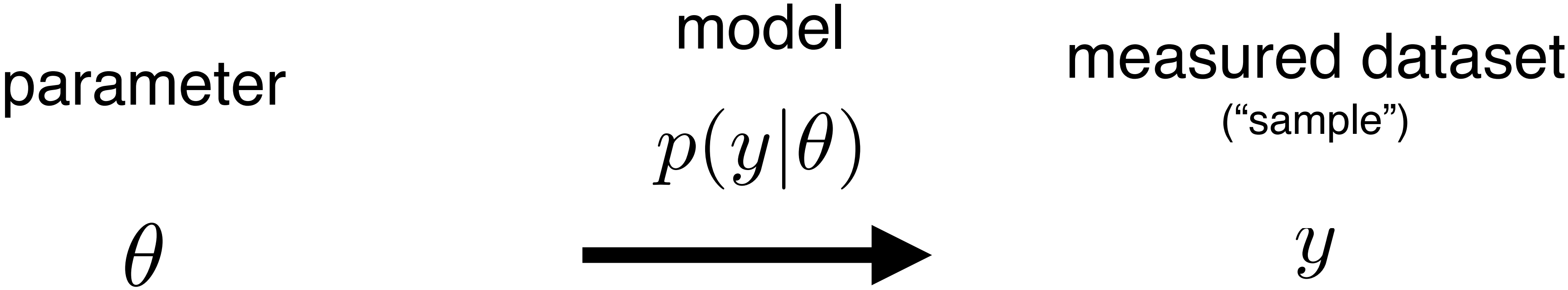
Maximum Likelihood

Jonathan Pillow

Mathematical Tools for Neuroscience (NEU 314)
Spring, 2016

lecture 16

Estimation



parameter

θ

model

$p(y|\theta)$

measured dataset

("sample")

y

An *estimator* is a function $f : \; y \longrightarrow \hat{\theta}$

- often we will write $\hat{\theta}(y)$ or just $\hat{\theta}$

Properties of an estimator

“expected” value
(average over draws of m)

bias: $b(\theta) = \mathbb{E}[\hat{\theta}] - \theta$

- “unbiased” if bias=0

variance: $\mathrm{var}(\theta) = \mathbb{E}\Big[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\Big]$

- “consistent” if bias and variance both go to zero asymptotically

mean squared error (MSE) $\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathrm{bias}^2 + \mathrm{variance}$

