Parameter Estimation

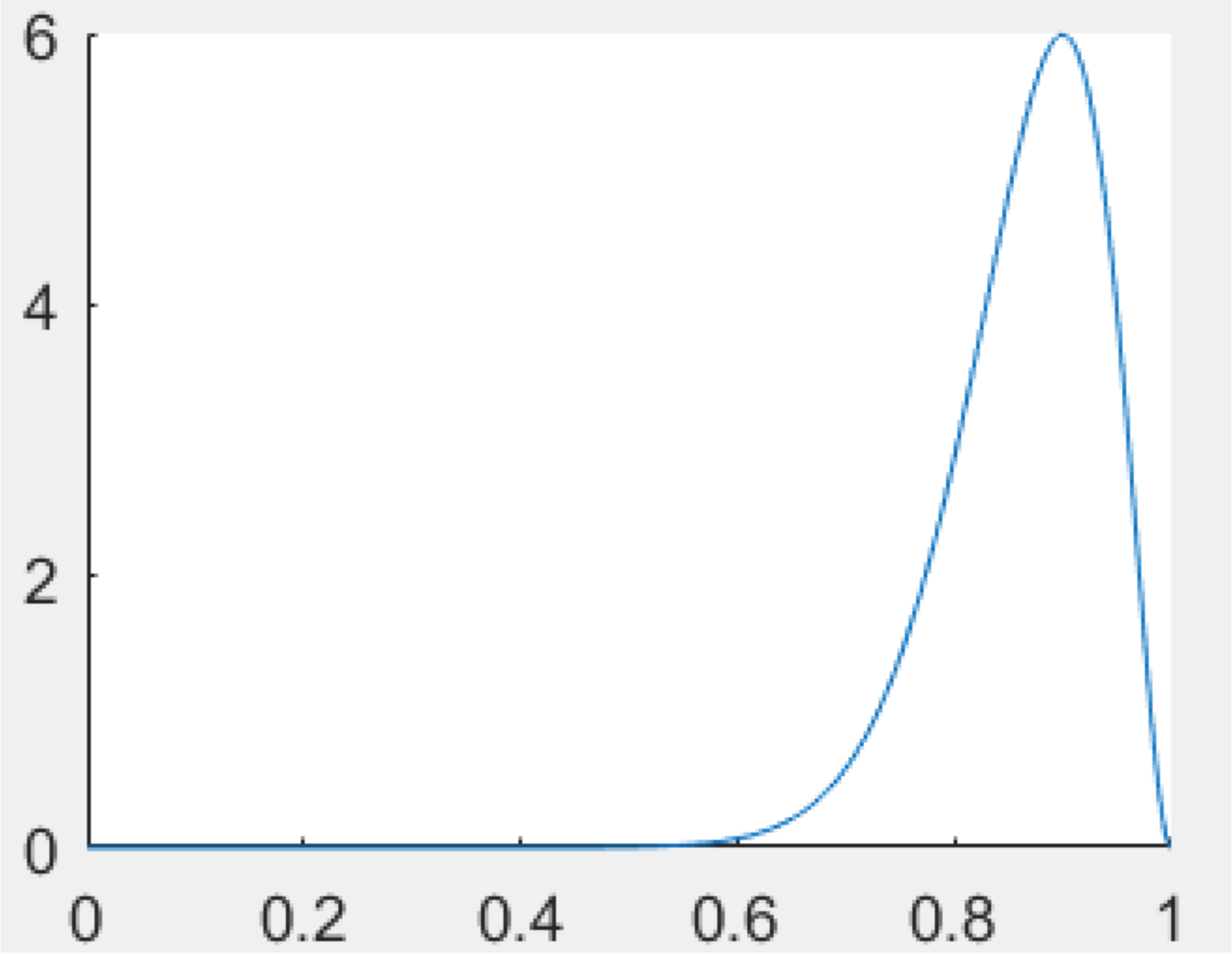
Parameters as Random Variables (Bayesian View)

- Experiment: flipping a coin

$$K \in \{\text{Heads}, \text{Tails}\}$$
$$P(K{=}Heads) = p$$
$$P(K{=}Tails) = 1 - p$$
p is unknown

- Suppose you tried 20 times and observed: 18 H and 2 T

- What you can say about p?
  - $0 < p < 1$ (strictly)
  - it is more likely that p is closer to 0.9
  - but other values of p, including 1/2 are not excluded...
- Bayes has proposed to assign probabilities to p considered as beliefs (the information that we have about p)



Bayes posterior of p (Beta distribution)

"Point" Estimates of Parameters

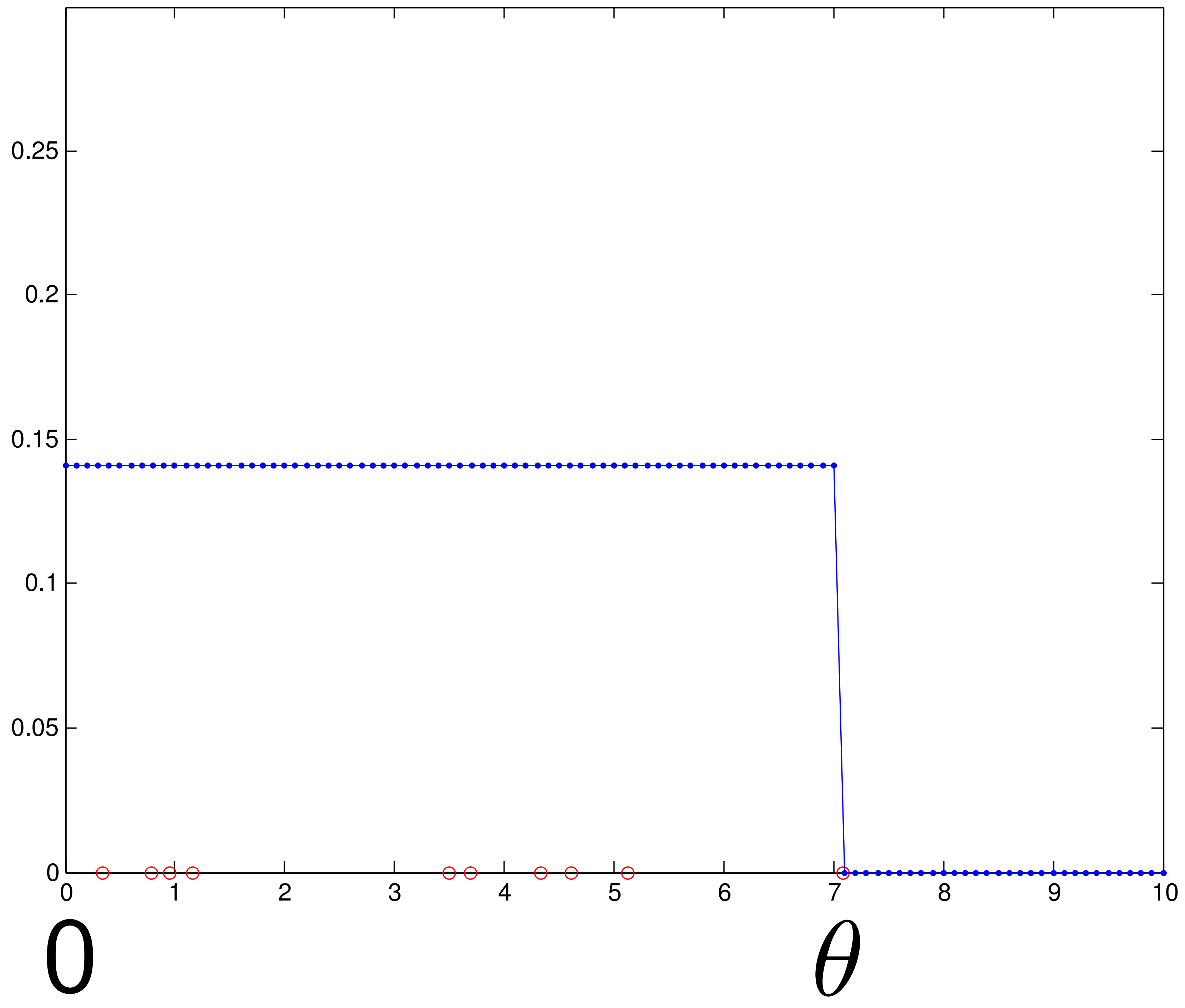
- Training data: $D = (x_1, x_2 \dots x_n)$
- Parametric model: $p(x; \theta)$
- Maximum Likelihood
  - θ - parameters
  - Estimate $\hat{\theta}_{ML} = \arg\max_\theta \prod_i p(x_i; \theta)$
  - Use: plug-in: $p(x; \hat{\theta}_{ML})$
- Maximum a Posteriori (MAP) and Minimum Mean Squared Error Estimate:
  - θ - random variable, prior $p(\theta)$
  - Likelihood: $p(D|\theta) = \prod_i p(x_i|\theta)$
  - Posterior: $p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$
  - MAP: $\hat{\theta}_{\text{MAP}} = \arg\max_\theta p(\theta|D)$
  - MMSE: $\hat{\theta}_{\text{MMSE}} = \arg\min_{\hat{\theta}} \int (\theta - \hat{\theta})^2 p(\theta|D) d\theta$
  - Use: plug-in: $p(x|\hat{\theta}_{\text{MAP}})$, $p(x|\hat{\theta}_{\text{MMSE}})$

ML: Uniform Distribution

- Training data: $D = (x_1, x_2, \ldots x_n)$,
- Assume x is uniform in $[0, \theta]$
- Want to estimate θ
  (we will consider θ_{ML}, θ_{MAP}, and Bayesian posterior $p(\theta|D)$).
- Density:

$$p(x;\theta) = \begin{cases} 1/\theta, & 0 \leq x \leq \theta; \\ 0, & \text{otherwise.} \end{cases}$$

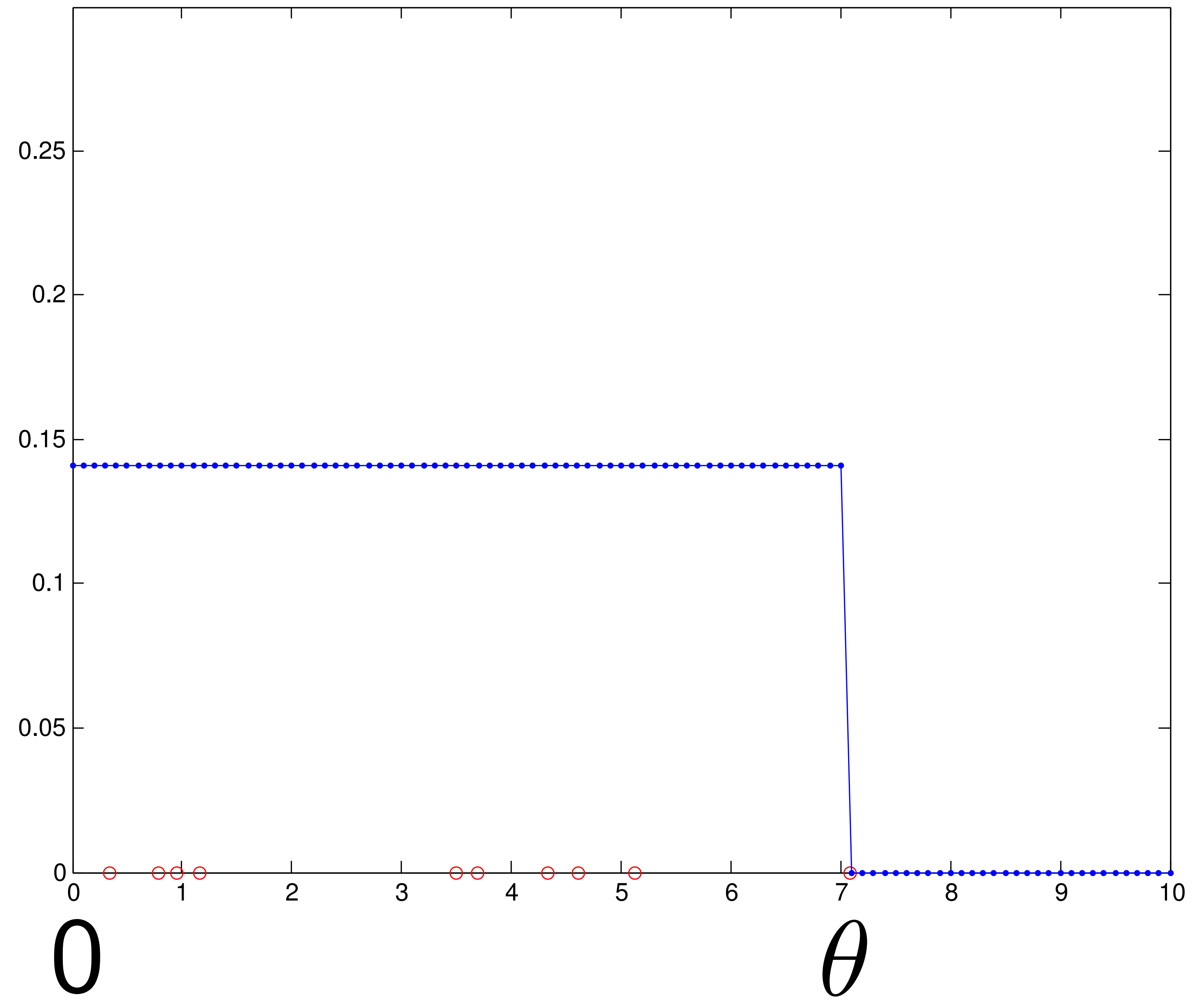
- ML estimate:



