ASEN 5044, Fall 2018

Statistical Estimation for Dynamical Systems

Lecture 14 [Special Topic Lecture #2]: Conditional PDFs and Bayes' Rule for PDFs

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Fri 10/05/2018

University of Colorado Boulder

Aerospace Engineering

Today…

- What is the meaning of a **conditional probability density function (conditional pdf)** for continuous random variables?

- How can we use Bayes' rule to find conditional pdfs?
  - Updating prior over random variable x with new info from observation y=c: how to go from p(x) → p(x|y=c) given **likelihood** relation p(y=c|x)?

- Grid-based evaluations of 1D/2D Bayesian posterior pdfs
  - Sanity checks for analytical/closed-form posterior pdfs
  - Approximations for analytically intractable/non-closed form posterior pdfs

Recall: Conditional PDFs, Factorization, Independence

- Definition of conditional pdf:

$p(y|x=d) \triangleq \dfrac{p(x=d,y)}{p(x=d)}$ for $p(x=d) \neq 0$

taking a slice of a joint pdf along some "unknown dim" while fixing other dim that's observed

$p(x|y=c) \triangleq \dfrac{p(x,y=c)}{p(y=c)}$ for $p(y=c) \neq 0$

- Conditional factorization of joint pdf:

Since $p(x,y) = p(x,y)\cdot\dfrac{p(x)}{p(x)} = p(x)\dfrac{p(x,y)}{p(x)} = p(x)\cdot p(y|x)\ \left(= p(x,y)\right)$

cond. pdf for y "as if" we knew x

& $p(x,y) = p(x,y)\dfrac{p(y)}{p(y)} = p(y)\dfrac{p(x,y)}{p(y)} = p(y)\cdot p(x|y)\ \left(= p(x,y)\right)$

cond. pdf x "as if" we know y

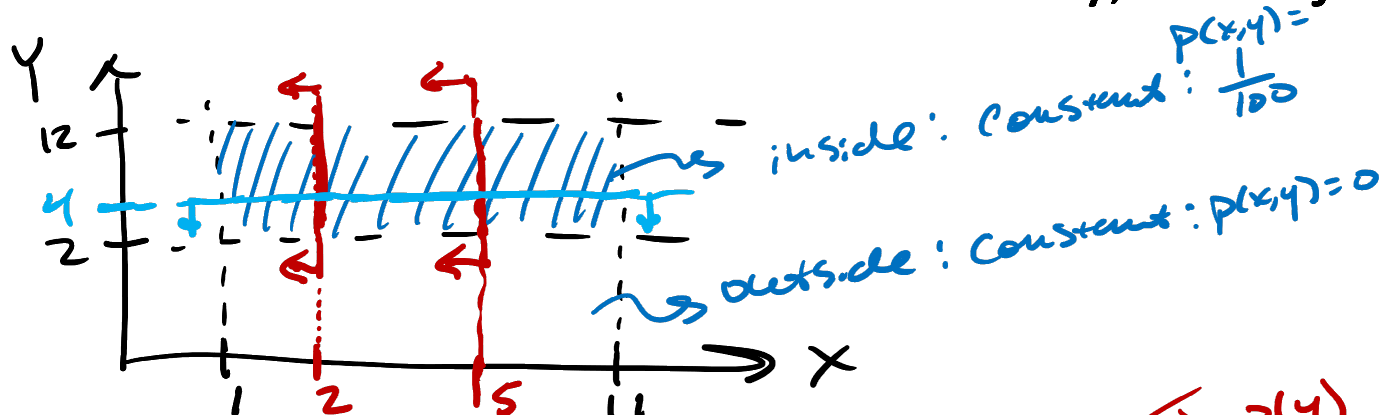
$\boxed{p(x,y) = \underbrace{p(x)\,p(y|x)}_{x\text{ prior factorization}} = \underbrace{p(y)\,p(x|y)}_{y\text{ prior factorization}}}$

- Independence:

If $x \perp\!\!\!\perp y$, then $p(x,y) = p(x)\,p(y)$ since $p(y|x) = p(y)$ & $p(x|y) = p(x)$

