EDS222 Week 6

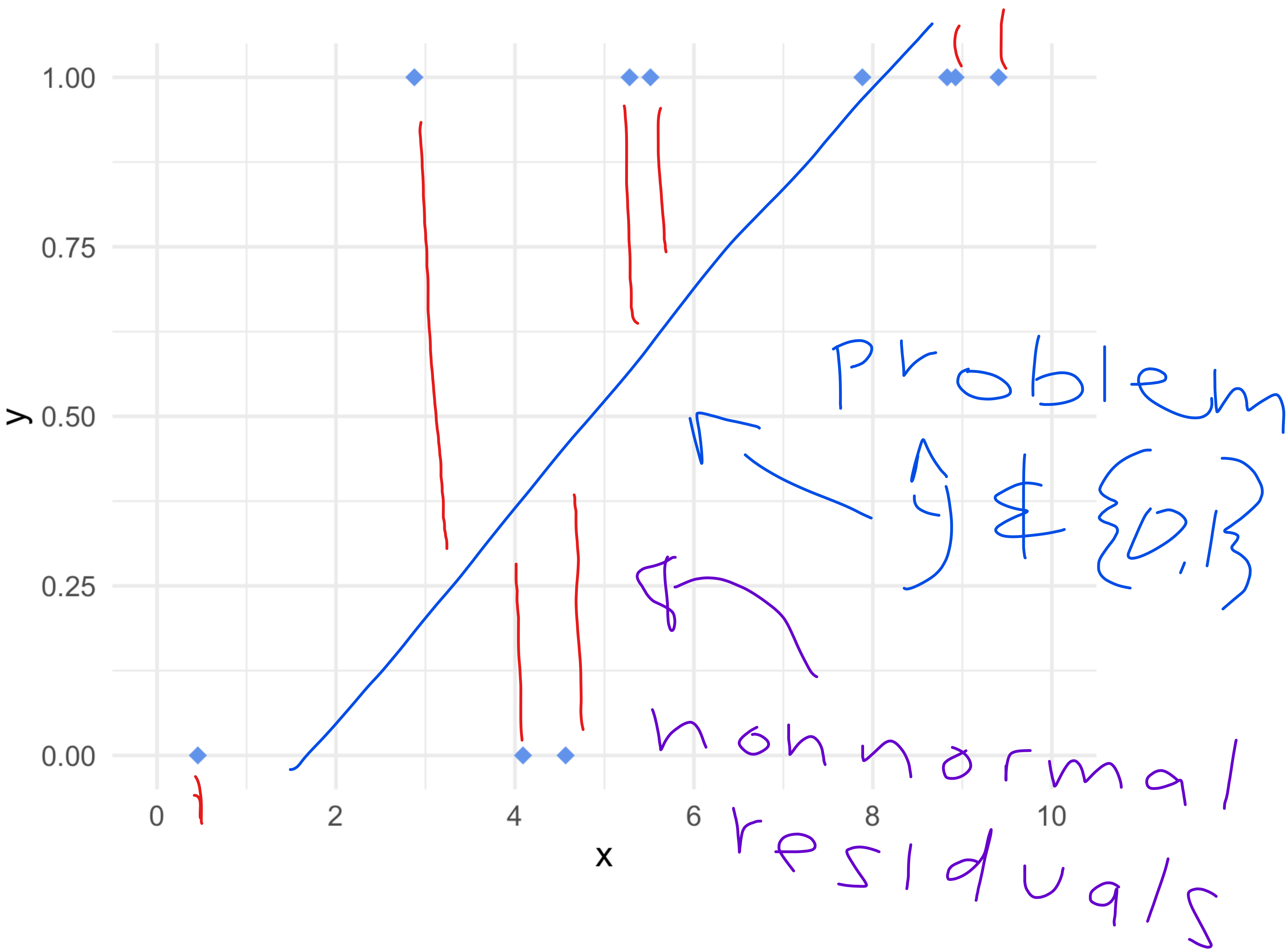
Modeling binary responses with *logistic regression*