ML: Uniform Distribution

- Training data: $D = (x_1, x_2, \dots x_n)$,
- Assume x is uniform in $[0, \theta]$
- Want to estimate θ
  (we will consider θ_{ML}, θ_{MAP}, and Bayesian posterior $p(\theta|D)$).
- Density:

$$p(x;\theta) = \begin{cases} 1/\theta, & 0 \le x \le \theta; \\ 0, & \text{otherwise.} \end{cases}$$



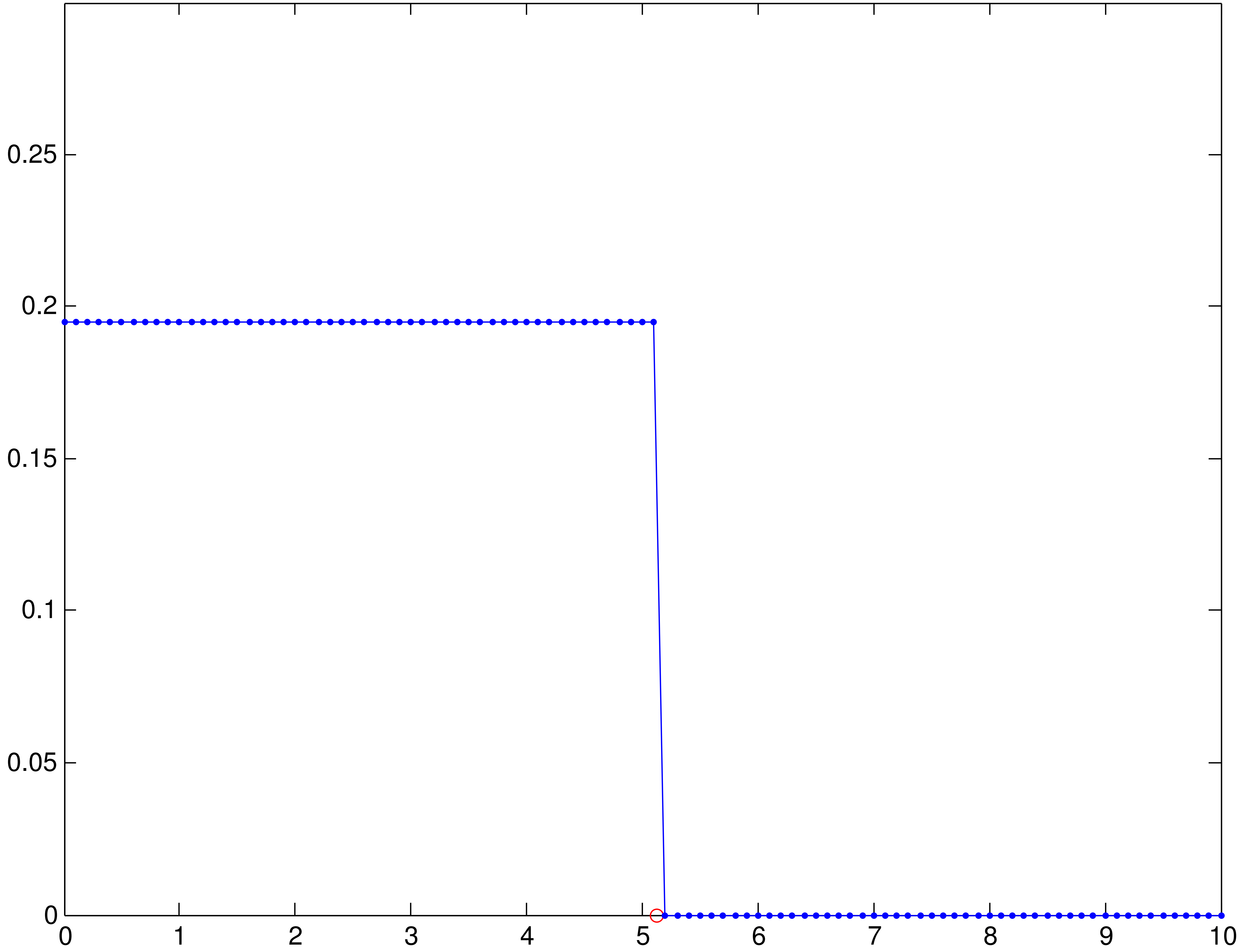
- ML estimate:

$$\max_\theta \prod_{i=1}^{n} \frac{1}{\theta} [\![x_i \le \theta]\!] \quad \rightarrow \quad \theta_{\mathrm{ML}} = \max_i x_i.$$

- Let us see how the distribution $p(x;\hat{\theta})$ changes as we get more training data $x_1, x_2, \dots x_n$.