Example 1: Conditioning on 2D Joint PDFs

Consider 2 scalar random vars x and y, with joint pdf $p(x,y)=U_x[1,11]\,U_y[2,12]$



→ easy to show that marginals are:

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy = \int_{-\infty}^{\infty} U_x[1,11]\cdot U_y[2,12]\,dy = U_x[1,11] \rightarrow \text{fxn of } x \text{ only}$$

$$\& \; p(y) = \int_{-\infty}^{\infty} p(x,y)\,dx = \int_{-\infty}^{\infty} U_x[1,11]\,U_y[2,12]\,dy = U_y[2,12] \rightarrow \text{fxn of } y \text{ only}$$

$(x \perp\!\!\!\perp y$ obviously$)$

- What is p(y|x=2)=?

$\cancel{p(x=2)}\cdot p(y)$

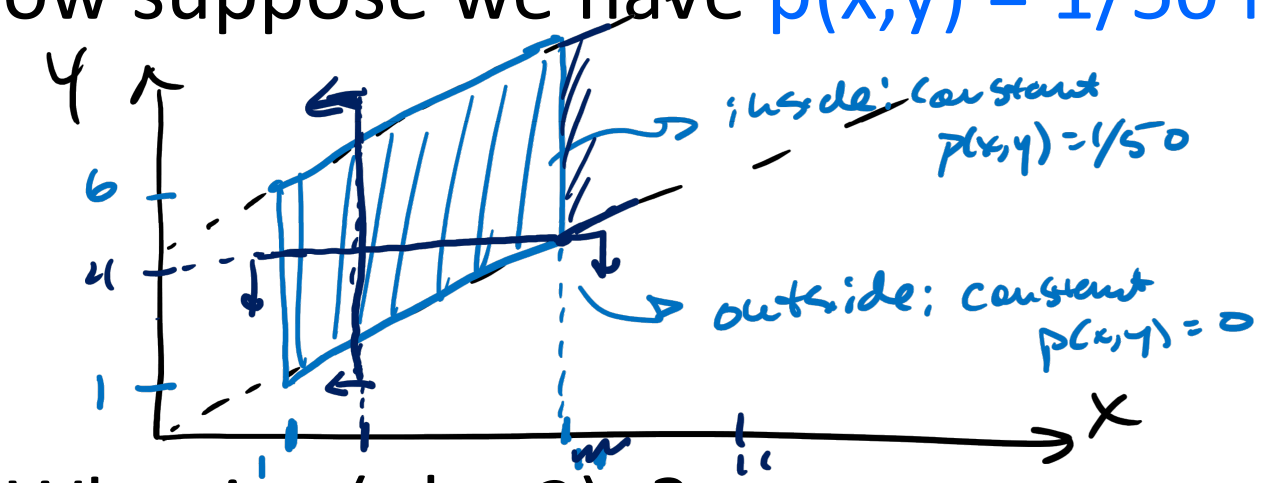
From def. of cond. pdf: $p(y|x=2) = \dfrac{p(x=2,y)}{\cancel{p(x=2)}} = \begin{cases} \dfrac{(1/100)}{U_x[1,11]|_{x=2}} \;\; [1/10], & \text{for } y\in[2,12] \\ 0, & \text{for } y\notin[2,12] \end{cases} = \begin{cases} 1/10 & \text{for } y\in[2,12] \\ 0, & \text{for o'wise} \end{cases}$

- p(y|x=5)=? → Likewise: get $p(y|x=5) = U_y[2,12]$ ⟹ $p(y) = U_y[2,12]$

- p(x|y=4)=? $p(x|y=4) = \dfrac{p(x,y=4)}{p(y=4)} = \dfrac{p(x)\cdot \cancel{p(y=4)}}{\cancel{p(y=4)}} = U_x[1,11]$ → conditioning says nothing about y given x or x given y (b/c $x\perp\!\!\!\perp y$)!

Example 2: Conditioning on 2D Joint PDFs

Now suppose we have p(x,y) = 1/50 for 1<x<11 and x<y<x+5 [slanted uniform slab]



can show that

$p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy = \mathcal{U}_x[1,11]$

& since $x \not\perp Y$,

$p(y) = \int_{-\infty}^{\infty} p(x,y)\,dx = (\ldots)$ something

- What is p(y|x=2)=?

