EDS222 Week 6

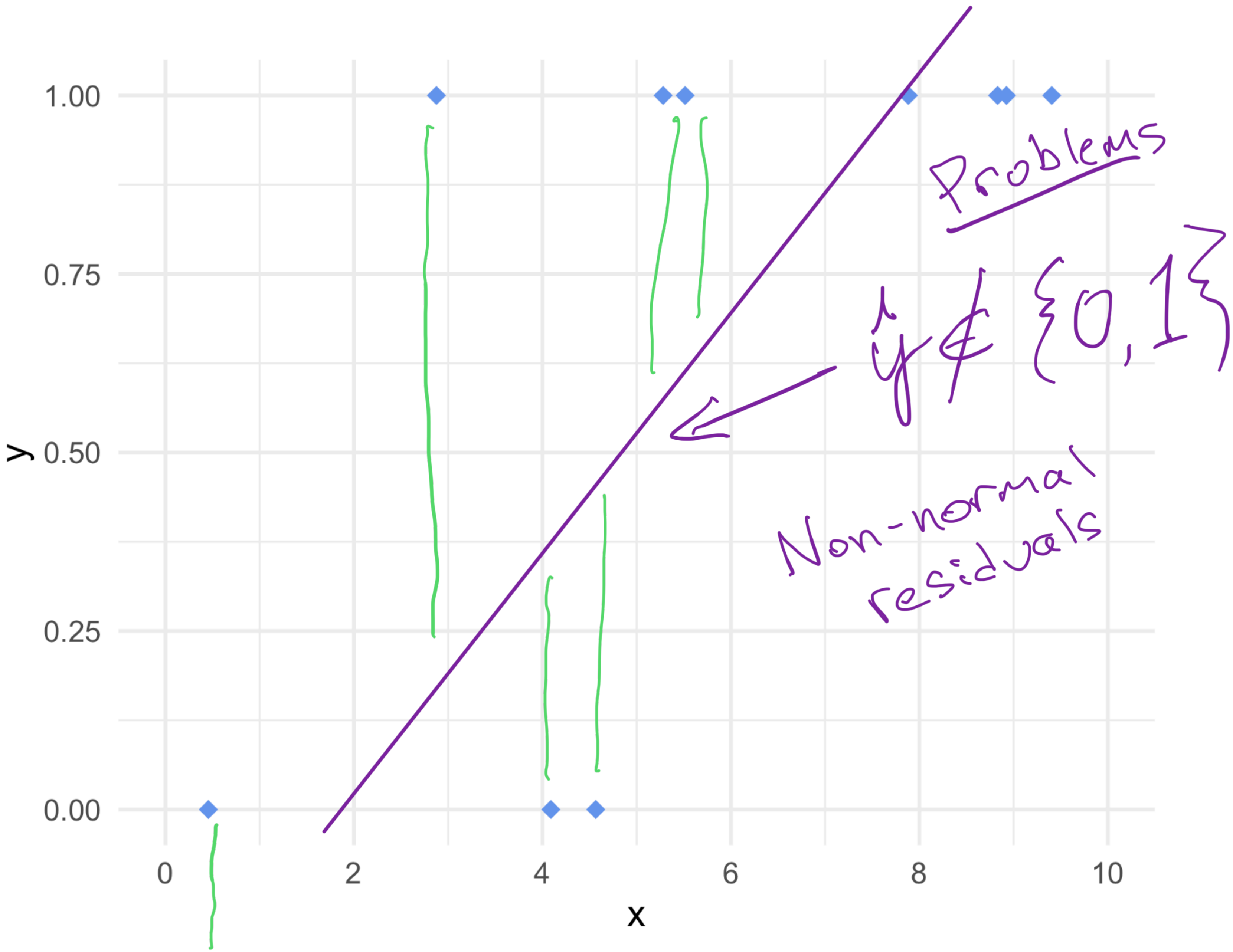
Modeling binary responses with *logistic regression*

