Computer vision: models, learning and inference

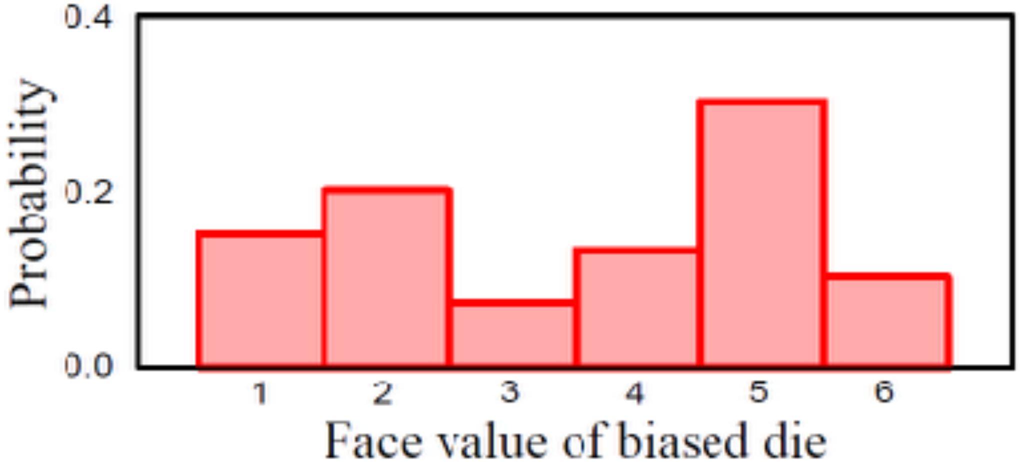
Chapter 2
Introduction to probability

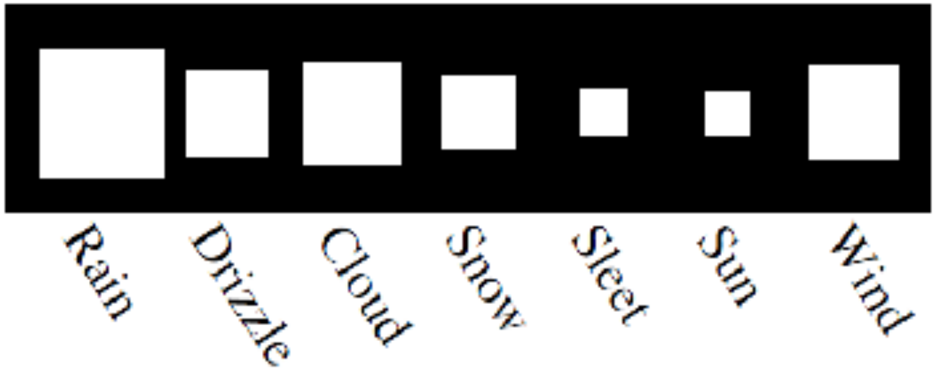
Please send errata to s.prince@cs.ucl.ac.uk

