^{\Delta N_{k\Delta}} \left(1-\lambda_k \Delta\right)$$

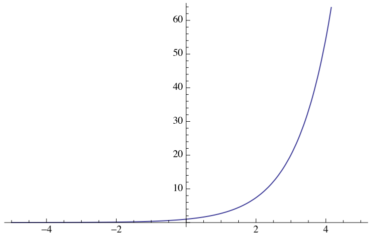
$$\approx \exp\left[\Delta N_{k\Delta} \log\left\{\lambda_k \Delta\right\} - \lambda_k \Delta\right]$$

Mean of the Poisson distribution at k-th bin: $\mu = E\left(\Delta N_{k\Delta}\right) = \lambda_k \Delta$

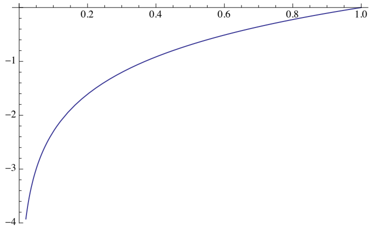
Modeling the mean response via a canonical link function:

$$\lambda_k \Delta = \exp\left[\beta_0 + \beta x_k\right] \qquad \log\left[\lambda_k \Delta\right] = \beta_0 + \beta x_k$$

Exp. function

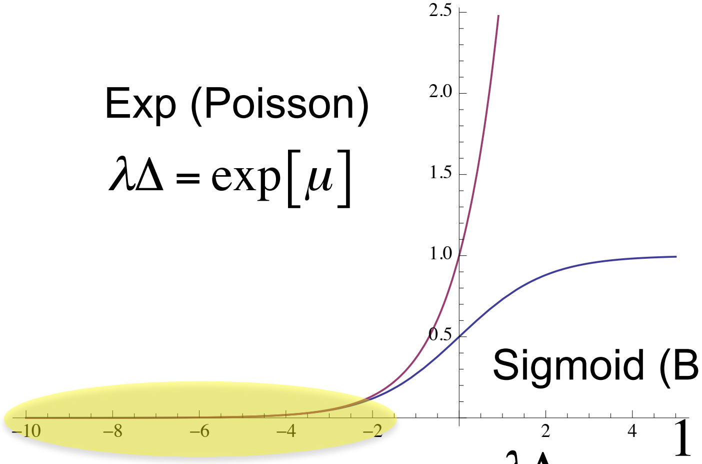


log function



Relation between Bernoulli and Poisson-GLM

Inverse link function



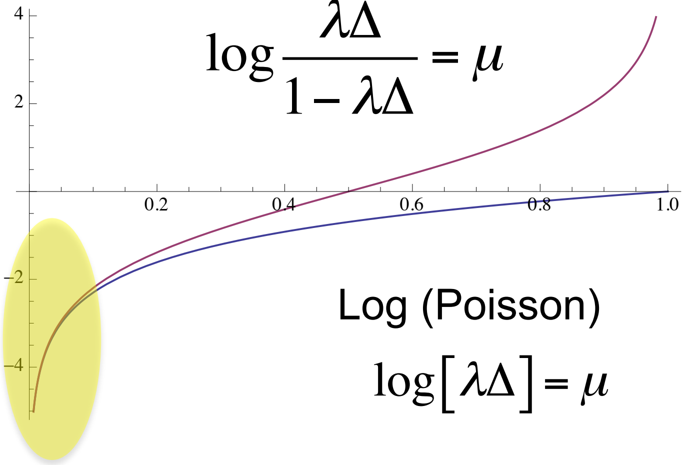
Exp (Poisson)

$$\lambda\Delta = \exp[\mu]$$

Sigmoid (Bernoulli)

$$\lambda\Delta = \frac{1}{1+e^{-\mu}}$$

Link function



Logit (Bernoulli)

$$\log\frac{\lambda\Delta}{1-\lambda\Delta} = \mu$$

Log (Poisson)

$$\log[\lambda\Delta] = \mu$$

misprint

If the mean λΔ is small, an exponential and sigmoid functions are nearly identical. Similarly, log and sigmoid functions are nearly identical for small λΔ.

Is the point process a Poisson?

We now recall that the number of spike count in a certain interval follows the Poisson distribution.

The spike count in a small bin is then approximated as

$$P\left(N_{\Delta}=n\right)=\frac{\left(\lambda\Delta\right)^{n}}{n!}e^{-\lambda\Delta}=\frac{\left(\lambda\Delta\right)^{n}}{n!}\left[1-\lambda\Delta+\frac{1}{2}\left(\lambda\Delta\right)^{2}+\cdots\right]$$

In particular,

$$P\left(N_{\Delta}=0\right)=1\left[1-\lambda\Delta+\frac{1}{2}\left(\lambda\Delta\right)^{2}+\cdots\right]=1-\lambda\Delta+o\left(\Delta\right)$$

$$P\left(N_{\Delta}=1\right)=\lambda\Delta\left[1-\lambda\Delta+\frac{1}{2}\left(\lambda\Delta\right)^{2}+\cdots\right]=\lambda\Delta+o\left(\Delta\right)$$

$$P\left(N_{\Delta}=2\right)=\left(\lambda\Delta\right)^{2}\left[1-\lambda\Delta+\frac{1}{2}\left(\lambda\Delta\right)^{2}+\cdots\right]=o\left(\Delta\right)$$

The probability of having a spike/no spike is an approximation of the Poisson count distribution for a small time bin.