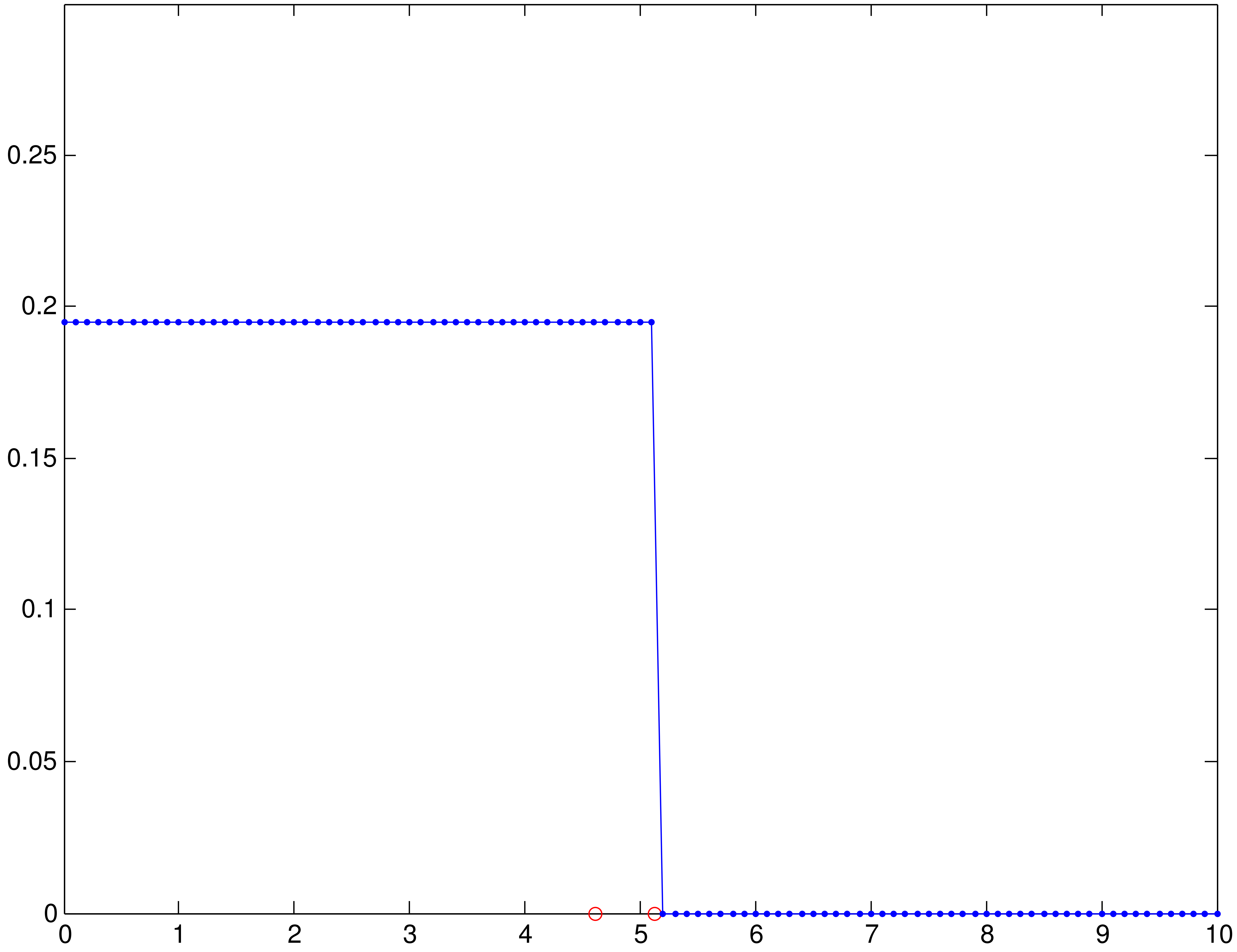
ML for Uniform Distribution

- Estimating interval bound

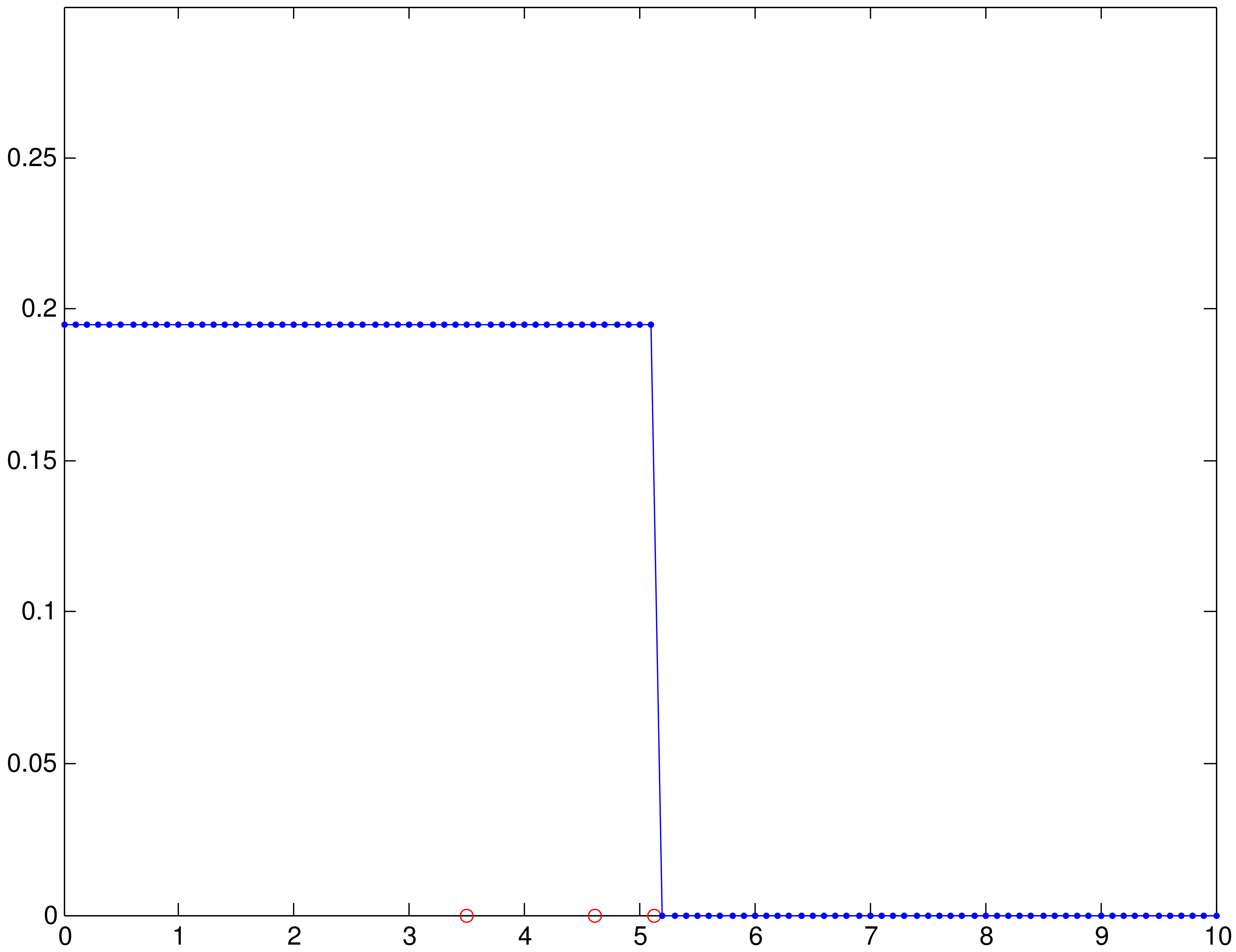


ML for Uniform Distribution

- Estimating interval bound

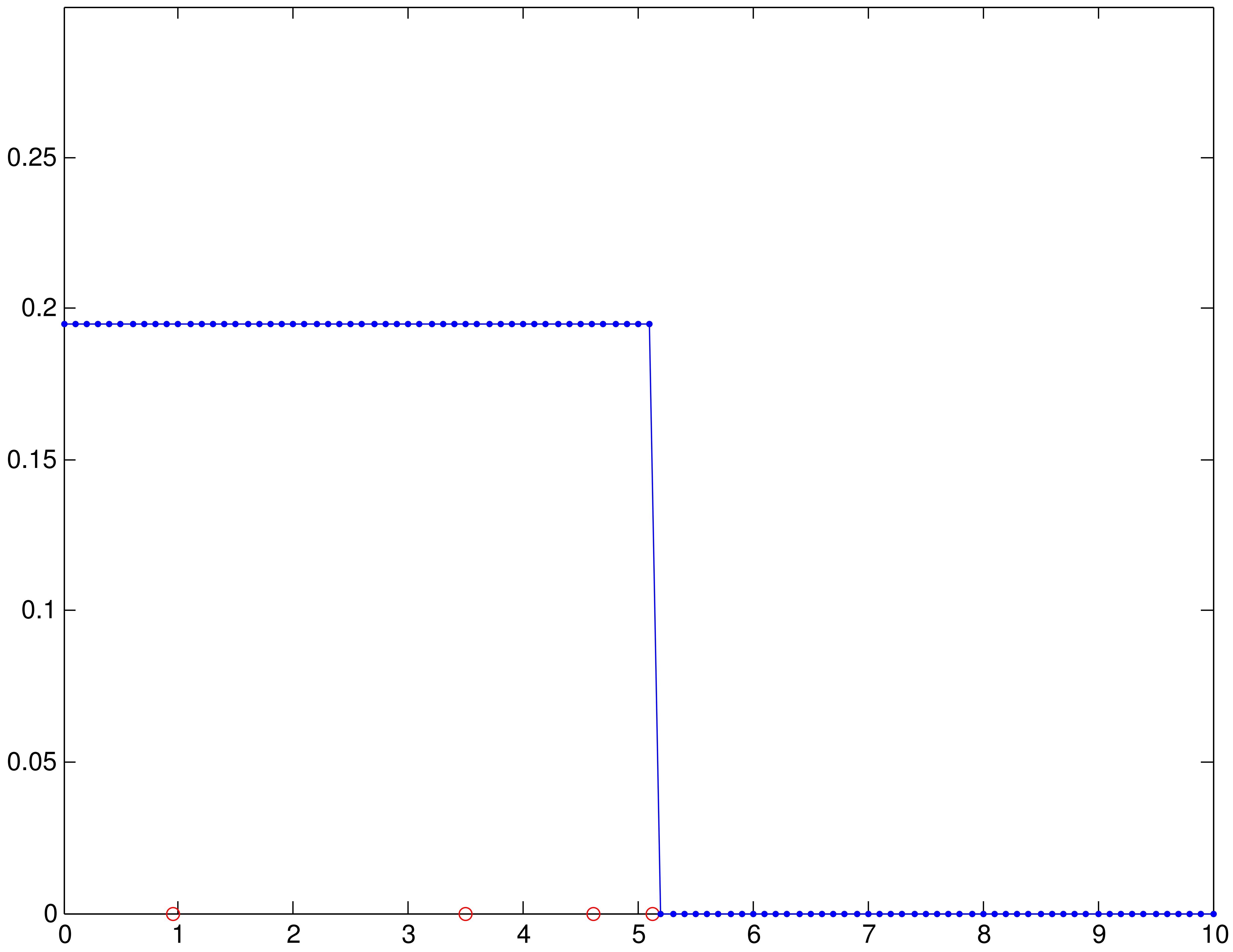


- Estimating interval bound

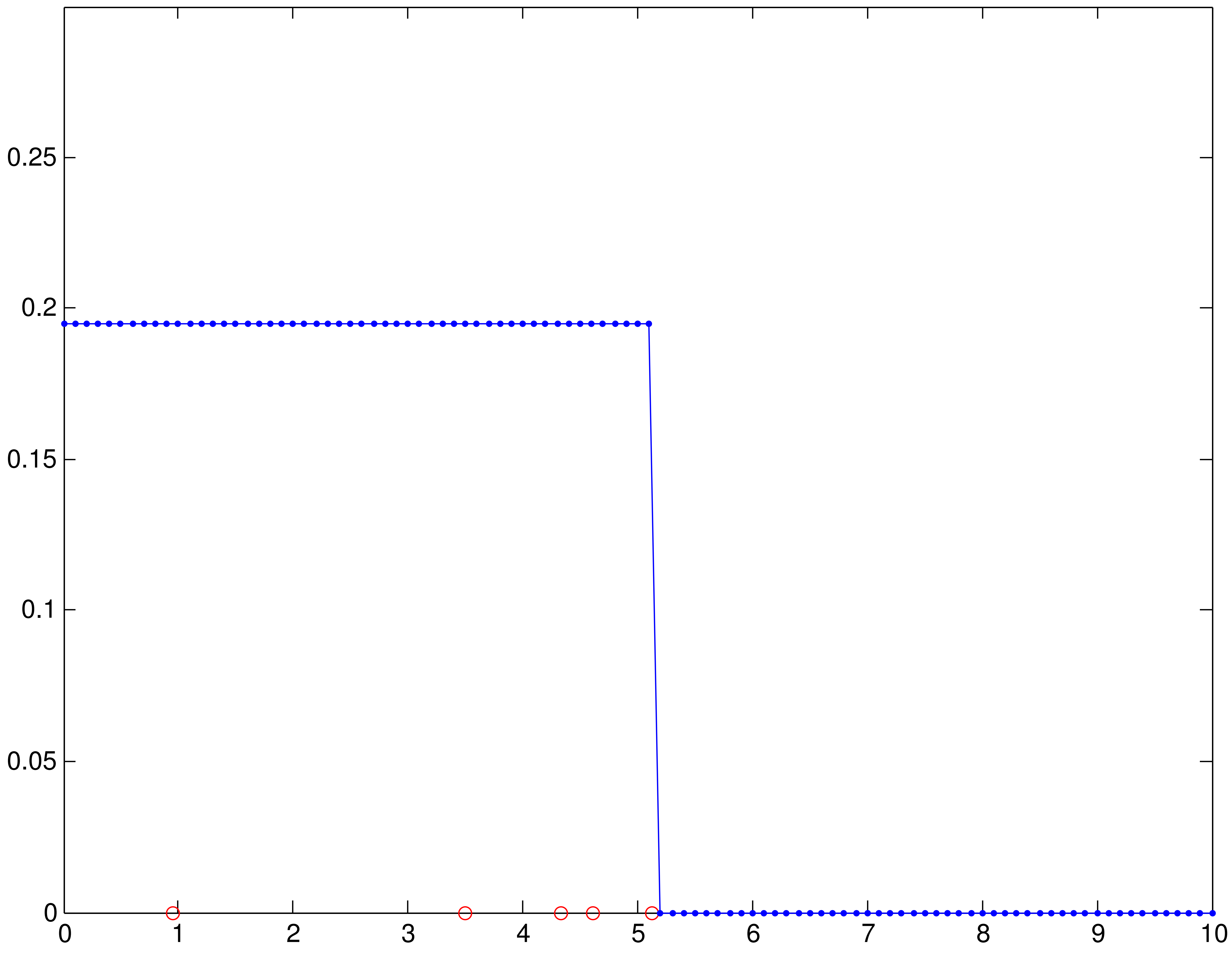


ML for Uniform Distribution

- Estimating interval bound

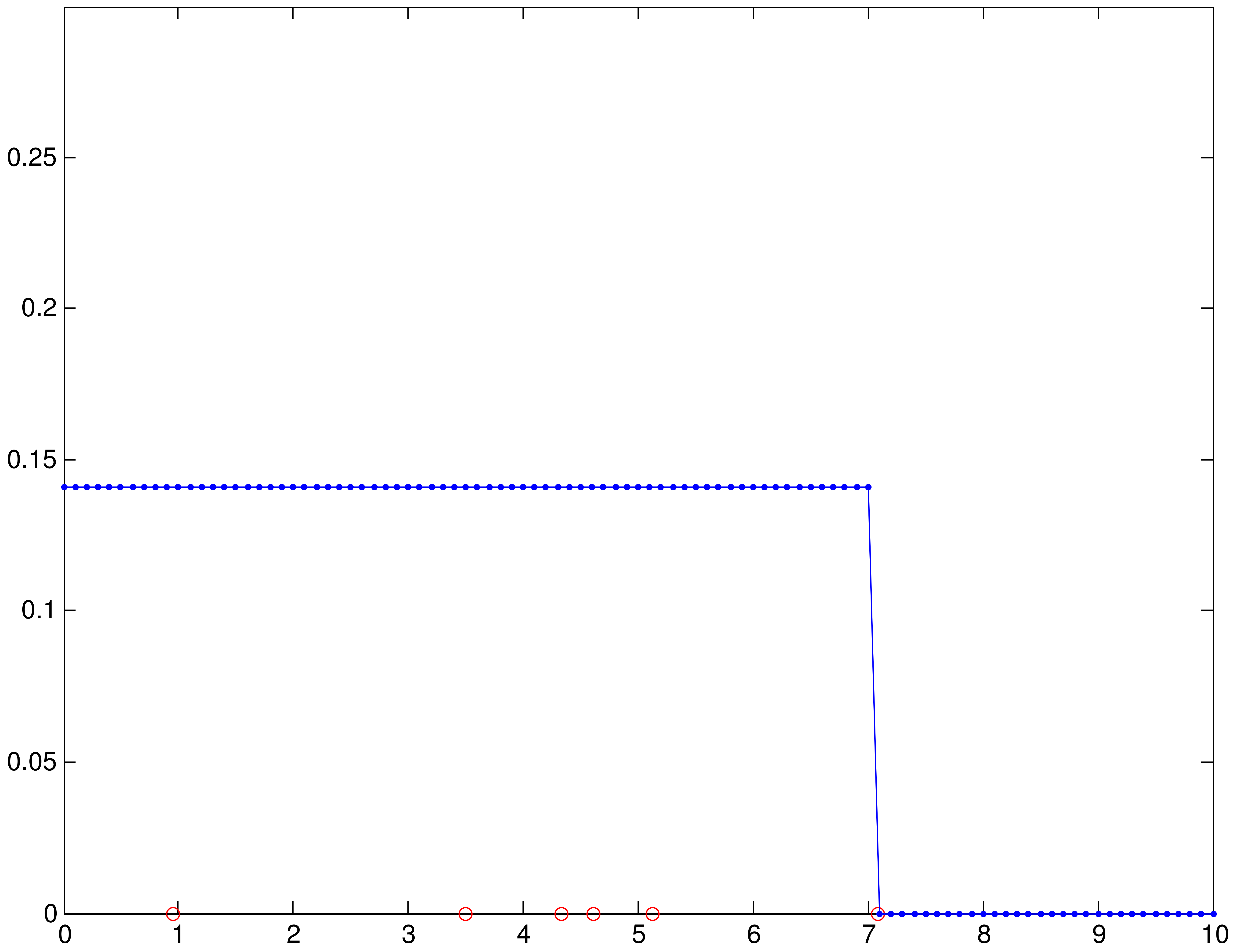


- Estimating interval bound



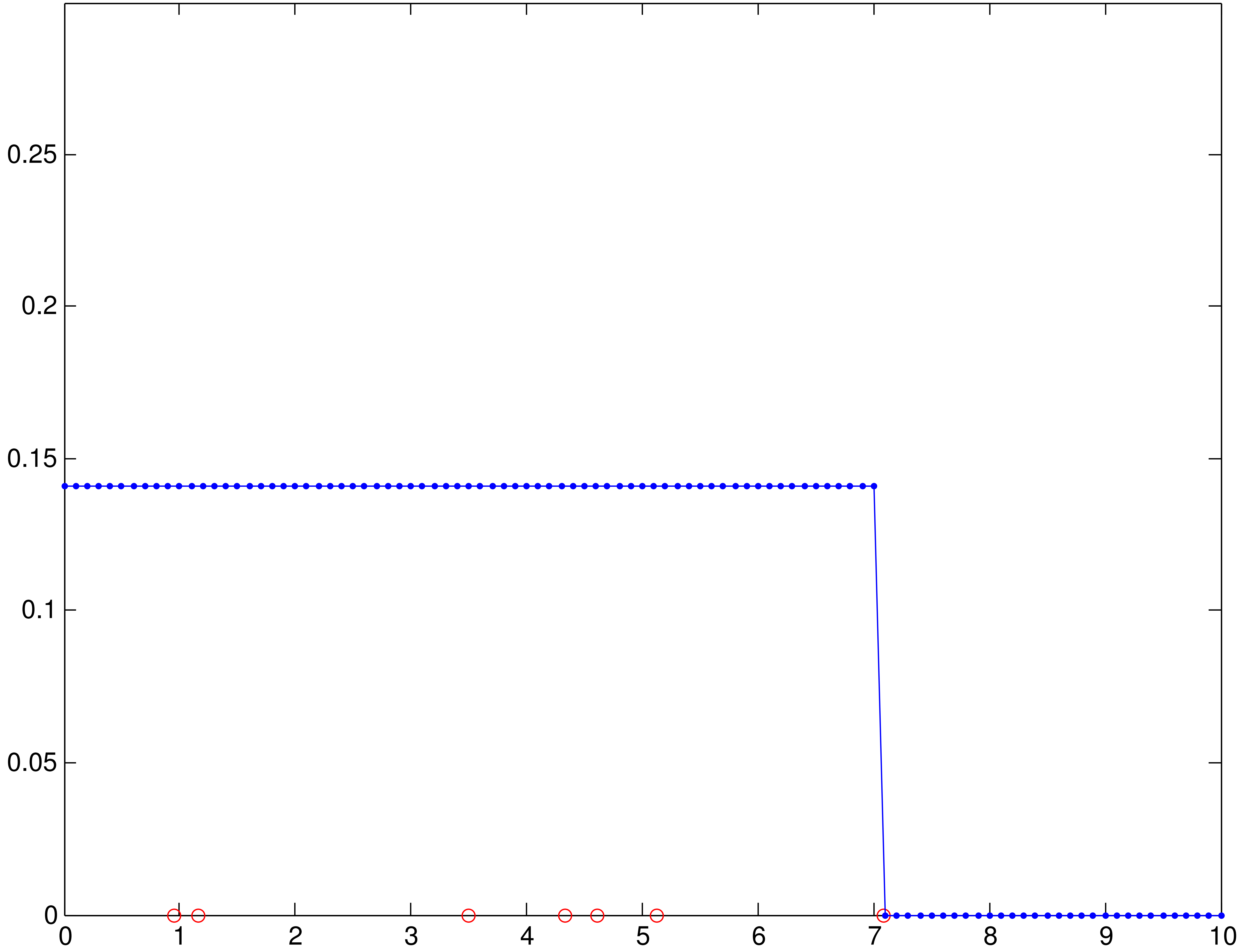
ML for Uniform Distribution

- Estimating interval bound



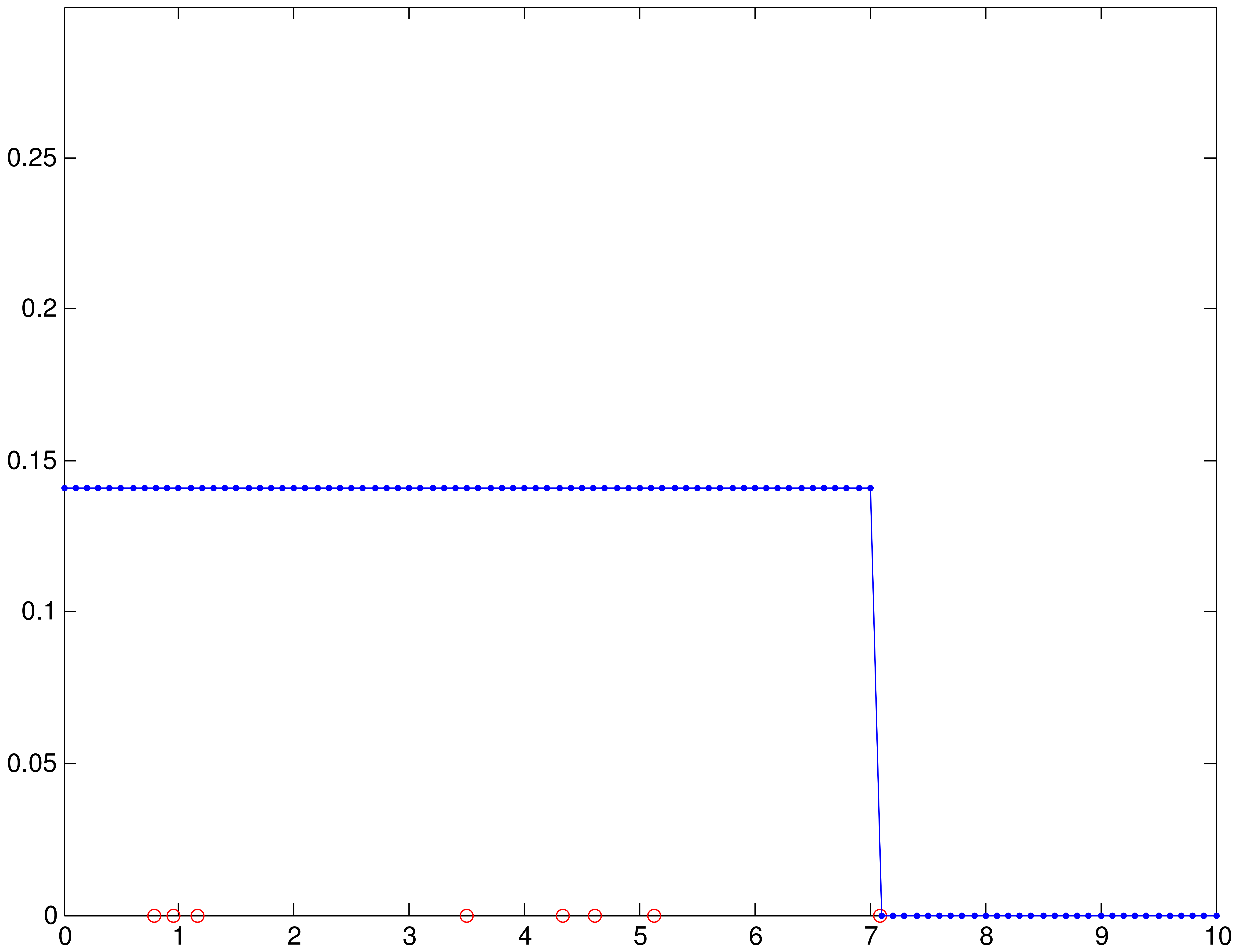
ML for Uniform Distribution

- Estimating interval bound



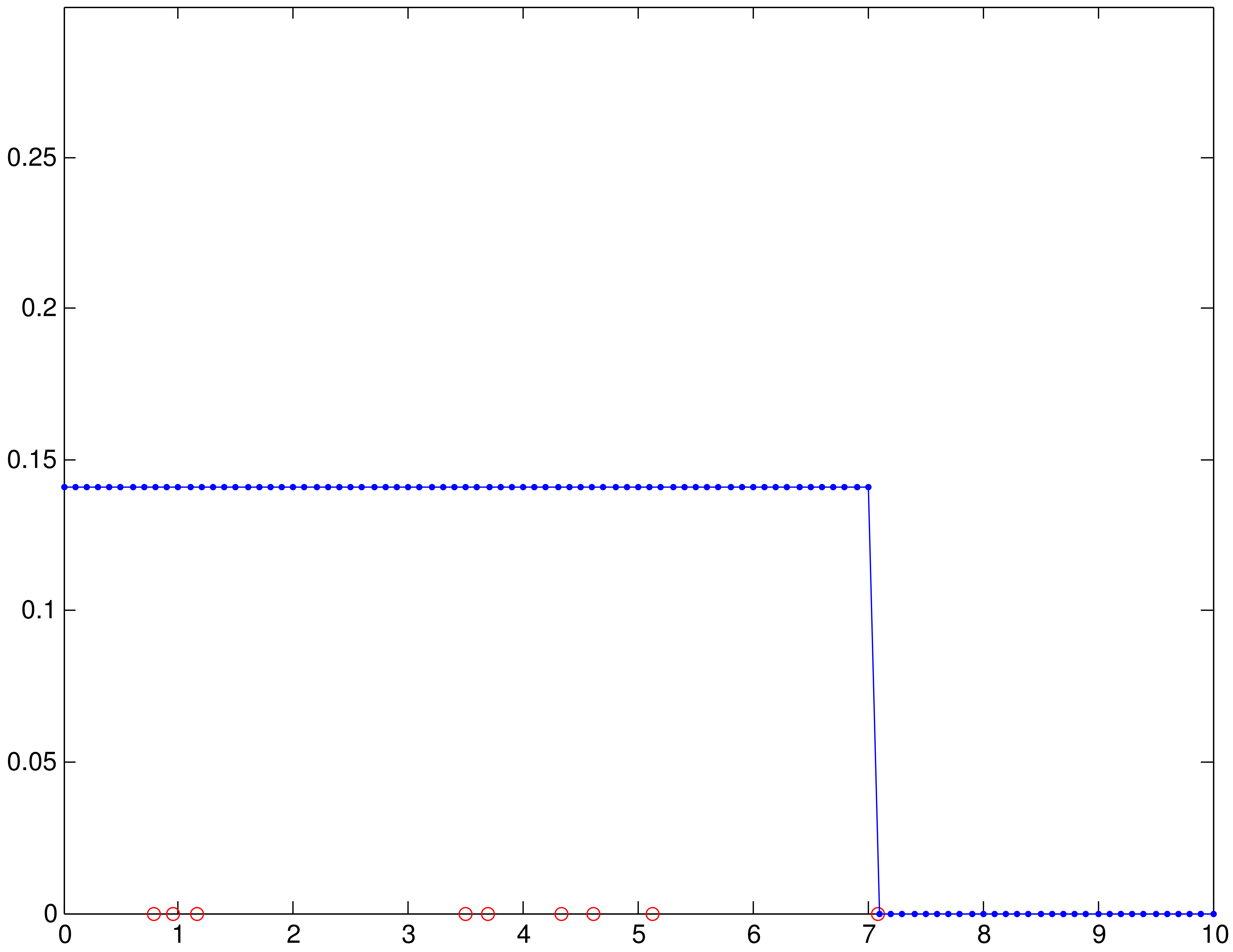
ML for Uniform Distribution

- Estimating interval bound



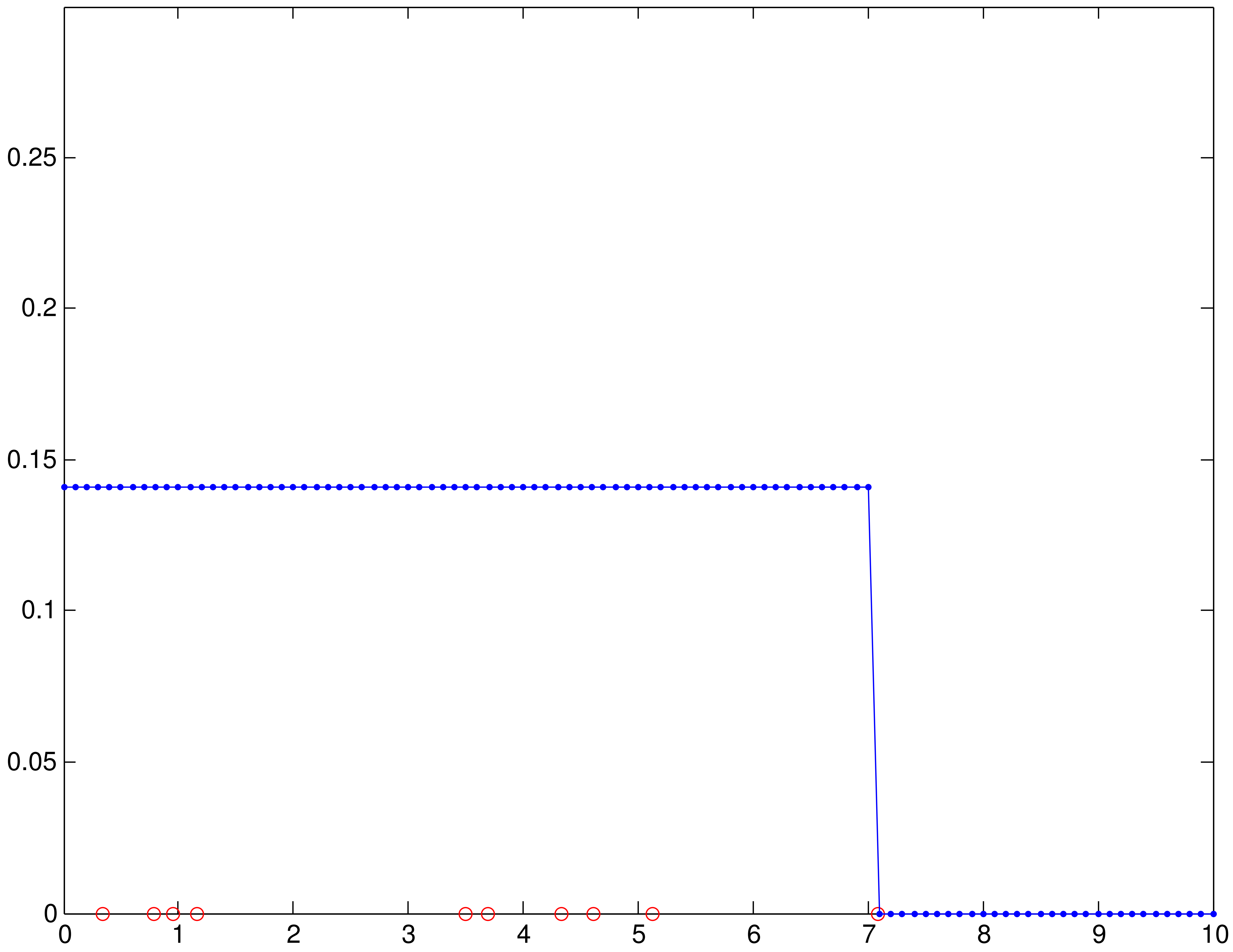
ML for Uniform Distribution