Random variables

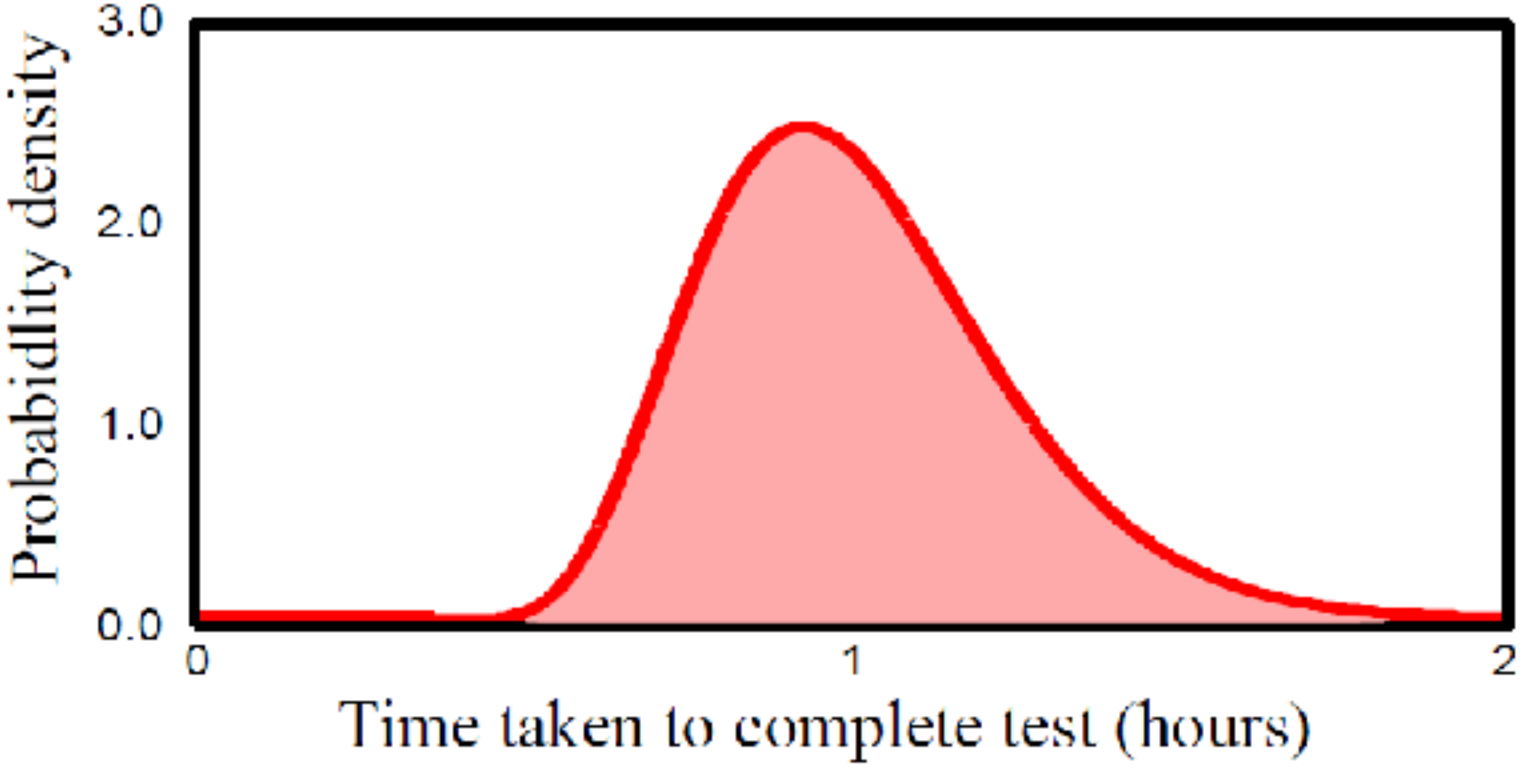
- A random variable x denotes a quantity that is uncertain
- May be result of experiment (flipping a coin) or a real world measurements (measuring temperature)
- If observe several instances of x we get different values
- Some values occur more than others and this information is captured by a probability distribution