Example 1: linear Poisson neuron

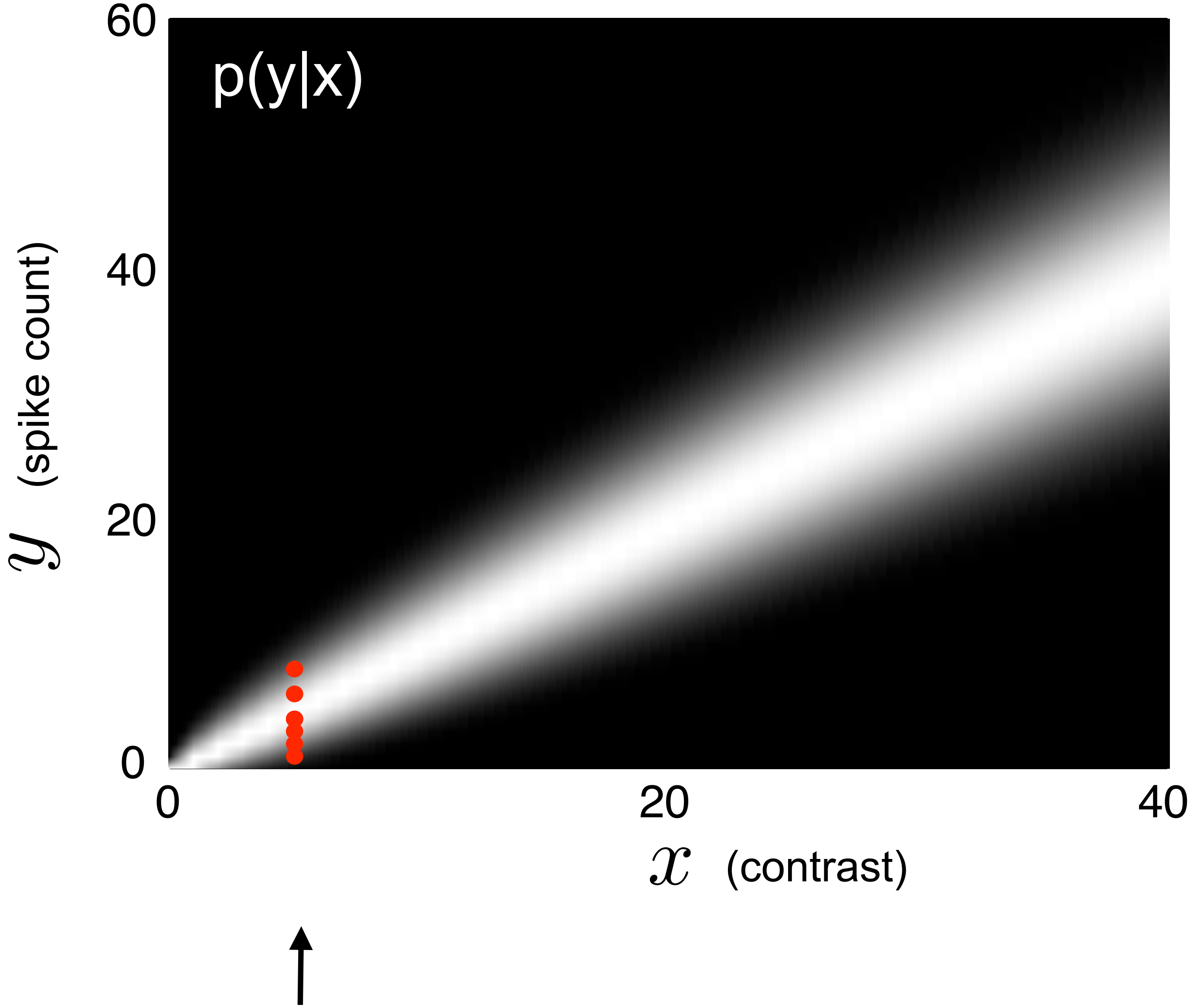
spike count $y \sim Poiss(\lambda)$

spike rate $\lambda = \theta x$

parameter stimulus

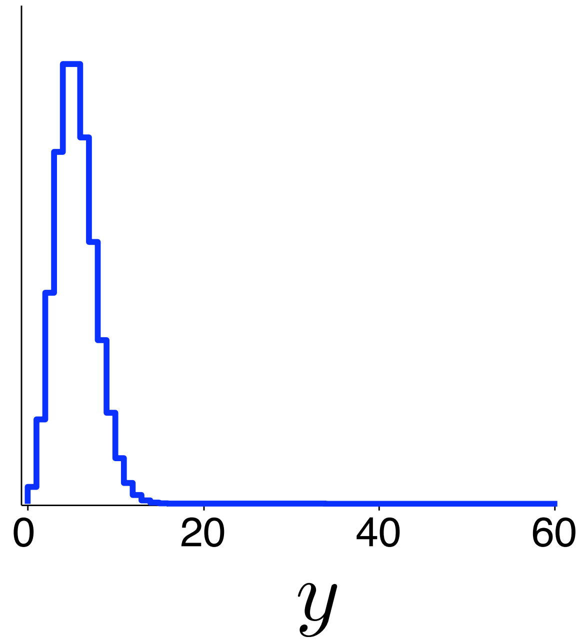
encoding model:
$$\begin{aligned} P(y|x,\theta) &= \frac{1}{y!}\lambda^y e^{-\lambda} \\ &= \frac{1}{y!}(\theta x)^y e^{-(\theta x)} \end{aligned}$$

