Introduction to Bayesian inference and generative models

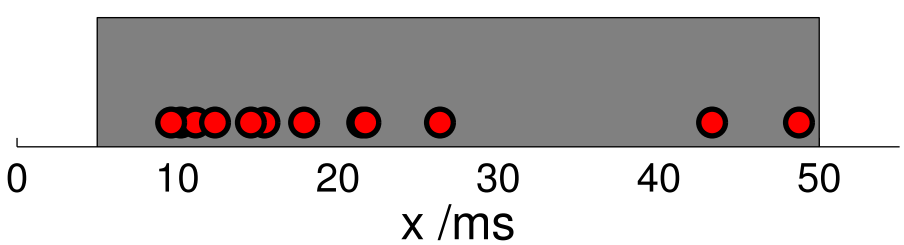
Dr. Richard E. Turner (ret26@cam.ac.uk)

Computational and Biological Learning Lab, Department of Engineering
University of Cambridge

Laboratory for Computational Vision, Center for Neural Science
New York University

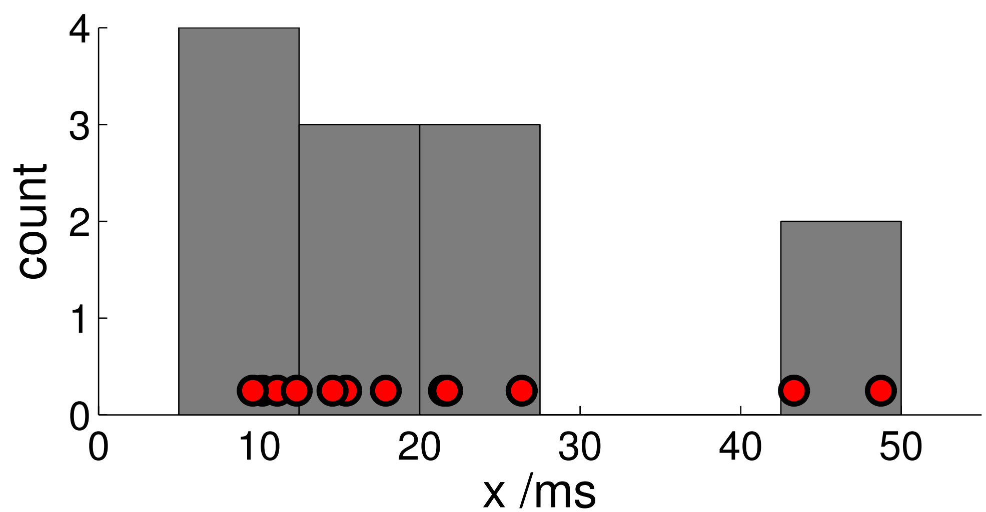
Question

- Collected inter-spike interval measurements, x, from a neuron
- x follows an exponential distribution with a characteristic time-scale λ, shifted to take account of the absolute refractory period of the neuron, 5ms long.
- ISIs over 50ms were not recorded (short trials used for data-collection)
- N ISIs are observed at $\{\mathrm{x}_1, \ldots \mathrm{x}_N\}$. What is λ?



Question

- Collected inter-spike interval measurements, x, from a neuron
- x follows an exponential distribution with a characteristic time-scale λ, shifted to take account of the absolute refractory period of the neuron, 5ms long.
- ISIs over 50ms were not recorded (short trials used for data-collection)
- N ISIs are observed at $\{\mathrm{x}_1, \ldots \mathrm{x}_N\}$. What is λ?



Ideas

Idea 1

- bin up into a histogram
  - where do we place the bins
- fit to density
  - what error measure do we minimise?

Idea 2

- construct an estimator e.g. the sample mean $\mu = \frac{1}{N}\sum_{n=1}^{N} \mathrm{x}_n$
  - which estimator should we choose? mean, variance, higher moments?
- relate to parameters via expectation of estimator e.g. $\mu \approx \langle x \rangle = f(\lambda)$
  - small sample effects can be problematic e.g. if $\mu > \frac{1}{2}(50+5)$ms

A less ad hoc method...probabilities as degrees of belief

Cox's axioms

- degrees of belief can be represented by real numbers
- take into account all evidence
- consistency: if things can be reasoned in more than one way, each must lead to the same answer
- equivalent states of knowledge $\rightarrow$ equivalent plausibility assignments

A less ad hoc method...probabilities as degrees of belief

Cox's axioms

- degrees of belief can be represented by real numbers
- take into account all evidence
- consistency: if things can be reasoned in more than one way, each must lead to the same answer
- equivalent states of knowledge $\rightarrow$ equivalent plausibility assignments

Conclusion: degrees of belief must follow the rules of probability.

A less ad hoc method...probabilities as degrees of belief

Cox's axioms

- degrees of belief can be represented by real numbers
- take into account all evidence
- consistency: if things can be reasoned in more than one way, each must lead to the same answer
- equivalent states of knowledge $\rightarrow$ equivalent plausibility assignments

Conclusion: degrees of belief must follow the rules of probability.

Product rule: $p(\lambda, \mathrm{x}) = p(\lambda|\mathrm{x})p(\mathrm{x}) = p(\mathrm{x}|\lambda)p(\lambda)$
Sum rule: $p(\lambda) = \sum_{\mathrm{x}} p(\lambda, \mathrm{x})$

A less ad hoc method...probabilities as degrees of belief

Cox's axioms

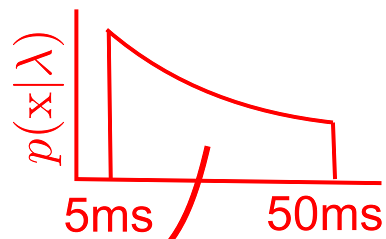
- degrees of belief can be represented by real numbers
- take into account all evidence
- consistency: if things can be reasoned in more than one way, each must lead to the same answer
- equivalent states of knowledge $\rightarrow$ equivalent plausibility assignments

Conclusion: degrees of belief must follow the rules of probability.

Product rule: $p(\lambda, \mathrm{x}) = p(\lambda|\mathrm{x})p(\mathrm{x}) = p(\mathrm{x}|\lambda)p(\lambda)$ $\leftarrow$**Bayes' Rule**
Sum rule: $p(\lambda) = \sum_{\mathrm{x}} p(\lambda, \mathrm{x})$ $\leftarrow$**marginalisation**

Mathematical solution

$$p(\mathrm{x}|\lambda) = \begin{cases} \frac{1}{Z(\lambda)} \exp\left(-\frac{\mathrm{x}}{\lambda}\right) & \text{if } 5\mathrm{ms} < \mathrm{x} < 50\mathrm{ms} \\ 0 & \text{otherwise} \end{cases}$$



$$Z(\lambda) = \int_{5}^{50} \exp\left(-\frac{\mathrm{x}}{\lambda}\right) \mathrm{dx} = \lambda \left(\exp\left(-\frac{5}{\lambda}\right) - \exp\left(-\frac{50}{\lambda}\right) \right)$$

