

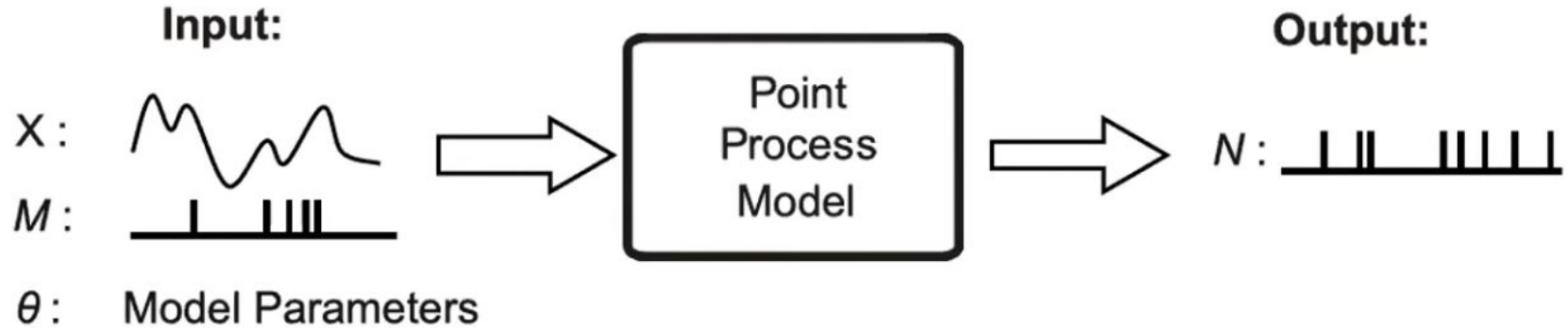
# Lecture 5: Spike Stimulus Analysis

# Conditional Intensity Function

Generalized rate function (instantaneous firing rate that depends on past history of spiking) that can describe any point process

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

Use Poisson regression to model point processes by parameterizing the conditional intensity function



$$\lambda(t) = e^{\beta_0 + \beta_1 x_1(t)}$$

Specify a simple linear predictor form

# Why do we need more explicit statistical models?

Measure how well the data is described by a particular model (how valid is the model?)

Compare models and test hypotheses (does one model fit the data better?)

Calculate confidence intervals (model uncertainty, a PETH/rate map only describes what happens on average)

Control for more than one effect (what are the contributions of position and speed? what is the contribution of bursting?)