$$\mathrm{mean}(y) = \theta x$$
$$\mathrm{var}(y) = \theta x$$

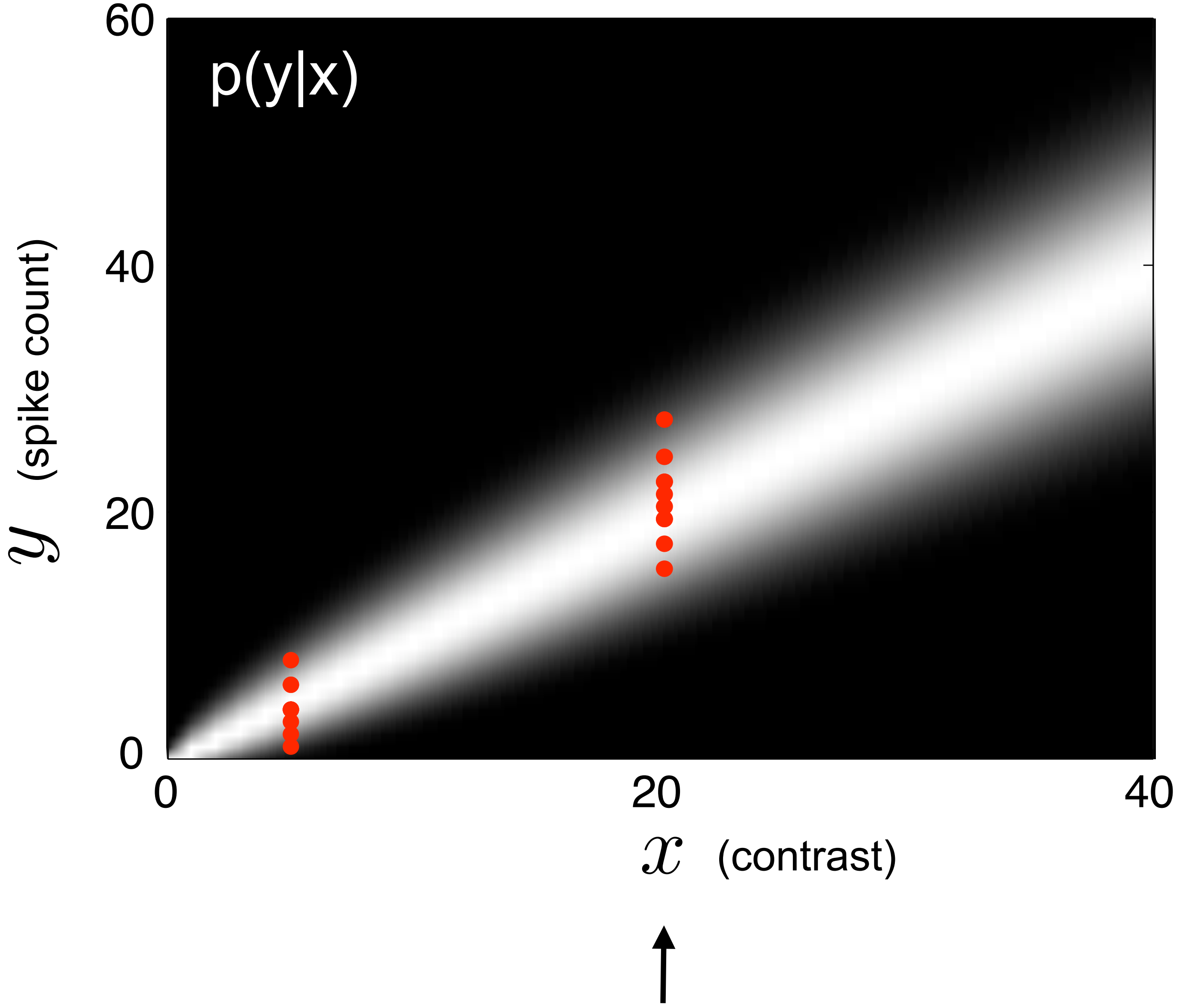


conditional distribution

$$p(y|x = 5)$$

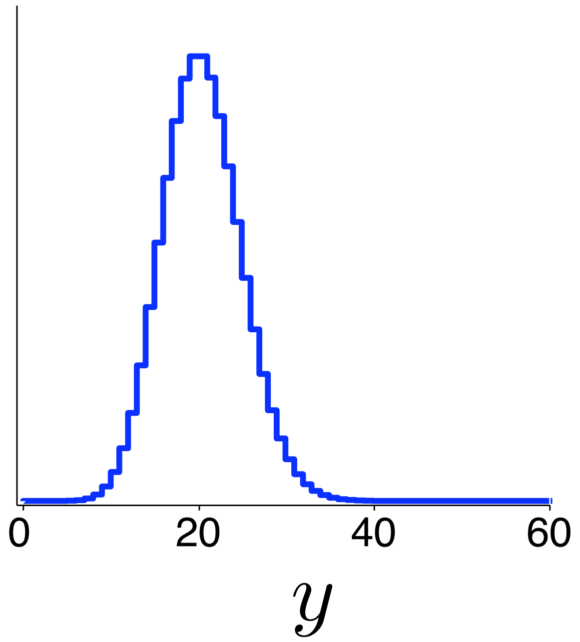


$$\mathrm{mean}(y) = \theta x$$
$$\mathrm{var}(y) = \theta x$$

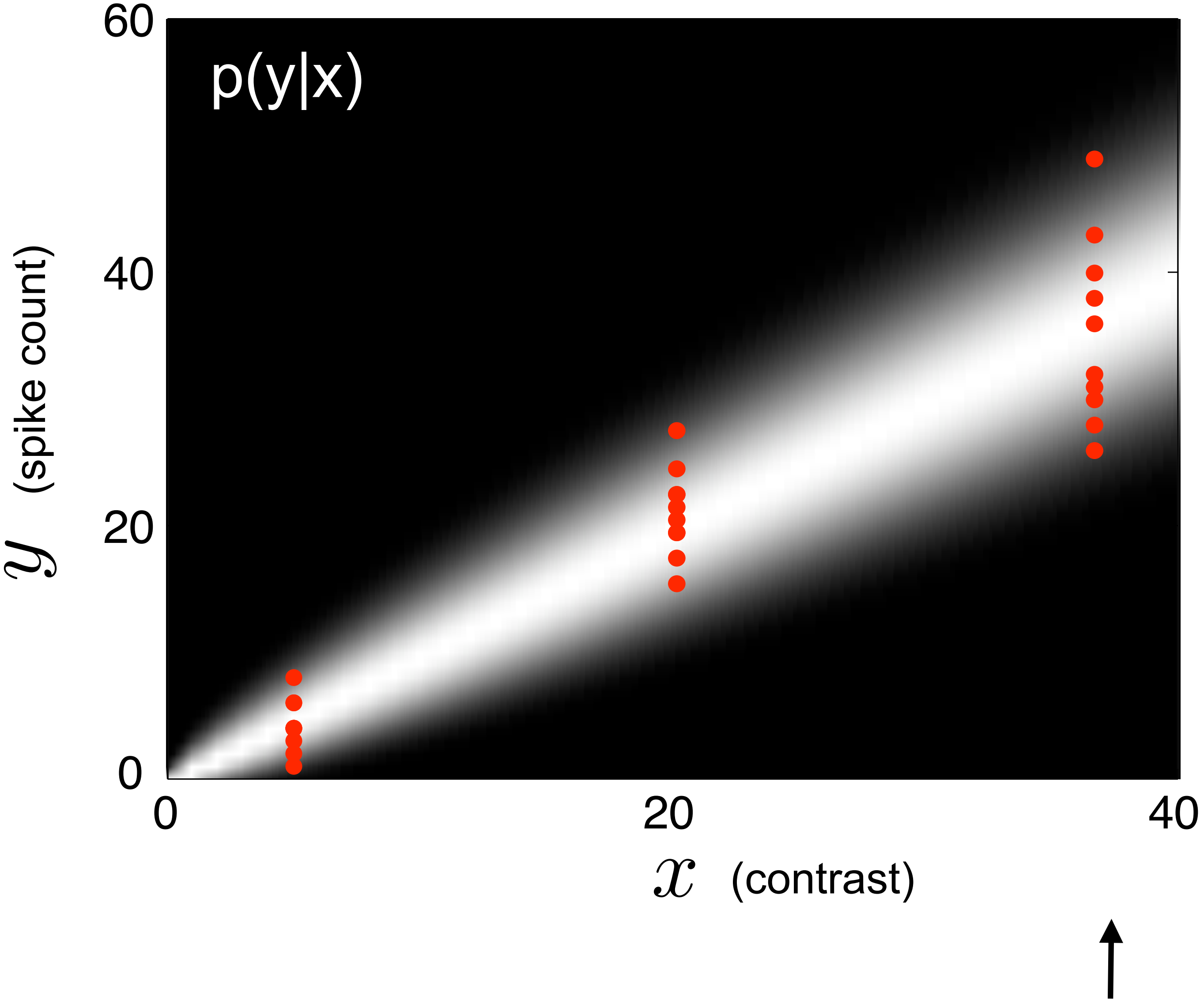