what

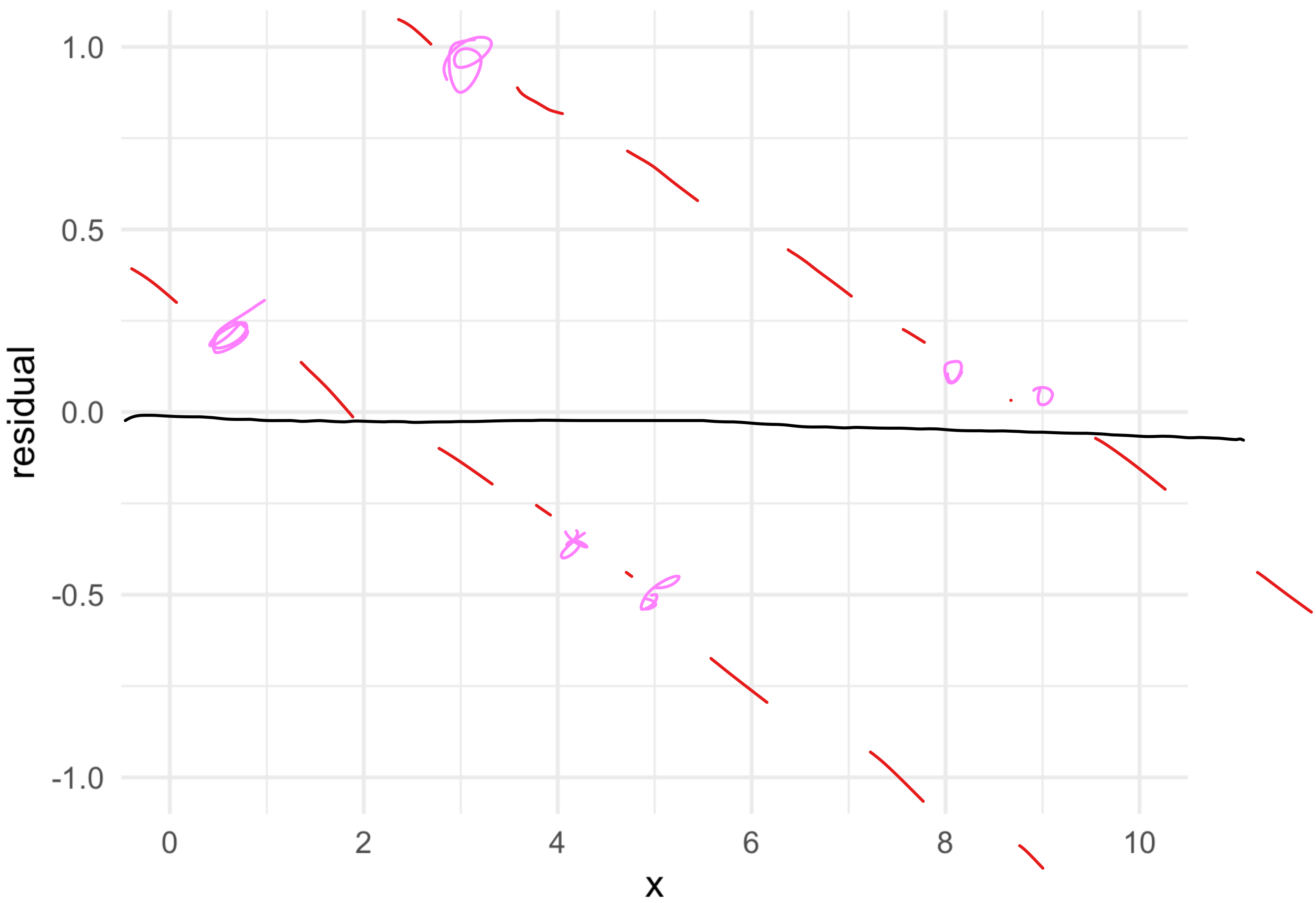
how

November 5, 2024

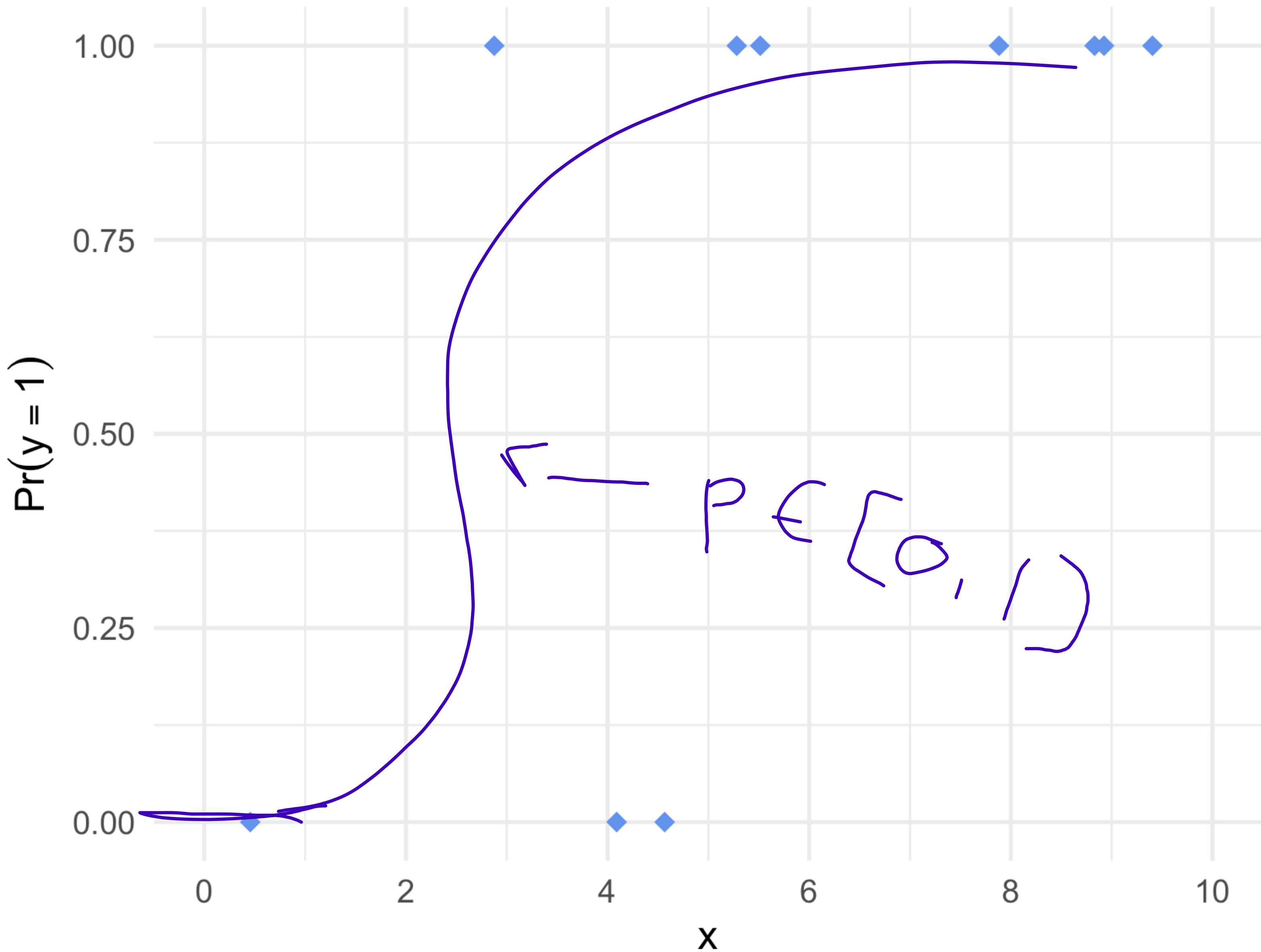
Modeling the unobserved



Modeling the unobserved



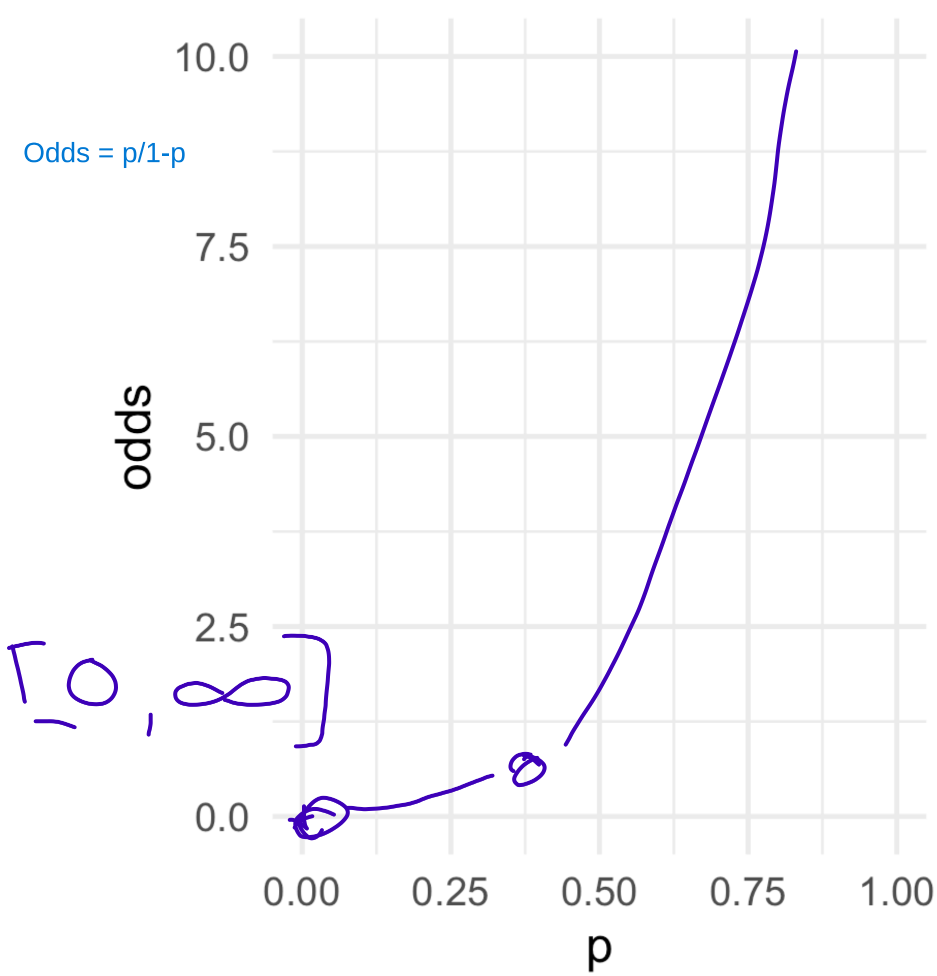