Making assumptions explicit

# Poisson Regression - Comparing and Evaluating Models

Method 1: Comparing AIC Values

Method 2: Likelihood Ratio Test for Nested Models

Method 3: Confidence Intervals for Individual Model Parameters

Method 4: KS-Test for Model Goodness-of-Fit

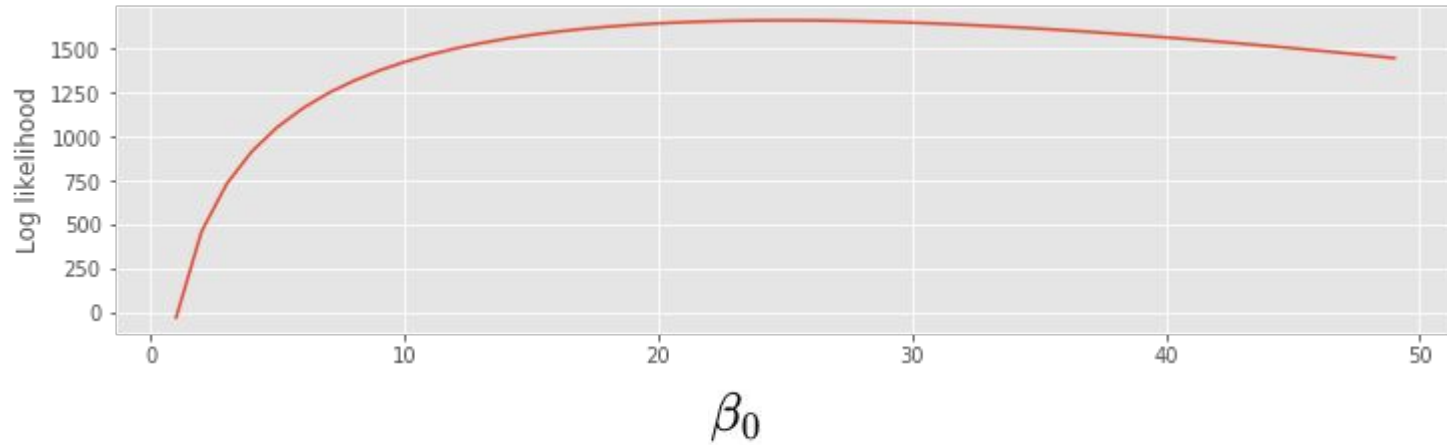
Method 5: Residual Analysis

It's not enough to just fit the model, we want to know how “good” of a model it is.

**We want to do iterative model building**

## Method 1: Comparing AIC Values

In fitting the regression model, you maximize the log likelihood of the parameters



Likelihood measures how well a particular parameter setting explains the data, assuming the model is true.

In each time bin, the neuron generates spikes according to a Poisson distribution with mean rate  $\lambda_i = \exp(x_i^\top \beta)$

We can score how well each set of parameters fits the data. For each time bin, the Poisson probability is:

$$P(k_i \mid \lambda_i) = \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!}$$

Likelihood measures how well a particular parameter setting explains the data, assuming the model is true.

We assume each time bin is independent, so the total probability of the spikes given the spike train is:

$$L(\beta) = \prod_i \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!}$$



Likelihood measures how well a particular parameter setting explains the data, assuming the model is true.

Since

$$\lambda_i = \exp(x_i^\top \beta)$$

we get

$$\log L(\beta) = \sum_i [k_i(x_i^\top \beta) - \exp(x_i^\top \beta) - \log(k_i!)]$$

# Method 1: Compare AIC values

If we want to compare two different models (say position and speed), we can use the AIC (Akaike Information Criterion).

$$\text{AIC} = -2 \log L(\theta_{ML}) + 2p$$

This is the maximum log likelihood of the model scaled and **penalized by the number of parameters**.

**Smaller AIC is better**

Asymptotically the same as **leave one out cross validation** (leave one data point out, compute log likelihood, leave next data point out, etc).

We can compare two models by looking at the difference in AIC (e.g speed vs. position)

# Example

**Model:**

$$y_t \sim \text{Poisson}(\lambda(t)\Delta t)$$

$$\log \lambda(t) = \beta_0 + f_{\text{pos}}(x(t)) + \beta_{\text{spd}} v(t)$$

- $f_{\text{pos}}(x)$ : position tuning (often many parameters: bins/splines)
- $\beta_{\text{spd}}$ : one coefficient for speed gain (simplest case)

# Multiple candidate models

- **M0**: baseline only

$$\log \lambda = \beta_0$$

- **Mpos**: position only

$$\log \lambda = \beta_0 + f_{\text{pos}}(x)$$

- **Mspd**: speed only

$$\log \lambda = \beta_0 + \beta_{\text{spd}}v$$

- **Mpos+spd**: position + speed

$$\log \lambda = \beta_0 + f_{\text{pos}}(x) + \beta_{\text{spd}}v$$

Compute the difference in AIC between each unique pair (Mspd vs. Mpos, Mpos vs. M0, etc).

Let's say:

**Mspd AIC**: 3344.79

**Mpos AIC**: 2708.77

**dAIC**: 636.01

Which model has better prediction?

# Method 2: Likelihood Ratio Test for Nested Models

What are Nested Models?