From discrete to continuous time

Likelihood of discrete-time conditional Poisson distributions:

$$P\left(\Delta N_{\Delta},\Delta N_{2\Delta},\cdots,\Delta N_{N\Delta}\right)=\prod_{k=1}^{K}\exp\left[\Delta N_{k\Delta}\log\left\{\lambda_k\Delta\right\}-\lambda_k\Delta\right]+o\left(\Delta^n\right)$$

$$=\Delta^n\exp\left[\sum_{k=1}^{K}\Delta N_{k\Delta}\log\lambda_k-\lambda_k\Delta\right]+o\left(\Delta^n\right)$$

Likelihood of a point process by continuous-time limit:

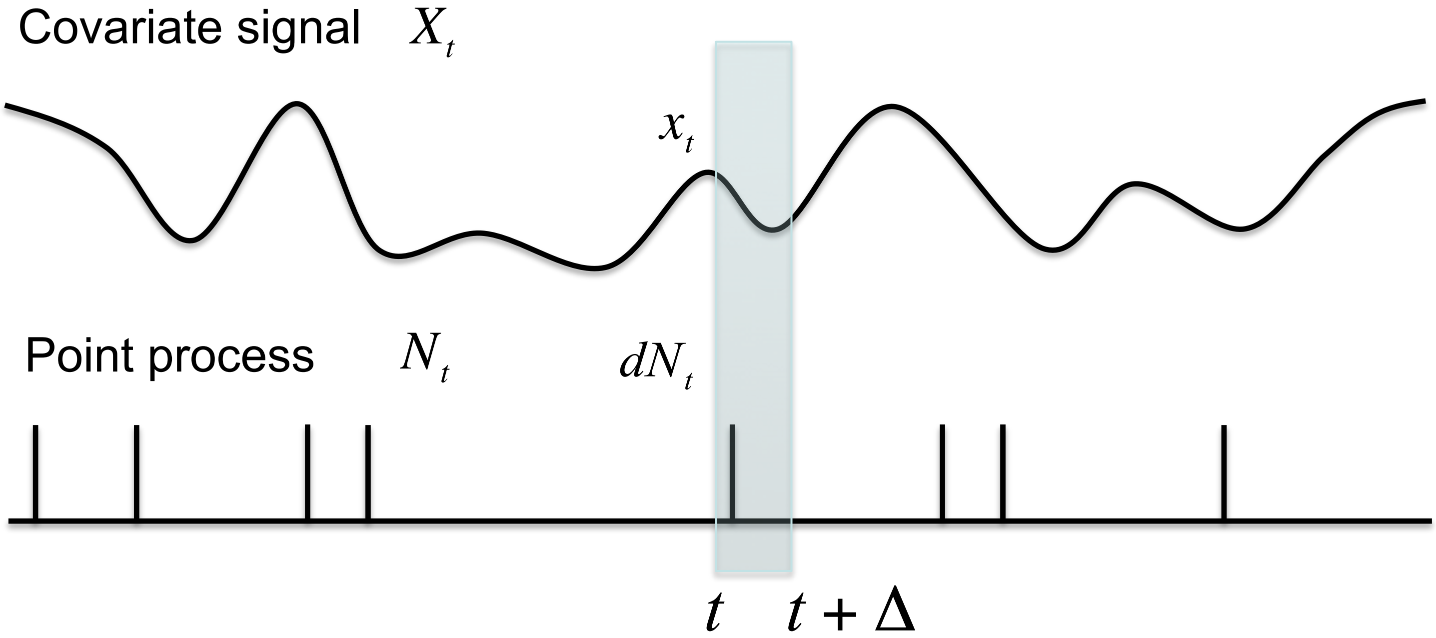
$$p\left(t_1,t_2,\ldots,t_n\cap N_T=n\right)=\lim_{\Delta\to 0}\frac{P\left(\Delta N_{\Delta},\Delta N_{2\Delta},\cdots,\Delta N_{N\Delta}\right)}{\Delta^n}$$

$$=\lim_{\Delta\to 0}\frac{\exp\left[\sum_{k=1}^{K}\Delta N_{k\Delta}\log\left\{\lambda_k\right\}-\sum_{k=1}^{K}\lambda_k\Delta\right]\Delta^n+o\left(\Delta^n\right)}{\Delta^n}$$

$$=\exp\left[\int_0^T\log\lambda_u\,dN_u-\int_0^T\lambda_u\,du\right]$$

Truccolo et al. J Neurophysiol (2005)

Point process



$$P\left(\text{a spike in } [t,t+\Delta) \mid H_t, X_t\right) = \lambda\left(t \mid H_t, X_t\right)\Delta + o\left(\Delta\right)$$

History of spiking activity

History of a covariate signal (Filtration)

$$H_t = \left\{t_1 < t_2 < \cdots < t_n \cap N_t = n\right\} \qquad X_t = \left\{x_u : 0 < u < t\right\}$$

Point process-GLM

Likelihood function:

$$p\left(t_1,t_2,\ldots,t_n \cap N_T = n\right) = \prod_{i=1}^{n} \lambda\left(t_i \mid X_{t_i}, H_{t_i}\right) \exp\left[-\int_0^T \lambda\left(u \mid X_u, H_u\right) du\right]$$

$$= \exp\left[\int_0^T \log \lambda\left(u \mid X_u, H_u\right) dN_u - \int_0^T \lambda\left(u \mid X_u, H_u\right) du\right]$$

Here $dN_t = \sum_{i=1}^{n} \delta\left(t - t_i\right) dt$

Modeling the conditional mean rate via a canonical link function.

$$E\left[\frac{dN_t}{dt}\right] = \lambda\left(t \mid H_t, X_t\right) = \exp\left[\beta_0 + \beta_1 f\left(X_t\right) + \beta_2 g\left(H_t\right)\right]$$

a log-link function

A canonical link function for a Poisson distribution.