Discrete Random Variables





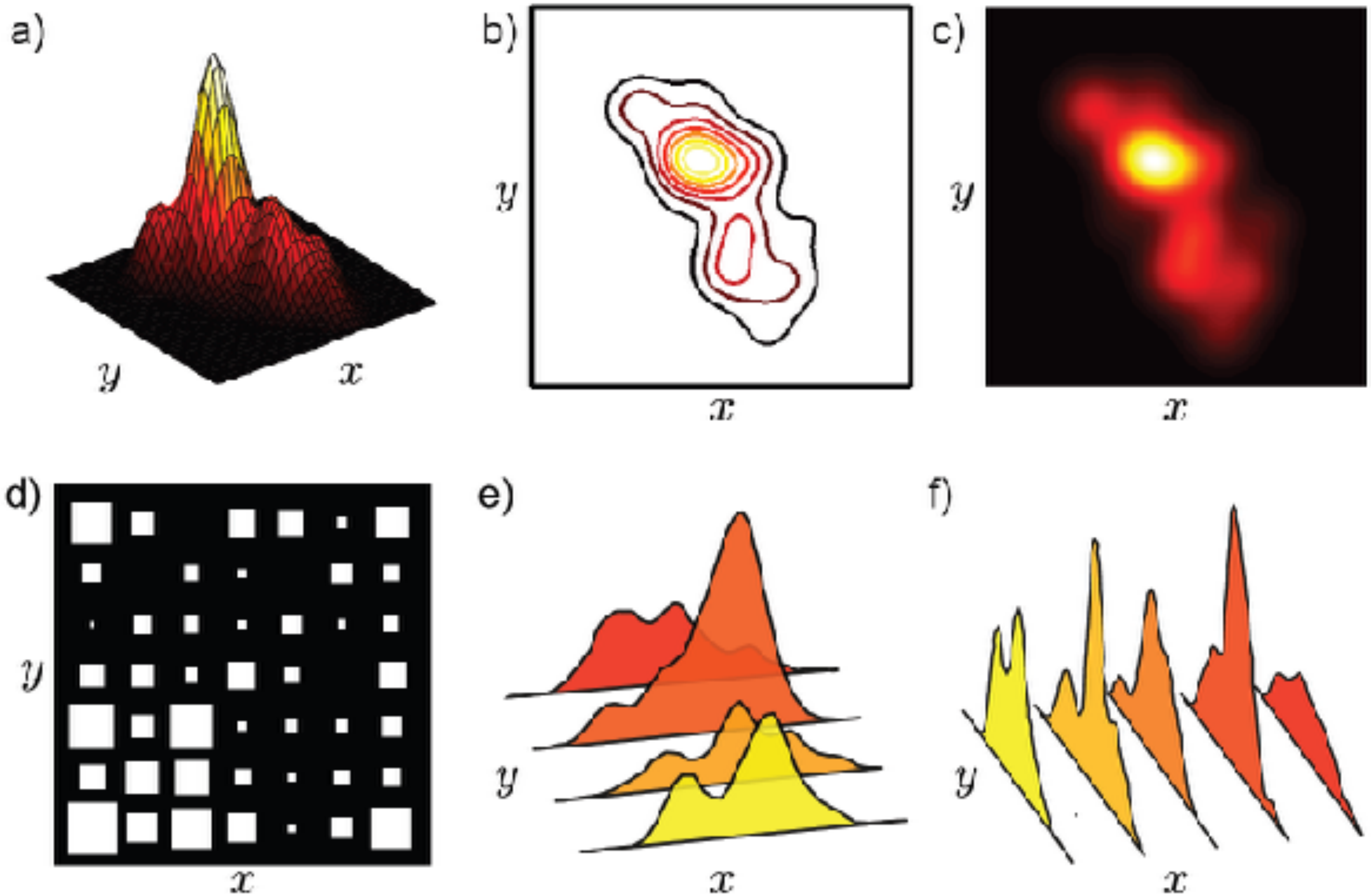
Continuous Random Variable



Joint Probability

- Consider two random variables x and y
- If we observe multiple paired instances, then some combinations of outcomes are more likely than others
- This is captured in the joint probability distribution
- Written as $Pr(x,y)$
- Can read $Pr(x,y)$ as “probability of x and y”