- One model contains the parameters of the other
- You can get Model 2 by setting some parameters of Model 1 to zero

Example:

Model 1 (full model):

$$\log \lambda = \beta_0 + \beta_{\text{pos}}(\text{position}) + \beta_{\text{speed}}(\text{speed})$$

Model 2 (reduced model):

$$\log \lambda = \beta_0 + \beta_{\text{pos}}(\text{position})$$

Model 2 is nested inside Model 1 because it is the same model with:

$$\beta_{\text{speed}} = 0$$

**Does speed explain spiking above and beyond position?**

## Method 2: Likelihood Ratio Test for Nested Models

Compare difference in deviance between the models.

The deviance measures **the difference between the log likelihood of a perfect (“saturated”) model that has a parameter for each data point and the log likelihood for your Poisson regression.**

This is a measure of how far away the fitted model is from the best possible fit.

## Method 2: Likelihood Ratio Test for Nested Models

Let:

- Smaller (reduced) model have  $n_1$  parameters
- Larger (full) model have  $n_2$  parameters
- $n_2 > n_1$

Compute the log-likelihood for each:

$$\ell_{\text{full}}, \quad \ell_{\text{reduced}}$$

**Test statistic:**

$$D = 2(\ell_{\text{full}} - \ell_{\text{reduced}})$$

## Method 2: Likelihood Ratio Test for Nested Models

$H_0$  : All additional parameters in the full model are 0       $D \sim \chi^2_{n_2 - n_1}$

$H_A$  : At least one additional parameter is nonzero

Decision Rule:

- If p-value is small, reject null hypothesis ( $H_0$ )
- Conclude the larger model significantly improves fit

Model 1 (full model):

$$\log \lambda = \beta_0 + \beta_{\text{pos}}(\text{position}) + \beta_{\text{speed}}(\text{speed})$$

Model 2 (reduced model):

$$\log \lambda = \beta_0 + \beta_{\text{pos}}(\text{position})$$



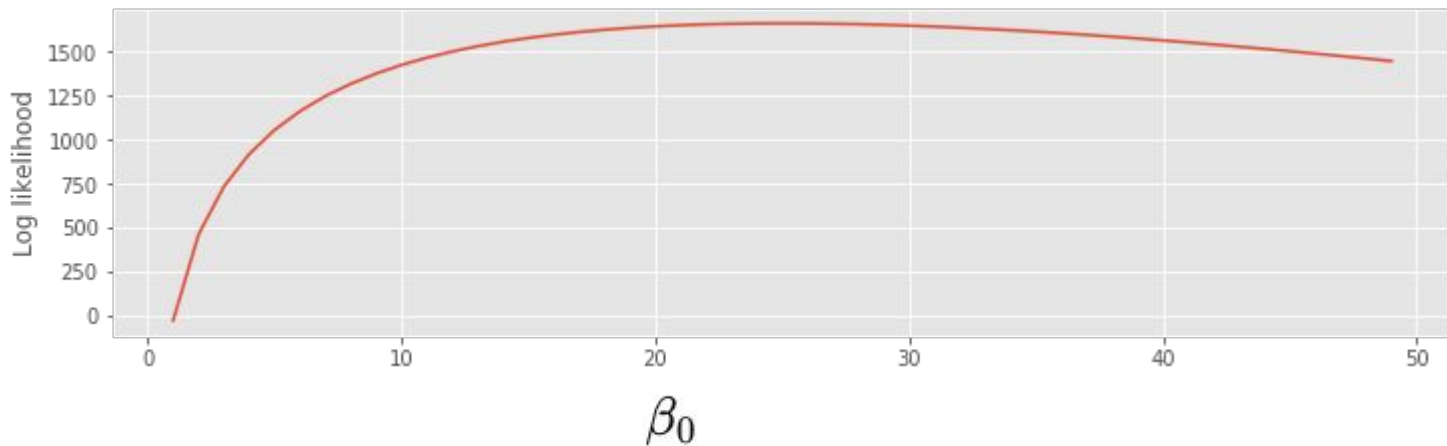
# Question

Can I compare the effect of position only versus speed only?