conditional distribution

$$p(y|x=20)$$

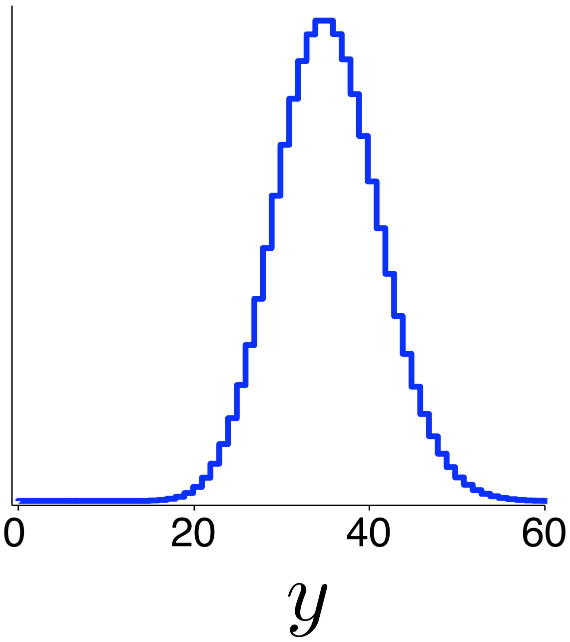


$$\mathrm{mean}(y) = \theta x$$
$$\mathrm{var}(y) = \theta x$$



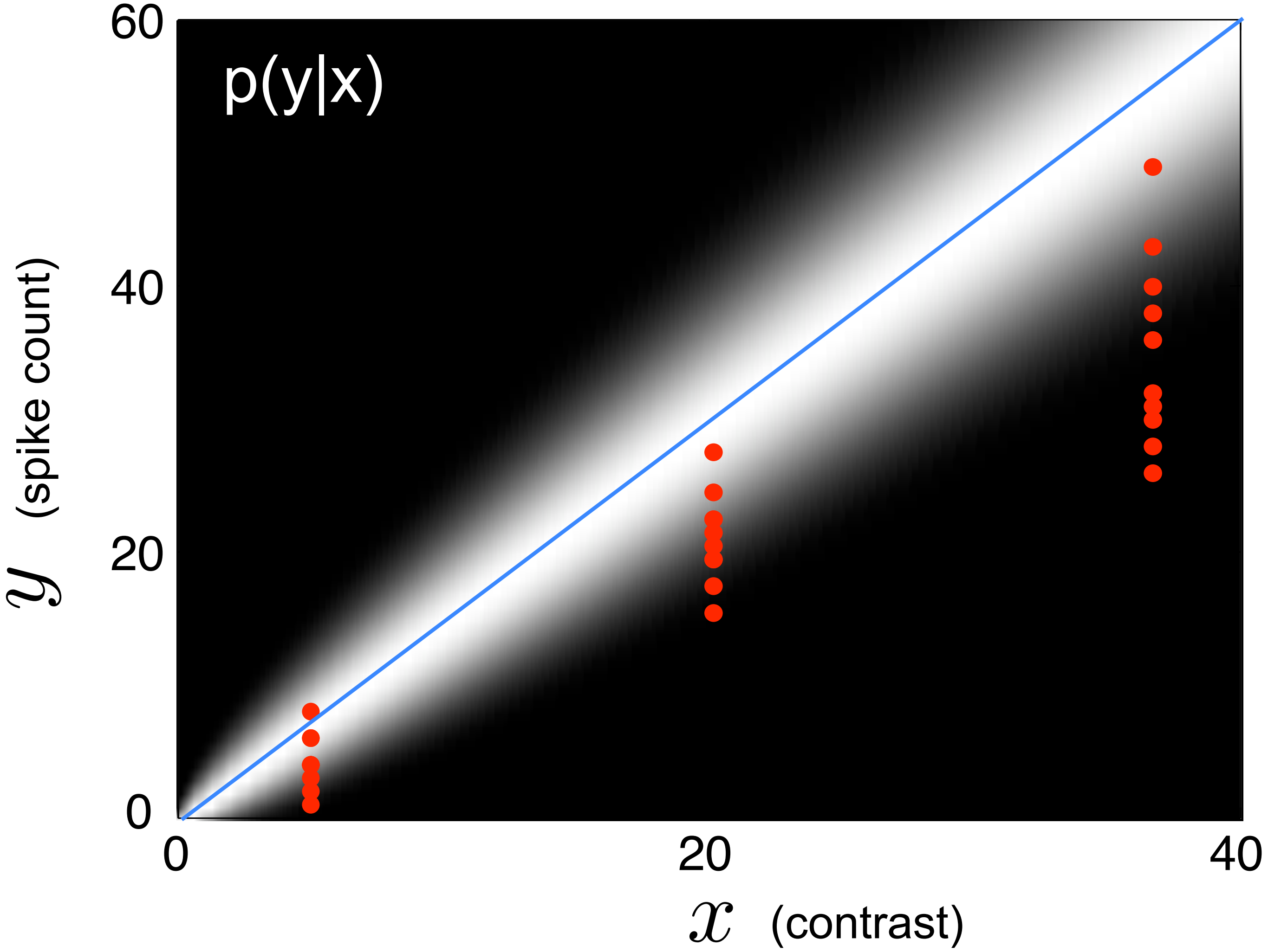
conditional distribution

$$p(y|x = 35)$$



Maximum Likelihood Estimation:

- given observed data (Y, X), find θ that maximizes $P(Y|X,\theta)$

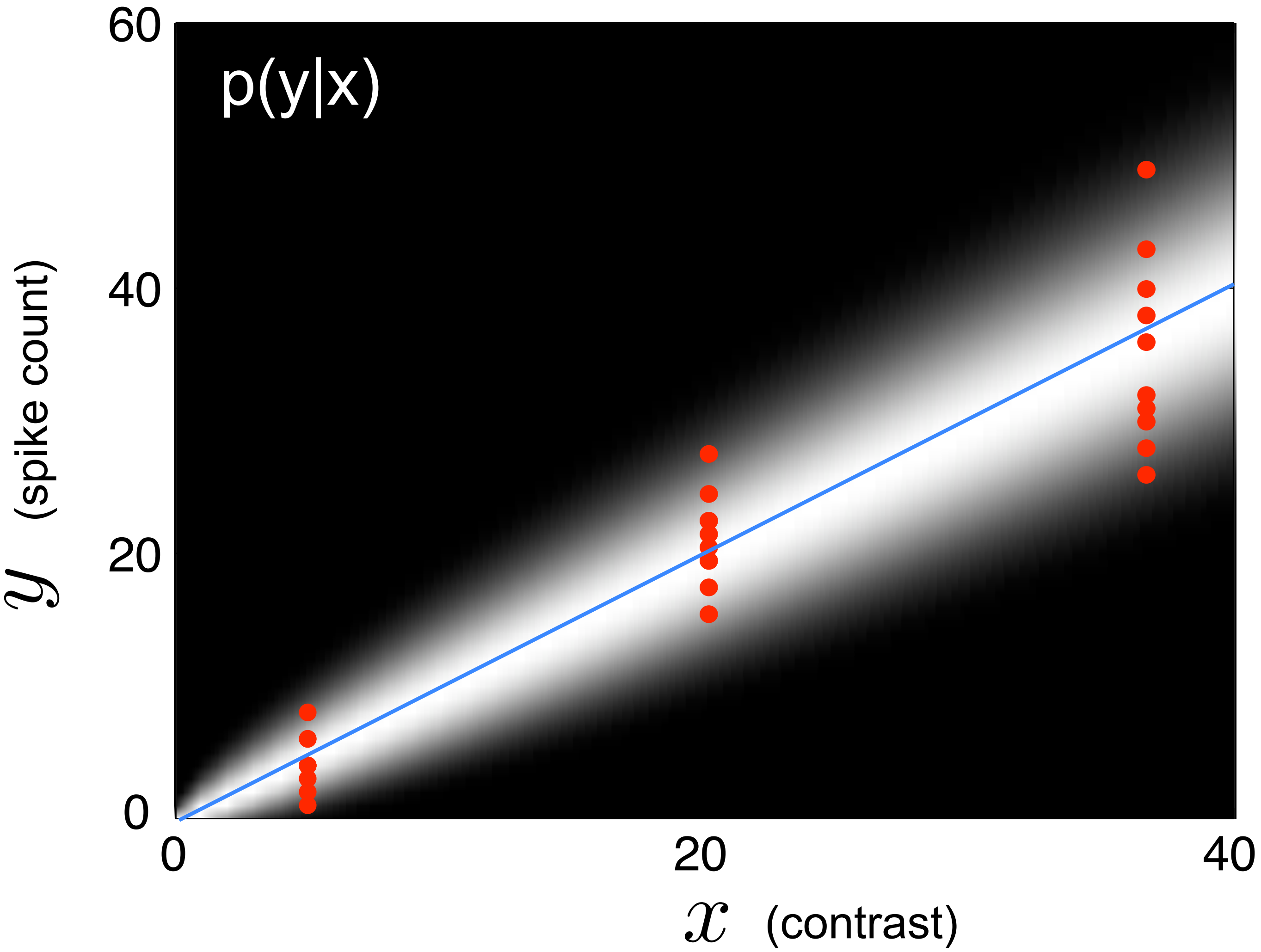


$y \sim \mathrm{Poiss}(\theta x)$

$\theta = 1.5$

Maximum Likelihood Estimation:

- given observed data (Y, X), find θ that maximizes $P(Y|X, \theta)$

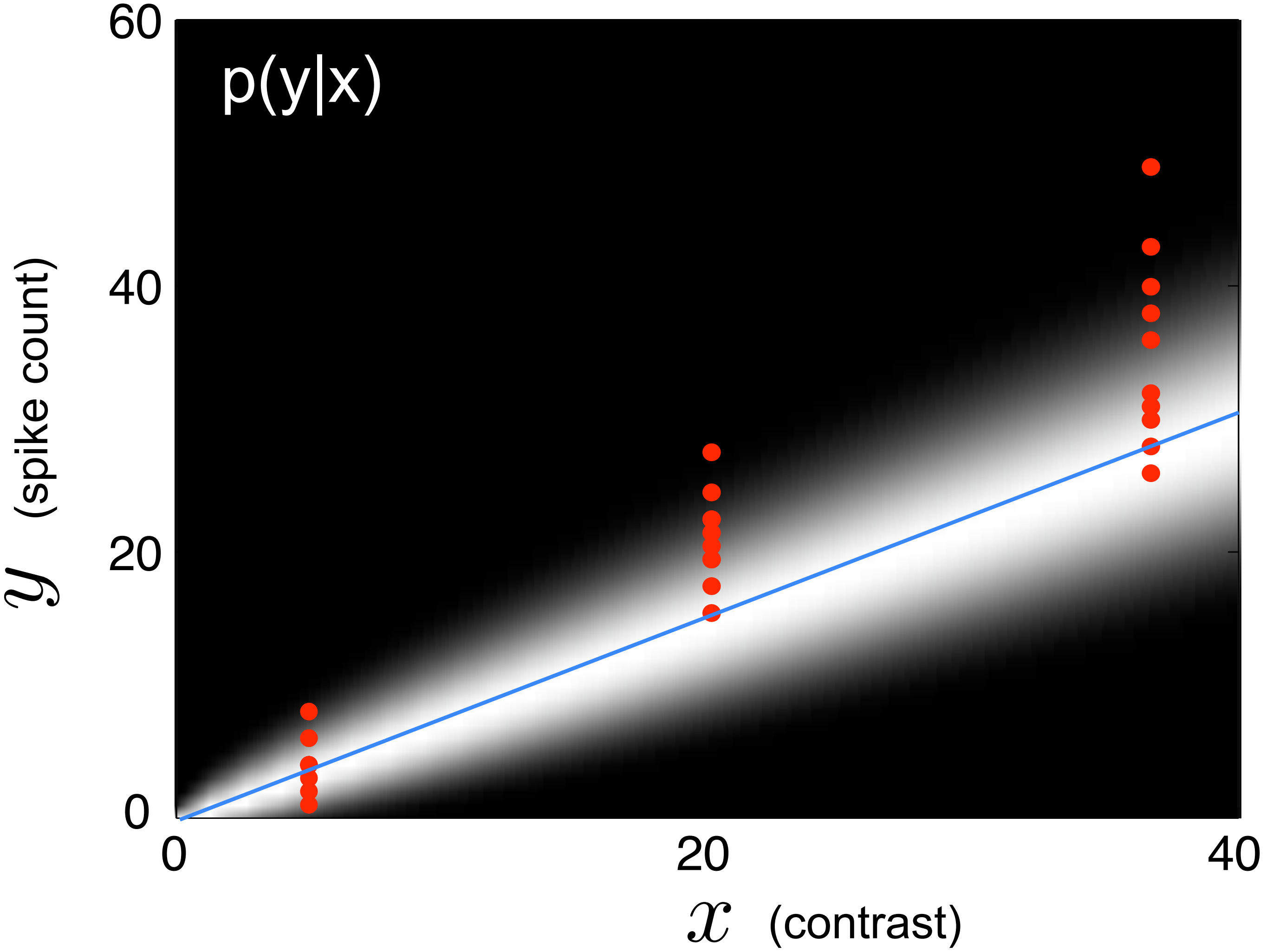


$y \sim \mathrm{Poiss}(\theta x)$

$\theta = 1$

Maximum Likelihood Estimation:

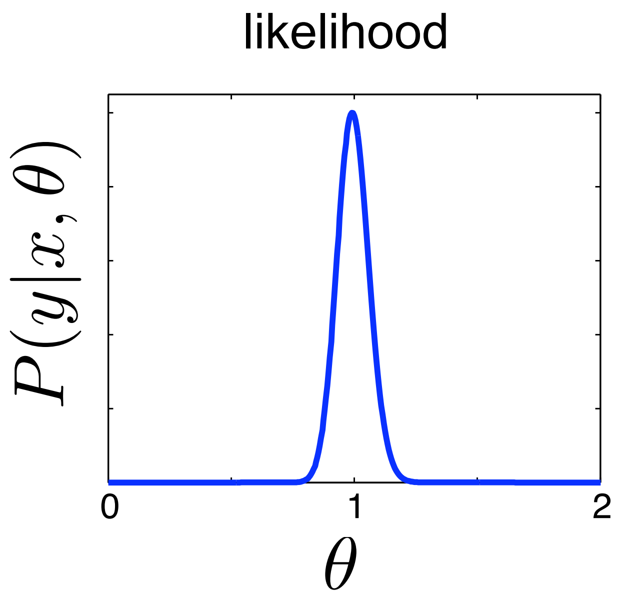
- given observed data (Y, X), find θ that maximizes $P(Y|X,\theta)$