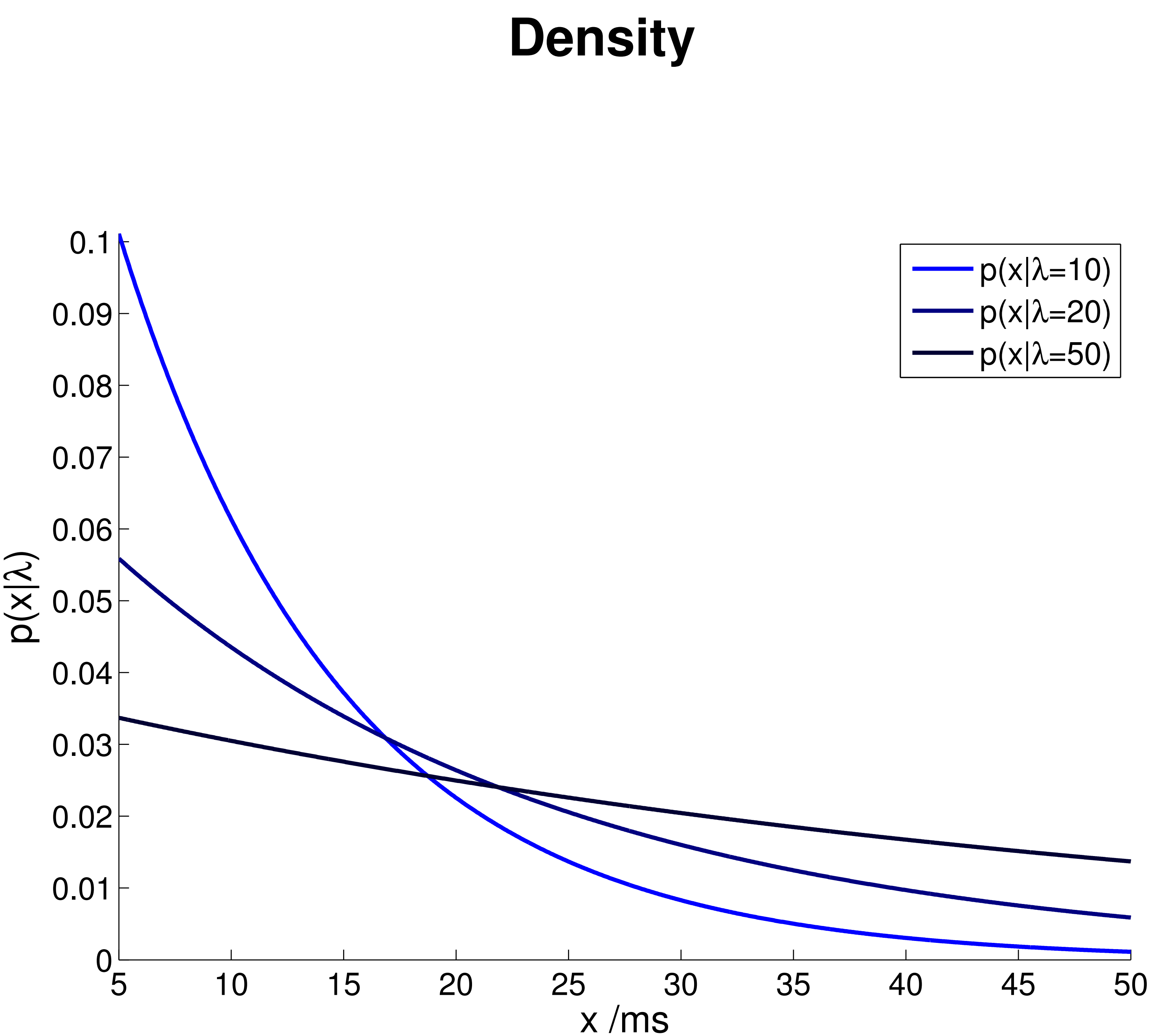
area must sum to one

what the data tell us (likelihood of parameters)

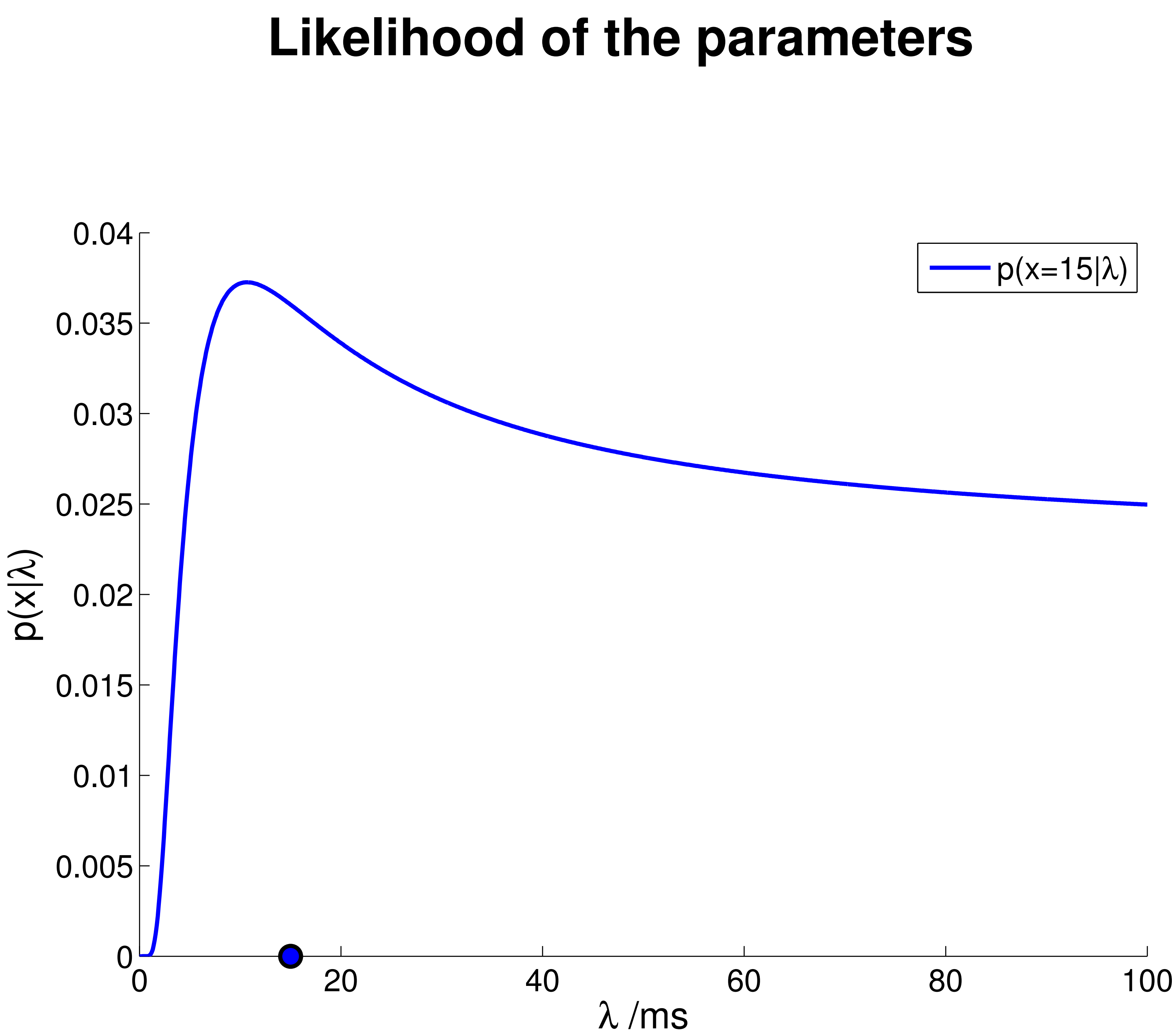
what we knew before hand (prior)

$$p(\lambda|\{\mathrm{x}_1 \ \ldots, \mathrm{x}_N\}) = \frac{p(\{\mathrm{x}\}|\lambda) p(\lambda)}{p(\{\mathrm{x}\})}$$

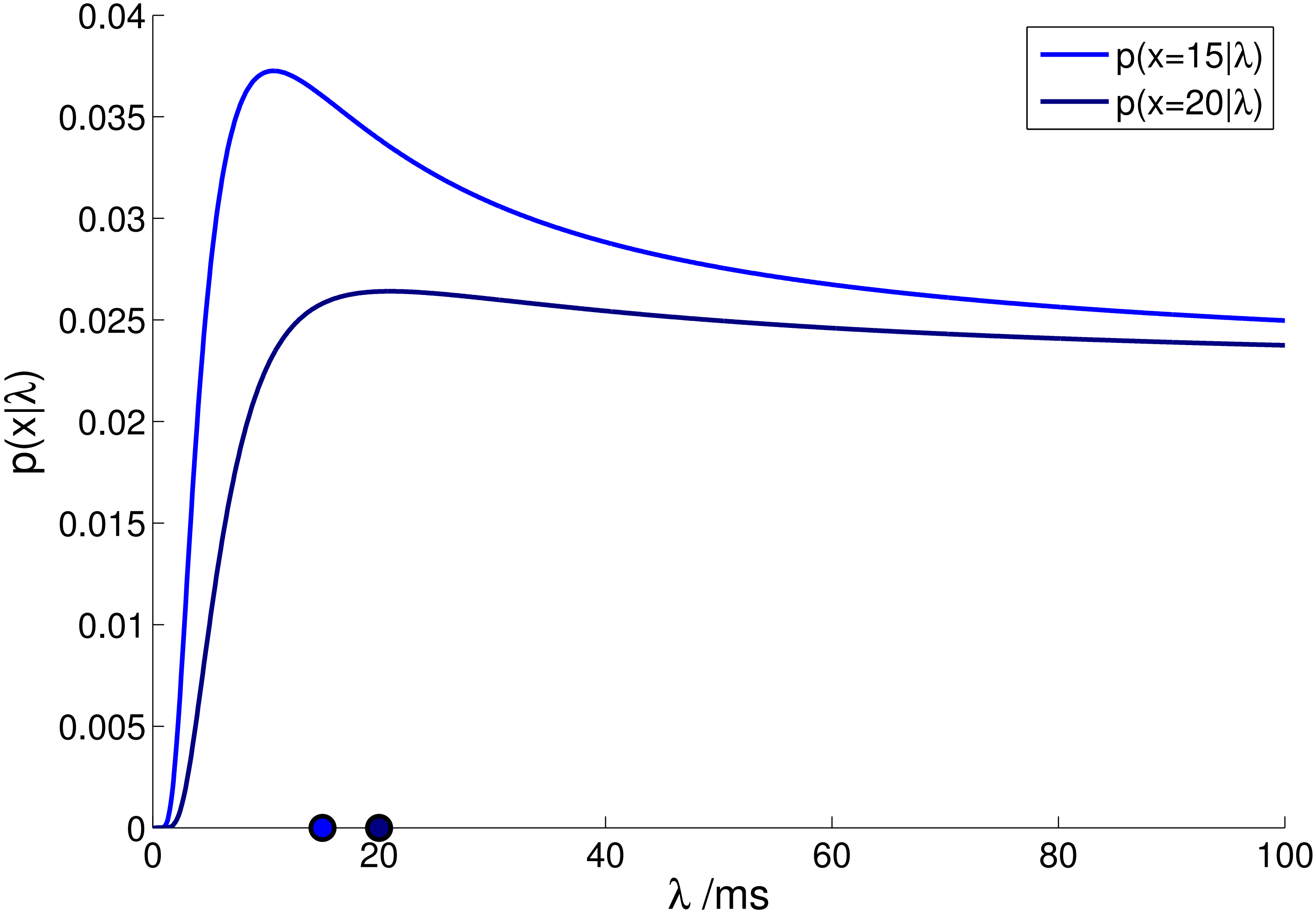
$$\propto \frac{1}{Z(\lambda)^N} \exp\left(-\frac{1}{\lambda} \sum_{n=1}^{N} \mathrm{x}_n \right) p(\lambda)$$