## Method 3: Confidence Intervals for Individual Model Parameters

We want to test directly for whether a single parameter contributes significantly to the model (contrast to before where we were looking at sets of parameters).

We can look at the log likelihood curvature and compute a confidence interval.



# Method 3: Confidence Intervals for Individual Model Parameters

We can check if this confidence interval contains zero.

The Wald statistic is:

$$Z = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)}$$

Under  $H_0$ :

$$Z \sim \mathcal{N}(0, 1)$$

# Comparing so far

## dAIC:

- **Compares:** two (or more) fitted models (not necessarily nested)
- **Goal:** pick the model expected to have better out-of-sample predictive performance (generalization)
- **What it trades off:** fit vs complexity
- **Interpretation:** lower AIC is better
- **Cons:** Can reward over-flexibility (penalty not that strong); you can pick the “best” among bad models

## Likelihood Ratio Test:

- **Compares:** nested models (reduced model is a special case of the full model)
- **Best for:** testing a **group of parameters** jointly (e.g., “does this whole covariate block matter?”)
- **Practical note:** Generally more robust (compared to Wald) because it doesn’t rely as heavily on local quadratic approximation around the maximum likelihood estimate
- **Tests:** “does adding these parameters improve fit?”

## Wald Test:

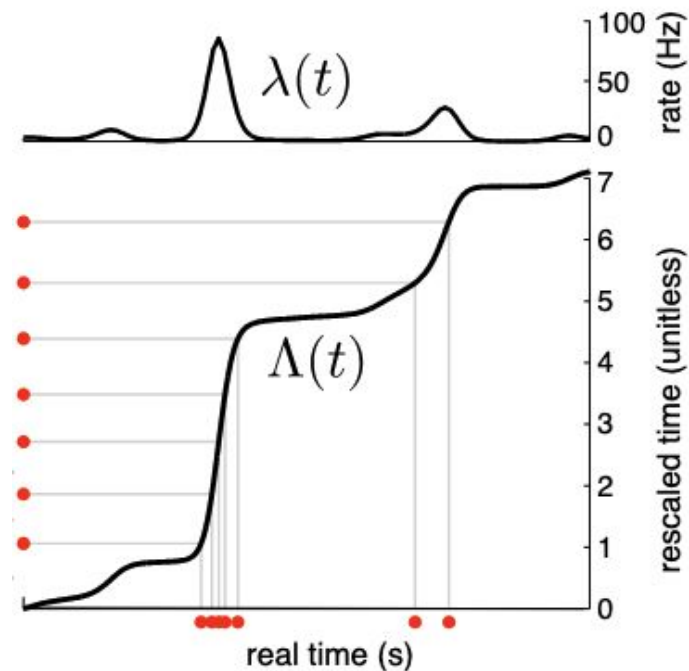
- **Tests:** individual coefficient(s) using the estimated SE at the maximum likelihood estimate
- **Idea:** Uses local curvature (estimated information) at the maximum likelihood estimate
- **Pros:** easiest to compute; you get it “for free” from the fitted model output
- **Cons:** can be unreliable when parameters are near boundaries, data are sparse, or the likelihood is skewed (not well-approximated by a quadratic)

# Method 4: KS Test for Model Goodness-of-Fit

## Key Idea

If we have the correct conditional intensity function, any point process can be transformed into a **homogeneous Poisson process of rate 1**.