- Estimating interval bound



ML for Uniform Distribution

- Estimating interval bound



Bayesian Inference: Uniform Distribution

- Training data: $D = (x_1, x_2, \ldots x_n)$,
- x is uniform in $[0, \theta]$
- θ is a random variable *apriori* θ is uniform in $[0, M]$, $M = 10$
- The *Posterior* distribution of θ when given the data is: $p(\theta \mid \mathcal{D})$:

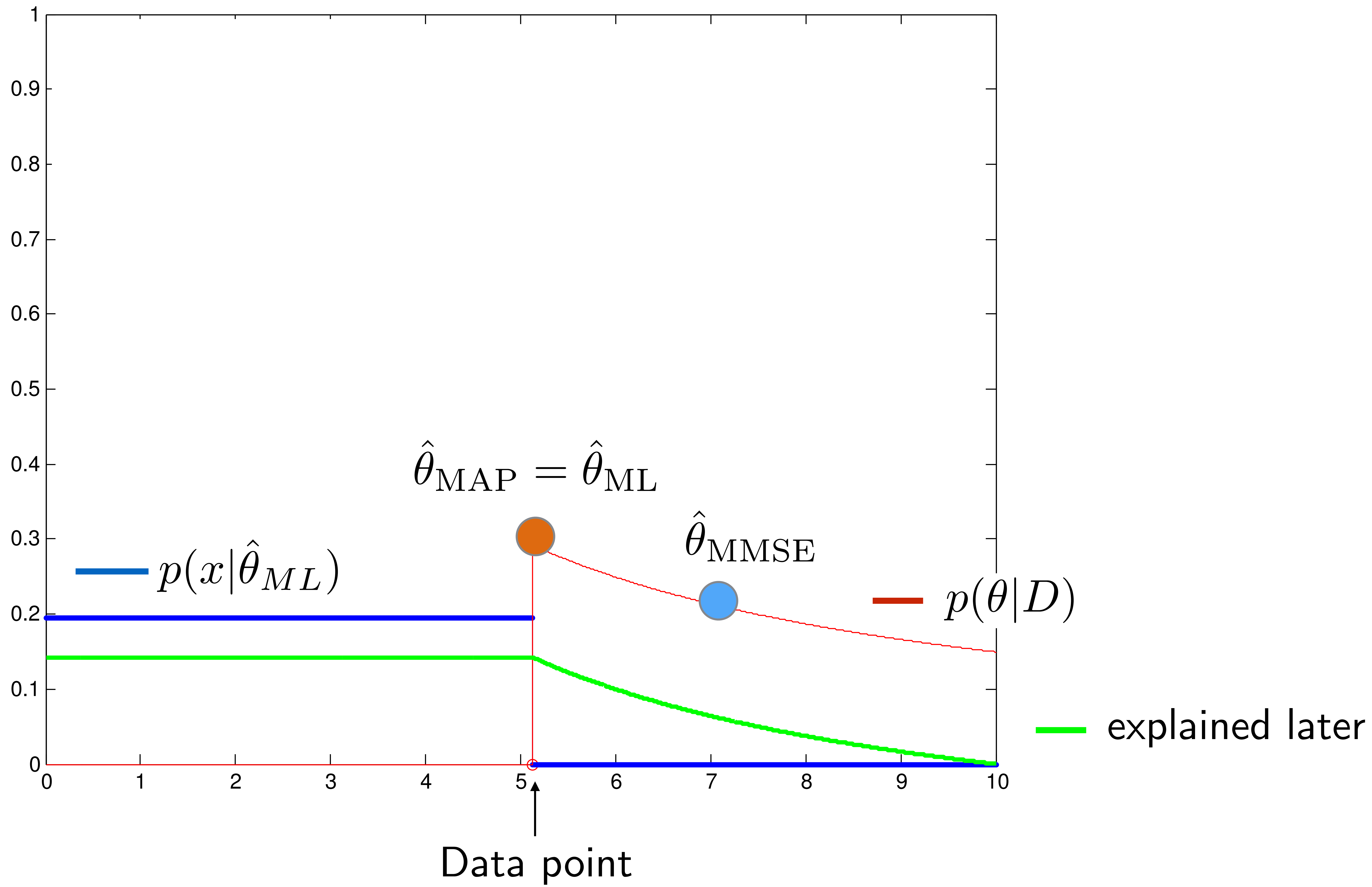
$$p(\theta \mid D) \propto \prod_{i=1}^{n} p(x_i \mid \theta) p(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} [\![x_i \leq \theta]\!] p(\theta)$$

$$p(\theta \mid D) = \frac{1}{Z} \frac{1}{\theta^n} [\![m \leq \theta \leq M]\!], \quad \text{where } m = \max_i(x_i).$$

- Z — normalization constant, not needed form MAP estimate.
- MAP (maximum a *posteriori*):

