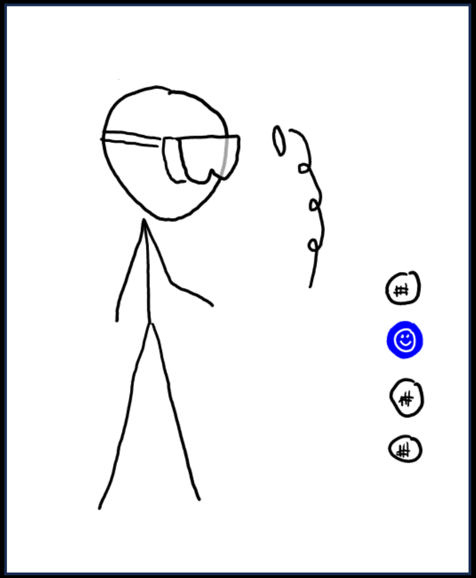
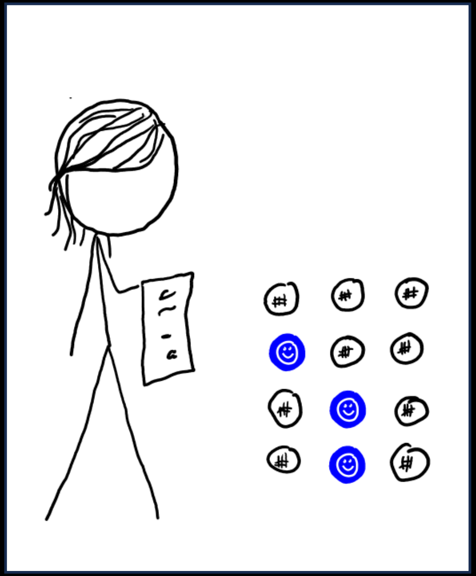


I tossed four coins
and got one head.

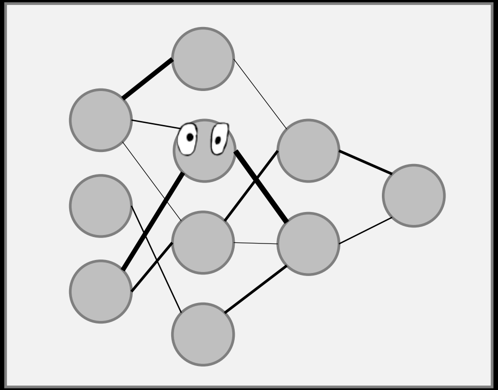
Using a $\text{Bin}(n, p)$ model, I estimate
the probability of heads is $\hat{p} = 25\%$



I tossed twelve coins
and got three heads.

Using a $\text{Bin}(n, p)$ model, I estimate
the probability of heads is $\hat{p} = 25\%$

But surely, the more data we have, the more confident we should be!

"This is a 40mph speed limit, with probability 98%."

Neural networks tell us *probabilities*, but they don't tell us their *confidence.*

No one has worked out how to extract confidences from neural networks. But, in Bayesian statistics, we do know how to …

# Baye's rule

Data from a population sample of 100,000 people:

| | test +ve | test -ve | total |
|---|---|---|---|
| **got COVID** | 376 | 24 | 400 |
| **not got COVID** | 996 | 98,604 | 99,600 |

What are these probabilities?

- $\mathbb{P}(\text{have COVID} \mid \text{test +ve})$
- $\mathbb{P}(\text{have COVID} \mid \text{test} -\text{ve})$

Let's rewrite this data as a probability model:

Let $X = 1_{\text{have COVID}}$ and let $Y = 1_{\text{test+ve}}$

```
X ~ Bin(1, 0.004)        400 / 100 000 = 0.004
if X == 1:
      Y ~ Bin(1, 0.94)   376 / 400 = 0.94
else:
      Y ~ Bin(1, 0.01)   996 / 99 600 = 0.01
```

$$\mathbb{P}(X = 1 \mid Y = 1)$$

$$= \frac{\mathbb{P}(X = 1)\,\mathbb{P}(Y = 1 \mid X = 1)}{\mathbb{P}(Y = 1)}$$

$$= \frac{0.004 \times 0.94}{0.004 \times 0.94 + 0.996 \times 0.01}$$

Reverend Thomas Bayes, 1701–1761

# Bayes's rule for random variables

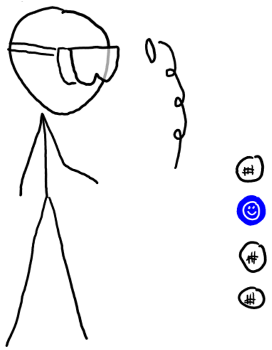
$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x)\,\mathbb{P}(Y = y \mid X = x)}{\mathbb{P}(Y = y)}$$

$$\Pr_X(x \mid Y=y) = \frac{\Pr_X(x)\ \Pr_Y(y \mid X=x)}{\Pr_Y(y)}$$

# Bayesianism

Whenever there's an unknown parameter, you should express your uncertainty about it by treating it as a random variable.

# By using random variables for unknown quantities, we can reason about confidence.



probability of heads, unknown

$\Theta$

↓

$X$

number of heads from 4 coin tosses

We don't know the value of $\Theta$, but we'll assume we know its distribution.

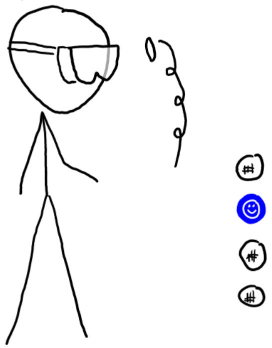
e.g. to express complete ignorance, $\Theta \sim \text{Uniform}[0,1]$

We observed $X = 1$

We can use Bayes's rule to work out how confident we are about the unknown parameter's value …

