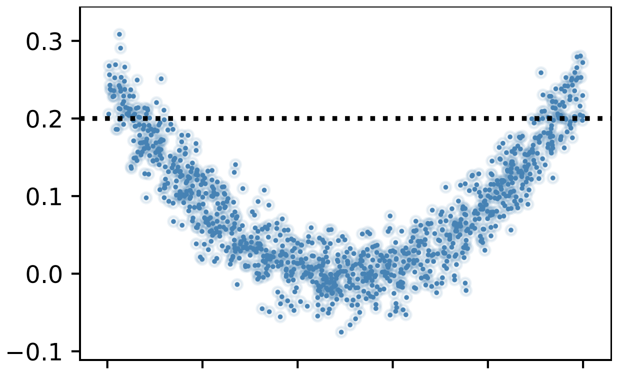



```
def rxy():
    x = numpy.random.uniform(0,1)
    y = numpy.random.normal(loc=x*(x-1)+1/4, scale=.03)
    return (x,y)
```

Question. Use Bayes's theorem to compute $\mathbb{P}(X > 0.9 \mid Y = 0.2)$.

Goal of today's lecture:

to understand how these four lines of code answer the question.

```
x_samp = numpy.random.uniform(size=10000)
Pr_y = scipy.stats.norm.pdf(0.2, loc=x_samp*(x_samp-1)+1/4, scale=0.03)
w = Pr_y / sum(Pr_y)
numpy.sum(w[x_samp > 0.9])
```

The elementary version of Bayes's rule doesn't work when there are continuous random variables. This version does:

Bayes's rule. For two random variables X and Y,

$$\mathrm{Pr}_X(x|Y=y) = \frac{\mathrm{Pr}_X(x)\mathrm{Pr}_Y(y|X=x)}{\mathrm{Pr}_Y(y)} \quad \text{when } \mathrm{Pr}_Y(y) > 0$$

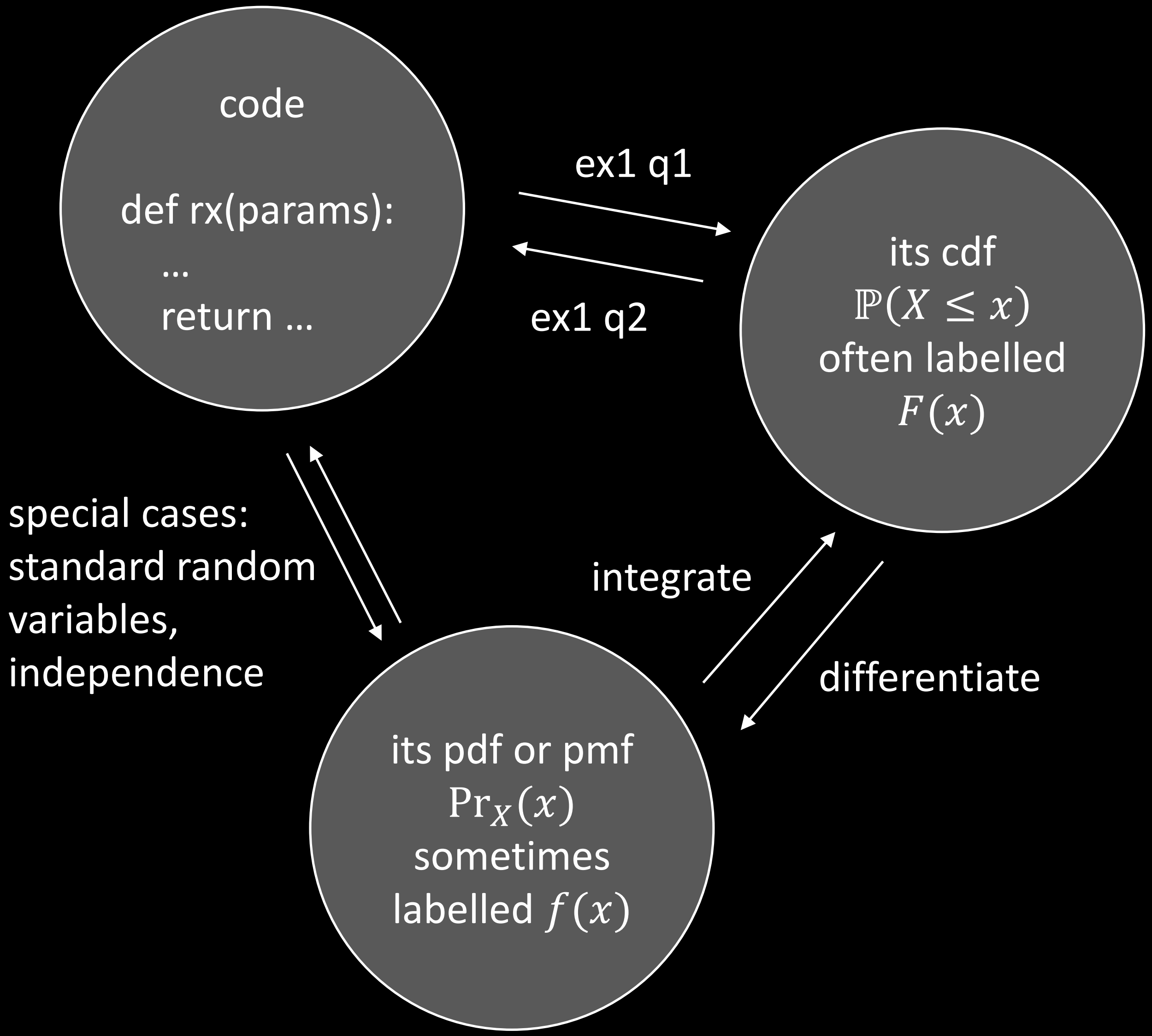
For a discrete random variable, $\mathrm{Pr}_X(x)$ is defined to be $\mathbb{P}(X = x)$.