$$\theta_{\mathrm{MAP}} = \operatorname*{argmax}_{\theta} p(\theta \mid D) \qquad \rightarrow \theta_{\mathrm{MAP}} = \max_i x_i = \theta_{\mathrm{ML}}.$$

Bayesian Estimation: Uniform Distribution



Bayesian Inference: Uniform Distribution

- Training data: $D = (x_1, x_2, \ldots x_n)$,
- x is uniform in $[0, \theta]$
- θ is a random variable *apriori* θ is uniform in $[0, M]$, $M = 10$

- "Bayesian inference of θ" $=$ computing (approximating) posterior distribution $p(\theta \mid \mathcal{D})$:

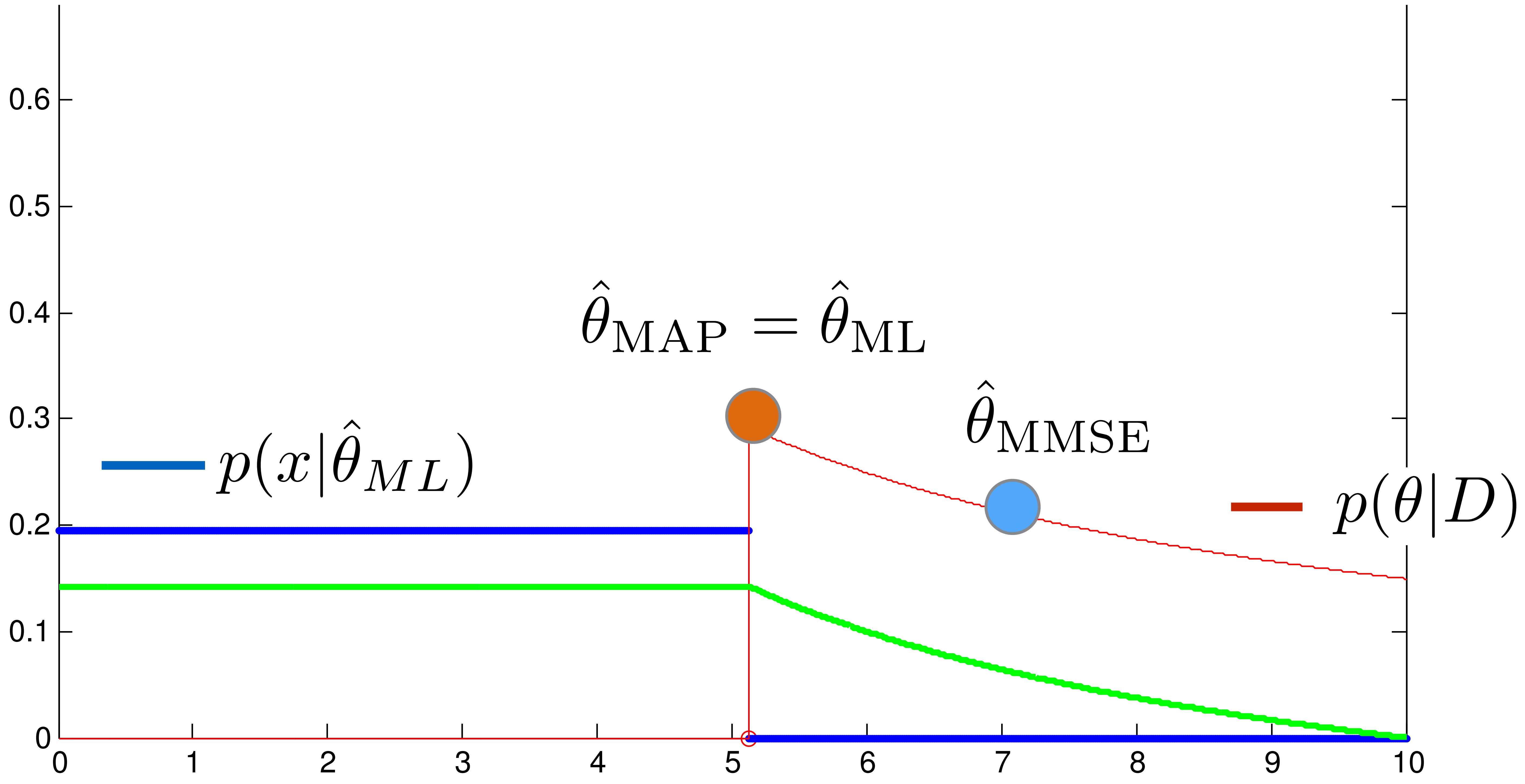
$$p(\theta \mid D) = \frac{1}{Z} \frac{1}{\theta^n} [\![m \leq \theta \leq M]\!], \quad \text{where } m = \max_i (x_i)$$

$$\int p(\theta \mid D) d\theta = 1 \quad \Rightarrow \quad Z = \int_m^M \frac{1}{\theta^n} d\theta = \frac{-1}{(n+1)\theta^{n+1}} \Big|_m^M = \frac{1}{n+1} \Big(\frac{1}{m^{n+1}} - \frac{1}{M^{n+1}} y \Big)$$