$\mathbb{P}(\Theta \in [20\%, 30\%] \mid X = 1) = 21\%$

# A more sophisticated way to reason about confidence is by using likelihood functions.

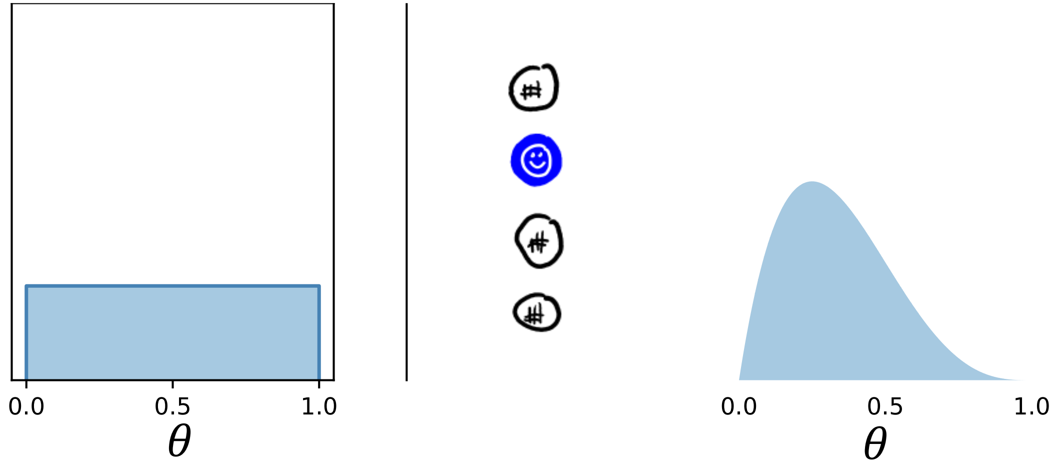


$\Theta \sim U[0,1]$

$\downarrow$

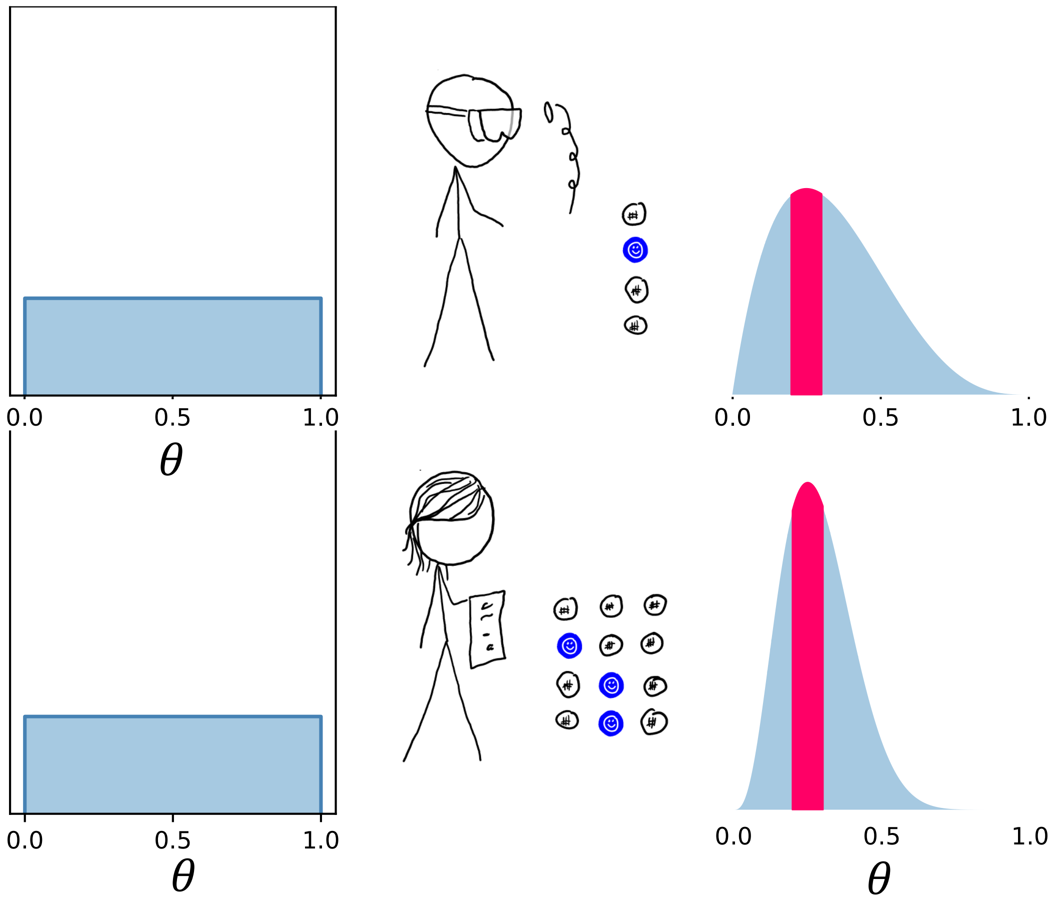
$X \sim \text{Bin}(n, \Theta)$

prior belief $\Pr_\Theta(\theta)$ + data $x$ → posterior belief $\Pr_\Theta(\theta | X = x)$



The data you see will affect your posterior belief about the parameter.

prior belief $\Pr_\Theta(\theta)$ + data $x$ → posterior belief $\Pr_\Theta(\theta | X = x)$




$\mathbb{P}(\Theta \in [.2, .3] \mid \text{data}) = 21\%$

$\mathbb{P}(\Theta \in [.2, .3] \mid \text{data}) = 33\%$

A tighter posterior distribution for $\Theta$ means we are more confident about its value.

# How does Bayes's rule apply to continuous random variables?

Let $X = 1_{\text{have COVID}}$
Let $Y = 1_{\text{test+ve}}$

What is the probability I have COVID, i.e. $X = 1$, if $Y = 1$?

Let $X = 1_{\text{have COVID}}$
Let $Y =$ amount of viral RNA in a PCR test **(CONTINUOUS)**

What is the probability I have COVID, for an amount $Y = y$?