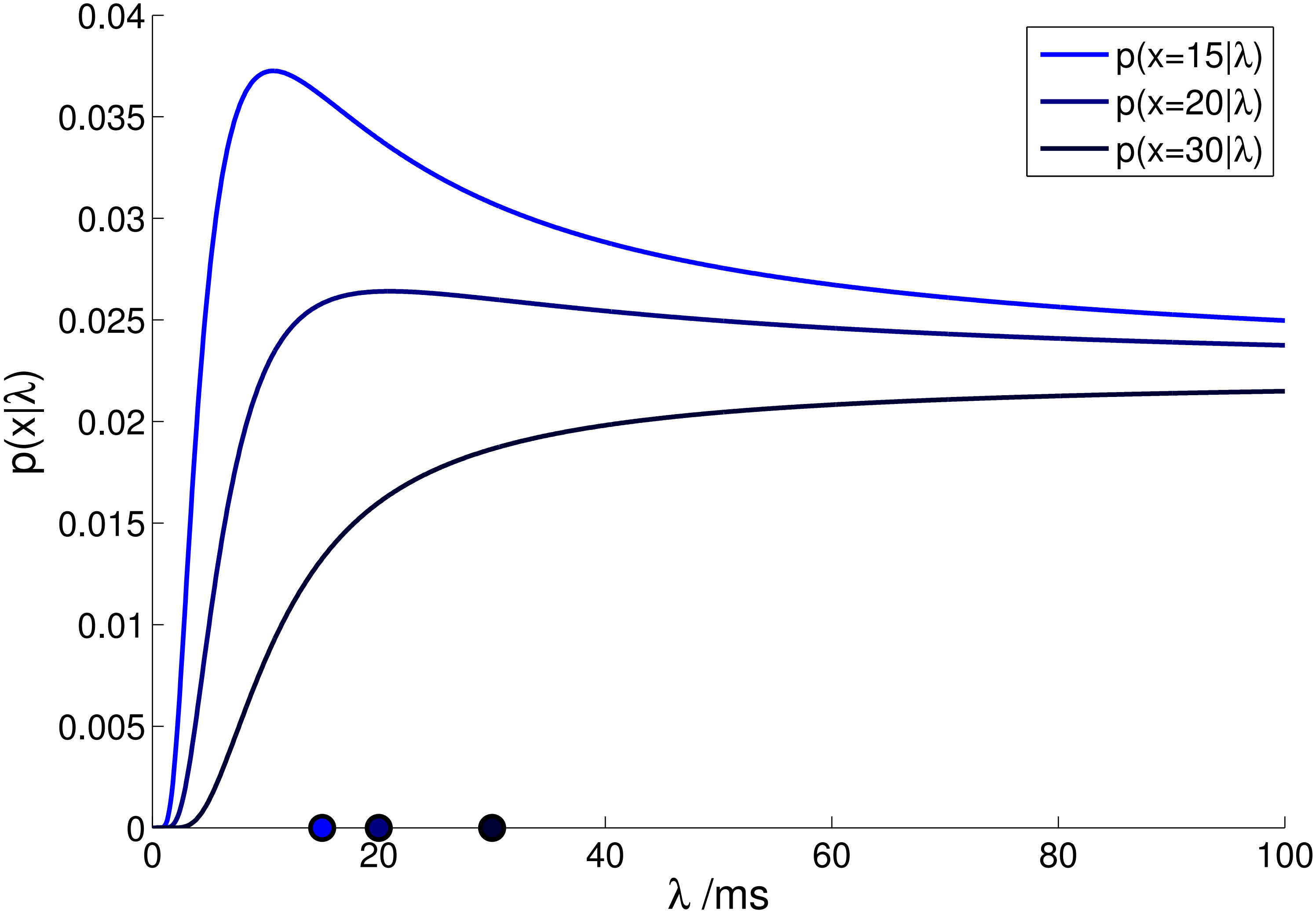
Density
p(x|λ=10)
p(x|λ=20)
p(x|λ=50)
p(x|λ)
x /ms
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
5
10
15
20
25
30
35
40
45
50


Likelihood of the parameters
p(x=15|λ)
p(x|λ)
λ /ms
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0
20
40
60
80
100