what

how

November 5, 2024

Modeling the unobserved

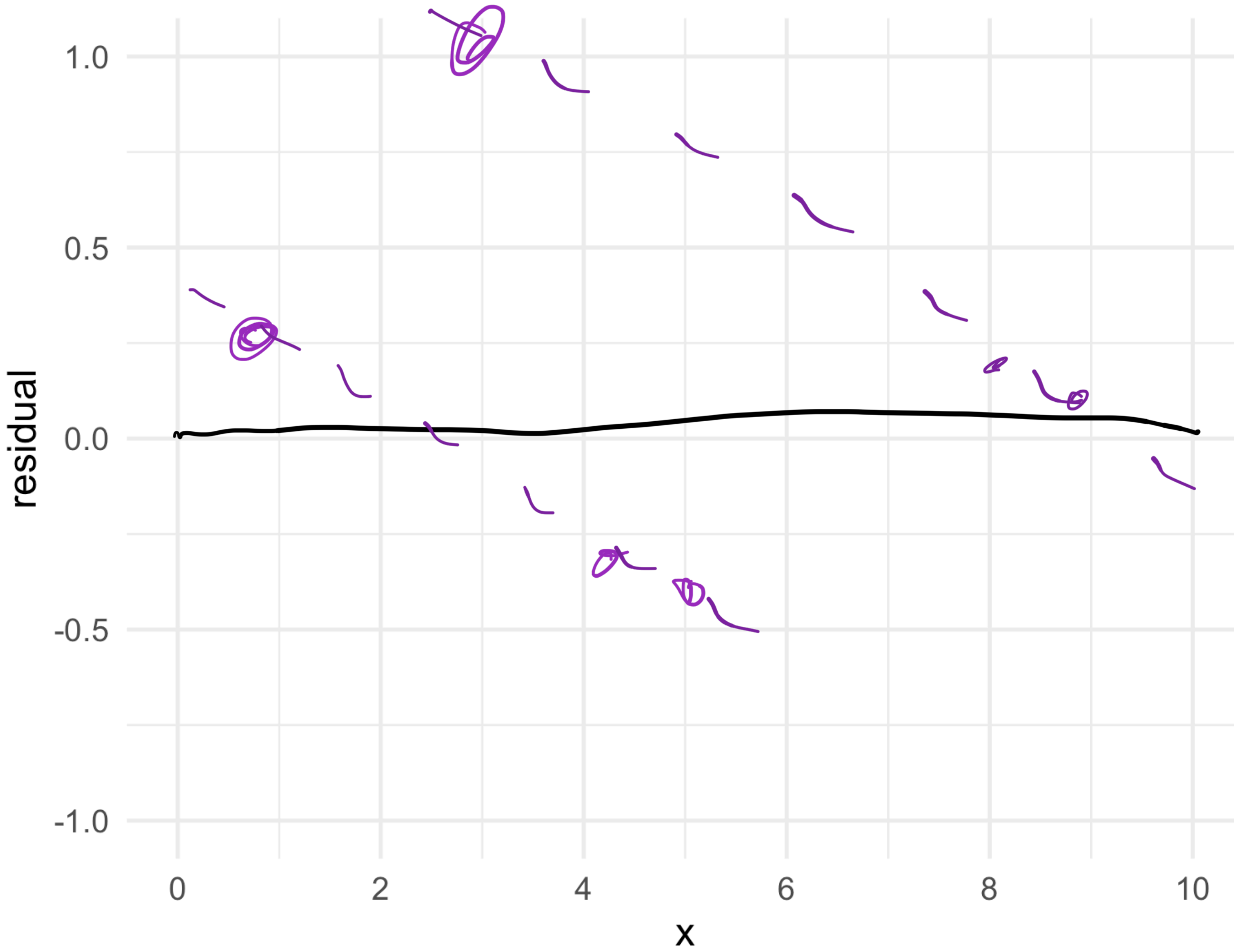


Modeling the unobserved

Key points

- Problem: OLS predictions are continuous, but our responses are 0's and 1's
- Another problem: the OLS assumption of normal errors is violated (see next slide)