$p(y|x=2) = \dfrac{p(x=2,y)}{p(x=2)} = \begin{cases} \dfrac{(1/50)}{\mathcal{U}_x[2,12]|_{x=2}} = \dfrac{(1/50)}{(1/10)}, & \text{for } x=2 \text{ \& } 2 \le y \le 7 \\ 0, & \text{o'wise} \end{cases} = \begin{cases} 1/5, & \text{for } y \in [2,7] \\ 0, & \text{o'wise} \end{cases}$

$\rightarrow$ so $\boxed{p(y|x=2) = \mathcal{U}_y[2,7]}$

$\rightarrow$ Likewise $p(y|x=5) = \dfrac{p(x=5,y)}{p(x=5)} = \overset{\text{math}}{(\ldots)} = \mathcal{U}_y[5,10] = \begin{cases} 1/5 & \text{for } y \in [5,10] \\ 0, & \text{o'wise} \end{cases}$

- p(x|y=4)=?

$p(x|y=4) = \dfrac{p(x,y=4)}{p(y=4)} = \begin{cases} \dfrac{(1/50)}{p(y=4)_{\leftarrow \text{some \#}}}, & \text{for } y=4 \text{ \& } x \in [1,4] \\ 0, & \text{otherwise} \end{cases} \rightarrow \mathcal{U}_x[1,4]$

only plausible values for x to get y=4

{ turns out: p(y=4) = 3/50 }

$\boxed{p(x|y=4) = \mathcal{U}_x[1,4]}$

Bayes' Rule for Finding Conditional PDFs

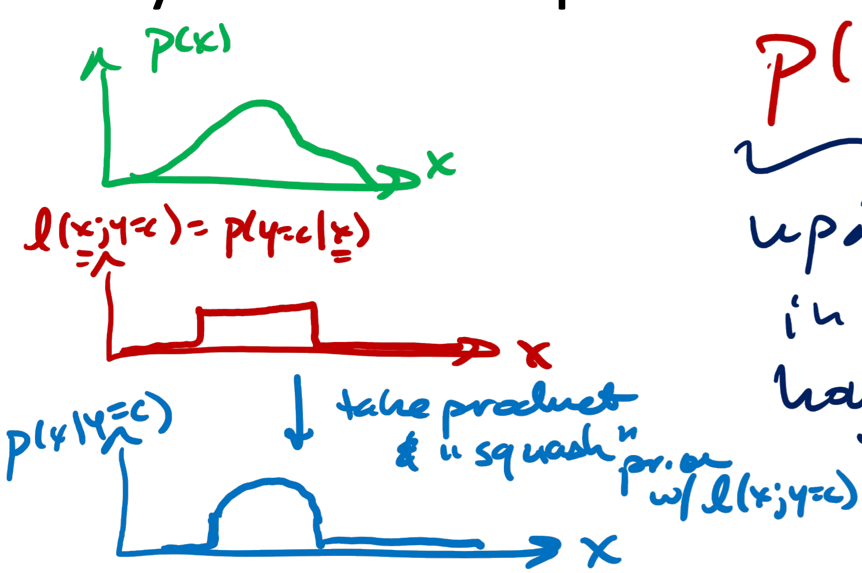
- Suppose we only get p(x,y) in factored form as p(x,y) = p(x)p(y|x)
  - i.e. **prior p(x)** times a **conditional likelihood p(y|x)**

what we know about x w/o knowing anything else

what we know about y "as if" we knew x (or: how likely x is to explain y)

- How to find p(x|y=c)? [i.e. p(y|x) says how to slice joint pdf along y given x – but how to slice joint pdf along x given y if we only know p(x) and p(y|x)?]