For a continuous random variable, $\mathrm{Pr}_X(x)$ is defined to be $\mathrm{pdf}(x)$.

What does $\mathrm{Pr}_X(y|X = x)$ mean?

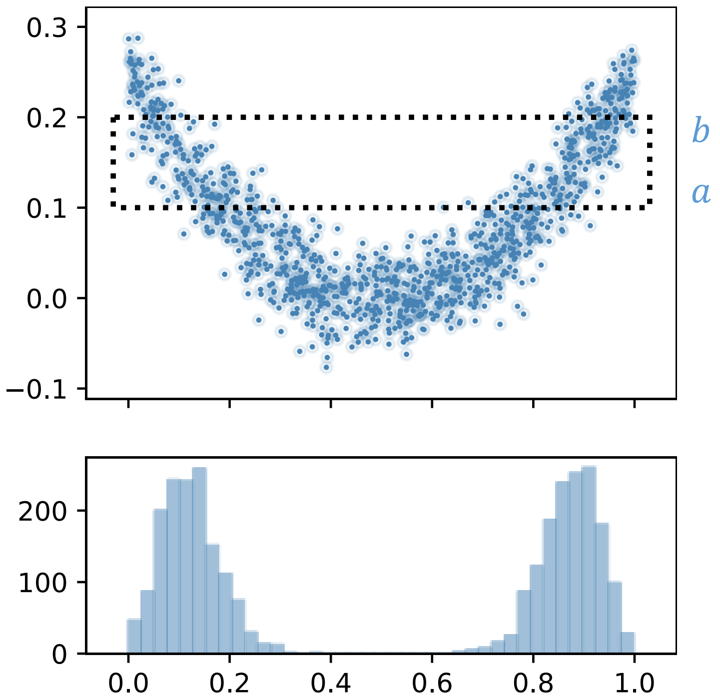
What is a random variable?

How can we describe a random variable?



3.5. Conditional random variables

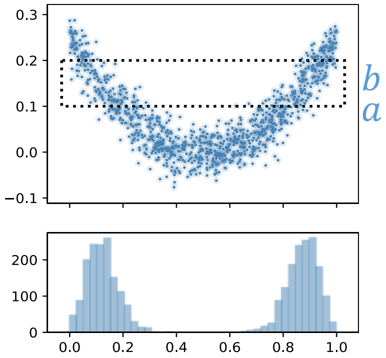
For a random variable X and an event C, we write $(X|C)$ for X *conditioned on* C.