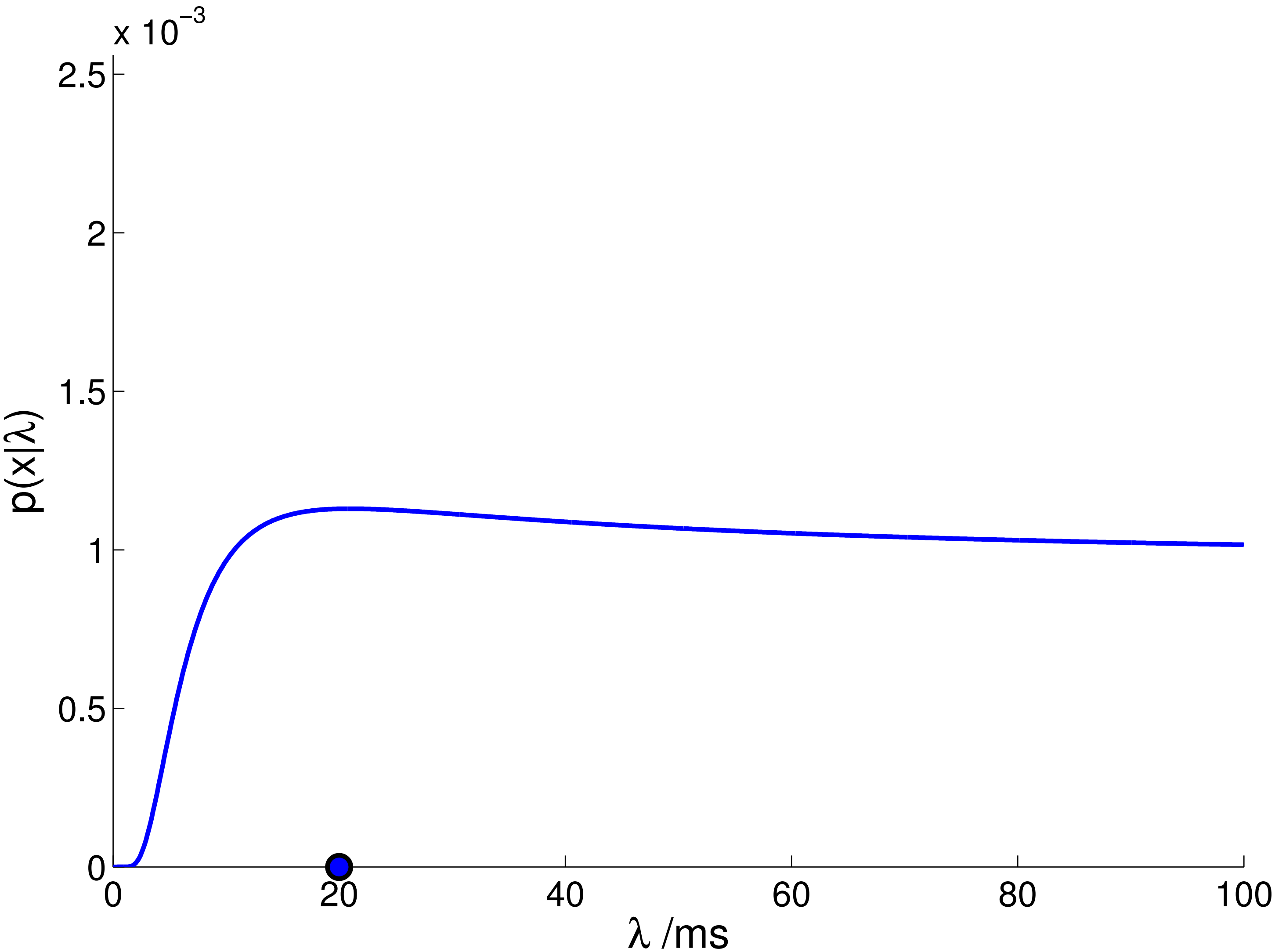
Likelihood of the parameters



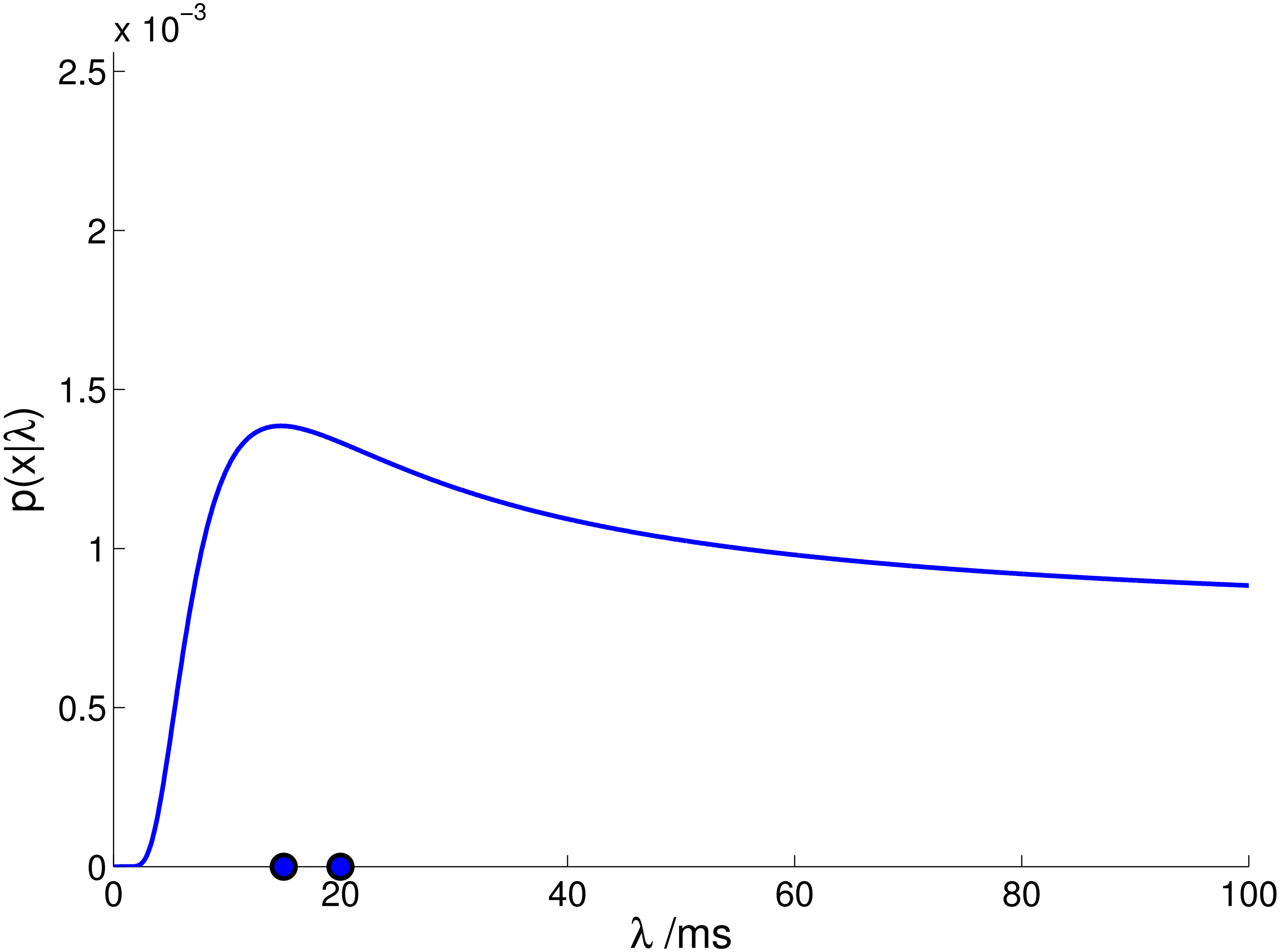
Likelihood of the parameters



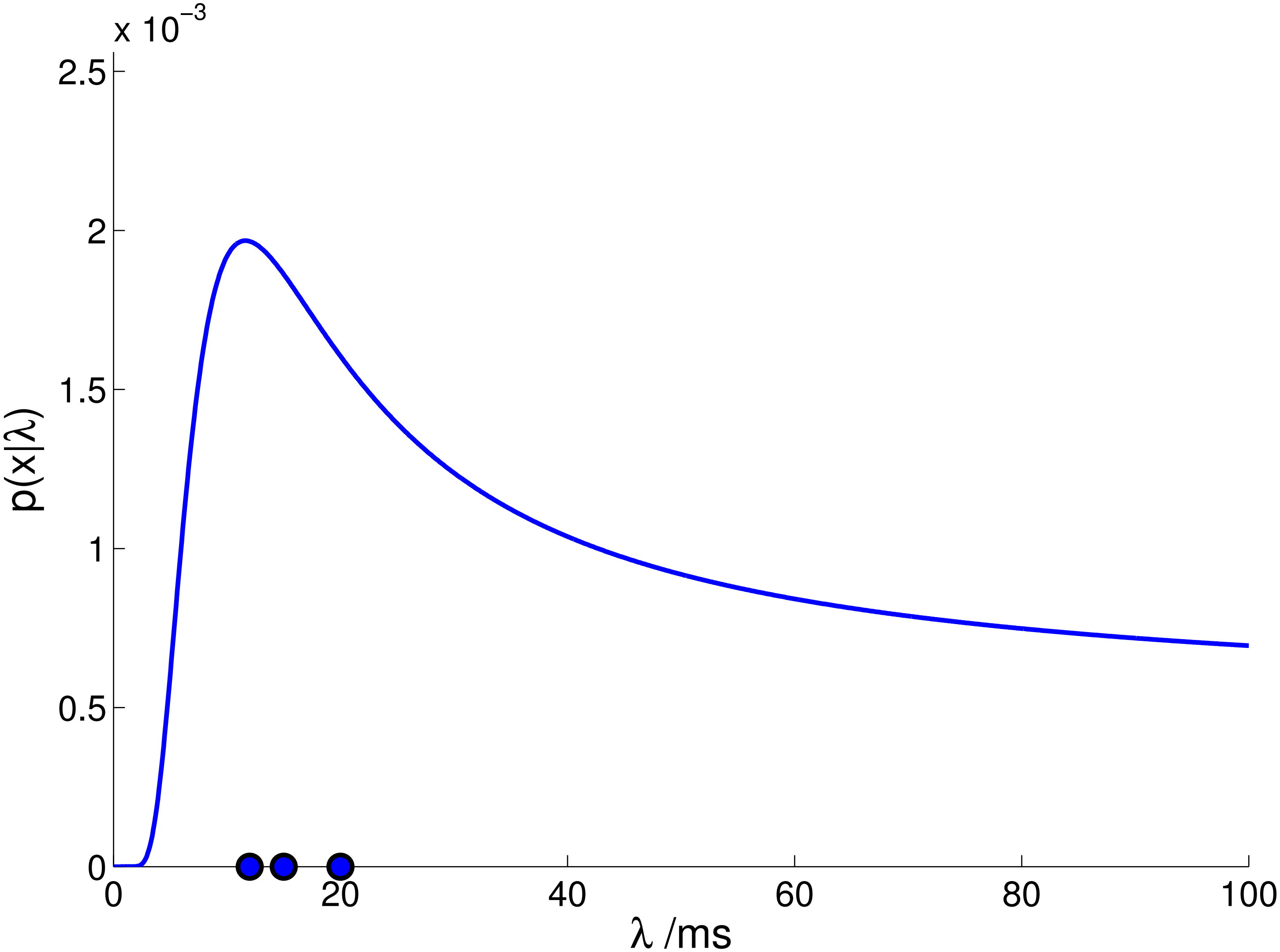
Posterior distribution: $p(\lambda|\mathrm{x}_1)$