Joint Probability

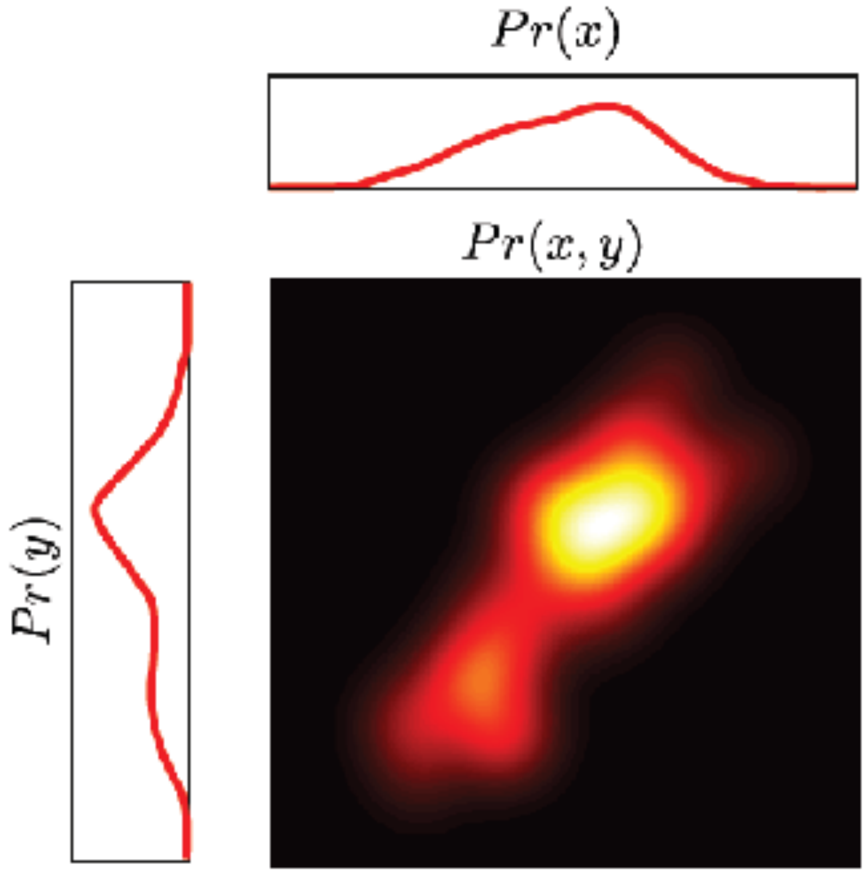


Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

$$Pr(x) = \int Pr(x,y)\,dy$$

$$Pr(y) = \int Pr(x,y)\,dx$$

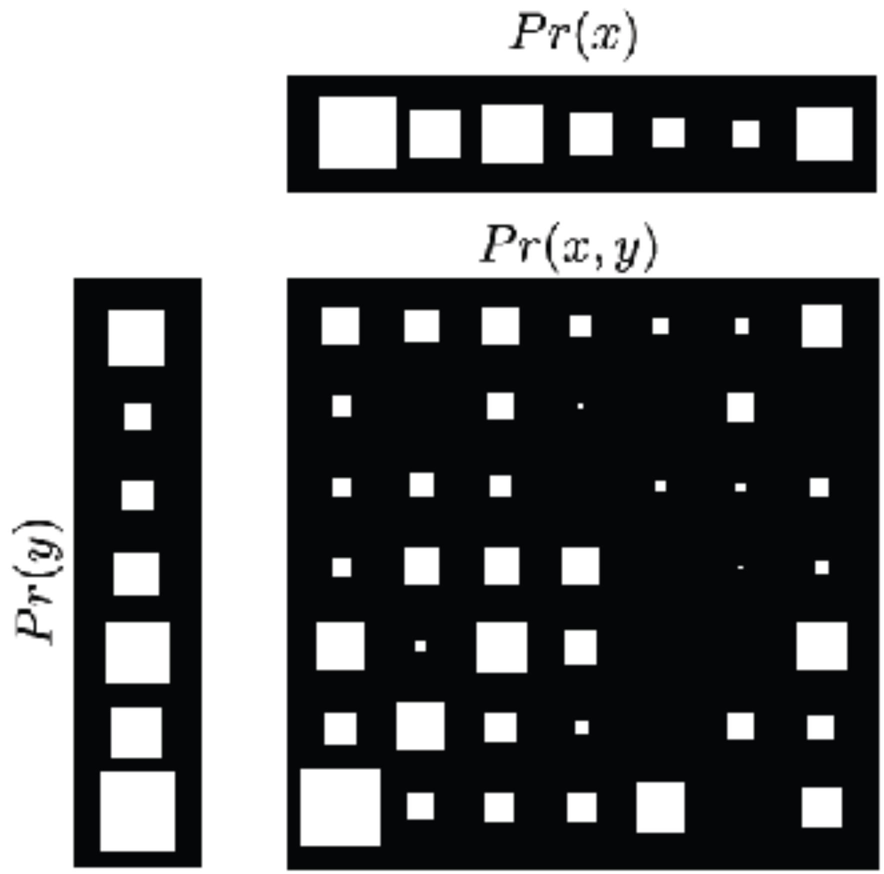


Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

$$Pr(x) = \sum_{y} Pr(x,y)$$

$$Pr(y) = \sum_{x} Pr(x,y)$$

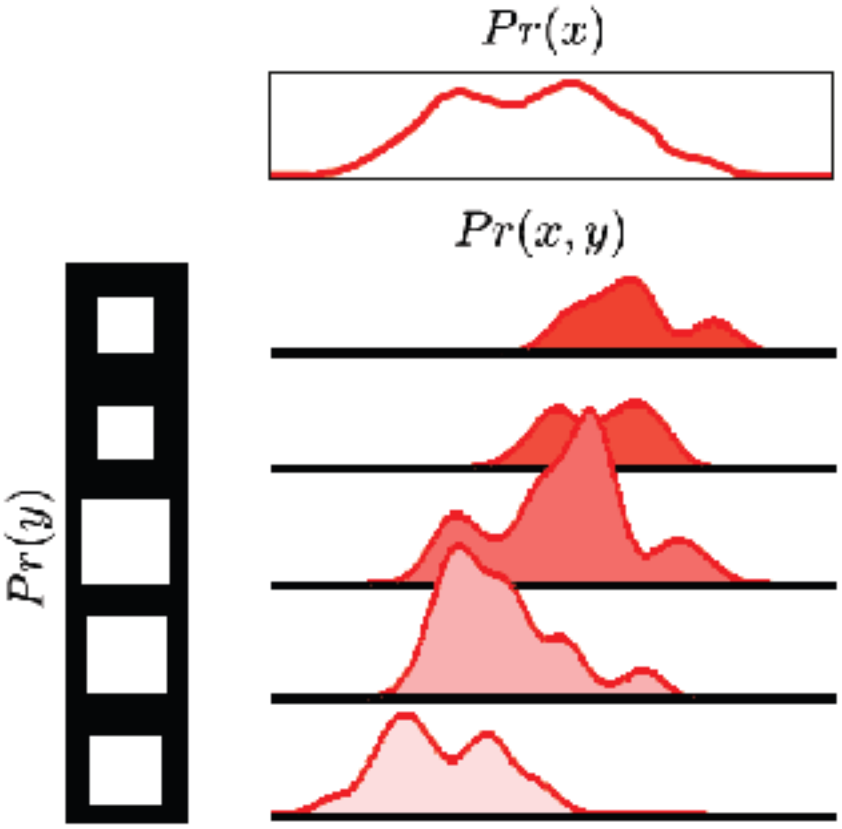


Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

$$Pr(x) = \sum_y Pr(x,y)$$

$$Pr(y) = \int Pr(x,y)\, dx$$



Marginalization