Modeling the unobserved

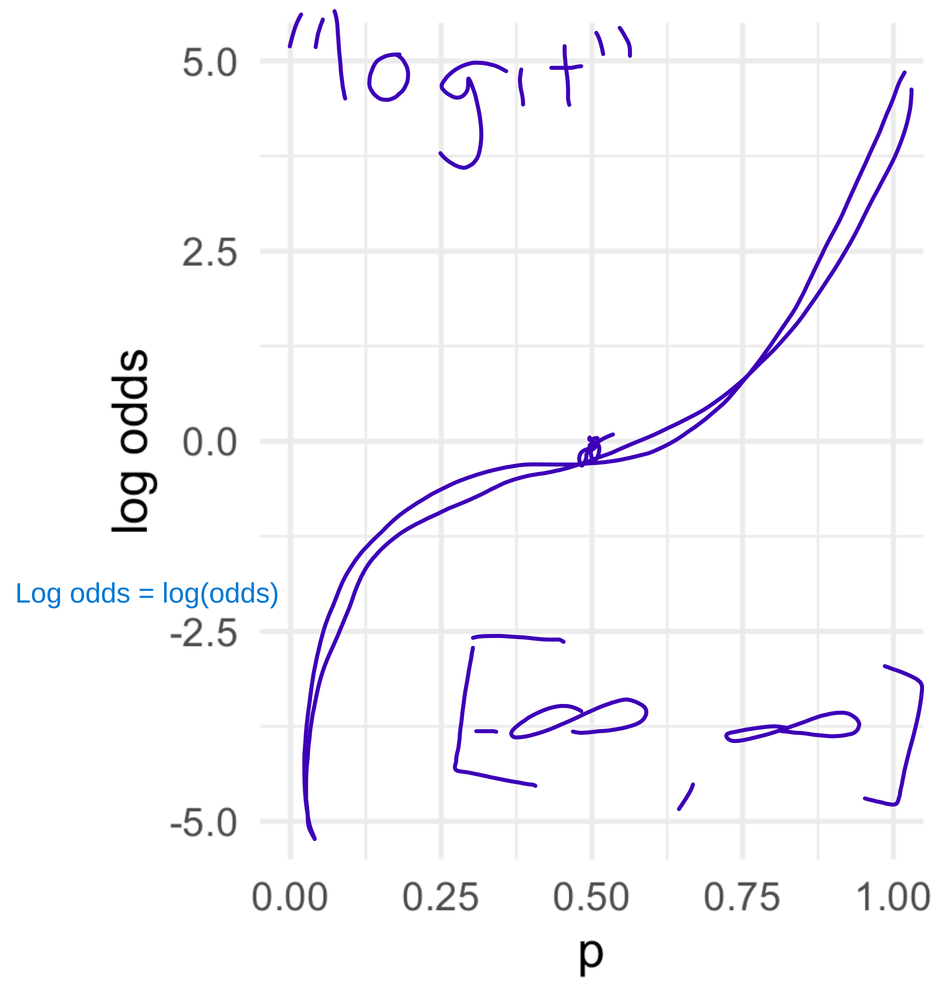


Property that the line satisfies (that ols was violating) is that it stay between zero and one

Link functions (logit)

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





Link functions (logit)

Logistic regression:

Y ~ Bernoulli (p)

Logit(p) = B0 + B1x OLS-ish

*there is still uncertainty in Bernoulli (p)

“Normal” regression:

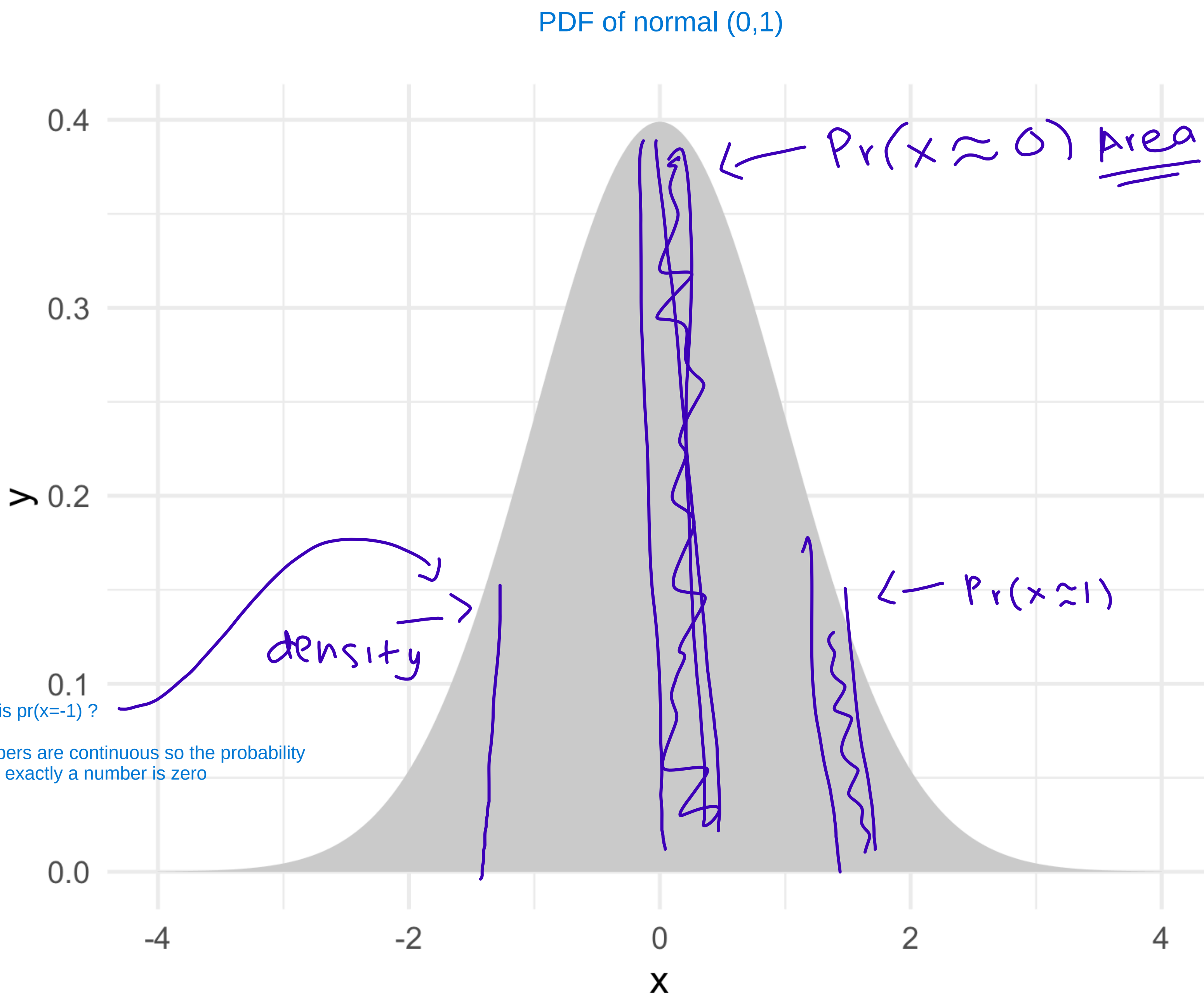
Y ~ normal (mu, Sigma)
Mu = B0 + B1x

(we ignored sigma)

Y = B0 + B1x + u

"Is distributed as"

Likelihood



Why isn't this pr(x=-1) ?

These numbers are continuous so the probability
that you get exactly a number is zero

Likelihood = PDF in reverse

$L(\mu, \sigma, x)$ = how likely $\underbrace{\mu, \sigma}_{\text{params.}}$ given $\underbrace{x}_{\text{data}}$

Let $\mu = 2$

$L(0,1,x=(-1.17,0,1.17))$?