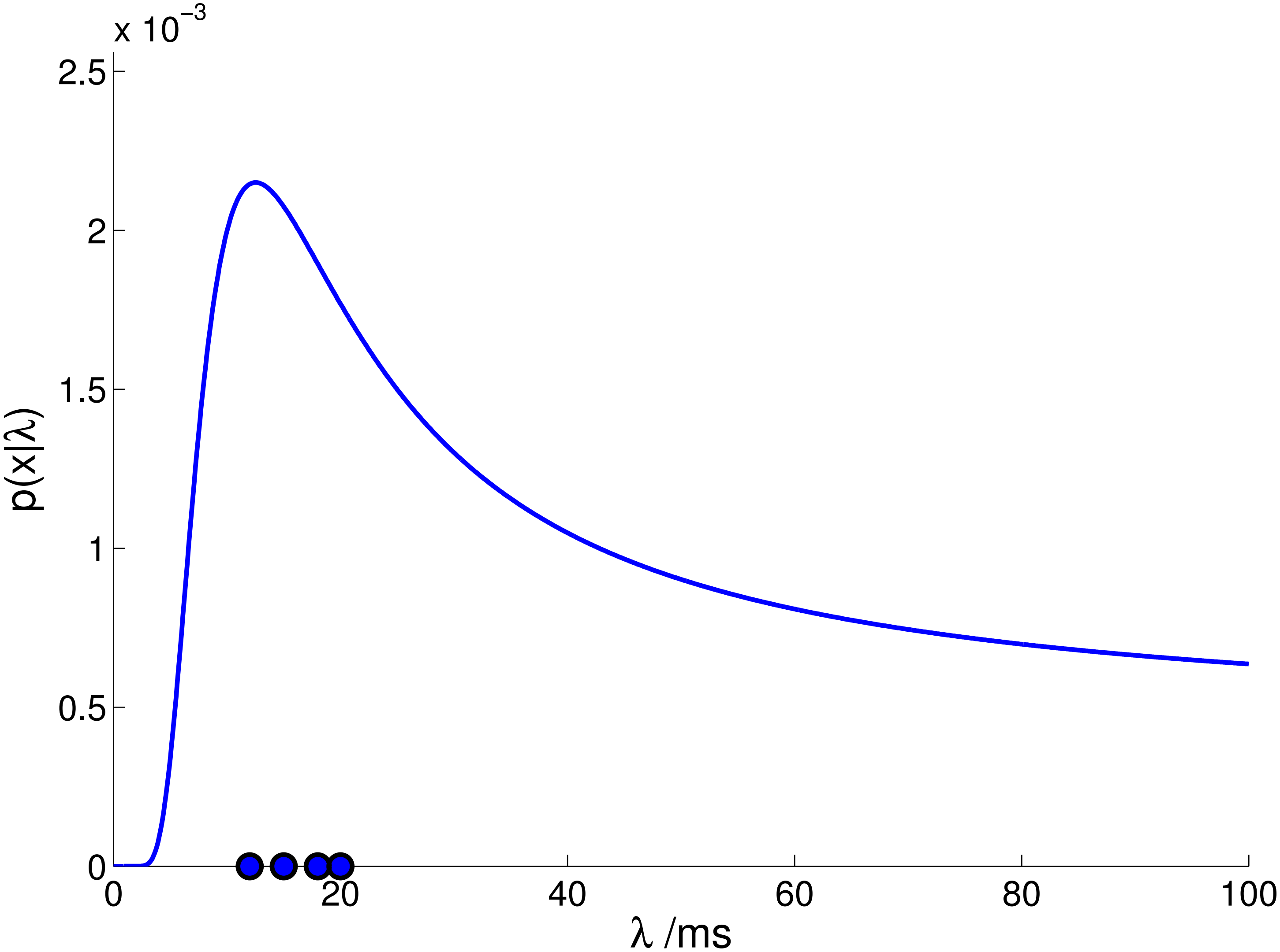
Posterior distribution: $p(\lambda|\mathrm{x}_1, \mathrm{x}_2)$



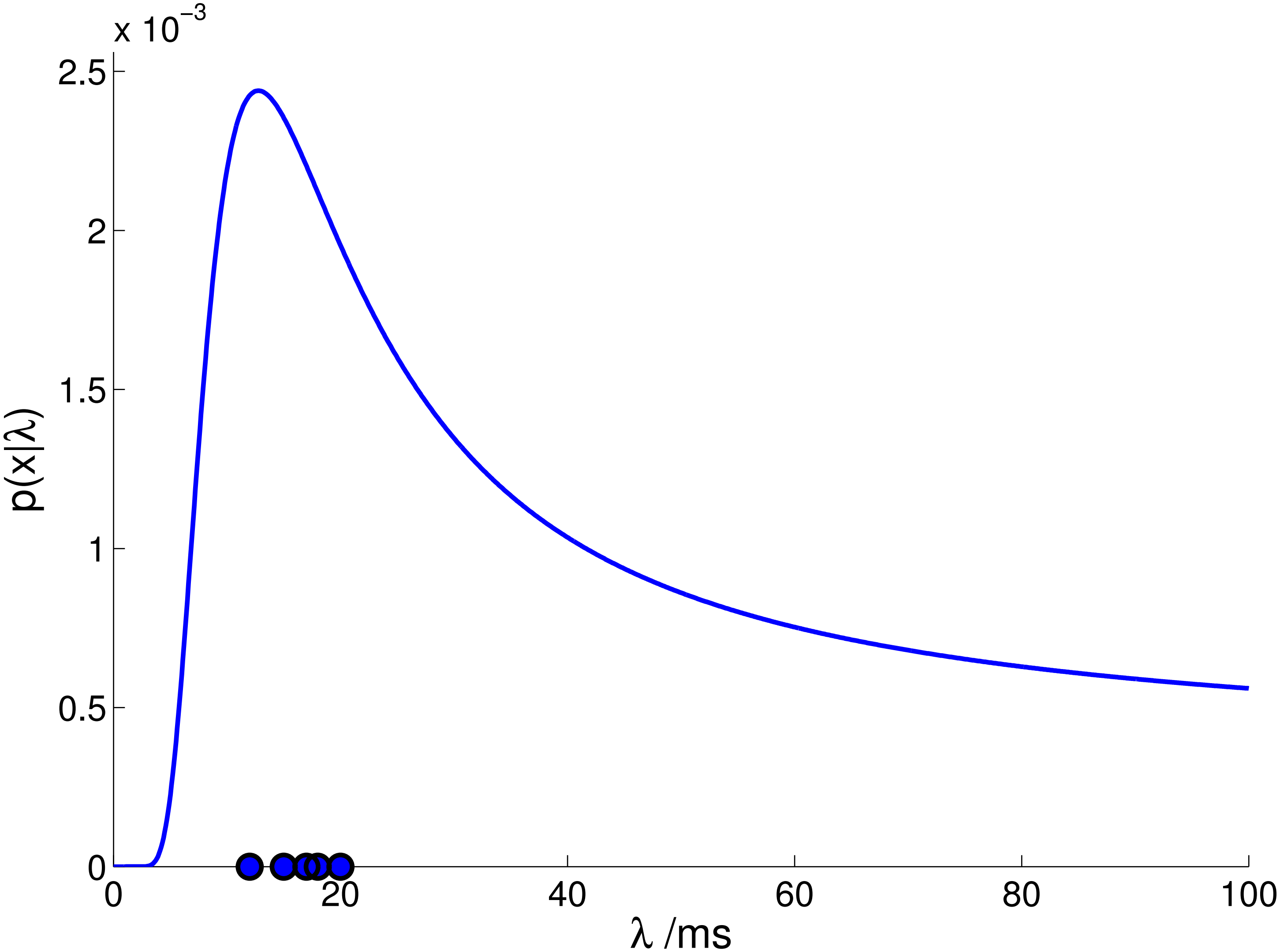
Posterior distribution: $p(\lambda|\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3)$