By Bayes's rule,

$$\mathbb{P}(X = 1 \mid Y = 1) = \frac{\mathbb{P}(X = 1)\,\mathbb{P}(Y = 1 \mid X = 1)}{\mathbb{P}(Y = 1)}$$

$$\mathbb{P}(X = 1 | Y = 2.1) = \frac{\mathbb{P}(X = 1)\,\mathbb{P}(Y = 2.1 | X = 1)}{\mathbb{P}(Y = 2.1)}$$

0

0

This version of Bayes's rule doesn't make sense for continuous random variables!

TODAY

## §5.1, 5.2. Bayes's rule done right

§4. Measuring how well a model fits the data (* non-examinable)

WEDNESDAY

## §6. Applying Bayes's rule computationally

Climate challenge

FRIDAY

## §8. Bayesianism

For questions or feedback, I'll be in the café after the lecture.

## Bayes's rule

For two **discrete** random variables $X$ and $Y$,

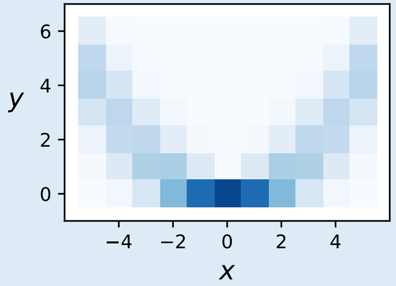
$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x)\mathbb{P}(Y = y | X = x)}{\mathbb{P}(Y = y)} \quad \text{when } \mathbb{P}(Y = y) > 0$$

For two **discrete or continuous** random variables $X$ and $Y$,

$$\Pr_X(x | Y = y) = \frac{\Pr_X(x)\Pr_Y(y | X = x)}{\Pr_Y(y)} \quad \text{when } \Pr_Y(Y) > 0$$

# Joint distribution






```
def rxy():
    x = np.random.randint(low=-5, high=6)  # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```

**The joint pmf of $(X, Y)$**

$\Pr_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ $= \mathbb{P}(X=x)\,\mathbb{P}(Y=y \mid X=x) = \frac{1}{11} \times \binom{6}{y} \left[\left(\frac{x}{6}\right)^2\right]^y \left[1-\left(\frac{x}{6}\right)^2\right]^{6-y}$

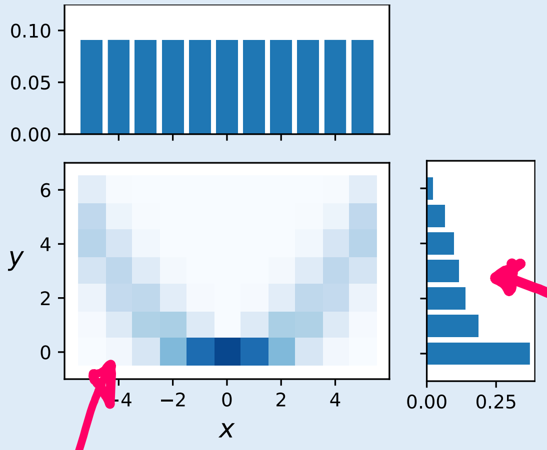
defn. of cond. prob.

**Code to plot the joint pmf**

```
xy_samp = [rxy() for _ in range(1000)]
plt.hist2d(xy_samp)
```

# Marginal random variables






```
def rxy():
    x = np.random.randint(low=-5, high=6)  # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```

**The joint pmf of $(X, Y)$**

$$\Pr_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$$

**The marginal of $Y$**

$$\Pr_Y(y) = \mathbb{P}(Y = y)$$
$$= \sum_x \mathbb{P}(X = x, Y = y) \quad \text{by the Sum Rule}$$
$$= \sum_x \Pr_{X,Y}(x,y)$$

**Code to plot the joint pmf**

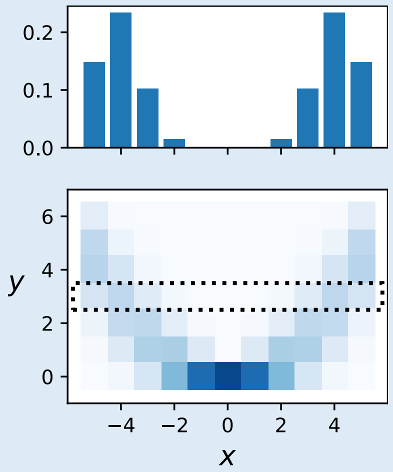
```
xy_samp = [rxy() for _ in range(1000)]
plt.hist2d(xy_samp)
```