We can recover probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

$$Pr(x) = \int Pr(x,y)\,dy$$

$$Pr(y) = \int Pr(x,y)\,dx$$

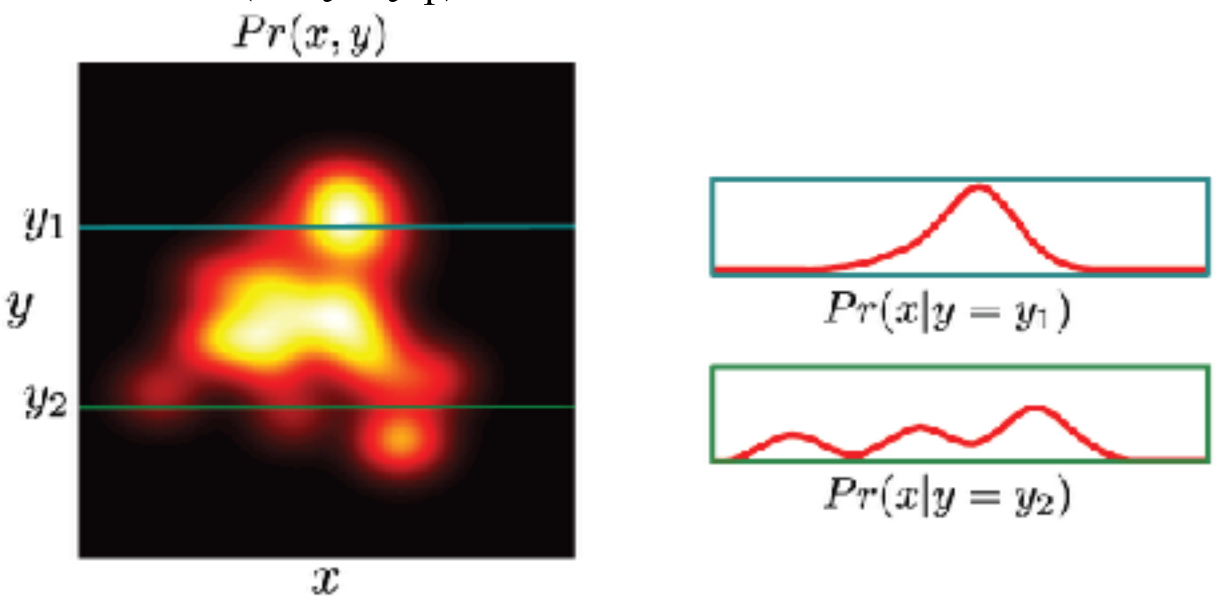
Works in higher dimensions as well – leaves joint distribution between whatever variables are left

$$Pr(x,y) = \sum_{w} \int Pr(w,x,y,z)\,dz$$

Conditional Probability

- Conditional probability of x given that $y=y_1$ is relative propensity of variable x to take different outcomes given that y is fixed to be equal to y_1.