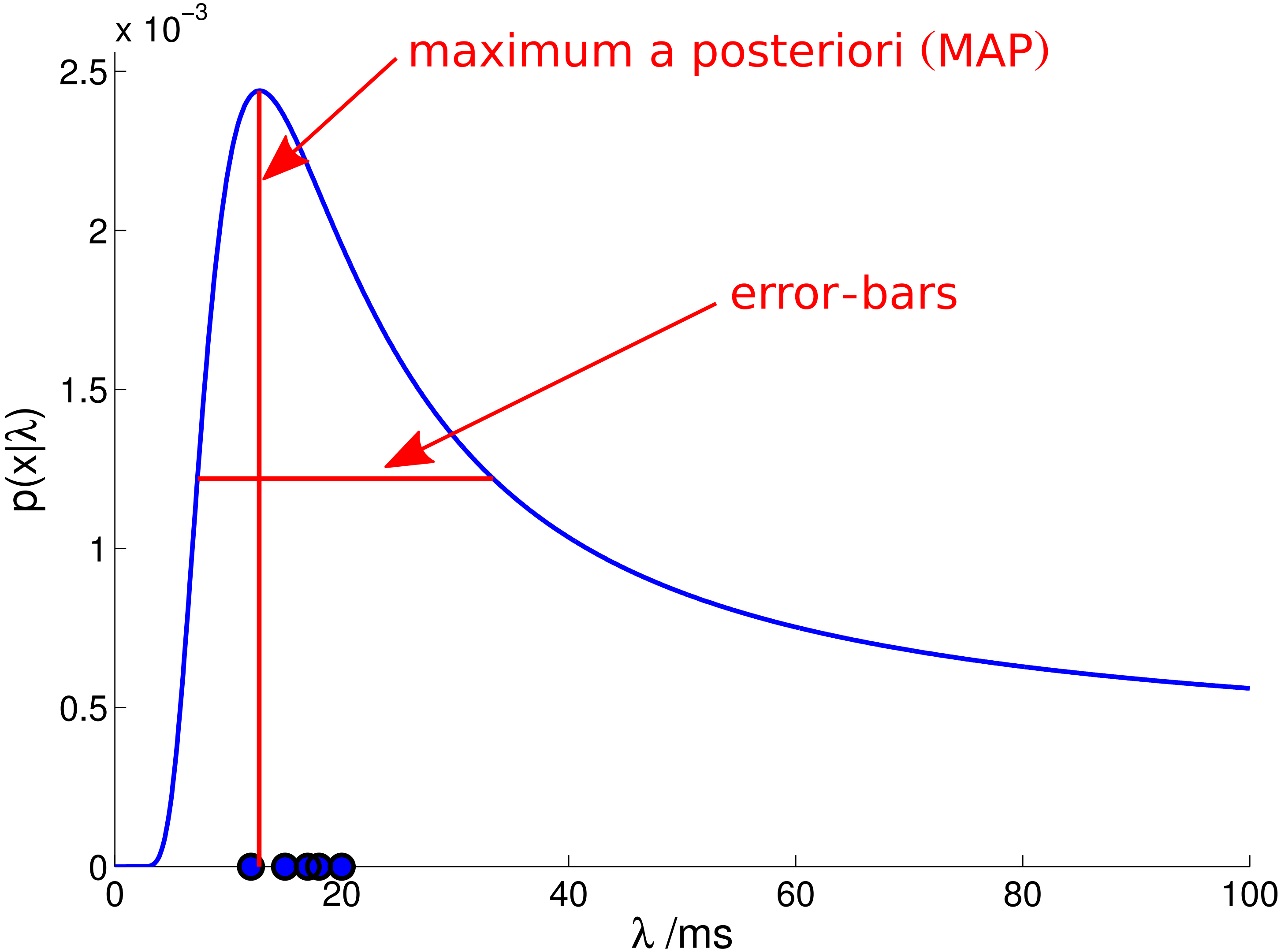
Posterior distribution: $p(\lambda|\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4)$



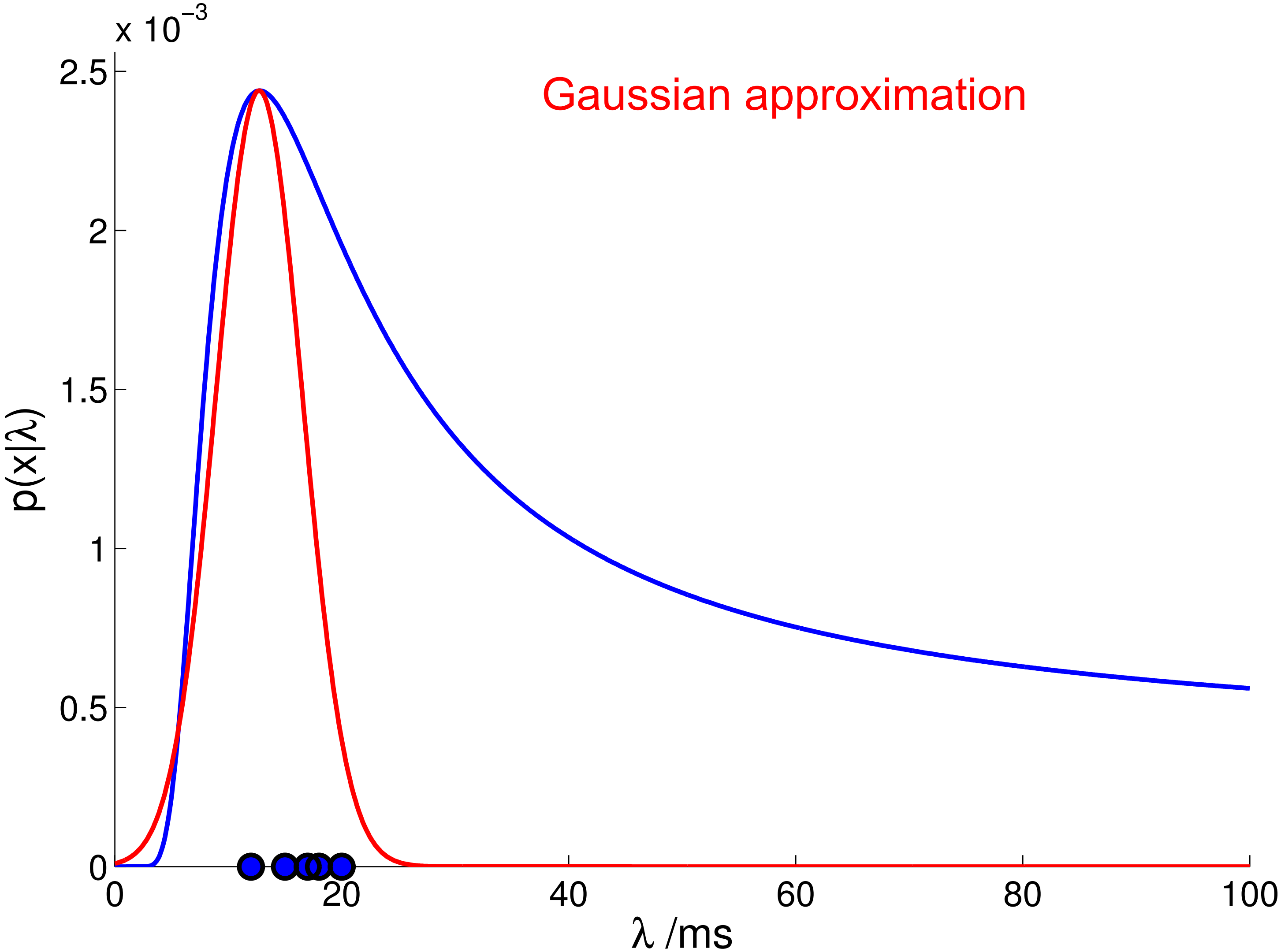
Posterior distribution: $p(\lambda|\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4, \mathrm{x}_5)$