Modeling the unobserved

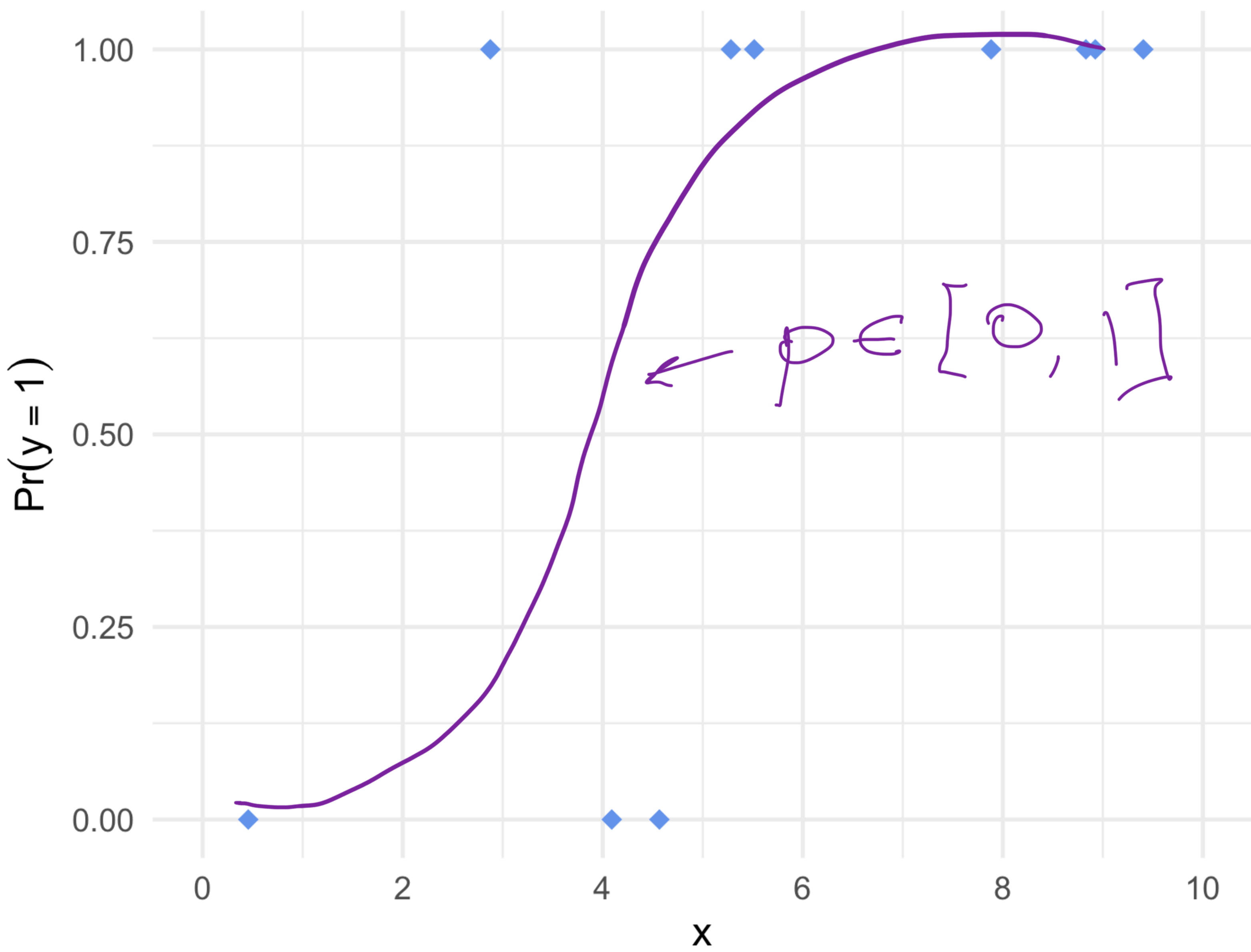


Modeling the unobserved

Key points

- The residuals fall along two parallel lines - definitely not normal
- If you call lm() you'll still get a line, it'll just be a bad line for these data. It's your responsibility to assess!