- Bayesian MMSE estimate of θ:

$$\theta_{\text{MMSE}} = \int \theta p(\theta | D) d\theta$$

Bayesian Estimation: Uniform Distribution

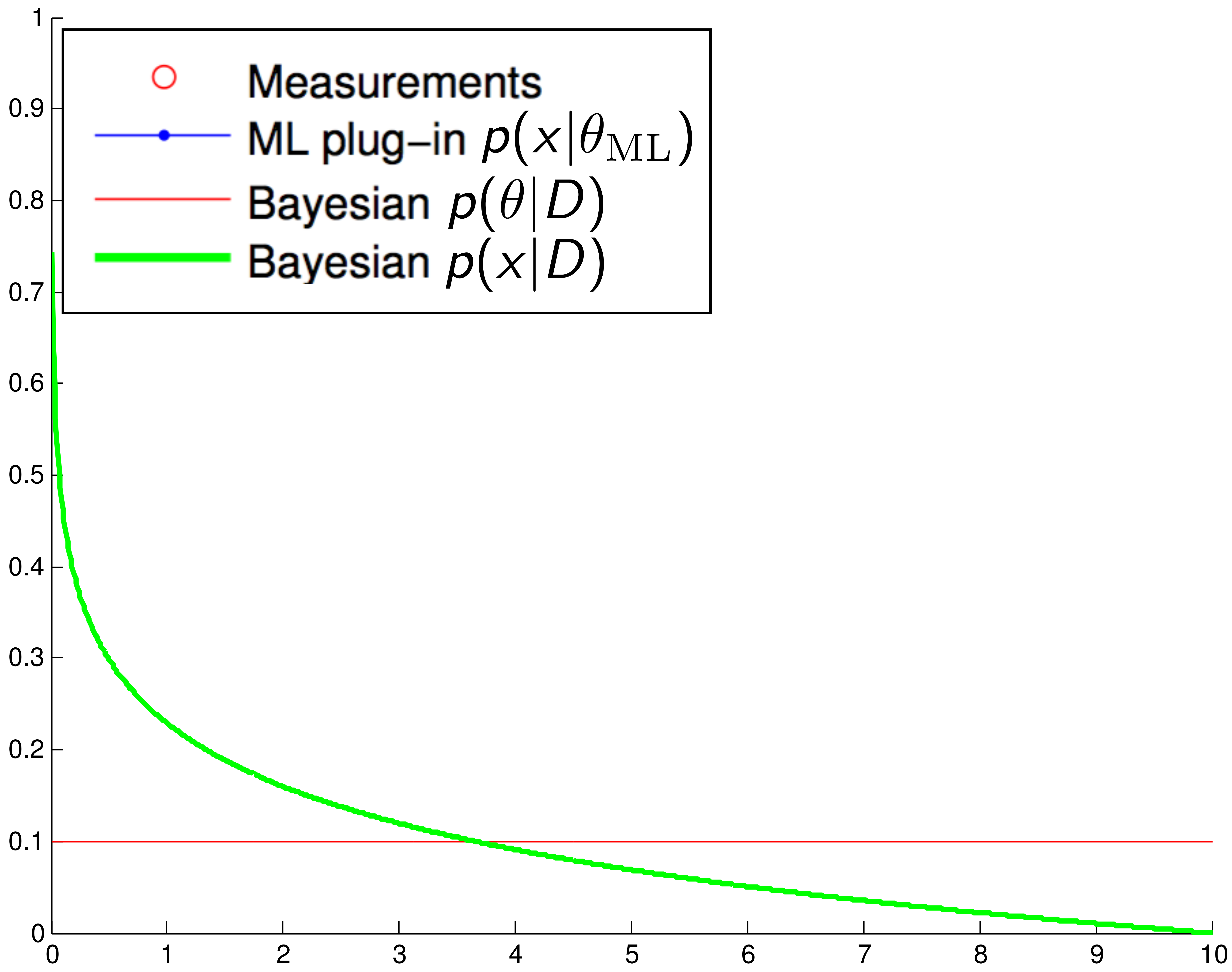


- posterior
- knowledge about parameters

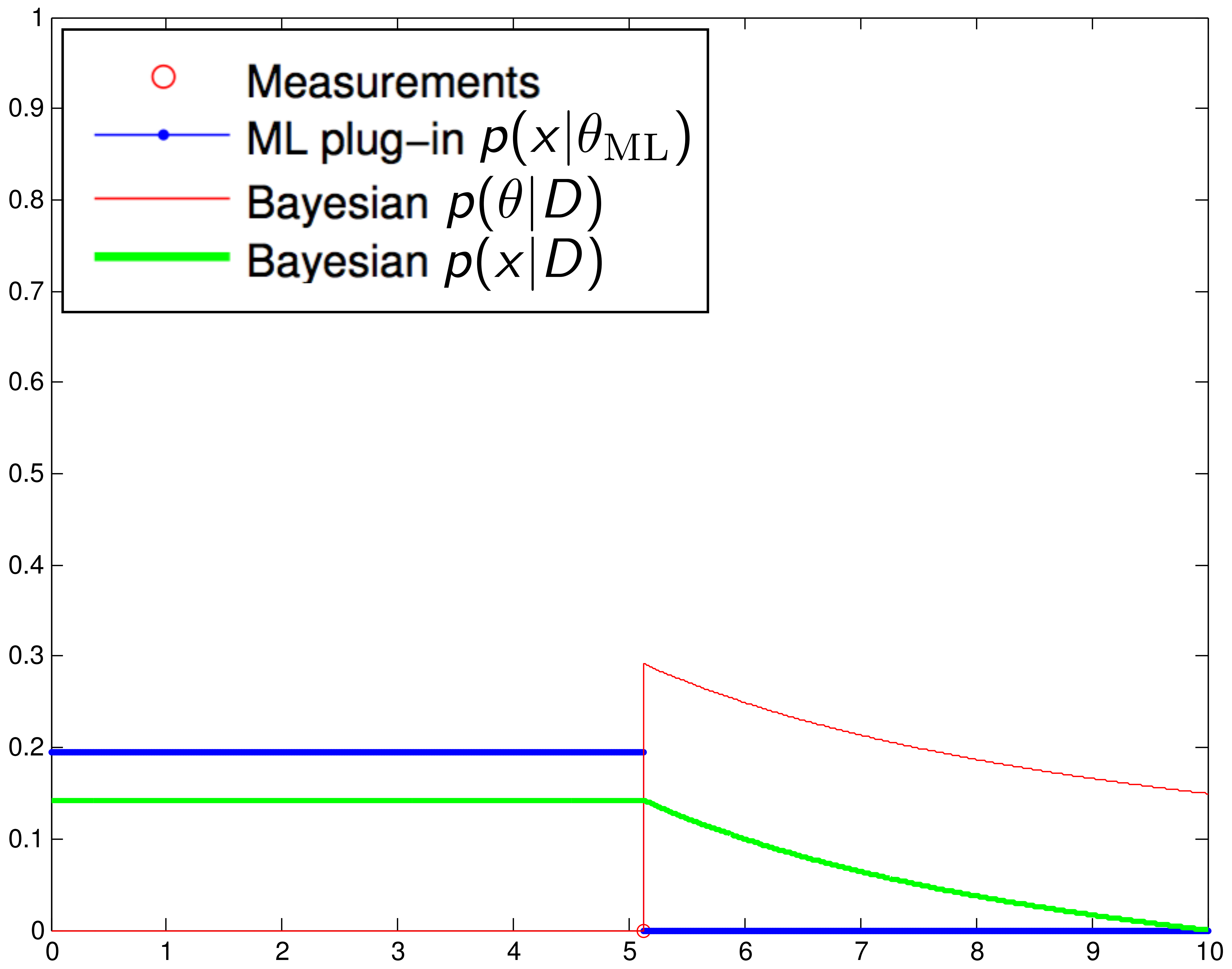
$p(x|D) = \int p(x|\theta)p(\theta|D)d\theta$

"predictive posterior": where new observations are expected to be found, compare to $p(x|\theta_{\mathrm{ML}})$

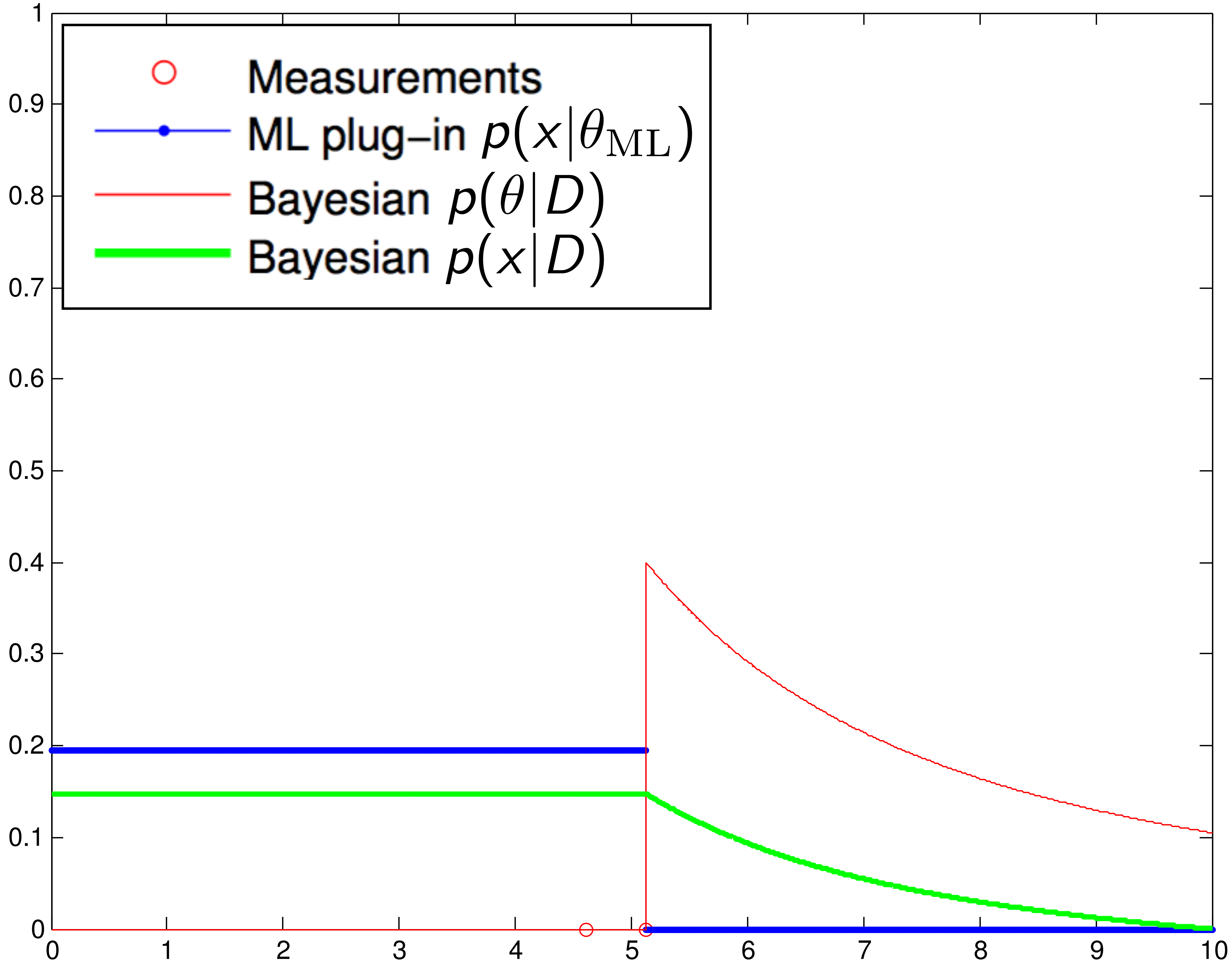
Bayesian Estimation: Uniform Distribution