Covariates for neural signals

The covariates can be added in a (generalized) linear manner.

$$\log \lambda\left(t \mid H_t, X_t\right) = \beta_0 + \beta_1\left[\text{Input stimulus}\right]$$

Log-link function

$$+\beta_1'\left[\text{Higher-order feature of stimulus}\right]$$

$$+\beta_2\left[\text{Spike history}\right]$$

$$+\beta_3\left[\text{Ensemble spike history}\right]$$

$$+\beta_4\left[\text{Local field potential}\right]$$

Maximum likelihood estimation

Likelihood

$$L(\beta \mid N_{0:T}) = \exp\left[\int_0^T \log \lambda\left(u \mid X_{t_i}, H_{t_i}\right) dN_u - \int_0^T \lambda\left(u \mid X_u, H_u\right) du\right]$$

where $\lambda(t \mid H_t, X_t) = \exp\left[x(t)\beta\right].$

Optimize parameters under the principle of maximizing the likelihood:

$$\beta_{\mathrm{MLE}} = \underset{\beta}{\arg\max} \log L(\beta \mid N_{0:T})$$

Newton-Raphson method:

$$l(\beta) = \log L(\beta \mid N_{0:T})$$

$$\beta^{j+1} = \beta^{j} - \nabla\nabla l\left(\beta^{j}\right) \cdot \nabla l\left(\beta^{j}\right)$$

Hessian Score function

Matlab: $b = \mathrm{glmfit}\left(X, y, '\,poisson'\right)$

Applications of point process-GLM

Here we review

A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects

Wilson Truccolo,[1] Uri T. Eden,[2,3] Matthew R. Fellows,[1] John P. Donoghue,[1] and Emery N. Brown[2,3]

J Neurophysiol 93: 1074–1089, 2005.

Modeling conditional intensity function

Conditional intensity function:

$$\lambda(t_k \mid N_{1:k}^{1:C}, \mathbf{x}_{k+\tau}, \theta) = \lambda_I(t_k \mid N_{1:k}, \theta_I)\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E)\lambda_X(t_k \mid \mathbf{x}_{k+\tau}, \theta_X)$$

Auto spike history | Ensemble history | Extrinsic covariate

Auto spike history **Expected to capture refractory effects, recovery periods, and oscillation.**

$$\lambda_I(t_k \mid N_{1:k}, \theta_I) = \exp\left\{ \gamma_0 + \sum_{n=1}^{Q} \gamma_n \Delta N_{k-n} \right\}$$

Determined by AIC

Ensemble history **Expected to capture lagged spike synchrony between target and other cells.**

$$\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E) = \exp\left\{ \beta_0 + \sum_{c} \sum_{r=1}^{R} \beta_r^c \Delta N_{k-r}^c \right\}$$

Expected to capture ensemble effect on slower time scale

$$\lambda_E(t_k \mid N_{1:k}^{1:C}, \theta_E) = \exp\left\{ \beta_0 + \sum_{c} \sum_{r=1}^{R} \beta_r^c \left(N_{k-(r-1)W}^c - N_{k-rW}^c\right) \right\}$$

Spike count in a window of size W.

Extrinsic covariate (hand velocity model)

Angle of hand velocity

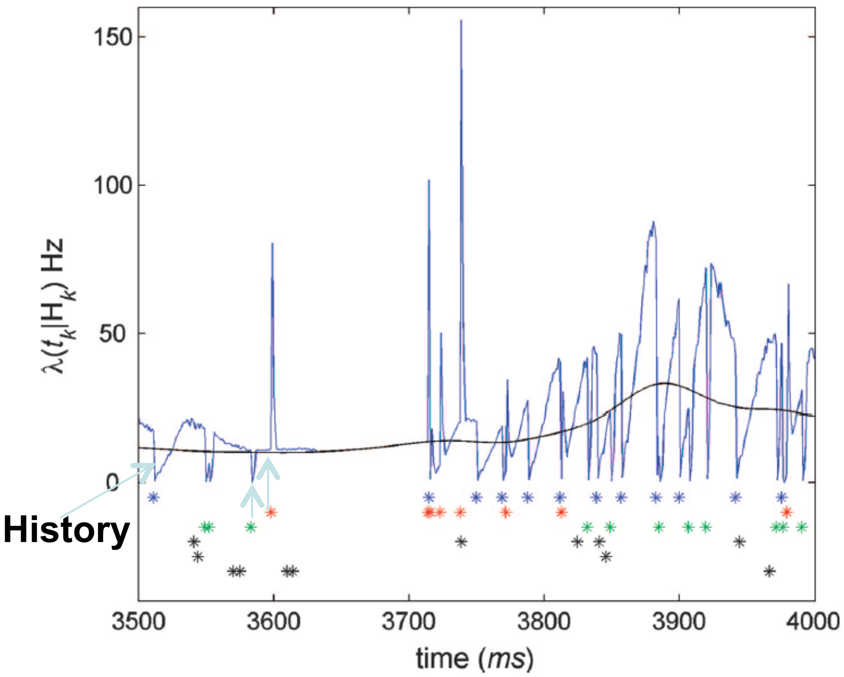
$$\lambda_X(t_k \mid \mathbf{x}_{k+\tau}, \theta_X) = \exp\{\alpha_0 + |V_{k+\tau}| \left[\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})\right]\}$$