- Bayes' rule for pdfs:



$$\underbrace{p(x|y=c)}_{\text{updates } p(x) \text{ in light of how well each } x \text{ "explains" } y=c} = \frac{p(x)\cdot p(y=c|x)}{\underbrace{\int_{-\infty}^{\infty} p(x)\, p(y=c|x)\, dx}_{=p(y=c)=\text{const. w.r.t. } x}} \propto p(x)\cdot p(y=c|x)$$

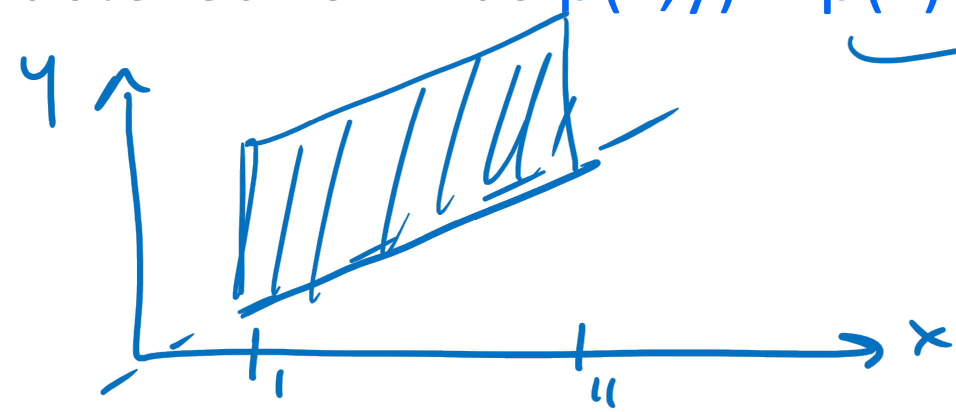
= fxn of x (w/ y=c fixed)

$$\propto \underbrace{p(x)}_{\text{prior}} \cdot \underbrace{\ell(x; y=c)}_{\text{likelihood "score"}}$$

Example 3: Conditioning on 2D Joint PDFs

Repeat p(x|y=4) for uniform slanted slab using Bayes' rule, where we can express joint pdf in factored form as $\underbrace{p(x,y) = p(x)\, p(y|x) = U_x[1,11]\;\; U_{y|x}[x,x+5]}$

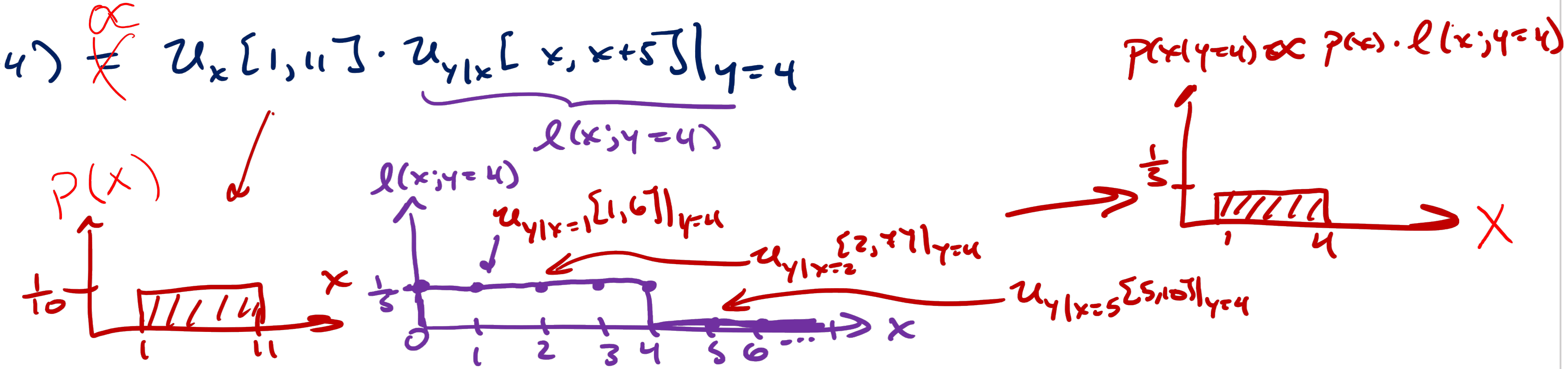
easy to show that this same joint pdf as before!



- p(x|y=4) = ?