For example, here's code to simulate $(X|Y \in [a,b])$

```
def rxy():
    x = numpy.random.uniform(0,1)
    y = numpy.random.normal(loc=x*(x-1)+1/4, scale=.03)
    return (x,y)

def rx_given_yrange(a,b):
    while True:
        x,y = rxy()
        if y>=a and y<=b:
            break
    return x
```




```
def rxy():
    x = numpy.random.uniform(0,1)
    y = numpy.random.normal(loc=x*(x-1)+1/4, scale=.03)
    return (x,y)

def rx_given_yrange(a,b):
    while True:
        x,y = rxy()
        if y>=a and y<=b:
            break
    return x
```

Maths notation to describe the code:

$X \sim \text{Uniform}[0,1]$
$Y = X(X-1) + 1/4 + N(0,\ .03^2)$

$(X \mid Y \in [a,b])$

	$\boldsymbol{X}$	$(\boldsymbol{X} \mid \boldsymbol{Y} \in [\boldsymbol{a}, \boldsymbol{b}])$
code	`random.uniform(0,1)`	`rx_given_yrange(a,b)`
sample space	$[0,1]$	$[0,1]$
parameters	none	a, b
$\text{pdf}(x) = \Pr_X(x)$	1	Once we've found the cdf, it's easy: $\text{pdf}(x) = (d/dx)\ \text{cdf}(x)$
$\text{cdf}(x) = \mathbb{P}(X \le x)$	x	We can try to calculate $\mathbb{P}(\text{rx\_given\_yrange(a,b)} \le x)$. It's tricky, but do-able. It turns out to be equal to $\mathbb{P}(X \le x \mid Y \in [a,b])$. In fact, mathematicians use this to define "conditional random variable"—they start with a definition of the cdf, they don't start with code.

The pdf is written $\Pr_X(x \mid Y \in [a,b])$.
But it'd be more logical to write $\Pr_{(X \mid Y \in [a,b])}(x)$.

The elementary version of Bayes's rule doesn't work when there are continuous random variables. This version does:

Bayes's rule. For two random variables X and Y,

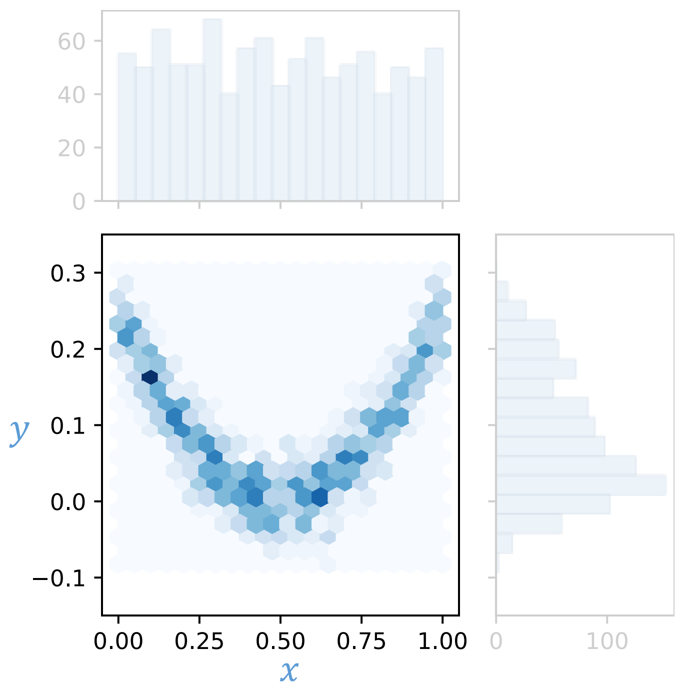
$$\Pr_X(x|Y=y) = \frac{\Pr_X(x)\Pr_Y(y|X=x)}{\Pr_Y(y)} \quad \text{when } \Pr_Y(y) > 0$$

$\Pr_X(y|X=x)$ is the density of a conditional random variable.