**Code to plot the marginal pmf**

```
y_samp = [y for (x,y) in xy_samp]   ← i.e. just throw away the x values
plt.hist(y_samp)
```

# Conditional random variables






```
def rxy():
    x = np.random.randint(low=-5, high=6)  # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```

**$X$ conditional on $Y = 3$**

$$\mathbb{P}(X = x | Y = 3) = \frac{\mathbb{P}(X = x, Y = 3)}{\mathbb{P}(Y = 3)} = \frac{\Pr_{X,Y}(x,3)}{\Pr_Y(3)}$$

$\text{pmf}_3(x)$ ″

i.e. take the Y=3 row, then rescale it to sum to 1

**We can think of "$X$ conditional on $Y = 3$" as a random variable …**

We've provided a valid probability mass function:

$\text{pmf}_3(\cdot) \geq 0 \qquad \sum_x \text{pmf}_3(x) = 1$

Sample space: $\Omega = \{-5, -4, \ldots, 4, 5\}$ same as for $X$.

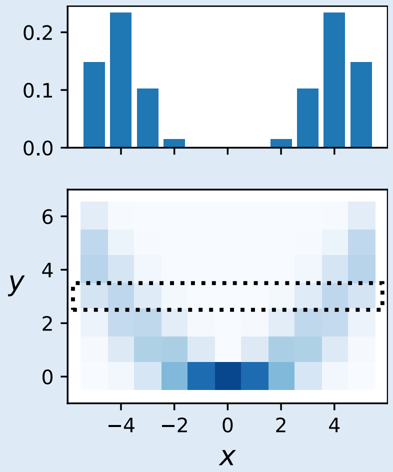
Code to generate values from it:

```
def rx_given_y():
    while True:
        x,y = rxy()
        if y == 3: break
    return x
```

```
def rx_given_y():
    Ω = {-5,...,5}
    p = [pmf(x) for x in Ω]
    return np.random.choice(Ω, p=p)
```

# Conditional random variables






```
def rxy():
    x = np.random.randint(low=-5, high=6)  # from -5 to +5 inclusive
    y = np.random.binomial(n=6, p=(x/6)**2)
    return (x,y)
```

$\Theta \sim U[0,1]$

↓

$X \sim Bin(4, \Theta)$

We *define* the **conditional random variable**, written $(X|Y=y)$, by specifying its likelihood:

$$\Pr_{(X|Y=y)}(x) = \frac{\Pr_{X,Y}(x,y)}{\Pr_Y(y)}$$

Taking the $Y=y$ row from joint pmf, rescale it

commonly written $\Pr_X(x \mid Y=y)$