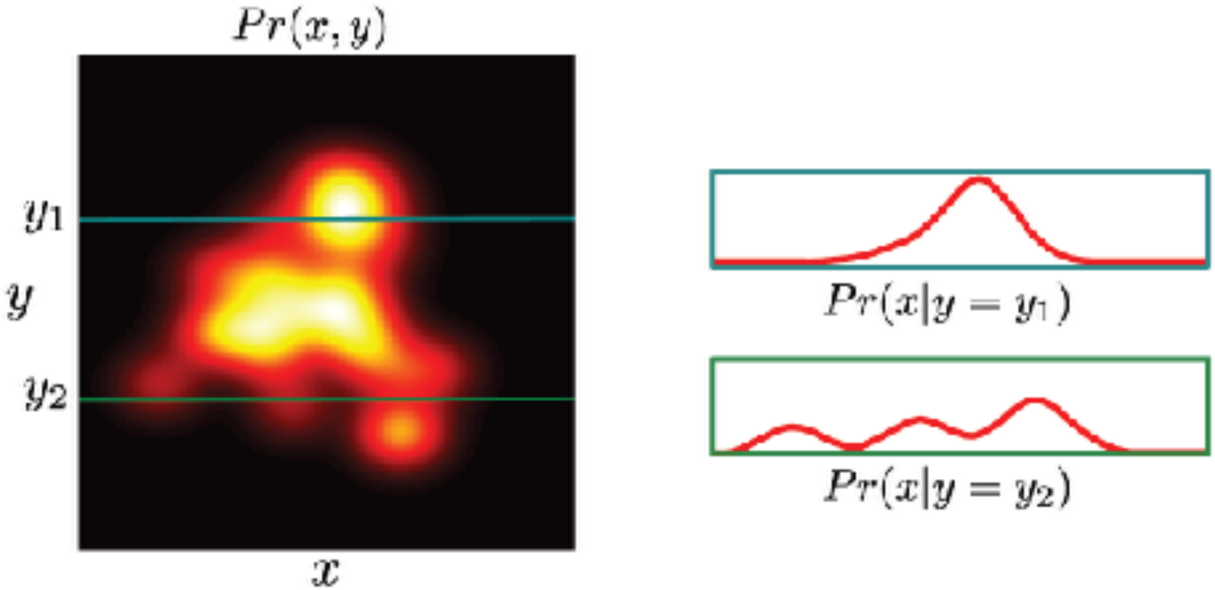
- Written as $Pr(x|y=y_1)$



Conditional Probability

- Conditional probability can be extracted from joint probability
- Extract appropriate slice and normalize

$$Pr(x|y=y^*) = \frac{Pr(x,y=y^*)}{\int Pr(x,y=y^*)dx} = \frac{Pr(x,y=y^*)}{Pr(y=y^*)}$$



Conditional Probability

$$Pr(x|y = y^*) = \frac{Pr(x, y = y^*)}{\int Pr(x, y = y^*)dx} = \frac{Pr(x, y = y^*)}{Pr(y = y^*)}$$

- More usually written in compact form

$$Pr(x|y) = \frac{Pr(x, y)}{Pr(y)}$$

- Can be re-arranged to give

$$Pr(x, y) = Pr(x|y)Pr(y)$$

$$Pr(x, y) = Pr(y|x)Pr(x)$$

Conditional Probability

$$Pr(x, y) = Pr(x|y)Pr(y)$$

- This idea can be extended to more than two variables

$$\begin{aligned}Pr(w, x, y, z) &= Pr(w, x, y|z)Pr(z)\\&= Pr(w, x|y, z)Pr(y|z)Pr(z)\\&= Pr(w|x, y, z)Pr(x|y, z)Pr(y|z)Pr(z)\end{aligned}$$

Bayes' Rule

From before:

$$Pr(x,y) = Pr(x|y)Pr(y)$$

$$Pr(x,y) = Pr(y|x)Pr(x)$$

Combining:

$$Pr(y|x)Pr(x) = Pr(x|y)Pr(y)$$

Re-arranging:

$$\begin{aligned} Pr(y|x) &= \frac{Pr(x|y)Pr(y)}{Pr(x)} \\ &= \frac{Pr(x|y)Pr(y)}{\int Pr(x,y)\, dy} \\ &= \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y)\, dy} \end{aligned}$$