Modeling the unobserved

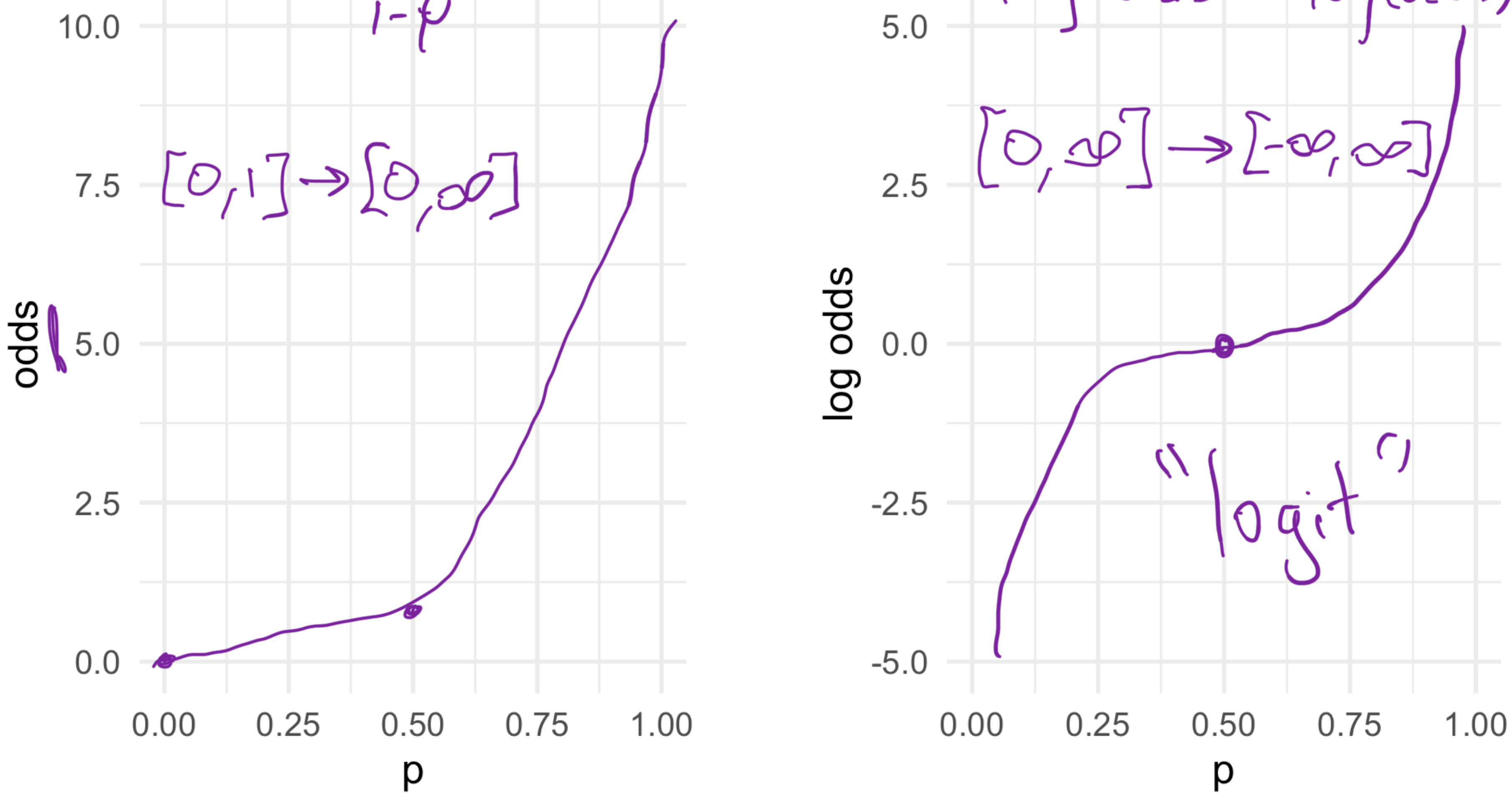


Link functions (logit)

Goal: $[0, 1] \to [-\infty, \infty]$

odds = $\frac{p}{1-p}$ success : failure

log odds = log(odds)



Link functions (logit)

Key points

- Goal: convert the range of probabilities [0,1] to all real numbers $[-\infty, \infty]$, so we can treat them as normal
- odds $= \dfrac{p}{1-p}$ i.e. the ratio of success to failure
    - odds $\in [0,\infty]$
    - All positive numbers - we're half way there
- log odds $= log(\text{odds}) = log(\dfrac{p}{1-p})$
    - log odds $\in [-\infty, \infty]$
    - All real numbers - we got it!
- We call log odds the "logit" transformation
    - The inverse of the "logit" is the "logistic", hence logistic regression

Link functions (logit)

Definitions

- Bernoulli(p)
  - A random variable that can take the value 0 or 1
  - p is the probability of the variable being 1
- ~
  - “Is distributed as”