Magnitude of hand velocity

offset τ=+150ms

Simulation (6 neurons)

Simulated conditional intensity of a blue neuron

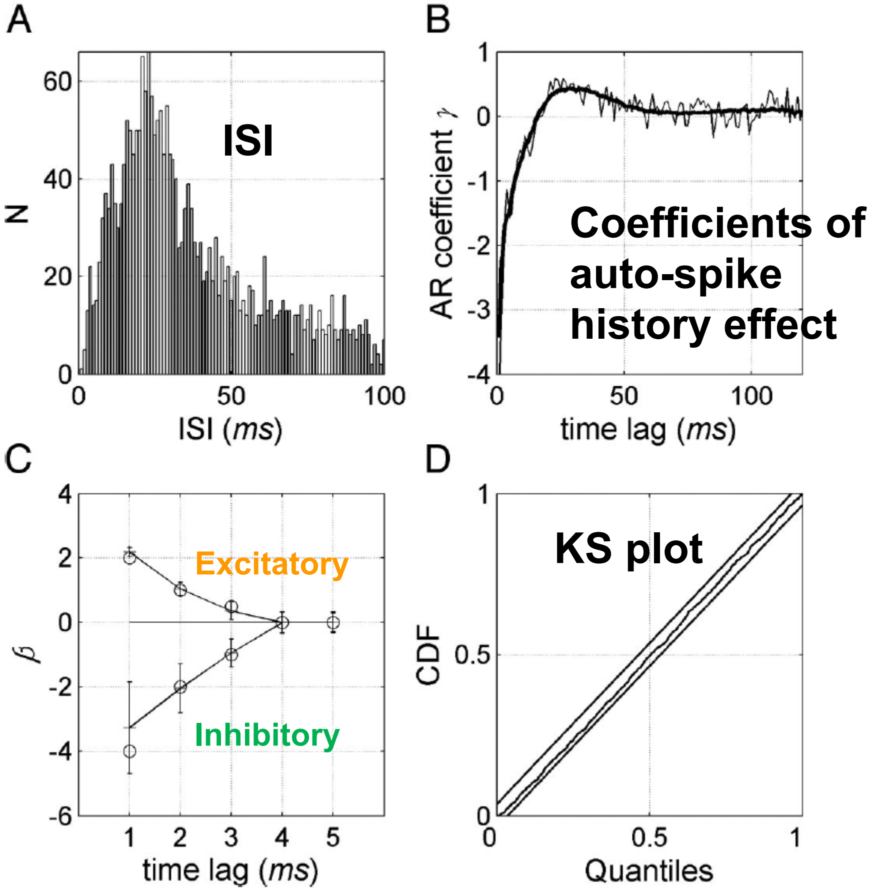


History

Excitatory

Inhibitory

Other neurons



ISI

Coefficients of auto-spike history effect

Excitatory

Inhibitory

KS plot

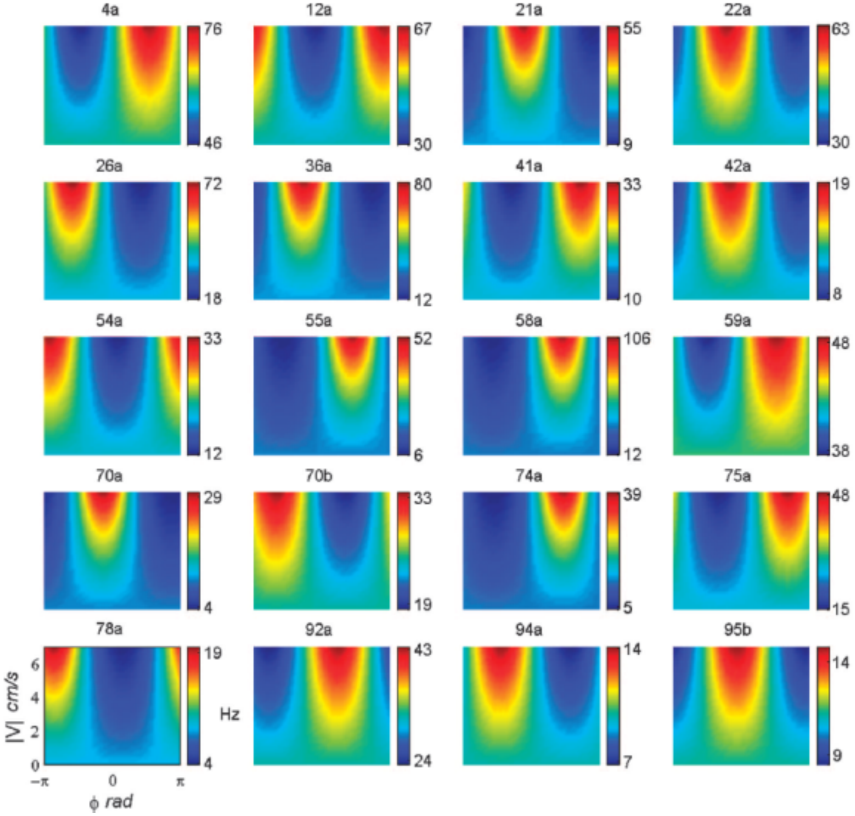
Application to MI spiking data

Two models were compared.