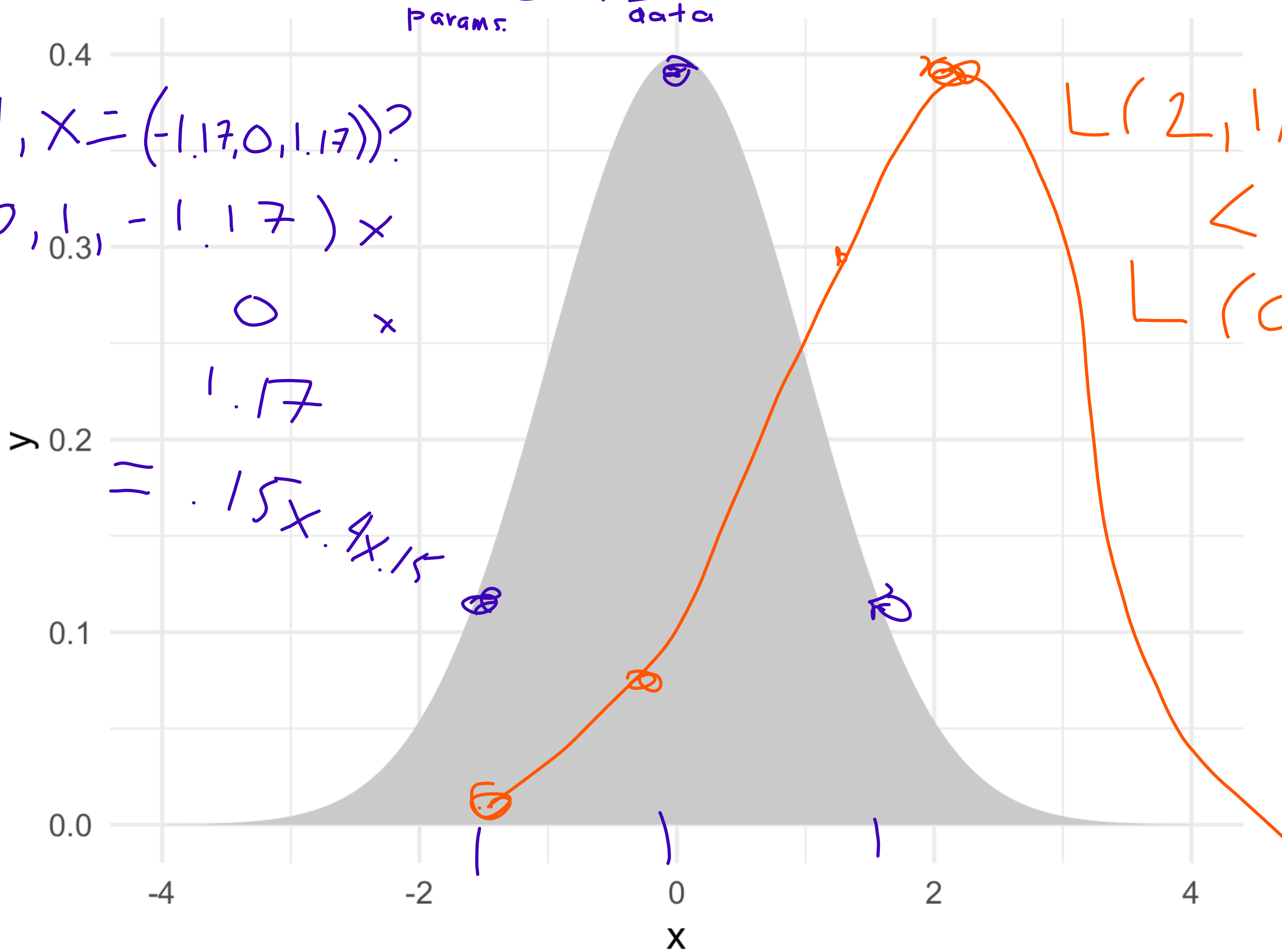
$= L(0,1,-1.17) \times$

$0 \times$

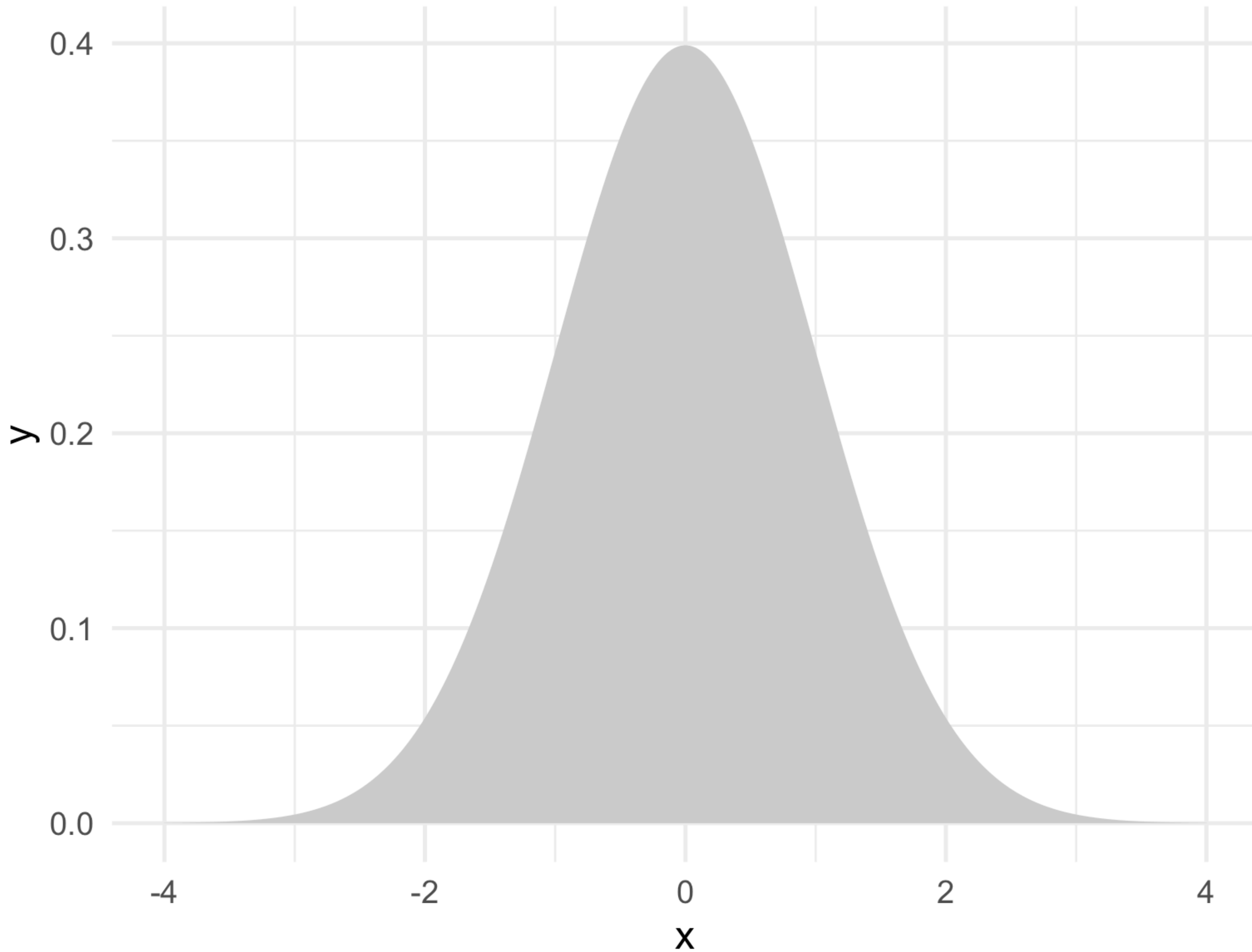