Bayes' Rule Terminology

Likelihood – propensity for observing a certain value of x given a certain value of y

Prior – what we know about y before seeing x

$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{\int Pr(x|y)Pr(y)\,dy}$$

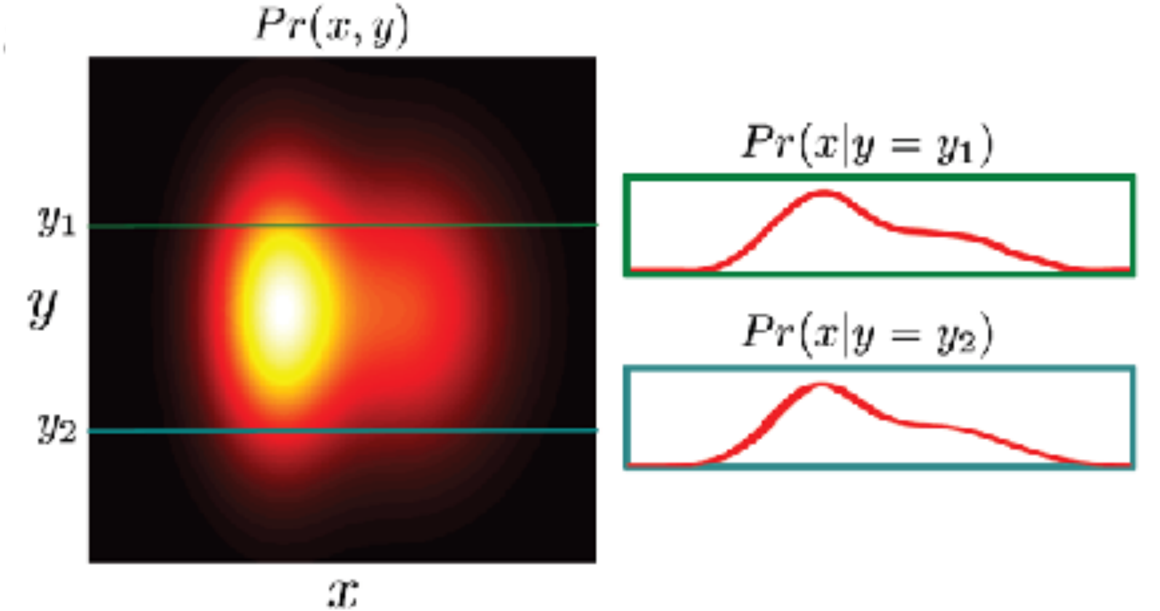
Posterior – what we know about y after seeing x

Evidence –a constant to ensure that the left hand side is a valid distribution

Independence

- If two variables x and y are independent then variable x tells us nothing about variable y (and vice-versa)

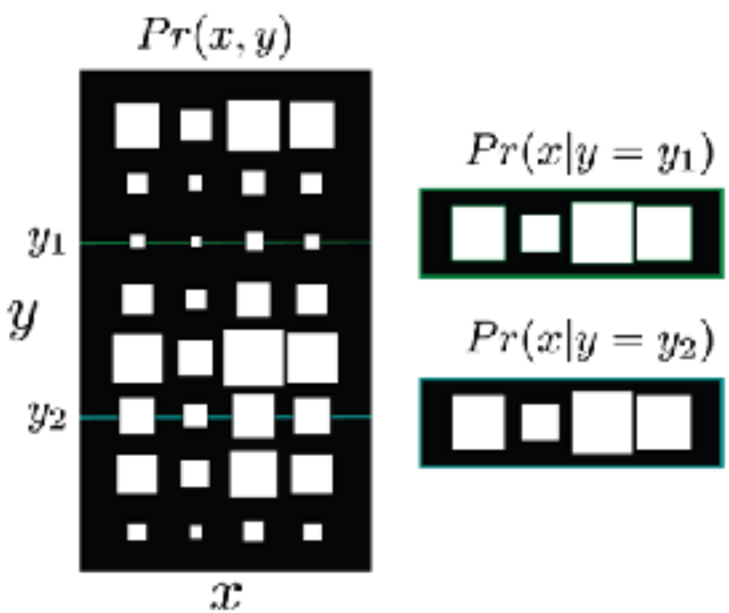
$$Pr(x|y) = Pr(x)$$
$$Pr(y|x) = Pr(y)$$



Independence

- If two variables x and y are independent then variable x tells us nothing about variable y (and vice-versa)