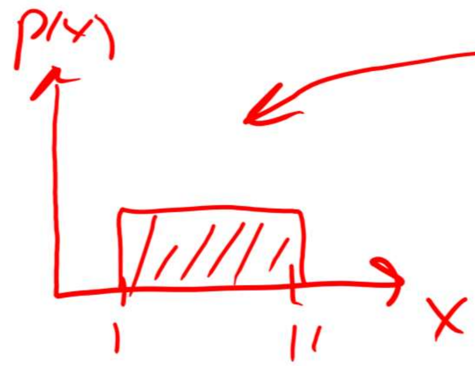
Bayes' rule: $p(x|y=4) = \dfrac{p(x)\cdot p(y=4|x)}{\int_{-\infty}^{\infty}(\ldots)\,dx} \propto p(x)\cdot p(y=4|x)$

$\longrightarrow p(x|y=4) \cancel{=} \propto U_x[1,11]\cdot \underbrace{U_{y|x}[x,x+5]}_{l(x;y=4)}\Big|_{y=4}$

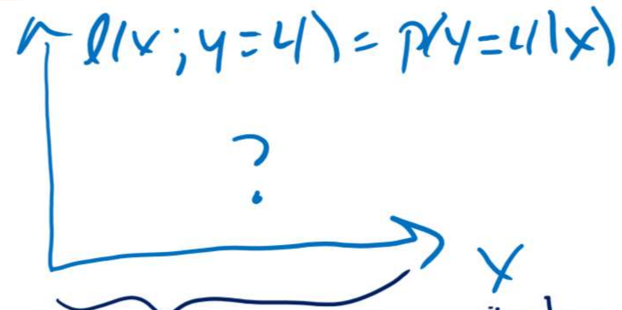


Example 4: Conditioning on 2D Joint PDFs

Consider more complex joint pdf, where $p(x,y) = p(x)\,p(y|x) = \underline{U_x[1,11]}\;\; U_{y|x}[\underline{x},\underline{(x+1)^2}]$