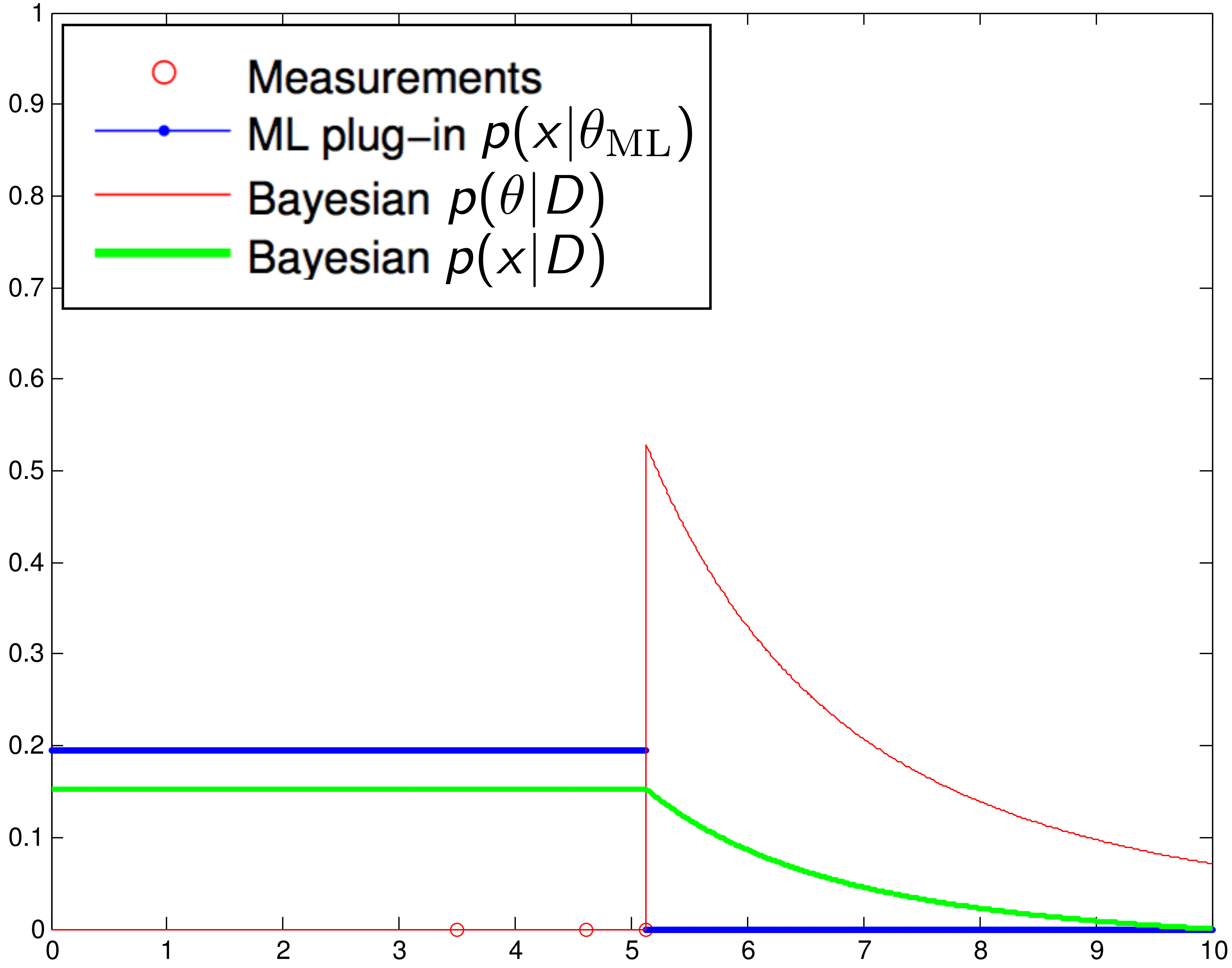
Bayesian Estimation: Uniform Distribution



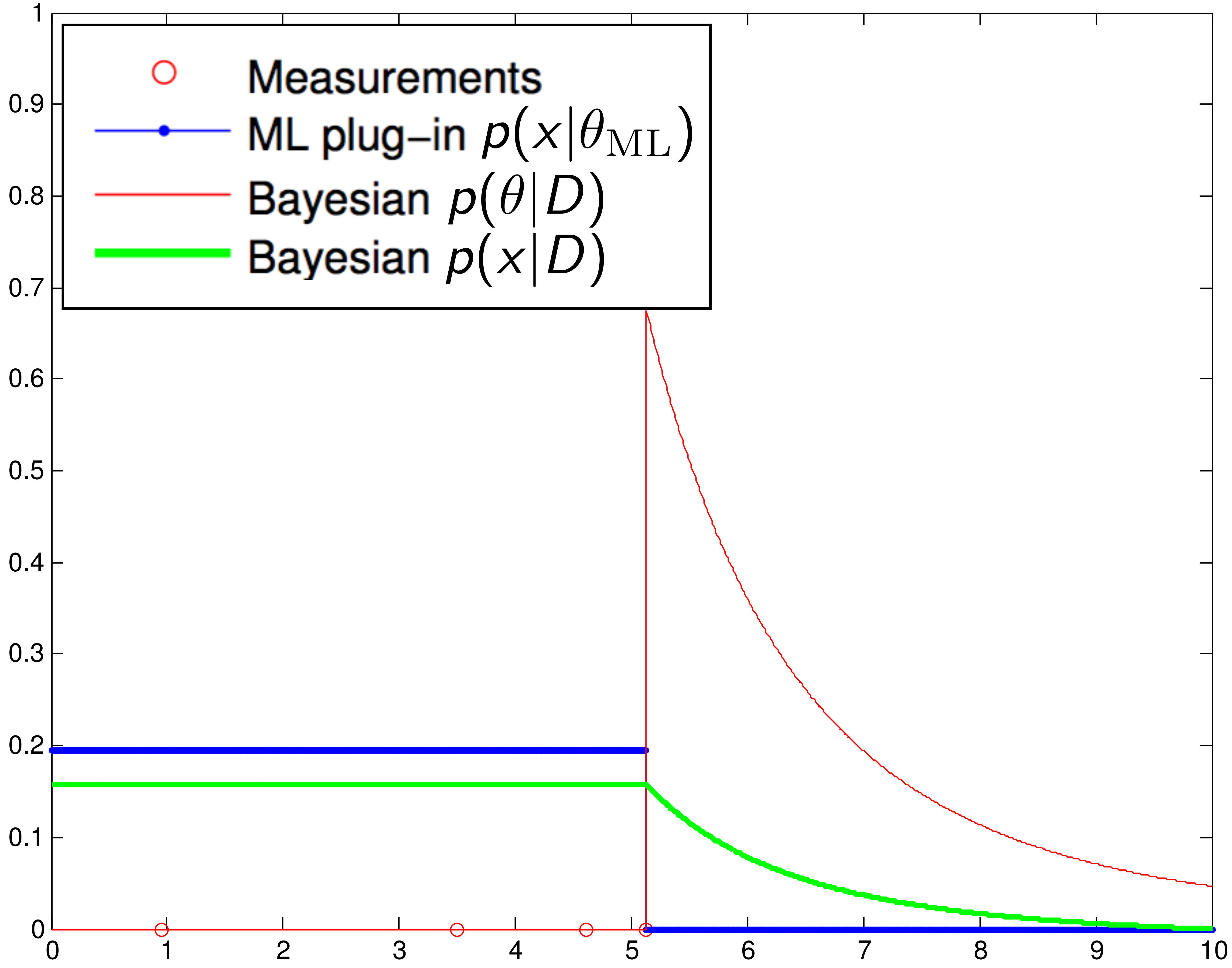
Bayesian Estimation: Uniform Distribution



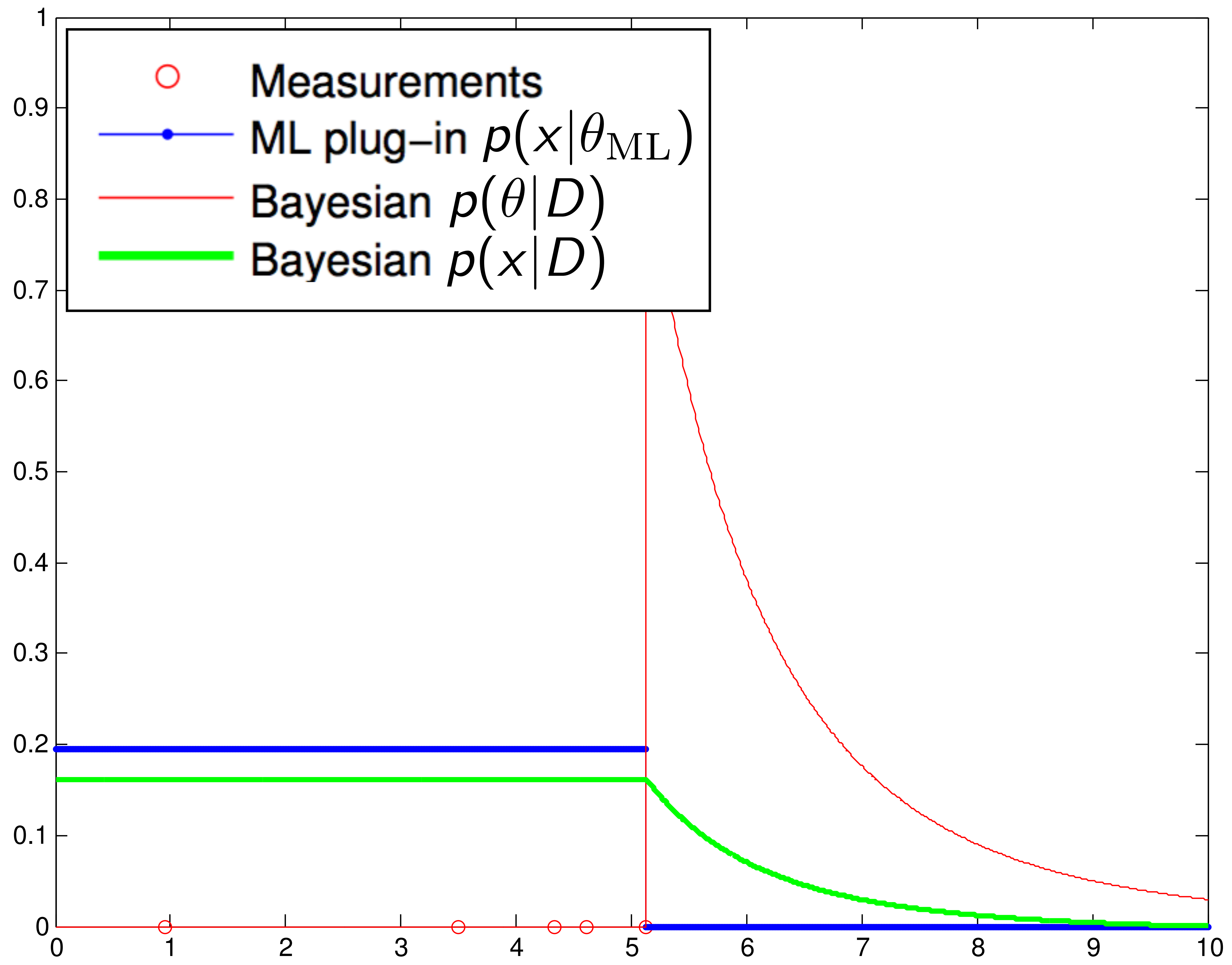
Bayesian Estimation: Uniform Distribution



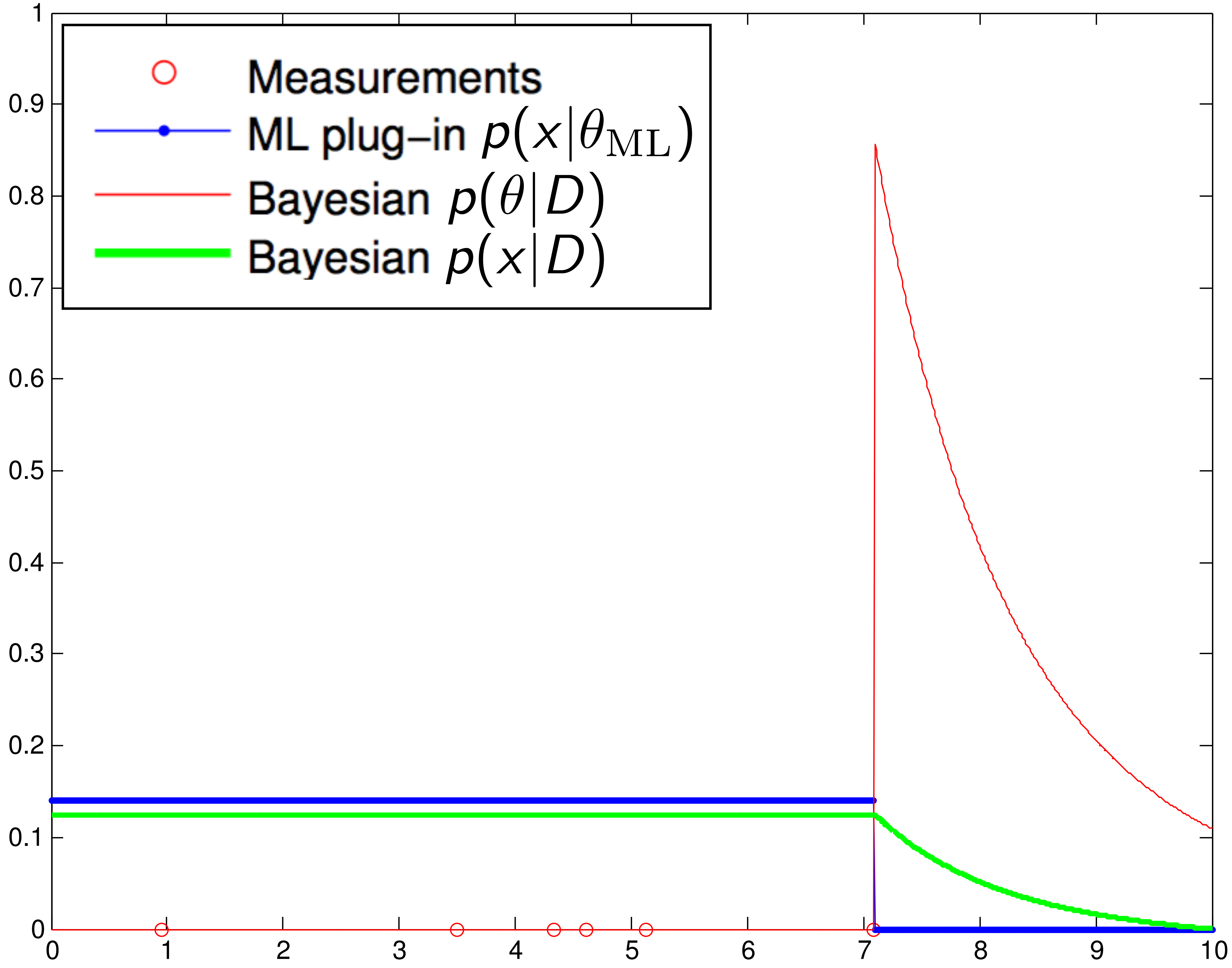
Bayesian Estimation: Uniform Distribution