This is known as the **Time Rescaling Theorem**



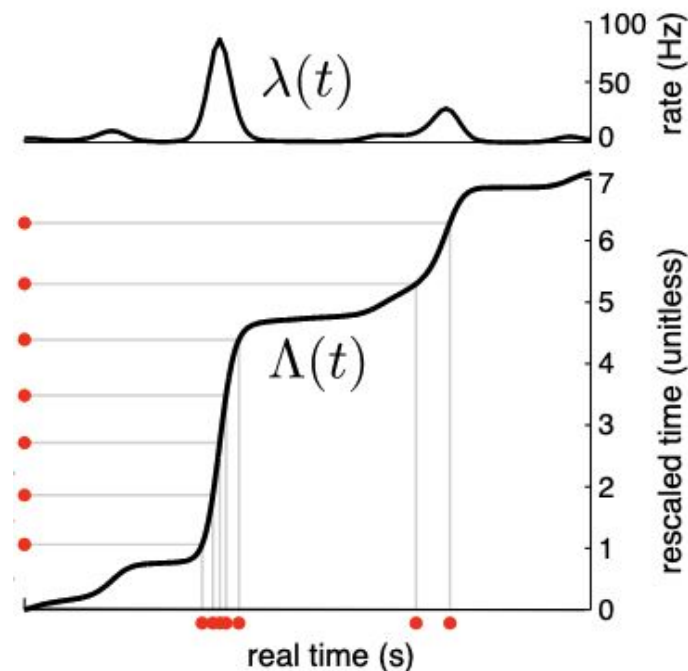
# Method 4: KS Test for Model Goodness-of-Fit

For each spike time  $t_i$ , compute the transformed time:

$$u_i = \int_{t_{i-1}}^{t_i} \lambda(s \mid \mathcal{H}_s) ds$$

This is the **integrated intensity between spikes**.

We stretch time in regions where the model predicts high rate, and compress time where it predicts low rate.



## Method 4: KS Test for Model Goodness-of-Fit

If the model is correct, the transformed ISIs ( $u$ ) should be distributed as exponential with rate 1.

$$u_i = \int_{t_{i-1}}^{t_i} \lambda(s \mid \mathcal{H}_s) ds \qquad u_i \sim \text{Exponential}(1)$$

For convenience, we can transform this further into a uniform distribution:

$$z_i = 1 - e^{-u_i} \sim \text{Uniform}(0, 1)$$

## Method 4: KS Test for Model Goodness-of-Fit

We test:

$$H_0 : z_i \sim \text{Uniform}(0, 1)$$

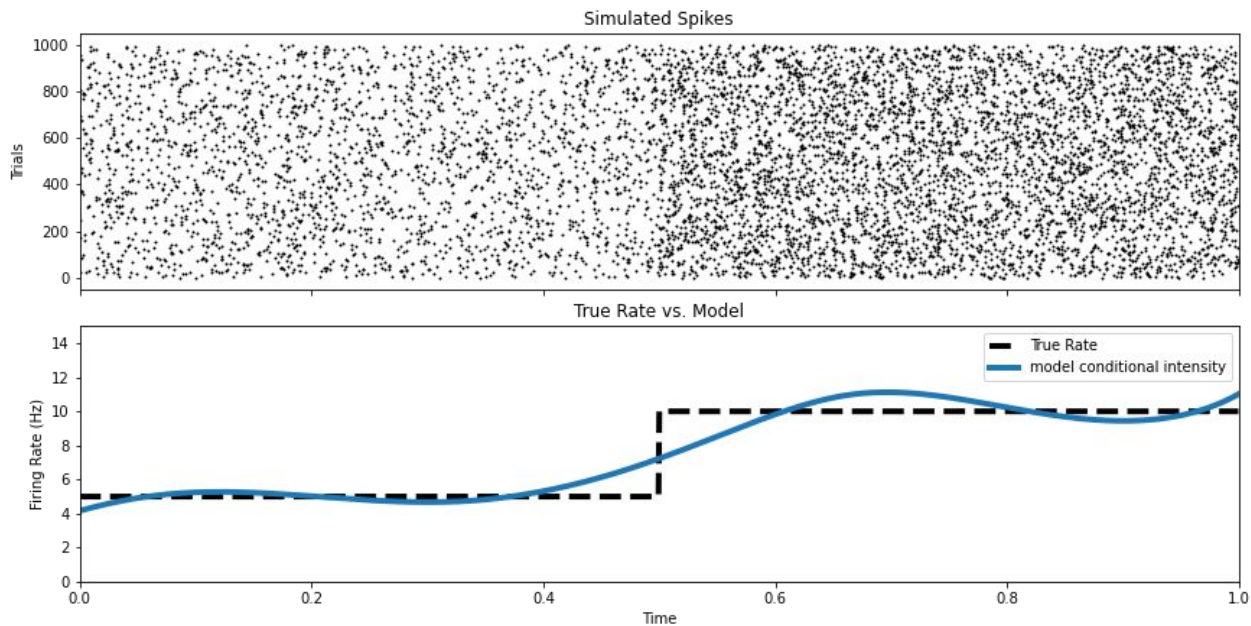
And get a p-value

This is known as a Kolmogorov-Smirnov Test (KS).

Alternatively we can also plot the  $z$  against the expected distribution and visually check for deviations. We can also look at their autocorrelation.



# Example



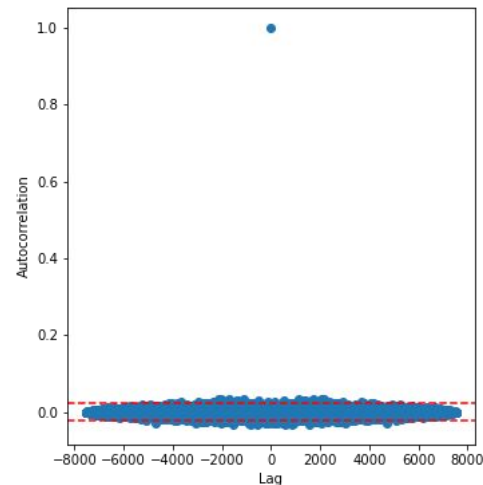
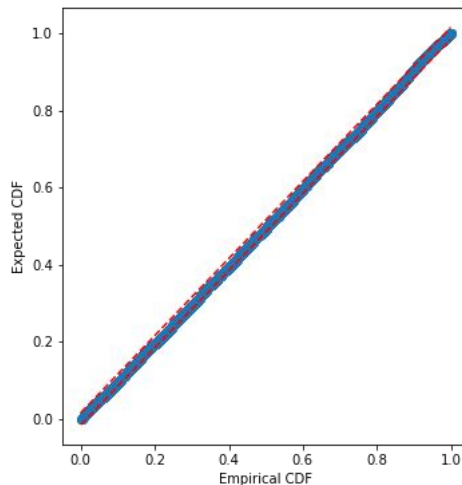
# Example