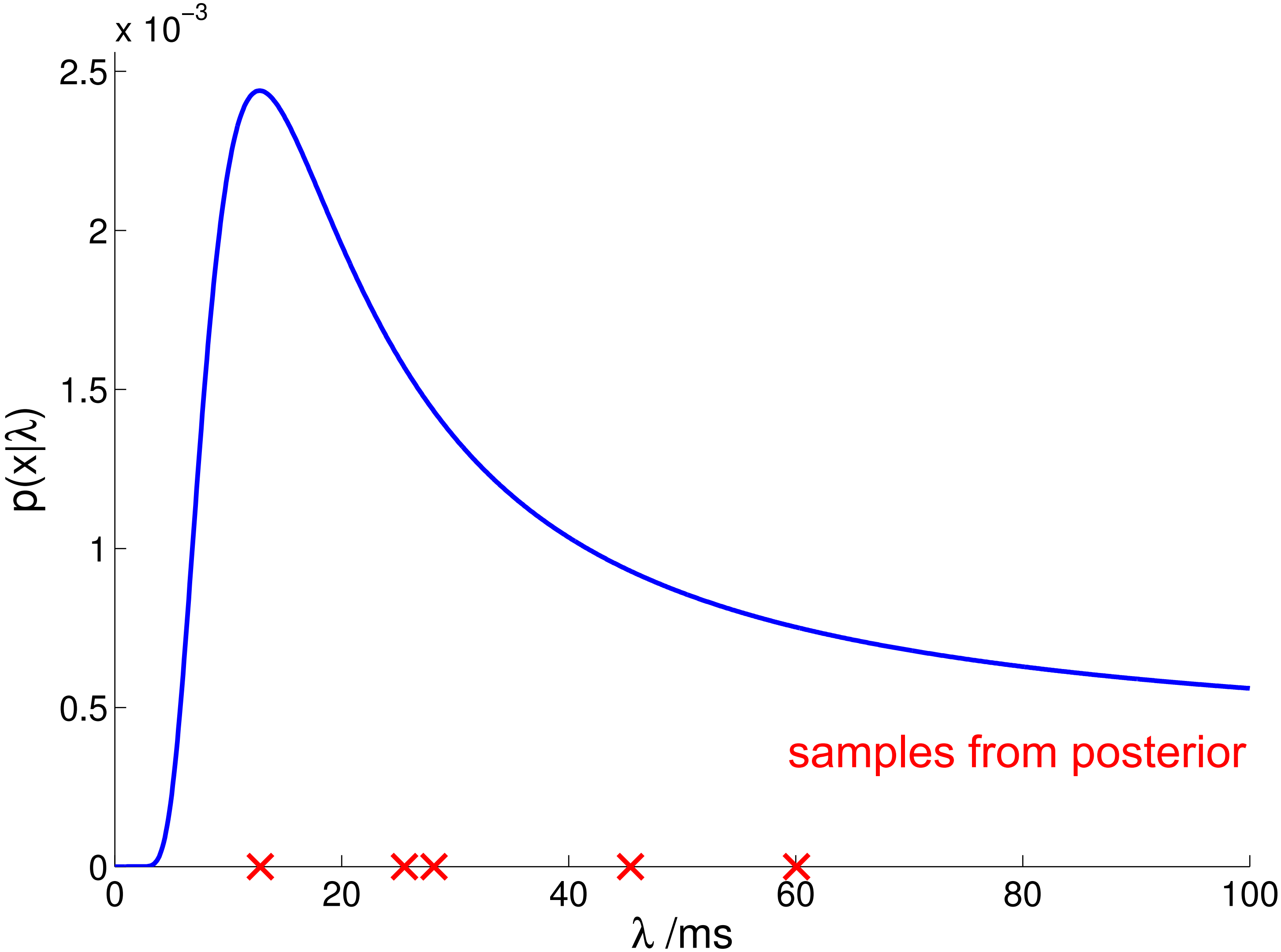
Summarising the posterior distribution



Summarising the posterior distribution



Summarising the posterior distribution



Question

- Record inter-spike interval measurements, x
- As before: absolute refractory of 5ms & ISIs above 50ms not recorded
- We know if the neuron is...
  - **quiescent**: x follows an exponential distribution with time-scale $\lambda_0 = 25$ms
  - **bursting**: x follows an exponential distribution with time-scale $\lambda_1 = 5$ms
- You observe a single ISI, $\mathrm{x} = 15$ms. Is the neuron in a bursting state?

Question

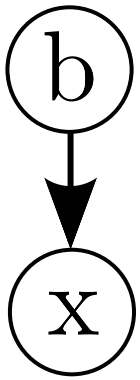
- Record inter-spike interval measurements, x
- As before: absolute refractory of 5ms & ISIs above 50ms not recorded
- We know if the neuron is...
  - **quiescent**: x follows an exponential distribution with time-scale $\lambda_0 = 25$ms
  - **bursting**: x follows an exponential distribution with time-scale $\lambda_1 = 5$ms
- You observe a single ISI, $\mathrm{x} = 15$ms. Is the neuron in a bursting state?

Intuition: should be close to 50:50

Mathematical solution

Introduce latent variable: $\mathrm{b} = 0$ not bursting
$\mathrm{b} = 1$ bursting

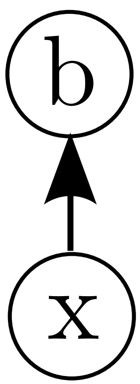
Generative model



$$p(\mathrm{b} = 0) = \tfrac{1}{2}$$

$$p(\mathrm{x}|\mathrm{b}, \lambda_0, \lambda_1) = \frac{1}{Z(\lambda_\mathrm{b})} \exp\left(-\frac{\mathrm{x}}{\lambda_\mathrm{b}}\right)$$

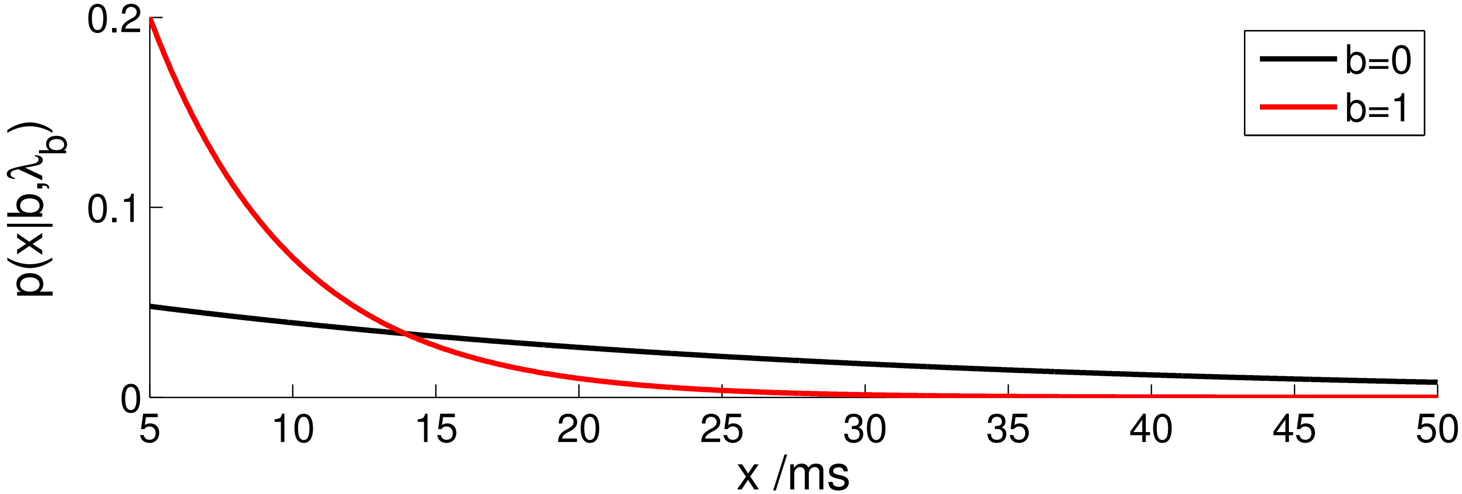
Recognition model: inference



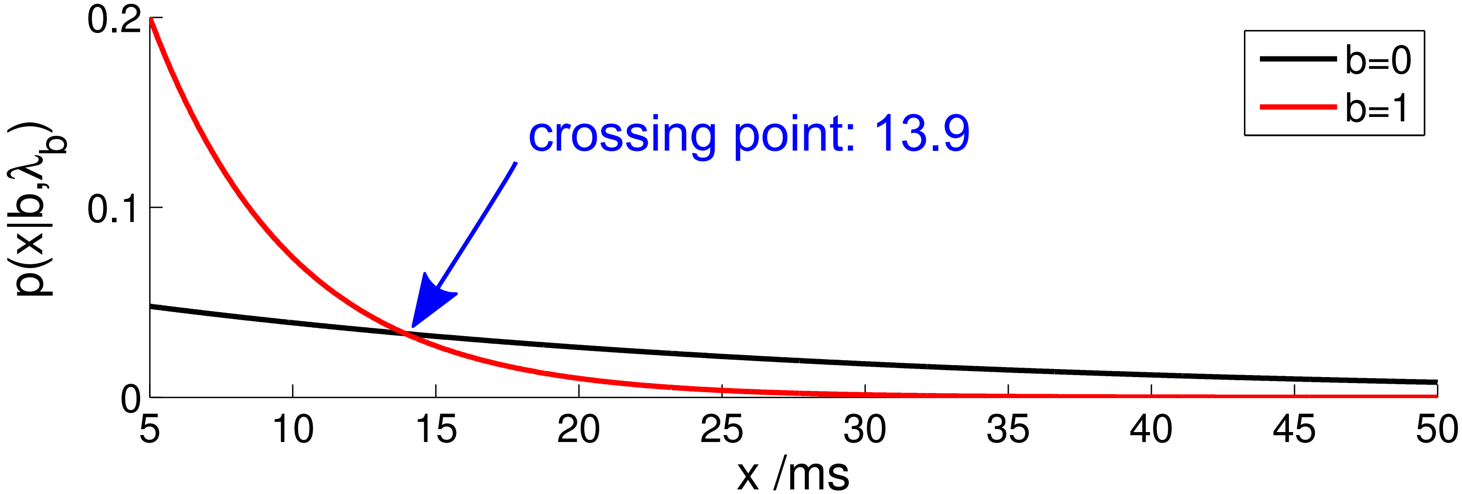
$$p(\mathrm{b} = 1|\mathrm{x}, \lambda_0, \lambda_1) = \frac{p(\mathrm{x}|\mathrm{b}{=}1, \lambda_1) p(\mathrm{b}{=}1)}{p(\mathrm{x}|\lambda_0, \lambda_1)}$$

$$= \frac{1}{1 + \frac{Z(\lambda_1)}{Z(\lambda_0)} \exp\left(-\mathrm{x}\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)\right)}$$