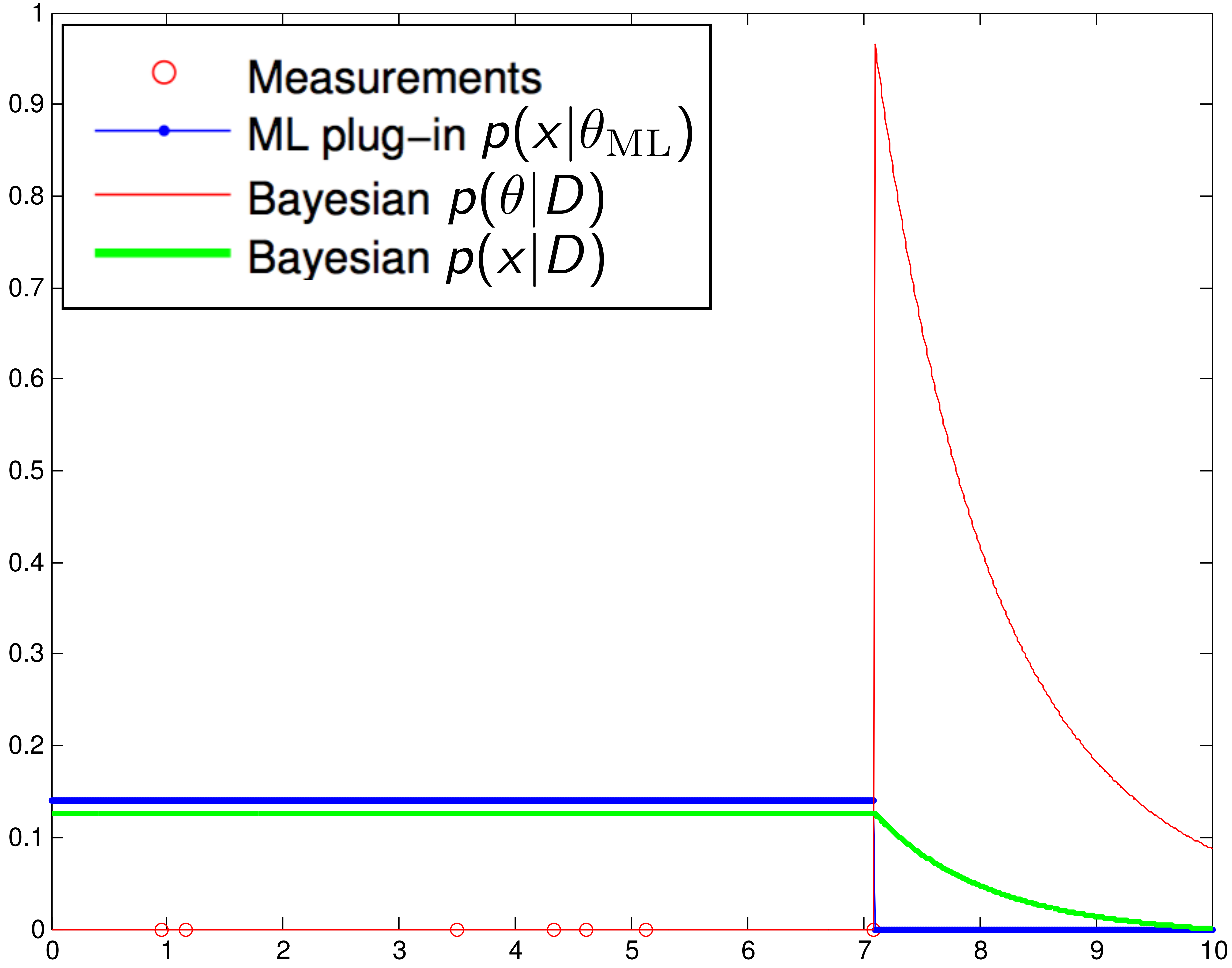
Bayesian Estimation: Uniform Distribution



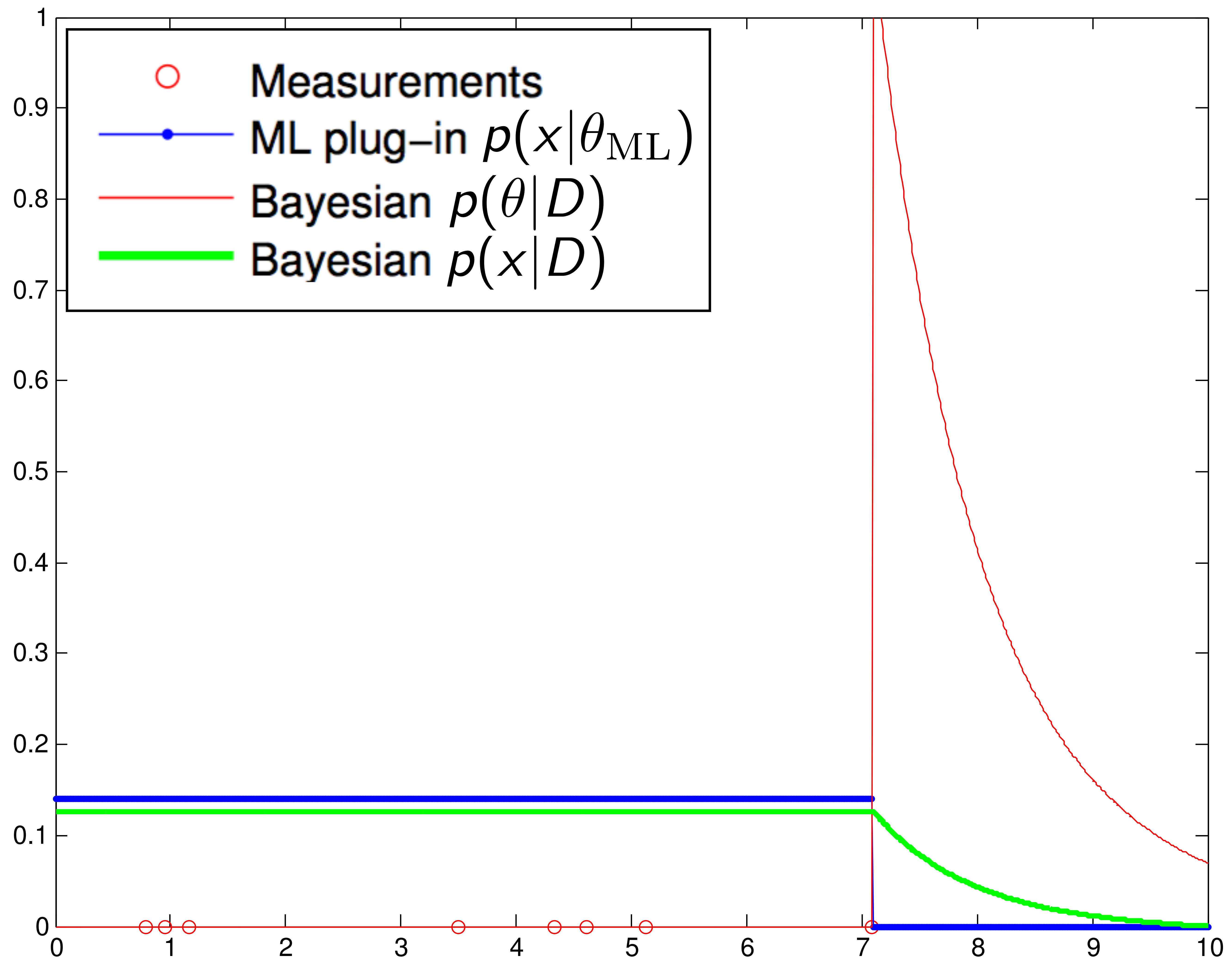
Bayesian Estimation: Uniform Distribution



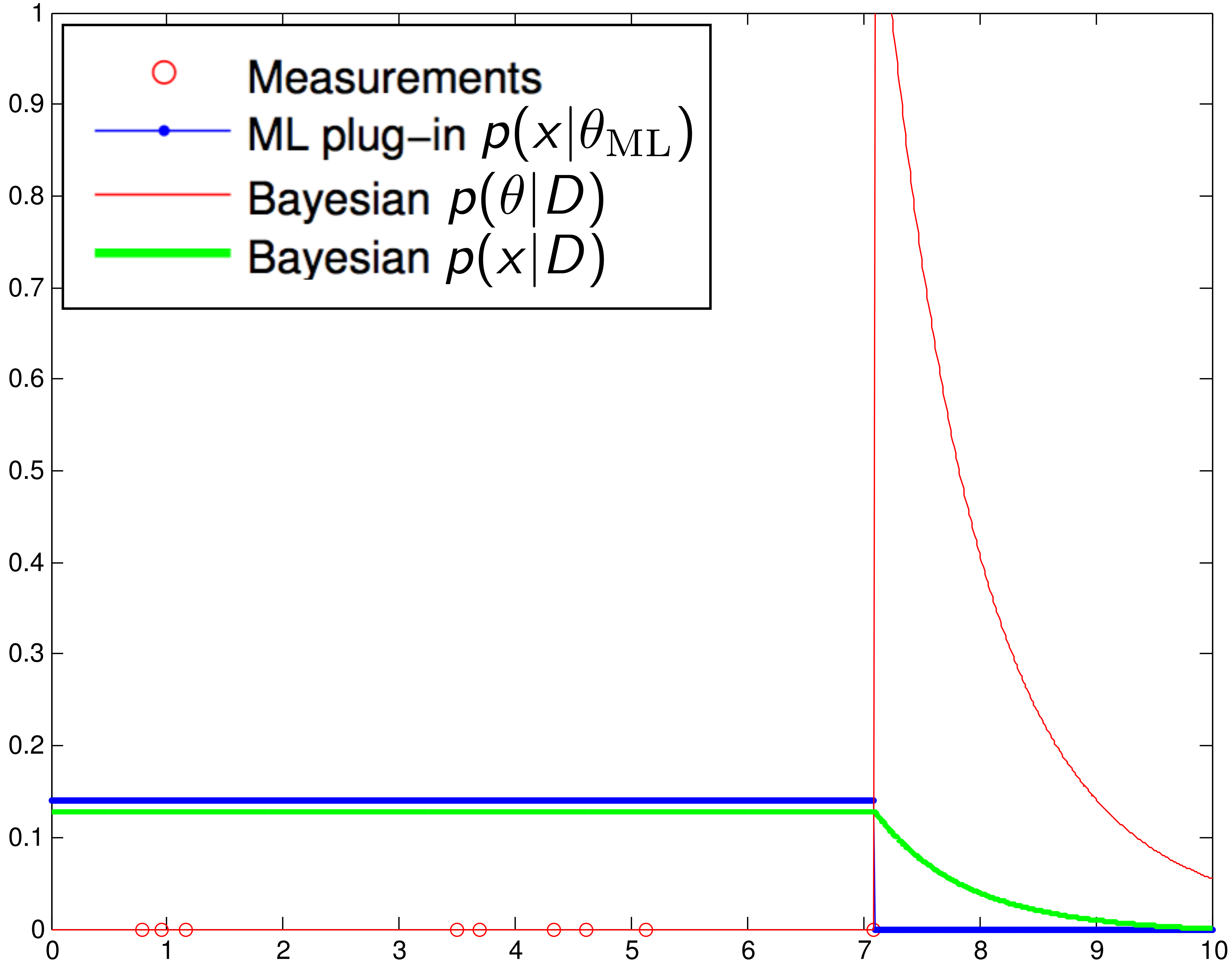
Bayesian Estimation: Uniform Distribution



Bayesian Estimation: Uniform Distribution



Bayesian Estimation: Uniform Distribution



Bayesian Estimation: Uniform Distribution