$$y \sim \mathrm{Poiss}(\theta x)$$

$$\theta = 0.5$$

Likelihood function: $P(Y|X,\theta)$ as a function of θ



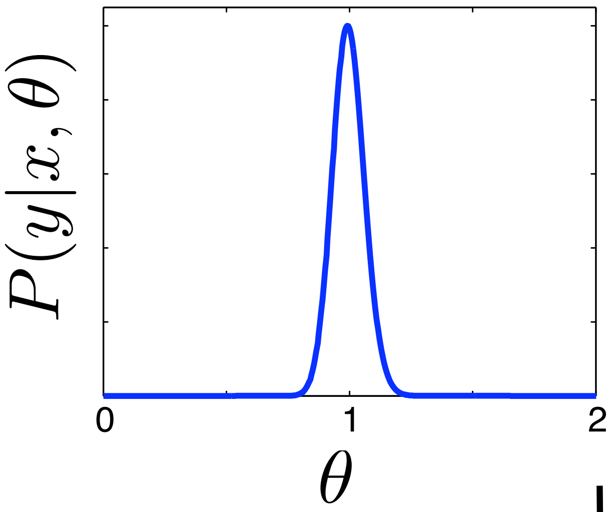
Because data are independent:

$$P(Y|X,\theta) = \prod_i P(y_i|x_i,\theta)$$

$$= \prod \frac{1}{y_i!}(\theta x_i)^{y_i} e^{-(\theta x_i)}$$

Likelihood function: $P(Y|X,\theta)$ as a function of θ

likelihood



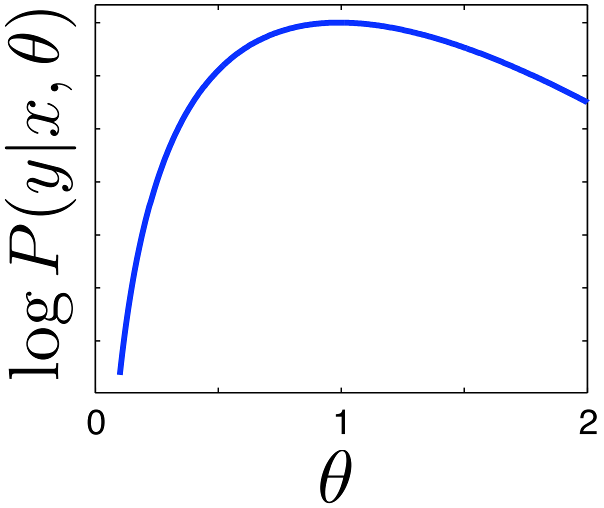
Because data are independent:

$$P(Y|X,\theta) = \prod_i P(y_i|x_i,\theta)$$

$$= \prod \frac{1}{y_i!}(\theta x_i)^{y_i} e^{-(\theta x_i)}$$

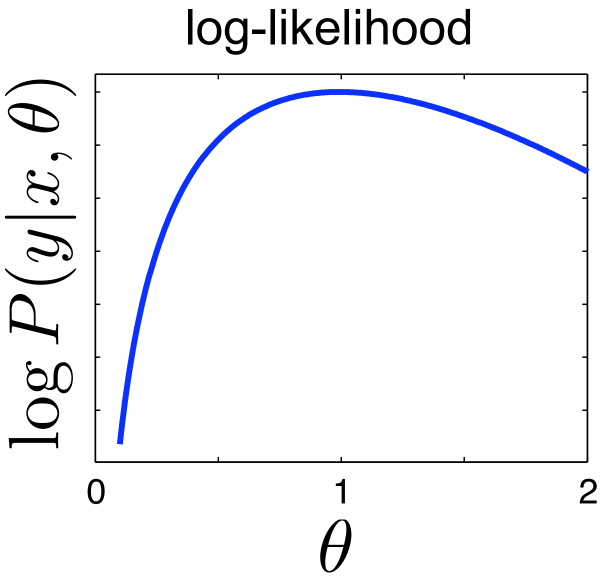
↓ log

log-likelihood



$$\log P(Y|X,\theta) = \sum_i \log P(y_i|x_i,\theta)$$

$$= \sum y_i \log\theta - \theta x_i + c$$



$$\log P(Y|X,\theta) = \textstyle\sum_i \log P(y_i|x_i,\theta)$$

$$= \sum y_i \log\theta - \theta x_i + c$$

$$= \log\theta(\sum y_i) - \theta(\sum x_i)$$

- Closed-form solution (exists for “exponential family” models)

$$\tfrac{d}{d\theta}\log P(Y|X,\theta) = \tfrac{1}{\theta}\sum y_i - \sum x_i = 0$$

$$\implies \quad \hat{\theta}_{ML} = \frac{\sum y_i}{\sum x_i}$$

Properties of the MLE (maximum likelihood estimator)

- consistent
  (converges to true θ in limit of infinite data)

- efficient
  (converges as quickly as possible,
  i.e., achieves minimum possible asymptotic error)