1.17

$= .15 \times .4 \times .15$

$L(2,1,X) < L(0,1,X)$



Likelihood



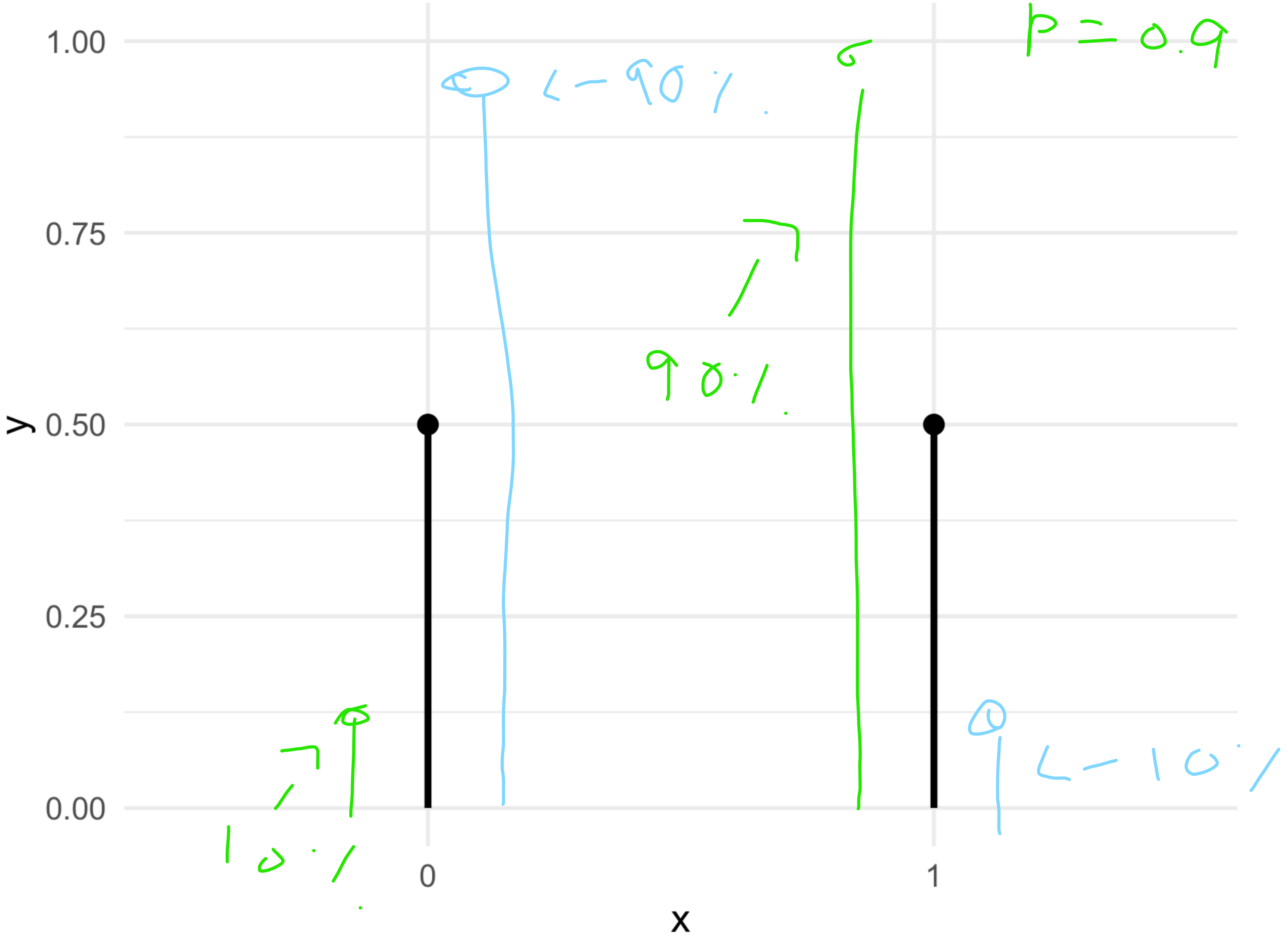