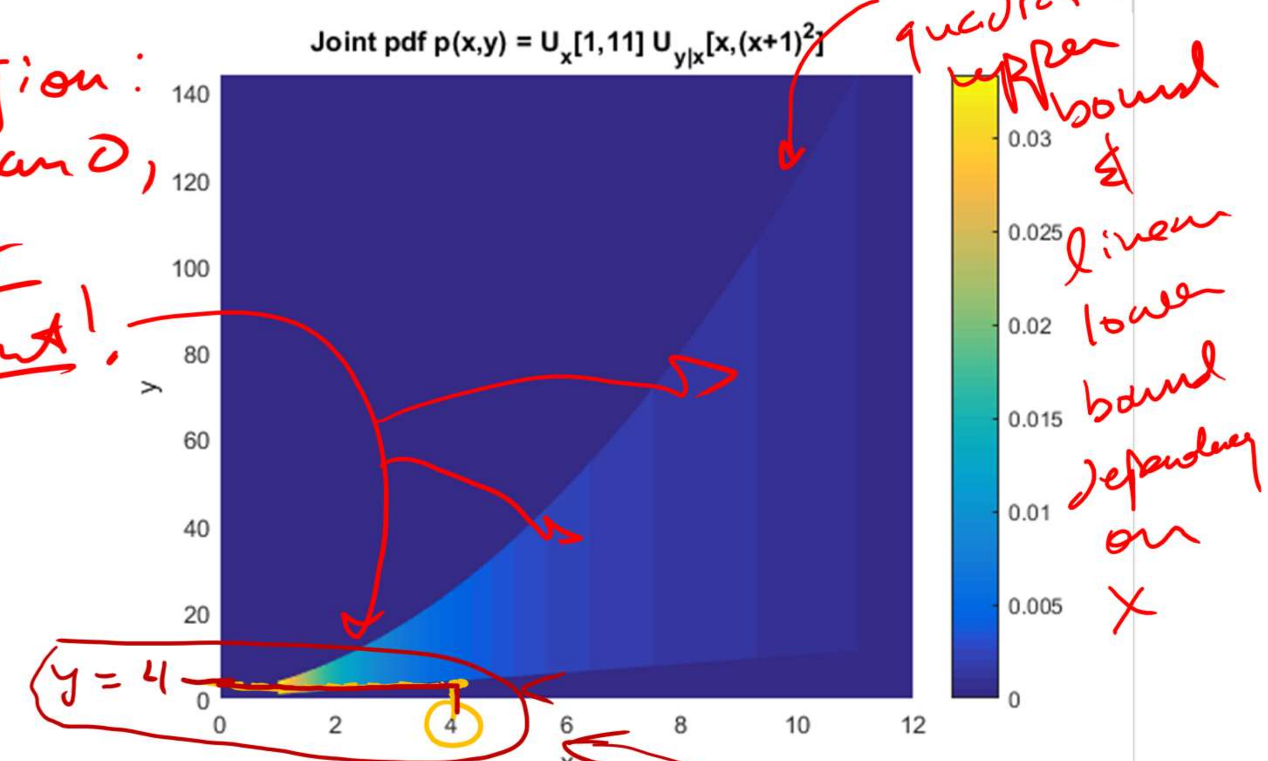


·



(tells us how likely it is that any possible x produced y=4 observation)

Inside region: greater than 0, but not constant!

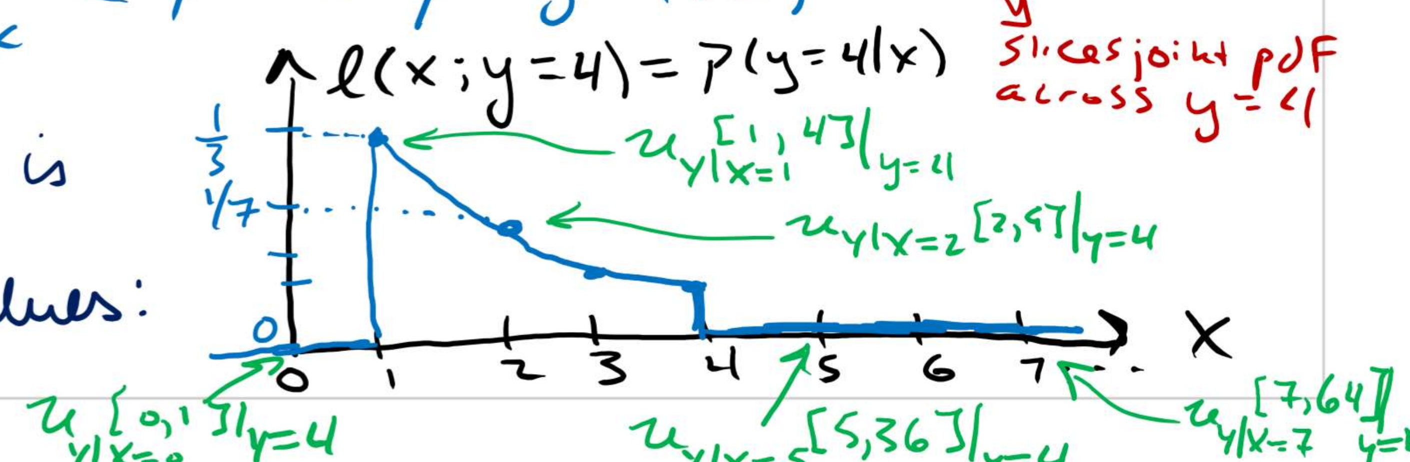


- p(x|y=4) = ?

using Bayes' rule again:

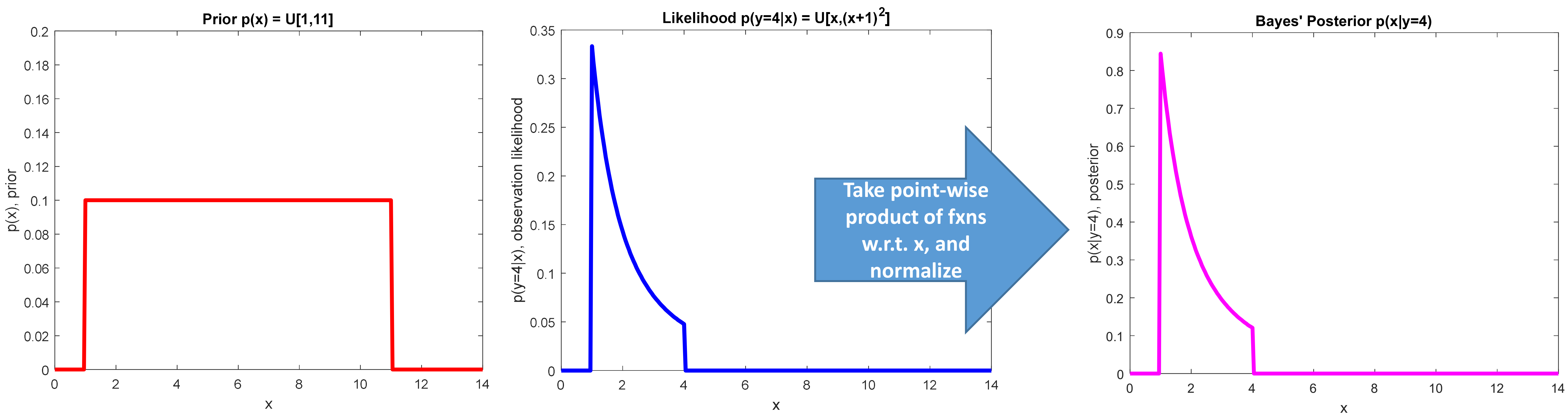
$$p(x|y=4) = \frac{p(x)\,p(y=4|x)}{\int_{-\infty}^{\infty} p(x)\,p(y=4|x)\,dx} \propto p(x)\cdot p(y=4|x)$$

→ likelihood fxn $p(y=4|x) = \ell(x;y=4)$ is much more complicated fxn of x in this case, e.g. consider typical values:



Evaluating p(x|y=c) for different observed c values

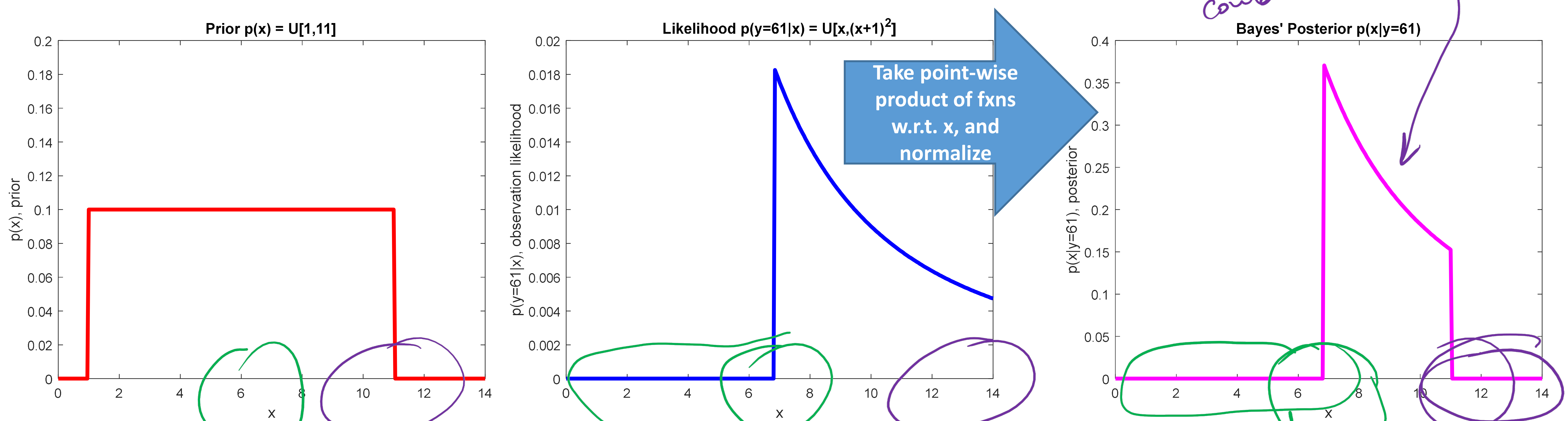
- Matlab code lec14_conditionalpdfBayesRuledemo.m: numerically evaluate the posterior pdf on a grid via Bayes' rule at different observed y=c values:
- y=4 results