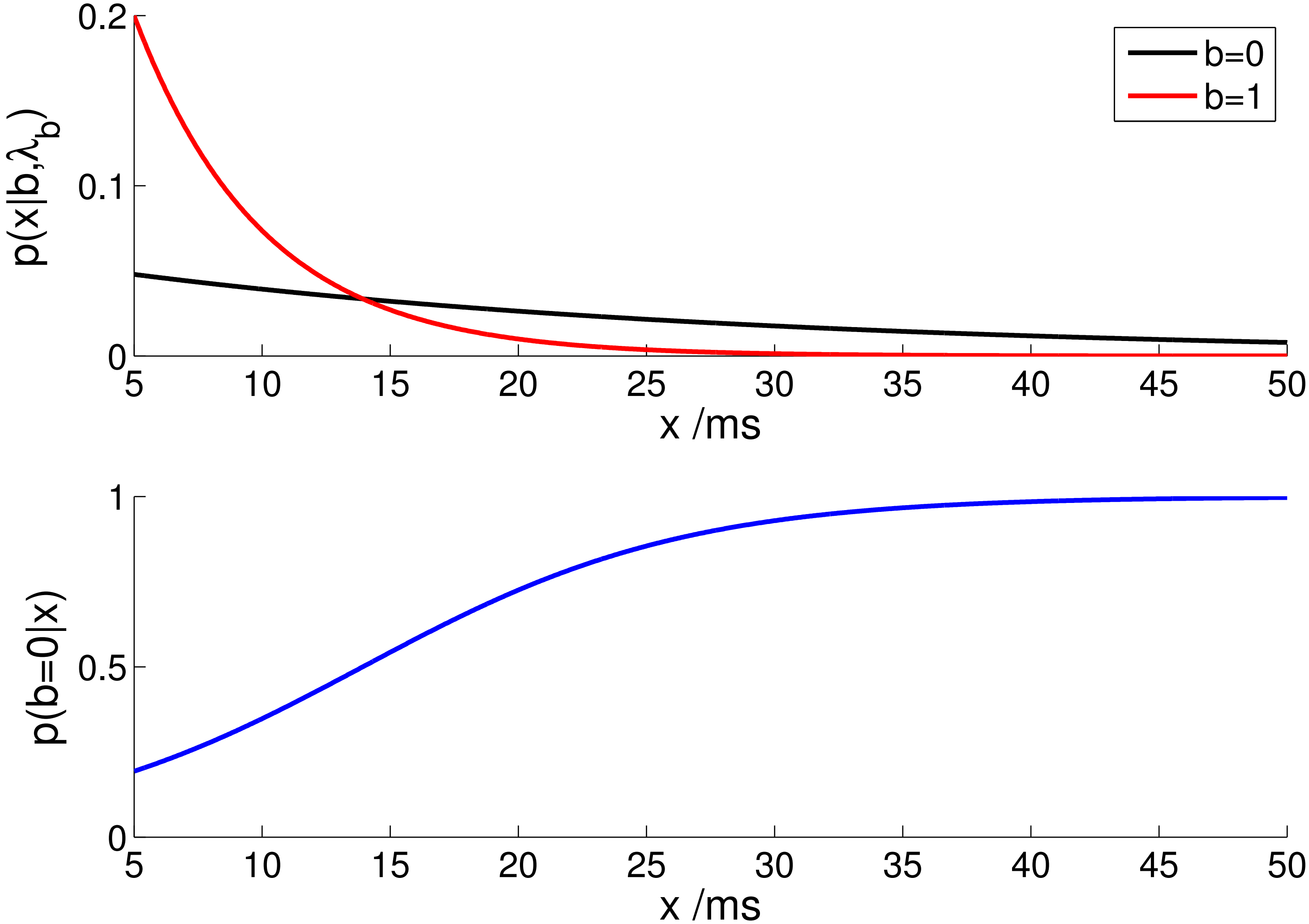
Graphical solution



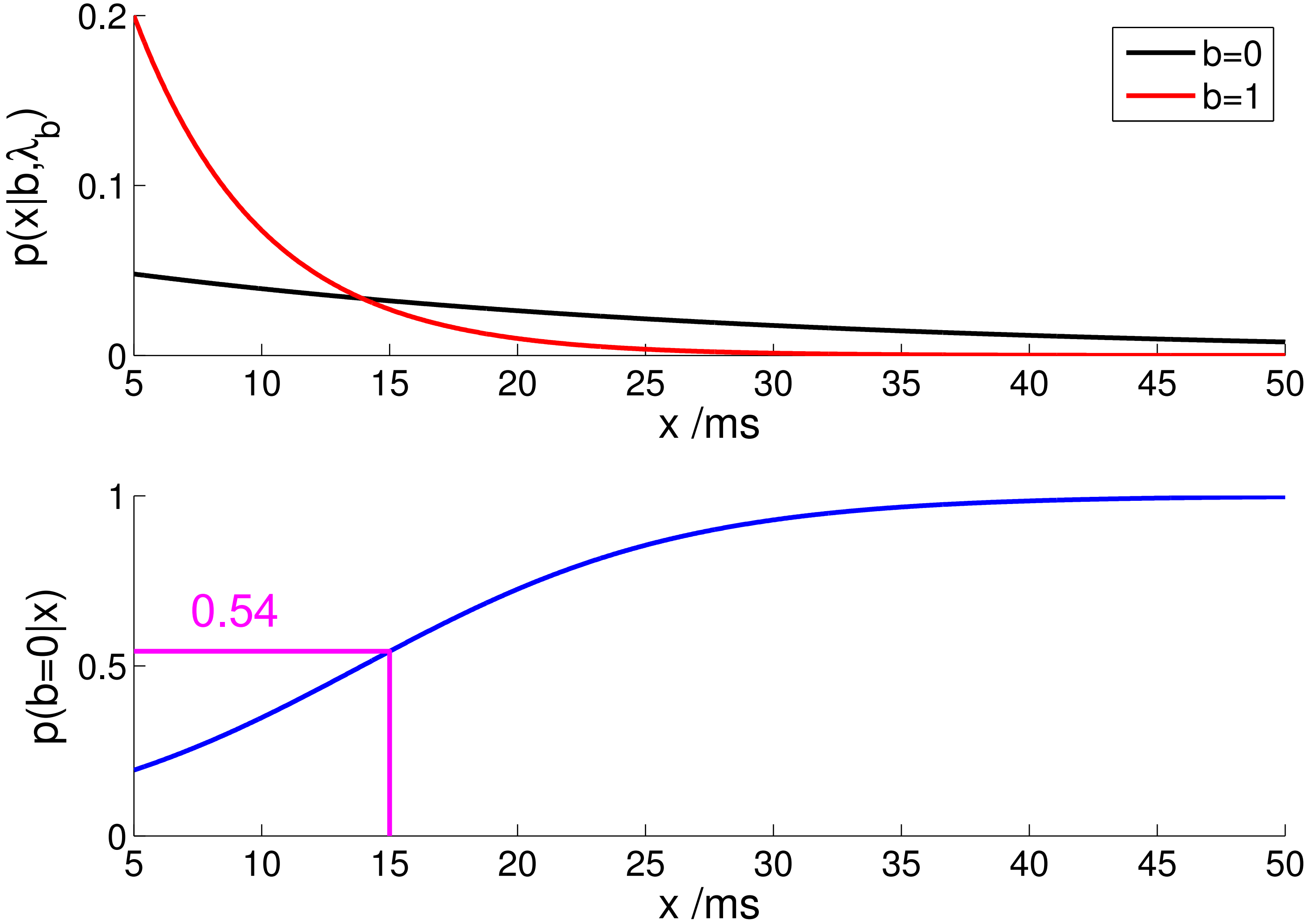
Graphical solution



Graphical solution



Graphical solution



Generative models in neuroscience

- data analysis (spike sorting, fMRI, etc.)
- ideal observer models in psychophysics
- neural encoding models
- neural decoding models
- **Bayesian Brain** - the brain is making inferences about the world using probabilistic calculus