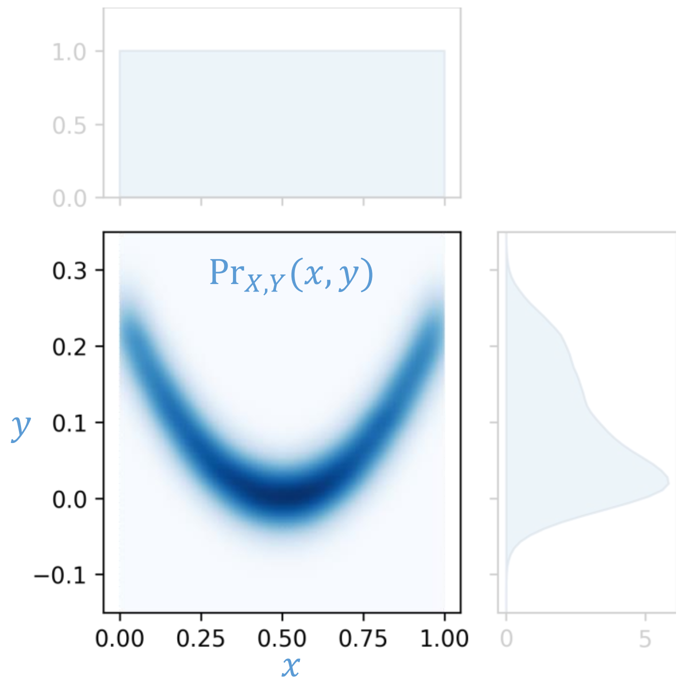
But what does it mean to condition on an event $\{X=x\}$ that has zero probability?

3.4. Random tuples

A pair of random variables (X, Y) can be described by their joint density, $\Pr_{X,Y}(x, y)$.



histogram of samples from `rxy()`



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

What *is* joint density?