  - Describes the distribution of a random variable
  - As opposed to =, which is an exact value

Link functions (logit)

Logistic regression:

$y \sim \text{Bernoulli}(p)$

$\text{logit}(p) = \beta_0 + \beta_1 x$

OLS ish

* There is still uncertainty in Bernoulli(p)

"Normal" regression:

$y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 x$

(we ignored σ)

$y = \beta_0 + \beta_1 x + u$

?

"is distributed as"

Link functions (logit)

Key points

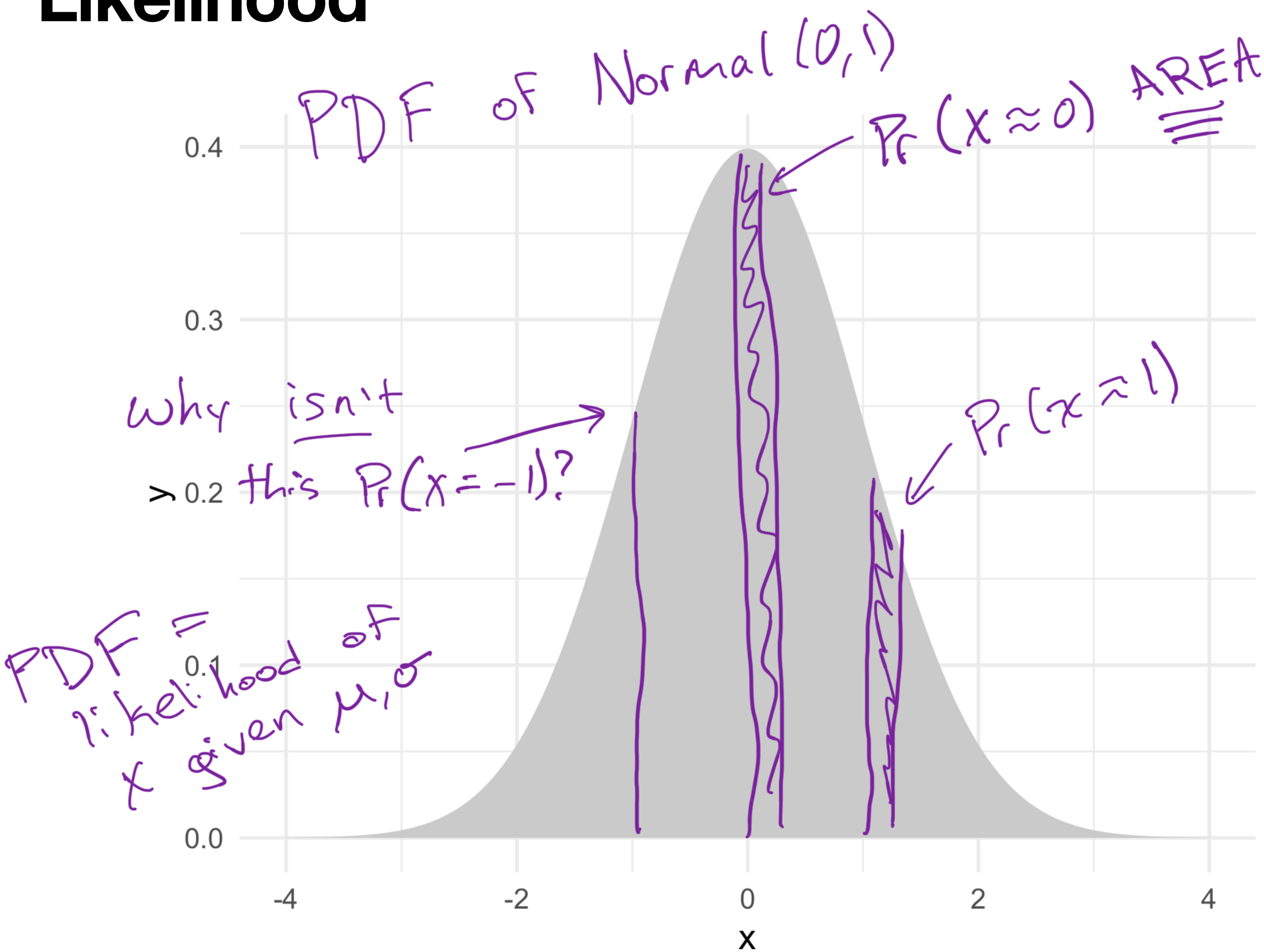
- Instead of modeling y directly, we model the *probability of y*
  - That part still looks pretty OLSish (after the logit transformation)
- This notation also describes the linear regression we've seen up until now, with a few changes
  - y *is distributed as* a normal variable with mean μ
  - No transformation of μ is necessary
- The uncertainty is still there even though we don't write it in the formula. It's implied by the "distributed as"