Example 2: linear Gaussian neuron

spike count $y \sim \mathcal{N}(\mu, \sigma^2)$

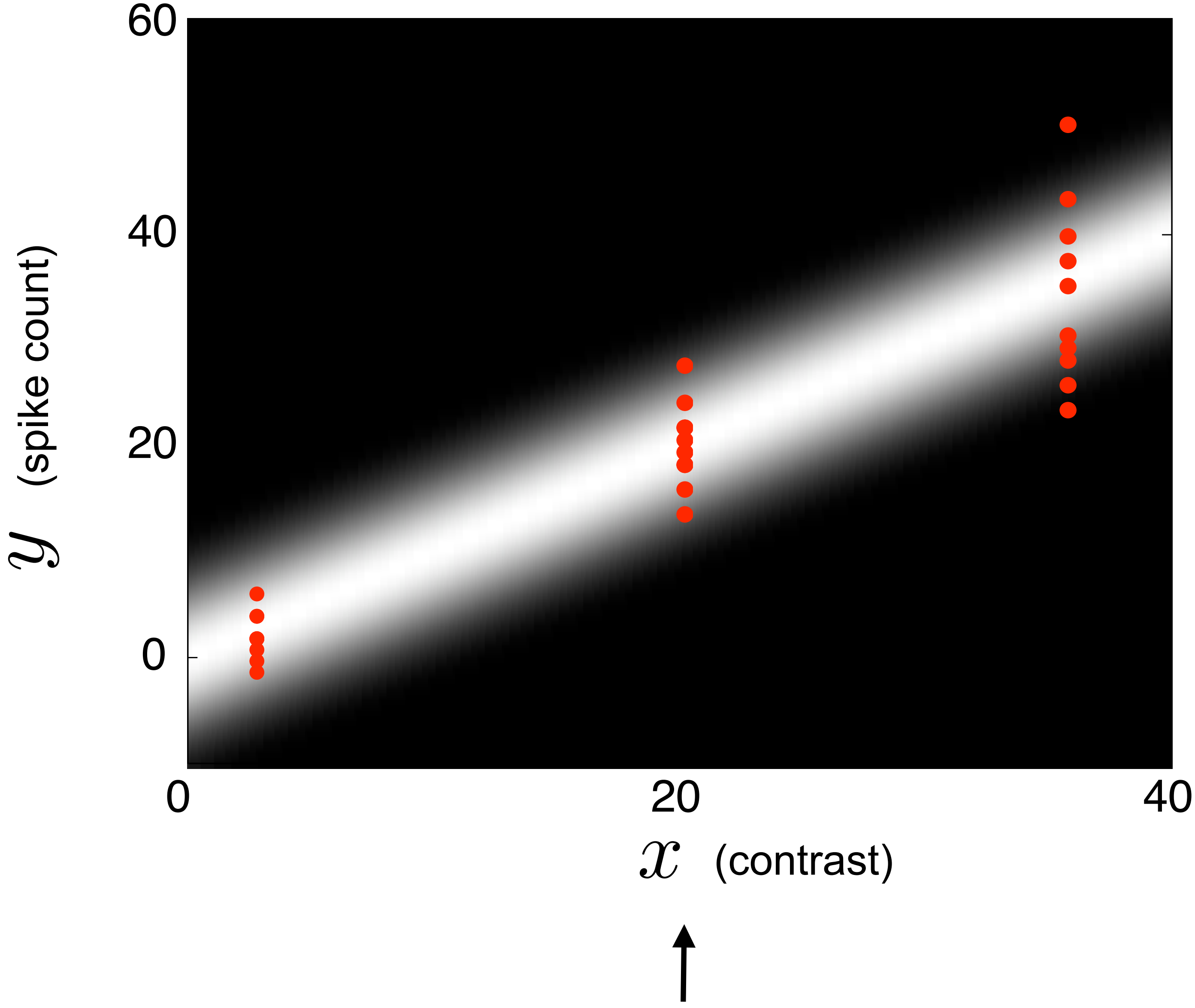
spike rate $\mu = \theta x$

parameter

stimulus

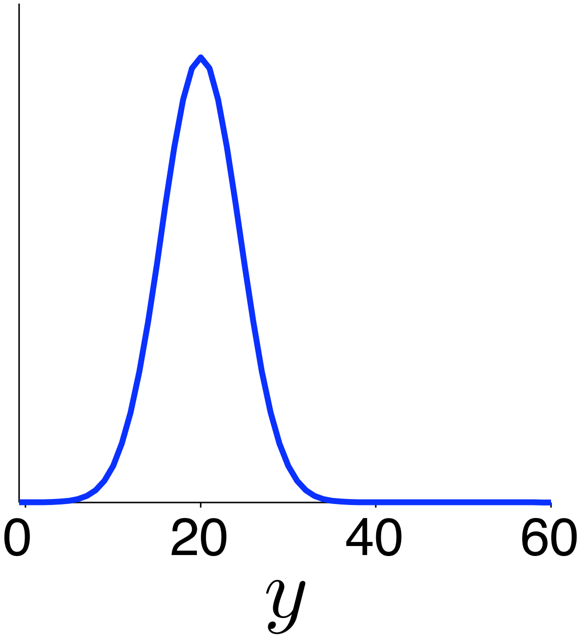
encoding model: $P(y|x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\theta x)^2}{2\sigma^2}}$

$$\mathrm{mean}(y) = \theta x$$
$$\mathrm{var}(y) = \sigma^2$$



encoding distribution

$$p(y|x = 20)$$



All slices have same width

Log-Likelihood $\log P(Y|X,\theta) = -\sum \frac{(y_i - \theta x_i)^2}{2\sigma^2} + c$

Differentiate and set to zero:

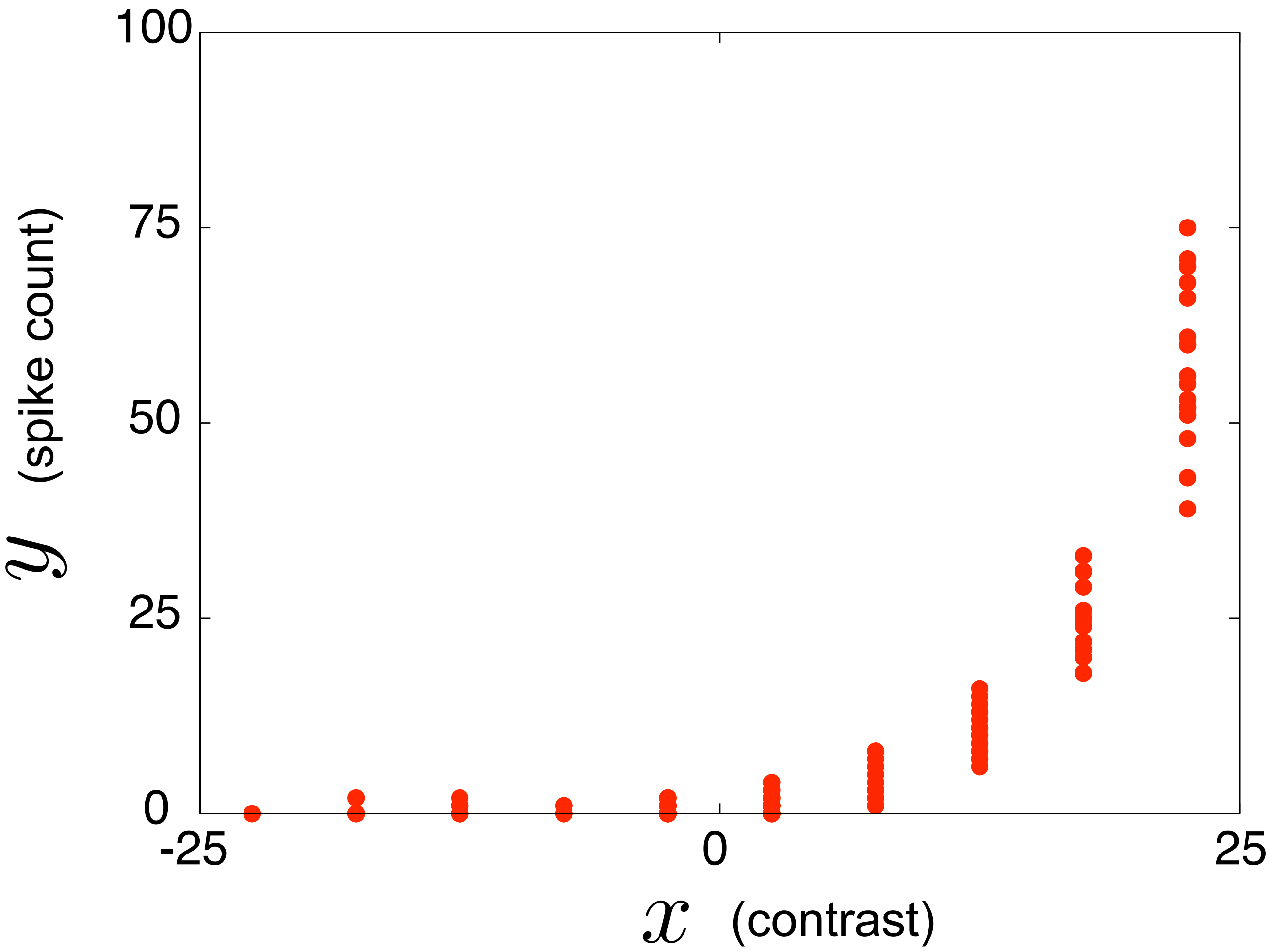
$$\frac{d}{d\theta} \log P(Y|X,\theta) = -\sum \frac{(y_i - \theta x_i)x_i}{\sigma^2} = 0$$