p.43+44

FOR DISCRETE RANDOM VARIABLES

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

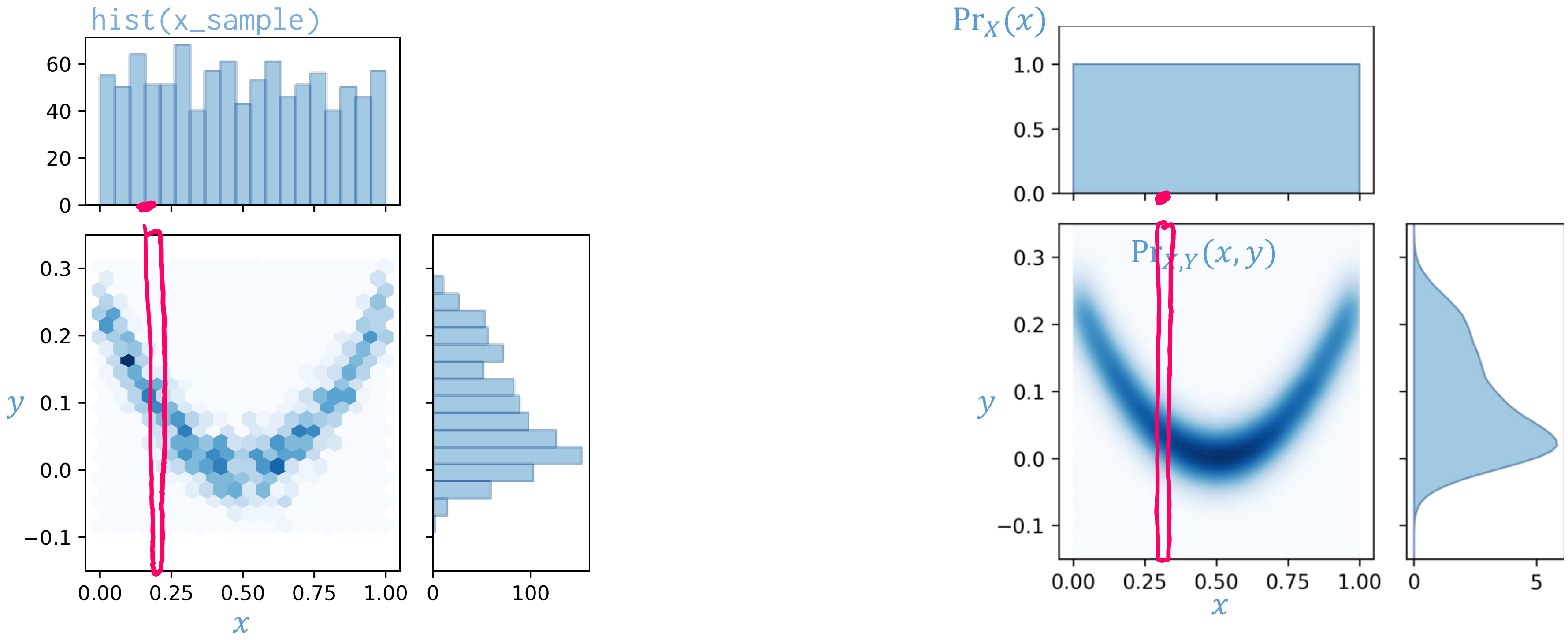
FOR CONTINUOUS RANDOM VARIABLES

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

$$\mathbb{P}\big((X, Y) \in A\big) = \int_{(x,y)\in A} \Pr_{X,Y}(x, y)\, dx\, dy$$

Marginalization

The joint density can be summarized to give the *marginal densities* $\Pr_X(x)$ and $\Pr_Y(y)$.



histogram of samples from `rxy()`

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

In code, we just pick out the marginal we want.

```
xy_sample = [rxy() for _ in range(1000)]
x_sample = [x for (x,y) in xy_sample]
```

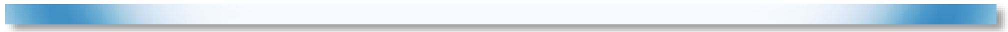
ie just ignore the y part of the sample

In maths, we integrate (or sum, for discrete r.v.)

$$\Pr_X(x) = \int_y \Pr_{X,Y}(x,y)\,dy$$

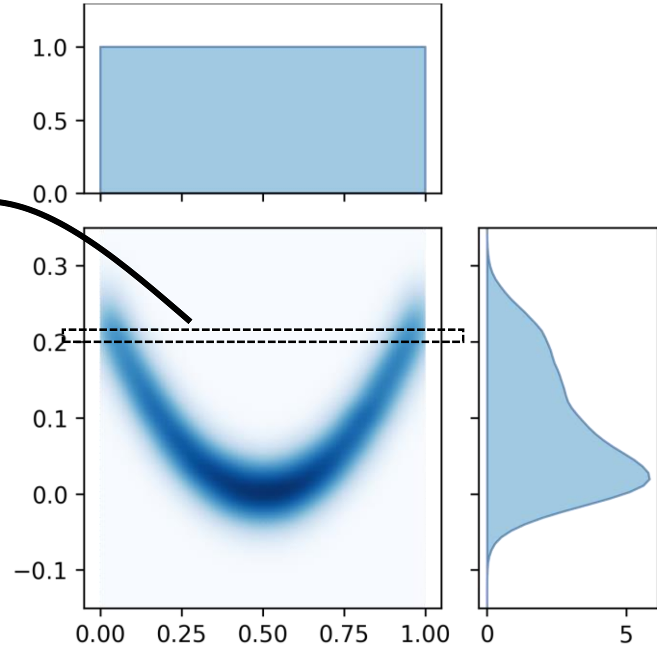
Conditional density

The conditional density of X given $\{Y = 0.2\}$ is $\quad \Pr_X(x \mid Y = 0.2) = \dfrac{\Pr_{X,Y}(x, 0.2)}{\Pr_Y(0.2)}$.



conditional density of $(X|Y = 0.2)$, written
$\Pr_X(x|Y = 0.2)$

We need to normalize it, so that this strip integrates to 1. (Otherwise it's not a valid density).



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

The correct normalization is clearly the marginal $\Pr_Y(0.2)$

In practice, how we find joint density?