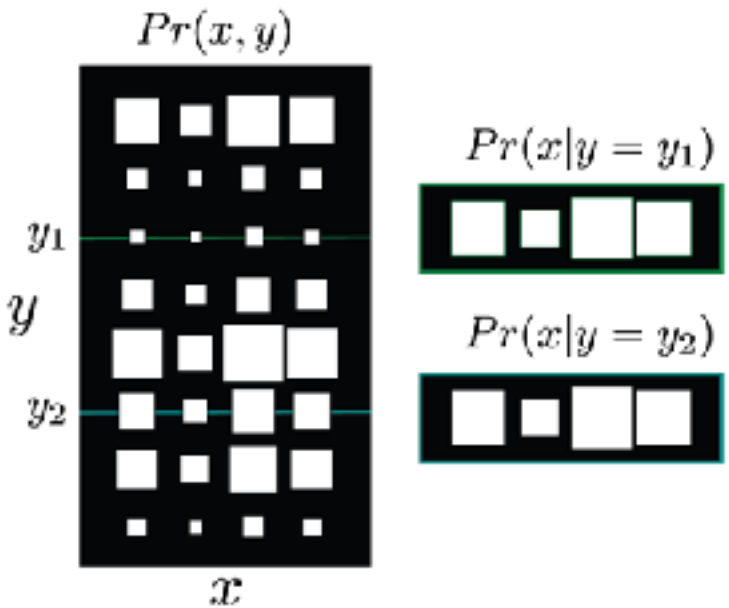
$$
\begin{aligned}
Pr(x|y) &= Pr(x) \\
Pr(y|x) &= Pr(y)
\end{aligned}
$$



Independence

- When variables are independent, the joint factorizes into a product of the marginals:

$$Pr(x, y) = Pr(x|y)Pr(y)$$
$$= Pr(x)Pr(y)$$



Expectation

Expectation tell us the expected or average value of some function $f[x]$ taking into account the distribution of x

Definition:

$$\mathrm{E}[f[x]] = \sum_{x} f[x] Pr(x)$$

$$\mathrm{E}[f[x]] = \int f[x] Pr(x) \, dx$$

Expectation

Expectation tell us the expected or average value of some function $f[x]$ taking into account the distribution of x

Definition in two dimensions:

$$\mathrm{E}[f[x,y]] = \int\int f[x,y]Pr(x,y)\,dx\,dy$$

Expectation: Common Cases

$$\mathrm{E}[f[x]] = \int f[x] Pr(x)\, dx$$

Function $f[\bullet]$	Expectation
x	mean, μ_x
x^k	k^{th} moment about zero
$(x-\mu_x)^k$	k^{th} moment about the mean
$(x-\mu_x)^2$	variance
$(x-\mu_x)^3$	skew
$(x-\mu_x)^4$	kurtosis
$(x-\mu_x)(y-\mu_y)$	covariance of x and y

Expectation: Rules

$$E[f[X]] \quad = \quad \int f[x] Pr(X = x) dx$$

Rule 1:

Expected value of a constant is the constant

$$\mathrm{E}[\kappa] = \kappa$$

Expectation: Rules

$$E[f[X]] \quad = \quad \int f[x]Pr(X=x)dx$$

Rule 2:

Expected value of constant times function is constant times expected value of function

$$\mathrm{E}[kf[x]] = k\mathrm{E}[f[x]]$$

Expectation: Rules

$$E[f[X]] \quad = \quad \int f[x]Pr(X=x)dx$$

Rule 3:

Expectation of sum of functions is sum of expectation of functions

$$\mathrm{E}[f[x]+g[x]] = \mathrm{E}[f[x]] + \mathrm{E}[g[x]]$$

Expectation: Rules

$$E[f[X]] = \int f[x] Pr(X = x) dx$$

Rule 4:

Expectation of product of functions in variables x and y
is product of expectations of functions if x and y are indepenc

$$\mathrm{E}[f[x]g[y]] = \mathrm{E}[f[x]]\mathrm{E}[g[y]] \qquad \text{if } x, y \text{ independent}$$

Conclusions

- Rules of probability are compact and simple
- Concepts of marginalization, joint and conditional probability, Bayes rule and expectation underpin all of the models in this book
- One remaining concept – conditional expectation – discussed later