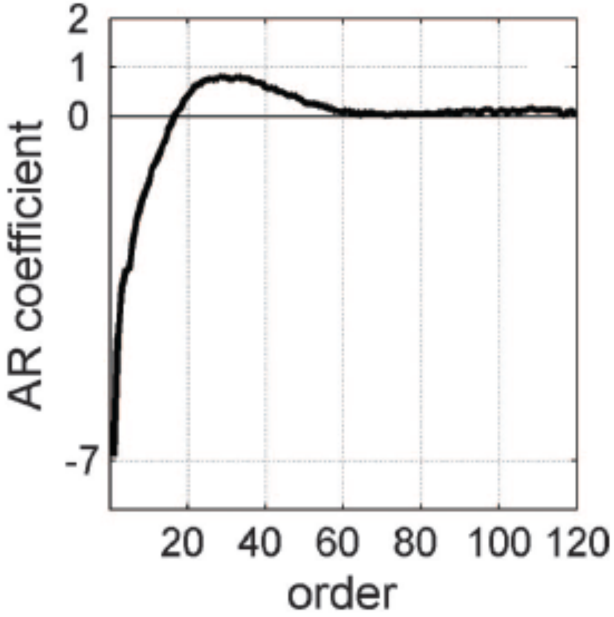
Velocity model

Velocity + Auto spike history model

Tuning curve of velocity model



Auto spike history

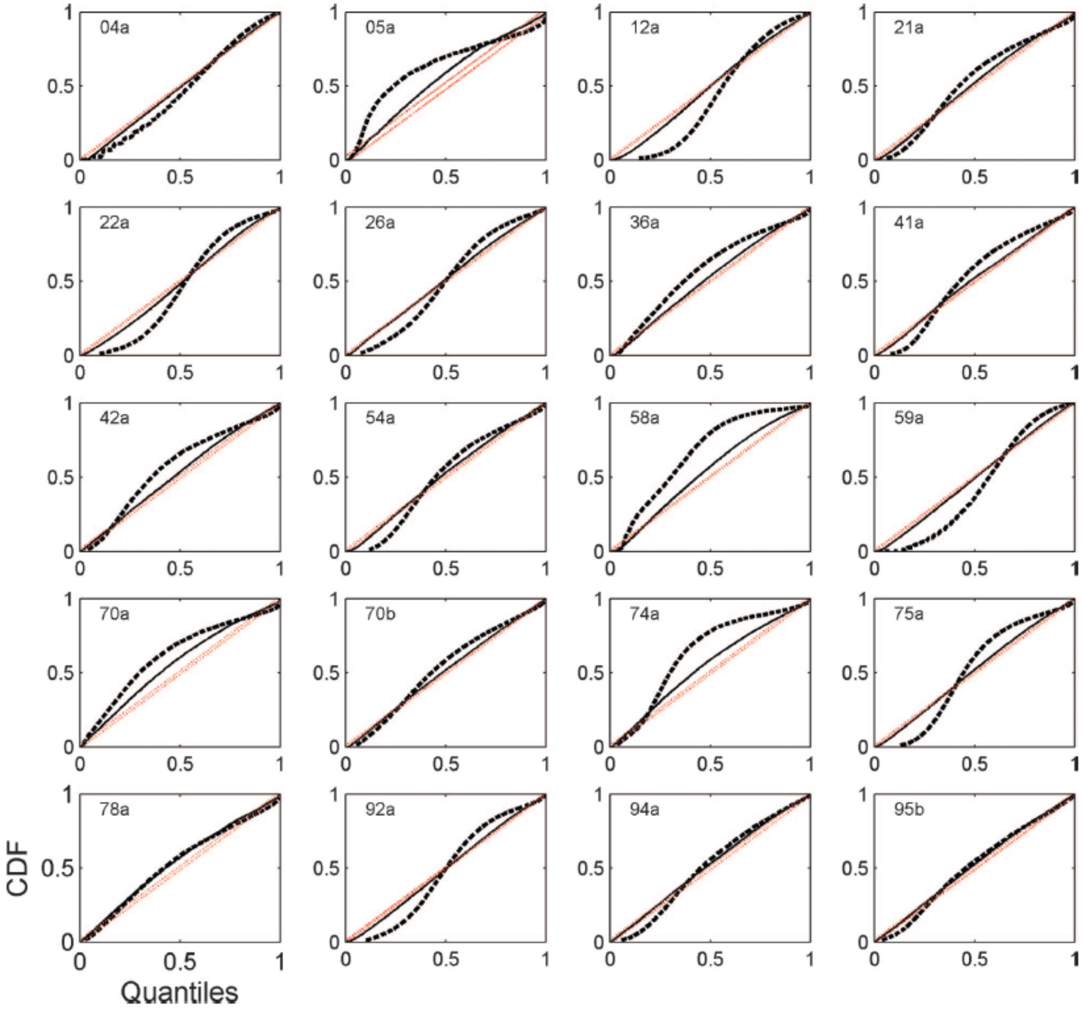


$$\lambda_I(t_k \mid N_{1:k}, \theta_I) = \exp\left\{ \gamma_0 + \sum_{n=1}^{Q} \gamma_n \Delta N_{k-n} \right\}$$

$$\lambda_X(t_k \mid \mathbf{x}_{k+\tau}, \theta_X) = \exp\{\alpha_0 + |V_{k+\tau}| [\alpha_1 \cos(\phi_{k+\tau}) + \alpha_2 \sin(\phi_{k+\tau})]\}$$