Likelihood

Definitions

- PDF
  - Probability density function
  - The *density* of probability for a random variable
  - Integrate it to get probability
- Likelihood
  - A quantitative measure of model fit
  - Has no direct interpretation in of itself
  - Useful for comparing models (e.g., different parameters)

Likelihood



PDF of Normal(0,1)
Pr(x ≈ 0) AREA
why isn't this Pr(x = -1)?
Pr(x ≈ 1)
PDF = likelihood of x given μ, σ
0.4
0.3
0.2
0.1
0.0
y
-4
-2
0
2
4
x

Likelihood

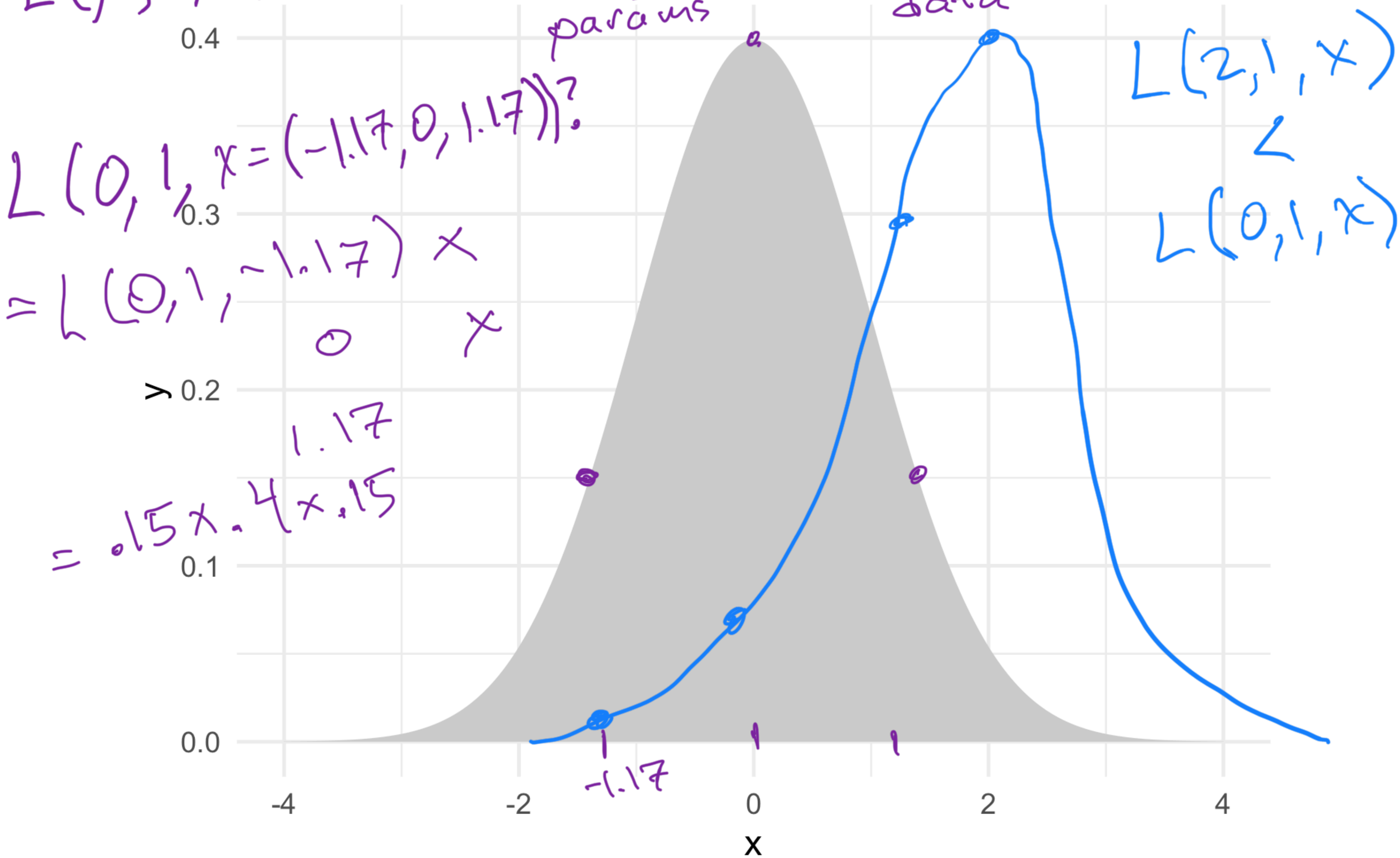
Key points

- Height of PDF tells us *how likely data are given parameters*
- The height of the PDF *is not* the probability of x taking a specific value!
  - Probability is the integral of the PDF
  - Area under the curve
  - A line has no width, so there's no area
- But the height of the PDF does tell us how likely the data are

Likelihood = PDF in reverse

Let $\mu = 2$

$L(\mu, \sigma, x) \equiv$ How likely $\underline{\mu, \sigma}$ (params) given $\underline{x}$ (data)



Likelihood

Key points

- Likelihood is the PDF in reverse
- *How likely are the parameters given the data*
- $$L(\mu, \sigma, x) = \prod_i PDF(x_i, \mu, \sigma)$$
  - The likelihood of our parameters (μ, σ) is the product of the PDF evaluated at the values of x

Likelihood

Key points

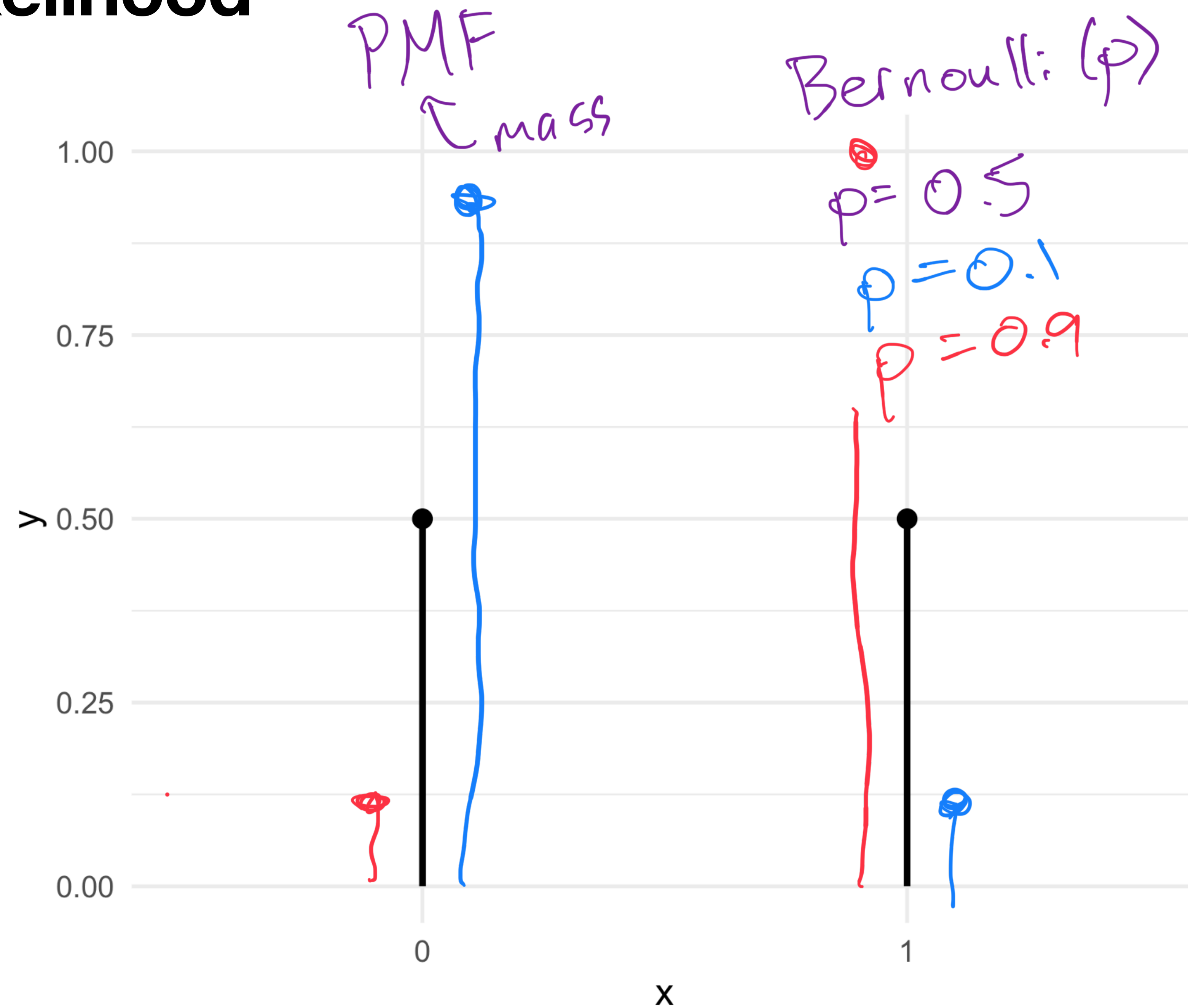
- For given data x, the likelihood L will change as we change the model parameters μ, σ
- That means there is a combination of parameters that gives us our most likely model
  - I.e. the *maximum likelihood* model