Evaluating p(x|y=c) for different observed c values

- Matlab code lec14_conditionalpdfBayesRuledemo.m: numerically evaluate the posterior pdf on a grid via Bayes' rule at different observed y=c values:
- y=61 results

These are the only plausible values for x which could have produced observation y = 61



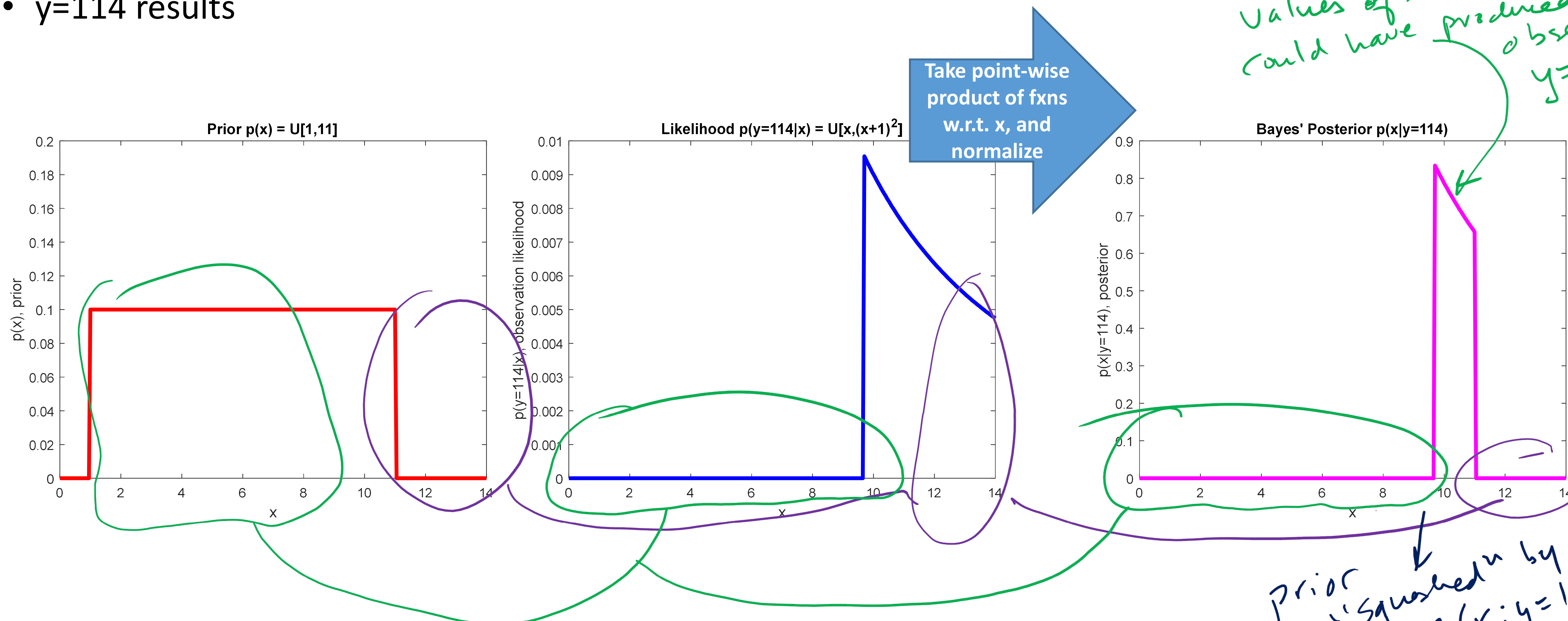
likelihood zeros this part out

prior p(x) zeros out this part

Evaluating p(x|y=c) for different observed c values

- Matlab code lec14_conditionalpdfBayesRuledemo.m: numerically evaluate the posterior pdf on a grid via Bayes' rule at different observed y=c values:
- y=114 results

only plausible values of x which could have produced observation y=114



prior "squashed" by $l(x; y=114)$