```
def rx_given_y():
    Ω = {-5,...,5}
    p = [pmf(x) for x in Ω]
    return np.random.choice(Ω, p=p)
```

# Recall: pdf and cdf for continuous random variables

### Definition of continuous RV

**Continuous random variable**

A random variable $X$ is continuous if there is a probability density function (PDF), $f(x) \geq 0$ such that for $-\infty < x < \infty$:

$$\mathbf{P}[a \leq X \leq b] = \int_a^b f(x)dx$$

To preserve the axioms that guarantee that $\mathbf{P}[a \leq X \leq b]$ is a probability, the following properties must hold:

$$0 \leq \mathbf{P}[a \leq X \leq b] \leq 1$$

$$\mathbf{P}[-\infty < X < \infty] = 1 \qquad \left(= \int_{-\infty}^{\infty} f(x)dx\right)$$

- Note: we also write $f(x)$ as $f_X(x)$.
- In continuous world, every RV has a PDF: its relative value wrt to other possible values.
- Integrate $f(x)$ to get probabilities.



For a continuous random variable $X$

$$\mathbb{P}(x_1 \leq X \leq x_2) = \int_{x=x_1}^{x_2} \Pr_X(x)\, dx$$

$$\Pr_X(x) = \frac{d}{dx}\mathbb{P}(X \leq x)$$

### Joint Distributions of Continuous Variables

**Definition**

Random variables $X$ and $Y$ have a joint continuous distribution if for some function $f : \mathbb{R}^2 \to \mathbb{R}$ and for all numbers $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$\mathbf{P}[a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2] = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x,y)\,dx\,dy.$$

The function $f$ has to satisfy $f(x,y) \geq 0$ for all $x$ and $y$, and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$. We call $f$ the joint probability density.

As in one-dimensional case we switch from $F$ to $f$ by differentiating (or integrating):

$$F(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)dx\,dy \quad \text{and} \quad f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$$



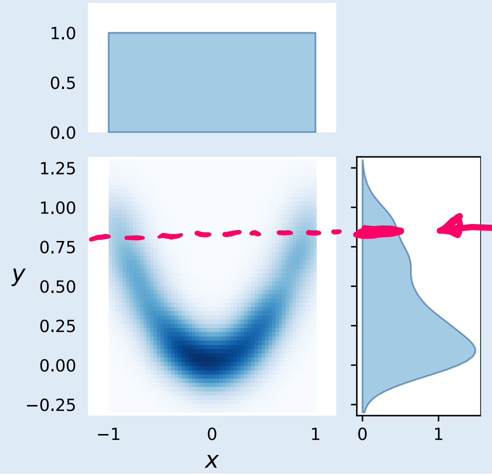
For a pair of continuous random variable $X$ and $Y$

$$\mathbb{P}(x_1 \leq X \leq x_2 \text{ and } y_1 \leq Y \leq y_2) = \int_{x=x_1}^{x_2}\int_{y=y_1}^{y_2} \Pr_{X,Y}(x,y)\, dx\, dy$$

$$\Pr_{X,Y}(x,y) = \frac{\partial^2}{\partial x\, \partial y}\mathbb{P}(X \leq x \text{ and } Y \leq y)$$

# Joint distribution and marginals (continuous case)






```
def rxy():
    x = np.random.uniform(-1,1)
    y = np.random.normal(loc=x**2, scale=0.1)
    return (x,y)
```

**The joint pdf of $(X, Y)$**

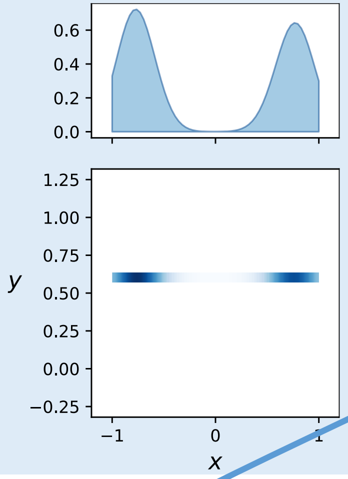
$\Pr_{X,Y}(x,y)$

**The marginal of $Y$**

$$\Pr_Y(y) = \int_x \Pr_{X,Y}(x,y)\ dx$$

# Conditional random variables (continuous case)