Early excess → model underestimates short ISIs

Late excess → model overestimates rate

Systematic curve shape → model mis-specification

Autocorrelated → unaccounted for temporal structure



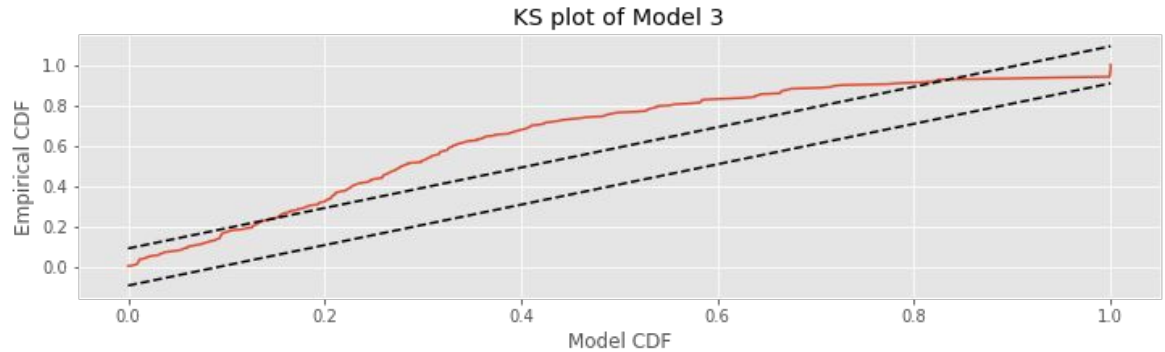
# Example

Early excess → model underestimates short ISIs

Late excess → model overestimates rate

Systematic curve shape → model mis-specification

Autocorrelated → unaccounted for temporal structure



## Method 5: Residual Analysis

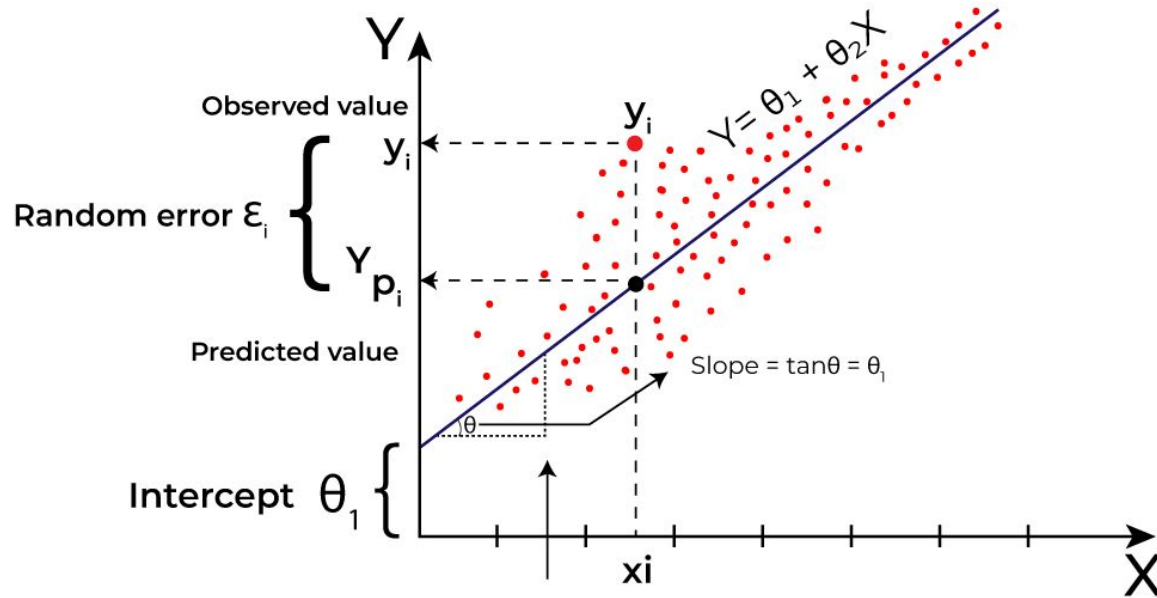
Understand what's going wrong with individual time points (when does the model not fit well)

We can use the conditional intensity function to compare what we observed and what we predict based on the conditional intensity function.

$R(t)$  = total observed no. of spikes at time  $t$  – total expected no. of spikes at time  $t$