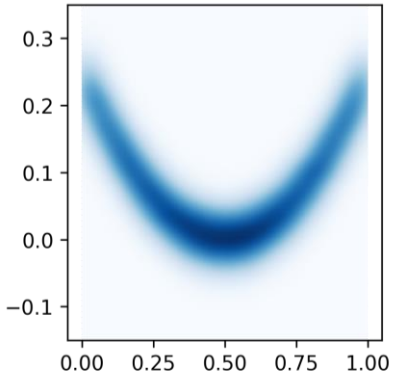
FOR INDEPENDENT RANDOM VARIABLES

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

FOR "TWO-STEP" RANDOM VARIABLES (i.e. "first generate X then use that value to generate Y")

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

and $\Pr_Y(y|X=x)$ is what we'd expect



```
x = numpy.random.uniform(0,1)
y = numpy.random.normal(loc=x*(x-1)+1/4, scale=.03)
```

Maths notation for Y: $Y \sim N\left(x(x-1)+\frac{1}{4},\ \sigma^2\right)$ where $\sigma = 0.03$

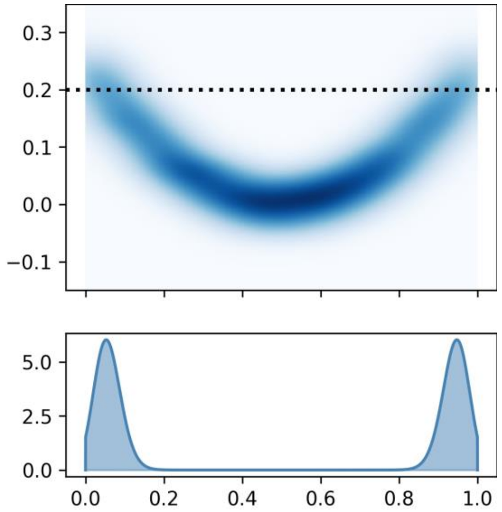
The joint density of (X,Y) is

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

$$= 1 \times \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\left(y-\left(x(x-1)+\frac{1}{4}\right)\right)^2/2\sigma^2}$$

Now we're equipped to answer the question —
if I'm told `y=0.2`, what can I deduce about `x`?

```
x = numpy.random.uniform(0,1)
y = numpy.random.normal(loc=x*(x-1)+1/4, scale=.03)
```



Use Bayes's rule:

$$Pr_X(x \mid Y=0.2) = \frac{Pr_{X,Y}(x,\ 0.2)}{Pr_Y(0.2)}$$ the definition of conditional density

a function of x

$$= \frac{Pr_X(x)\ Pr_Y(0.2 \mid X=x)}{Pr_Y(0.2)}$$ the joint density formula for a two-step random variable pair

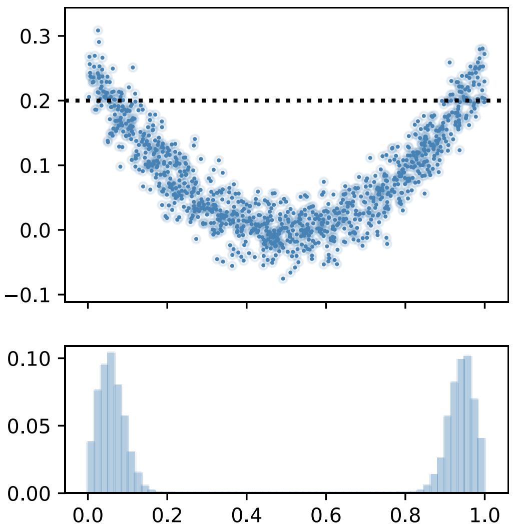
$$= \frac{1}{Pr_Y(0.2)} \times 1 \times \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-[0.2-(x(x-1)+1/4)]^2/2\sigma^2}$$

$$= K\, e^{-[x(x-1)+\frac{1}{4}-0.2]^2/2\sigma^2}$$

gather all terms that don't involve x into a constant factor.

a function of x

We could (with great effort) find K using the "densities sum to one" rule — K is whatever it has to be to make this density function integrate to 1, $K = \dfrac{1}{\int_x e^{-[\dots]^2/2\sigma^2}\,dx}$.



If we're told `y=0.2`, what can we deduce about `x`?

We can write down a formula for $\Pr_X(x|Y=0.2)$ using Bayes's rule.

Can we calculate the normalizing constant?
Can we calculate the cdf? expected value?
Can we sample from it?
…