Maximum-Likelihood Estimator: $\hat{\theta}_{ML} = \frac{\sum y_i x_i}{\sum x_i^2}$

("Least squares regression" solution)

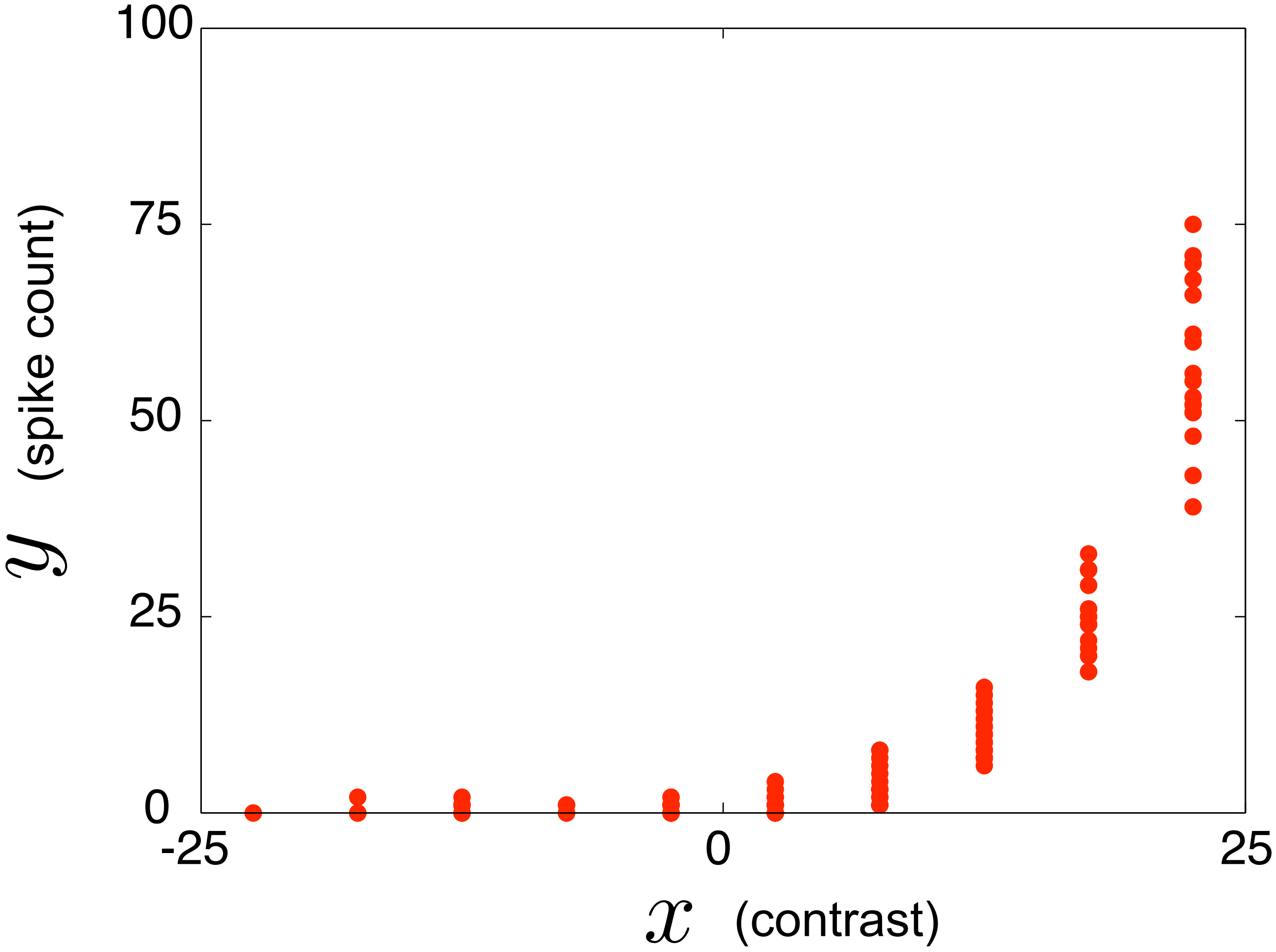
(Recall that for Poisson, $\hat{\theta}_{ML} = \frac{\sum y_i}{\sum x_i}$)

Example 3: unknown neuron



What model would you use to fit this neuron?

Example 3: unknown neuron



More general setup: $y \sim Poiss(\lambda)$

$\lambda = f(\theta x)$

for some nonlinear function f