Test of goodness-of-fit: KS-plot

KS plot



$$z_j = 1 - \exp\left\{ -\int_{u_j}^{u_{j+1}} \lambda(t \mid H(t), \hat{\theta})\mathrm{d}t \right\}$$

is expected to be independent uniformly distributed random variables in an unit interval.

- - - - - - -

Velocity model

———

Velocity + Auto-spike history model

———

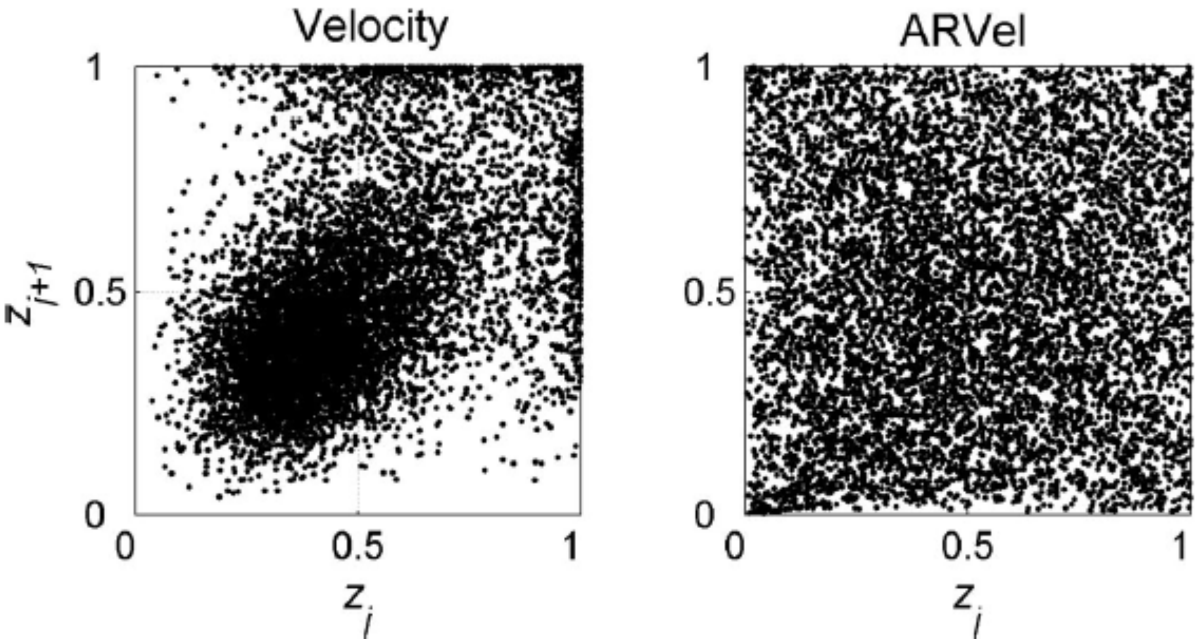
95% confidence bound of KS test

Test of independence for rescaled ISIs

If the conditional intensity model completely explains the data,

$$z_j = 1 - \exp\left\{ -\int_{u_j}^{u_{j+1}} \lambda(t \mid H(t), \hat{\theta}) \mathrm{d}t \right\}$$

is expected to be **independently** and **uniformly** distributed random variables in an unit interval.



There is dependency between zj and zj+1.

There is no dependency between zj and zj+1.

Model selection by AIC

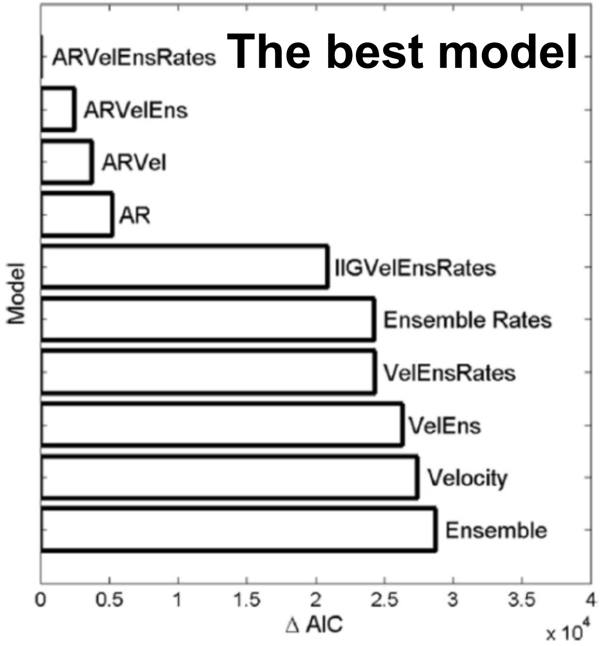
Akaike information criterion

$$\mathrm{AIC} = -2\log p\left(\mathrm{DATA} \mid \beta\right) + 2 \times \dim\left(\beta\right)$$

Parameters of the model

Model dimension
(Number of parameters)

The smaller the AIC, the better the predictive ability of the model is.