Likelihood

Definitions

- PMF
  - Probability mass function
  - Like a PDF, but for discrete variables
  - Because the variable is discrete, the height of the PMF is the probability that the variable takes that exact value

Likelihood

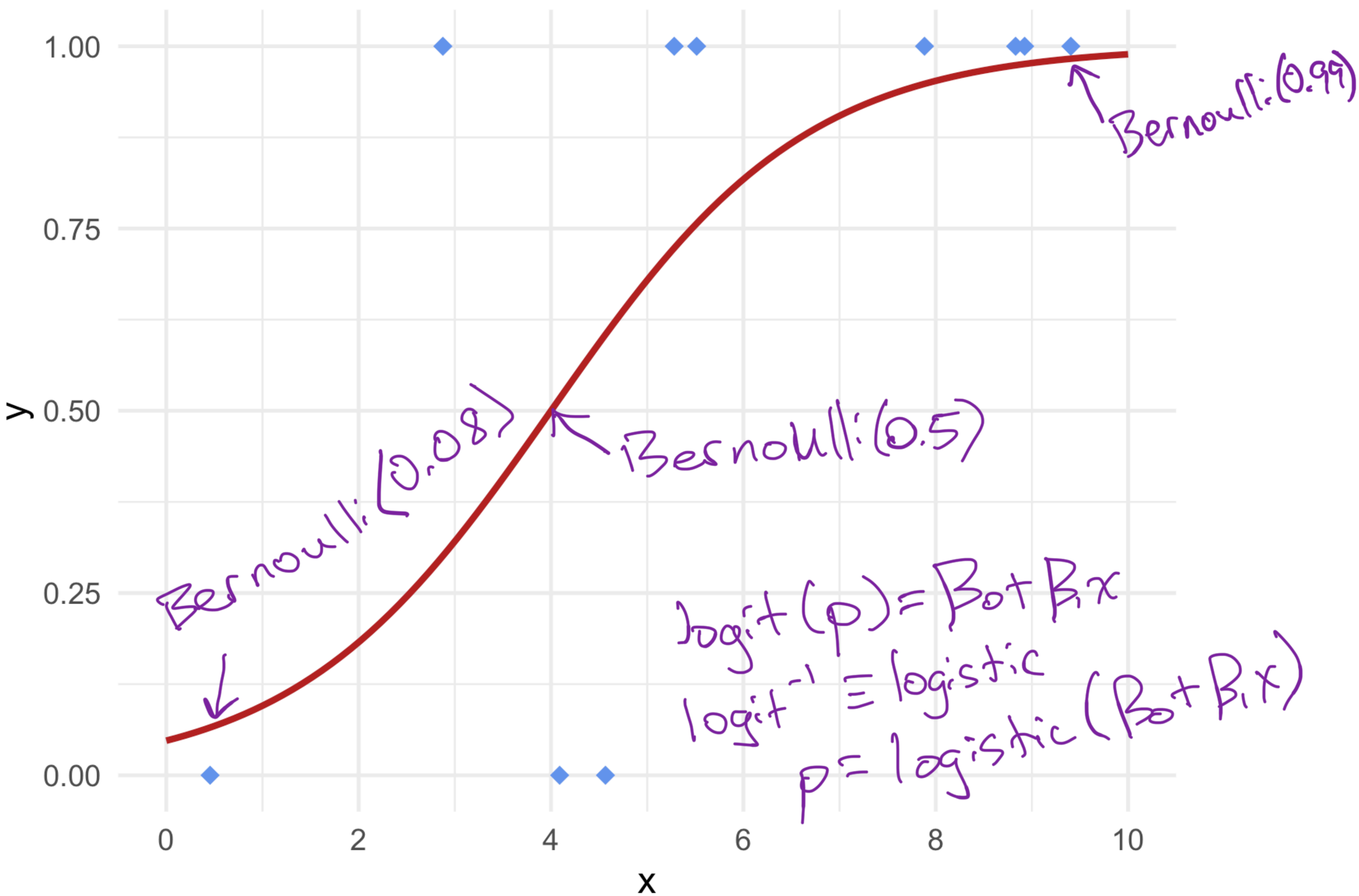


Likelihood

Key points

- The PMF of the Bernoulli has two peaks because a Bernoulli variable can be either 0 or 1
- In other words, given $y \sim Bernoulli(p)$:
  - The value of the PMF at y=1 is p
  - The value of the PMF at y=0 is 1-p