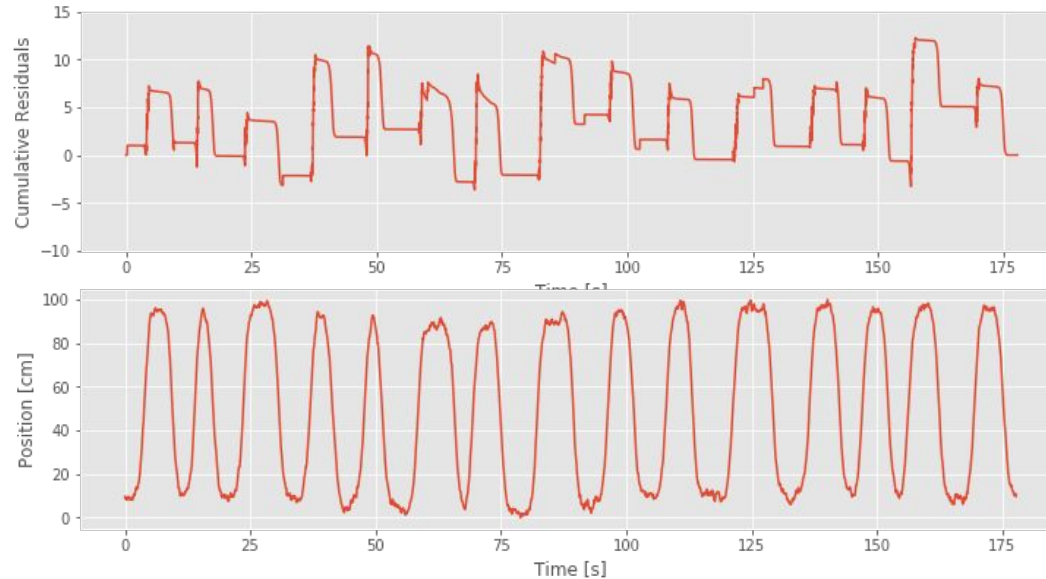
$$R(t_k) = \sum_{i=1}^k \Delta N_i - \lambda(t_i) \Delta t,$$

There is something similar in linear regression



# Method 5: Residual Analysis

If model fits well, residuals should be zero mean and uncorrelated with any covariates.



$$R(t) = \text{total observed no. of spikes at time } t - \text{total expected no. of spikes at time } t$$

# Poisson Regression - Comparing and Evaluating Models

Method 1: Comparing AIC Values

Method 2: Likelihood Ratio Test for Nested Models

Method 3: Confidence Intervals for Individual Model Parameters

Method 4: KS-Test for Model Goodness-of-Fit

Method 5: Residual Analysis

# Model Refinement Cycle

**Fit a simple baseline model** (few covariates, clear interpretation)

**Interpret the parameters:** what tuning / gain / history effect does the model predict?

**Diagnose misfit**

- **Residual structure:** systematic errors vs time, position, speed, etc.
- **Time-rescaling / KS plot:** does the model capture spike timing statistics?

**Propose an improvement** (add covariate, nonlinearity, interaction, spike-history term)

**Fit the refined model and re-interpret parameters**

**Re-check diagnostics** (residuals + KS again)

**Compare to the simpler model**

- **LRT** (nested: did the added terms help?)
- **$\Delta$ AIC / CV** (prediction vs complexity)

**Conclude or iterate**

- If improvement is real + diagnostics look good → keep model, interpret biology
- If still failing → the failure suggests what structure is missing



# Model refinement cycle and statistics embody the core of science

- Propose a hypothesis (a model)
- Fit it to data
- Test predictions and compare alternatives
- Reject/repair what fails
- Repeat

**Poisson Regression and  
Point Process Theory  
provides a structured  
principled way to do  
science**

# Some challenges

As we record from more and more neurons, how do we apply these at scale?

Some methods can be summarized easily (i.e. single number summaries)

Spot check?

# Spike Stimulus Analysis Summary

## **Basic visualizations & descriptive stats**

Raster, PETH, autocorrelation (what patterns exist?)

## **Point-process view of spiking from neurons**

Spikes as stochastic events in time; define an instantaneous rate  $\lambda(t)$

## **Poisson regression (GLM) as a model of $\lambda(t)$**

Fit how covariates (position, speed, stimulus, history) shape firing

## **Model checking & refinement**

Residual structure + time-rescaling/KS  $\rightarrow$  revise model  $\rightarrow$  compare (LRT/ $\Delta$ AIC)

$\rightarrow$  conclude