```
def rxy():
    x = np.random.uniform(-1,1)
    y = np.random.normal(loc=x**2, scale=0.1)
    return (x,y)
```

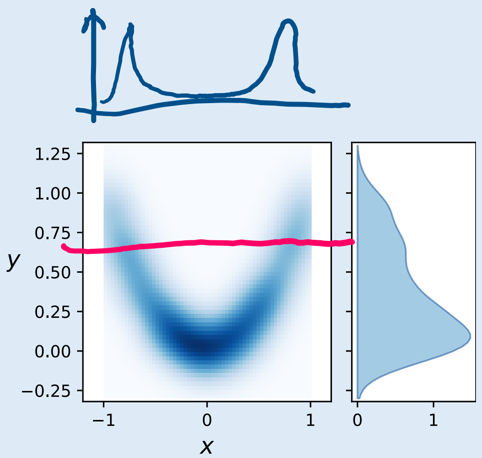
Take the $Y = 0.6$ slice of the joint pdf, then rescale it so it integrates to 1 i.e. so we get a legitimate pdf.

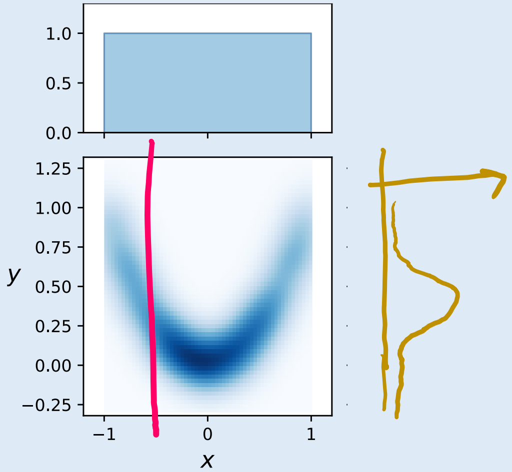
**We define the conditional random variable** $(\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y})$ by specifying its likelihood:

$$\Pr_X(x|Y = y) = \frac{\Pr_{X,Y}(x, y)}{\Pr_Y(y)}$$

# Bayes's rule






$$\Pr_X(x \mid Y=y) = \frac{\Pr_{X,Y}(x,y)}{\Pr_Y(y)}$$

$$\Pr_Y(y \mid X=x) = \frac{\Pr_{X,Y}(x,y)}{\Pr_X(x)}$$

$$\Pr_X(x \mid Y=y) = \frac{\Pr_{X,Y}(x,y)}{\Pr_Y(y)} = \frac{\Pr_X(x)\,\Pr_Y(y \mid X=x)}{\Pr_Y(y)}$$

Bayes's rule is true for any pair of random variables $X, Y$.
It's only useful in "sequential models" i.e. when the question tells us $\Pr_X(x)$ and $\Pr_Y(y|X=x)$.

## Bayes's rule for discrete or continuous random variables

For two random variables $X$ and $Y$,

$$\Pr_X(x|Y=y) = \frac{\Pr_X(x)\,\Pr_Y(y|X=x)}{\Pr_Y(y)} \quad \text{when} \quad \Pr_Y(y) > 0$$

In practice, we use it as

$$\Pr_X(x|Y=y) = \kappa\,\Pr_X(x)\,\Pr_Y(y|X=x)$$

$\Pr_{(x|Y=y)}(x)$

constant that doesn't involve $x$

then figure out $\kappa$ so that $\Pr_X(\cdot\,|Y=y)$ is a legitimate likelihood function

$\int_x \Pr_X(x|Y=y)\,dx = 1$ or $\sum_x \Pr_X(x|Y=y) = 1$

ie so that $\int_x \Pr_{(X|Y=y)}(x)\,dx = 1.$

## Exercise 5.2.1

Consider the pair of random variables $(X, Y)$ generated by

```
def rxy(σ):
    x = np.random.uniform(-1,1)
    y = np.random.normal(loc=x**2, scale=σ)
    return (x,y)
```

Or, in maths notation,

$$X \sim U[-1,1], \qquad Y \sim N(X^2, \sigma^2)$$

Calculate $\Pr_X(x \mid Y = y)$.

$$\Pr_X(x) = \frac{1}{2} \quad \text{since } X \sim U[-1,1]$$


$$\Pr_Y(y|X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-x^2)^2/2\sigma^2}$$

$$\Pr_X(x|Y = y) = \kappa \Pr_X(x) \Pr_Y(y|X = x) = \kappa \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-x^2)^2/2\sigma^2}$$

function of $x$

$$= \kappa' e^{-(y-x^2)^2/2\sigma^2}$$ where $\kappa'$ has non-$x$ terms.

$$= \kappa' e^{-(x^2-y)^2/2\sigma^2}$$ to remind me it's a function of $x$.

$$\int_{-1}^{1} \kappa' e^{-(x^2-y)^2/2\sigma^2}\, dx = 1$$

$$\Rightarrow \quad \kappa' = \frac{1}{\int_{-1}^{1} e^{-(x^2-y)^2/2\sigma^2}\, dx}$$

= <yuck />