AR: Auto-spike history
The models with the auto-spike history significantly decrease AIC.

Point process residual analysis

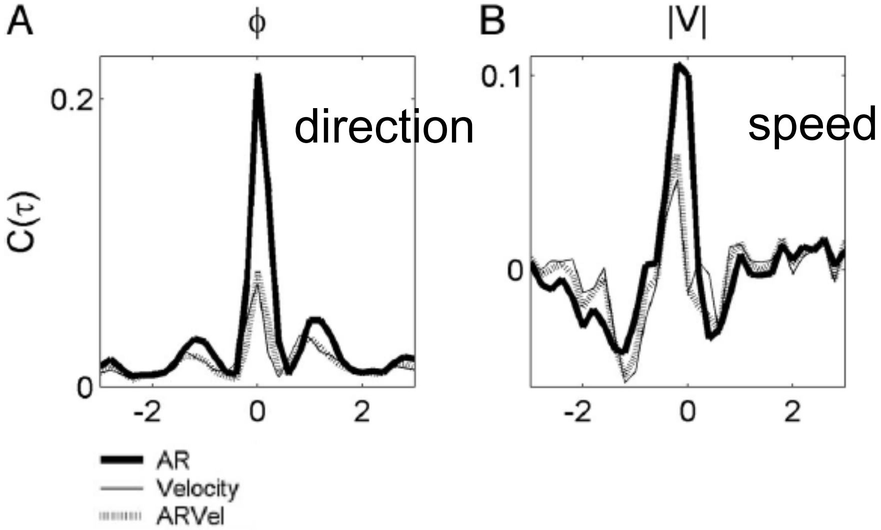
Point process residual

$$M(t_k) = \sum_{i=k-B}^{k} \Delta N_i - \int_{t_{k-B}}^{t_k} \lambda(t \mid H(t), \hat{\theta})dt$$

The difference between the actual spike count and spike count predicted by the model.

If the CIF completely characterizes response characteristics to covariates (stimulus/motor output), the residual should have no dependency with the covariates.

Cross-correlation function between residual and covariates (hand movement).



Inclusion of velocity signal in a CIF greatly reduced the correlation.

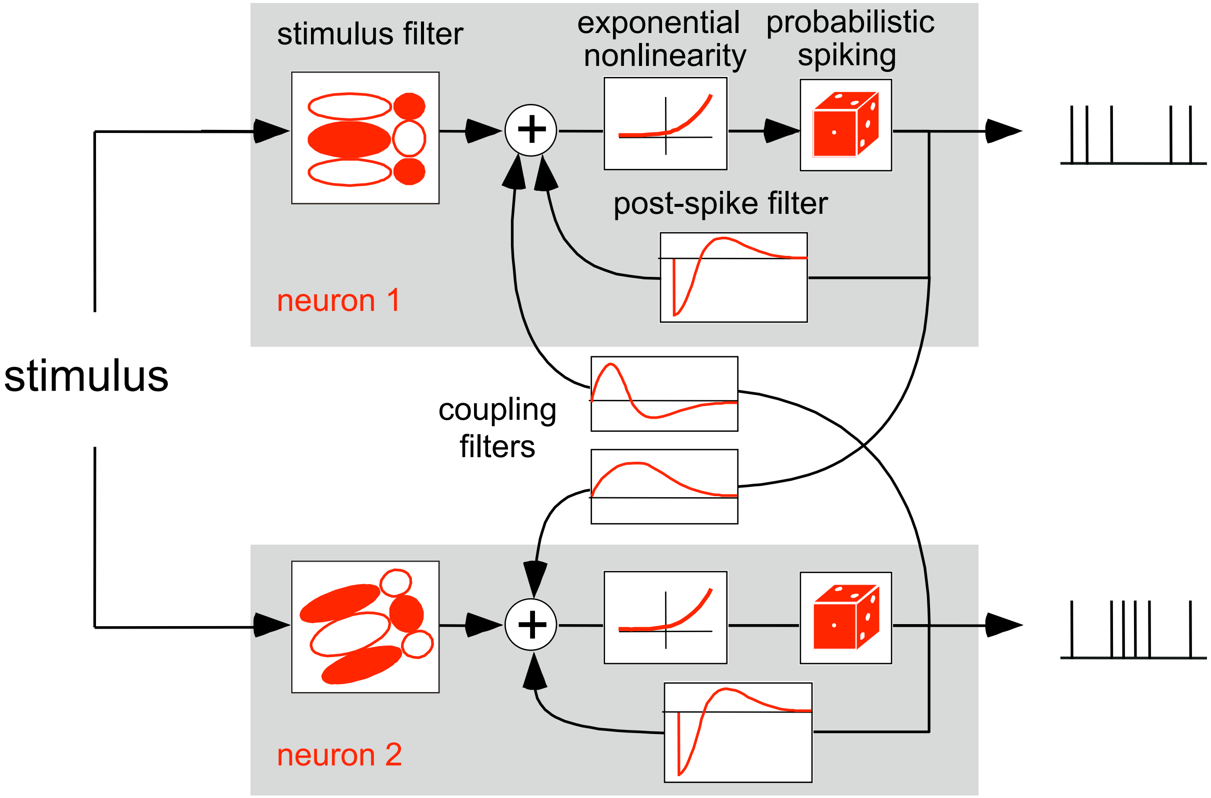
However, the auto-spike history did not further reduce the correlation.