Likelihood



1.00
0.75
0.50
0.25
0.00
y
0
2
4
6
8
10
x
Bernoulli(0.99)
Bernoulli(0.5)
Bernoulli(0.08)
Bernoulli
logit(p) = β0 + β1x
logit⁻¹ = logistic
p = logistic(β0 + β1x)

Likelihood

Key points

- The logistic regression curve describes *how p changes with respect to x*
- The likelihood of our model:

$$y \sim Bernoulli(p)$$

$$logit(p) = \beta_0 + \beta_1 x$$

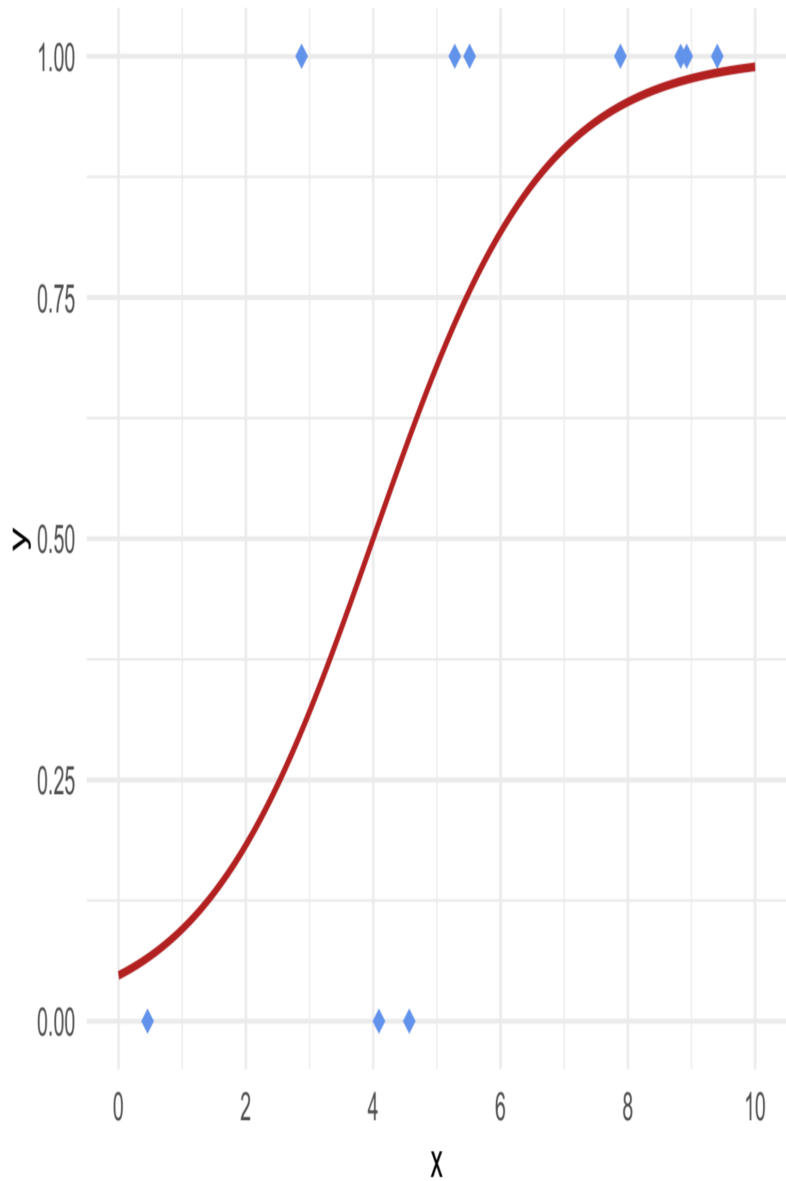
- Is therefore:

$$L(\beta_0, \beta_1, x) = \prod_i PMF(p_i, x_i)$$

$$PMF(p_i, x_i) = \begin{cases} p_i & x_i = 1 \\ 1 - p_i & x_i = 0 \end{cases}$$

- In other words, the likelihood goes up when *y* and *p* are "aligned" (y=1, p>0.5 OR y=0, p<0.5)
- Changing β_0, β_1 won't change *x* or *y*, but it will change *p*.

Coefficient estimation



The process for calculating likelihood is therefore:

1. Nominate some coefficients β_0, β_1
2. Calculate $logit(p) = \beta_0 + \beta_1 x$
3. Invert the logit to get p $p = logit^{-1}(logit(p))$
4. Get the PMF value for each point (based on *p* and *y*)
5. Multiply them all together to get the likelihood

You want the coefficients that give you the maximum likelihood.

Coefficient estimation

Live coding example

Review

1. **Modeling the unobserved**
   Model the *underlying probability*, not the data directly
2. **Link functions**
   Use a *link function* (logit) to transform the parameters of a non-normal distribution (Bernoulli)
3. **Coefficient estimation**
   Say goodbye to SSE, embrace the power of *likelihood* for coefficient estimation