AR component does not carry additional statistical information about hand velocity.

There is a component that was not captured by the point process-GLMs.

We will review

Point process-GLM for the analysis of retinal ganglion cells



Pillow et al (2008). Nature, 454(7207), 995–999.

What we learned

1

- **Exponential family of distributions and generalized liner model.**

2

- **Discrete-time likelihood of conditional Bernoulli and Poisson distributions**.

3

- **GLM framework for a continuous-time point process.**

4

- Review of a paper by Truccolo et al. Focused on methods for model validation **(KS-plot, AIC, residual analysis).**

Course overview

1

- Introduction and a **Poisson** point process

2

- **Renewal** and **non-Poisson** processes

3

- **Point process-GLM**: Stimulus and a point process (Encoding)

4

- Inference for a Poisson process: **Spike-rate estimation**

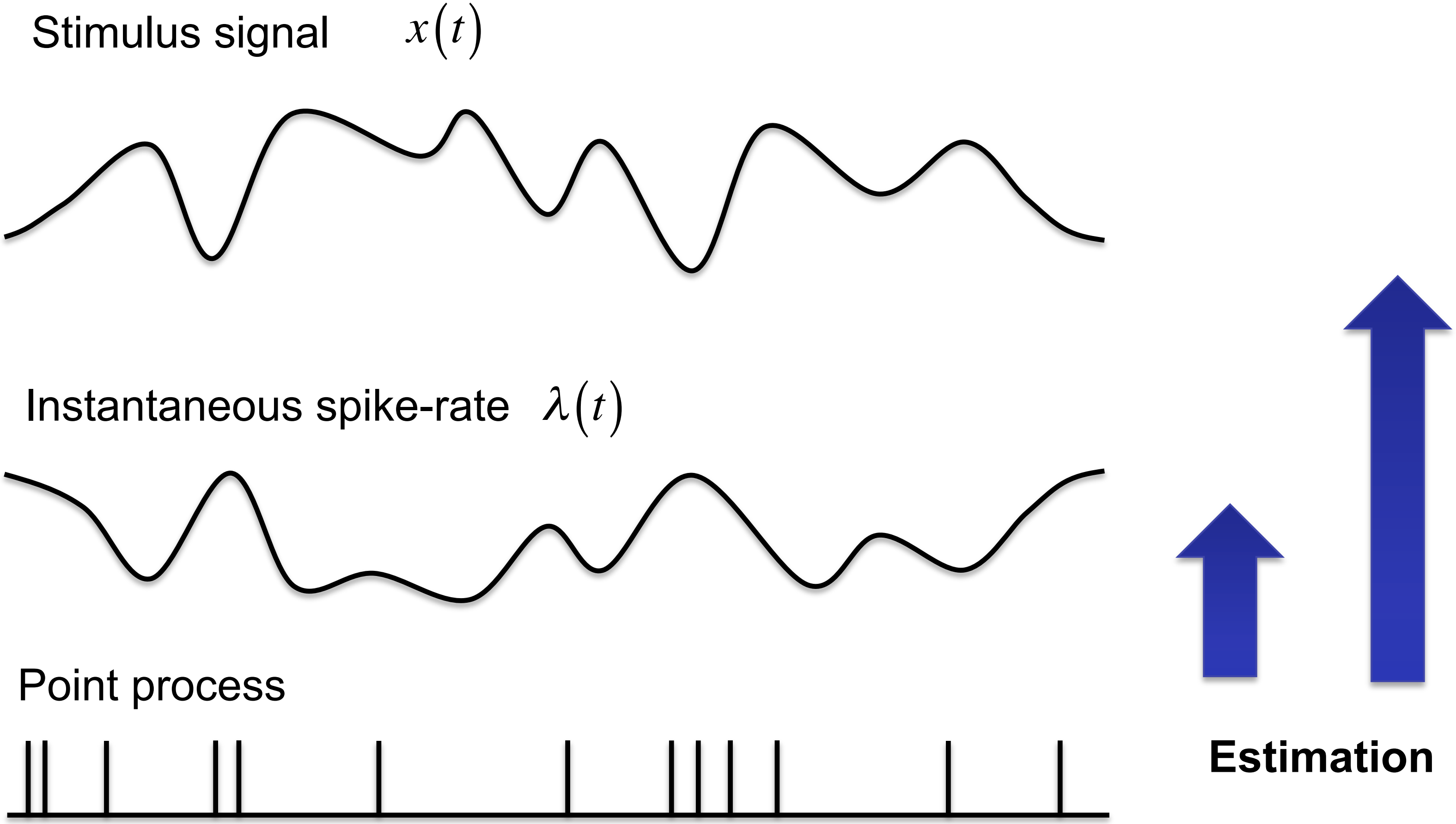
5

- **State-space** model and a **point process filter** (Decoding)

6

- Paper reviews by SCS 1st year students

From the feedback of the last lecture



Stimulus signal $x(t)$

Instantaneous spike-rate $\lambda(t)$

Point process

Estimation