Likelihood

PMF - probability mass function

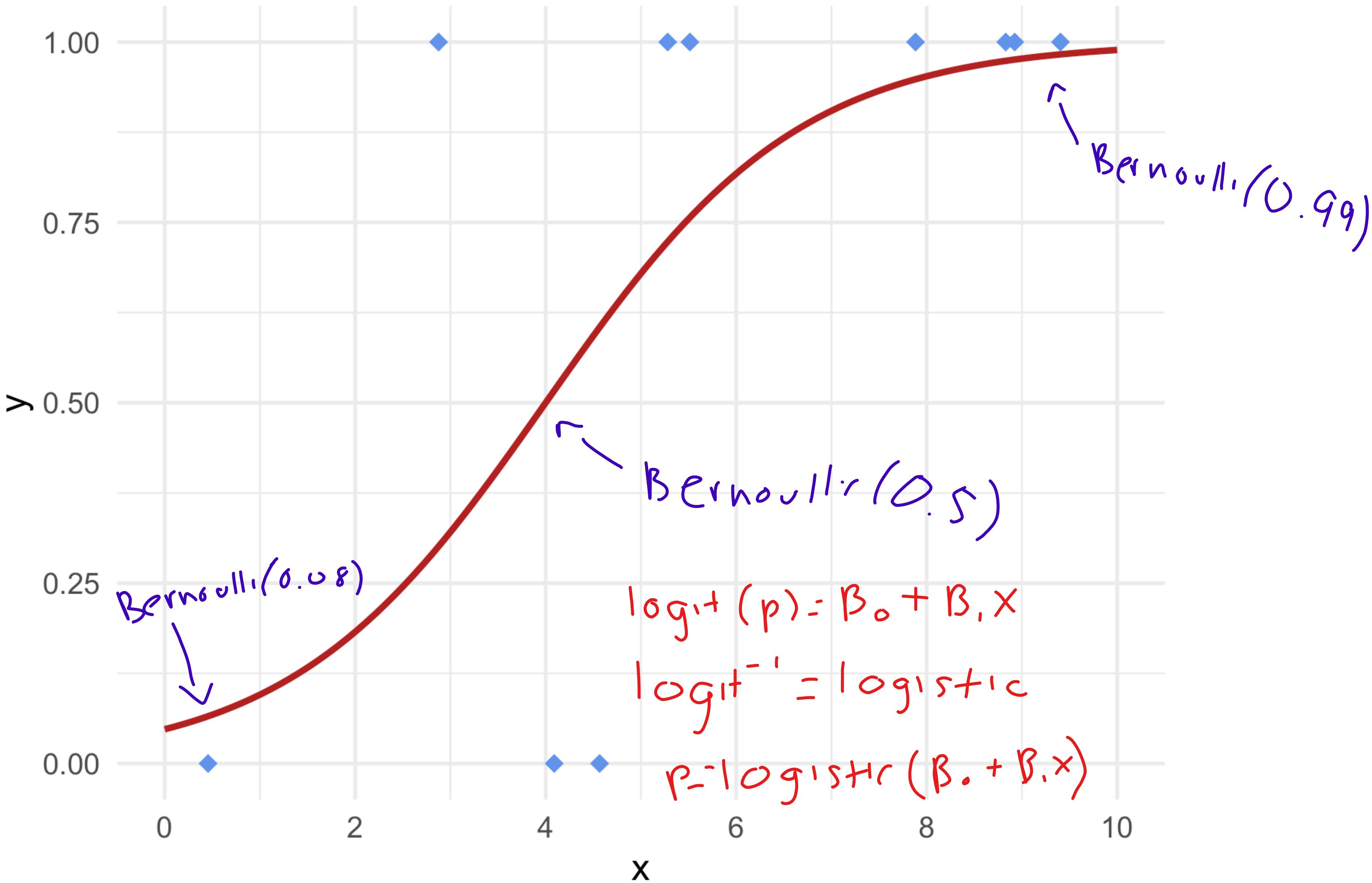
Bernoulli(p)
p = 0.5

p = 0.1

p = 0.9

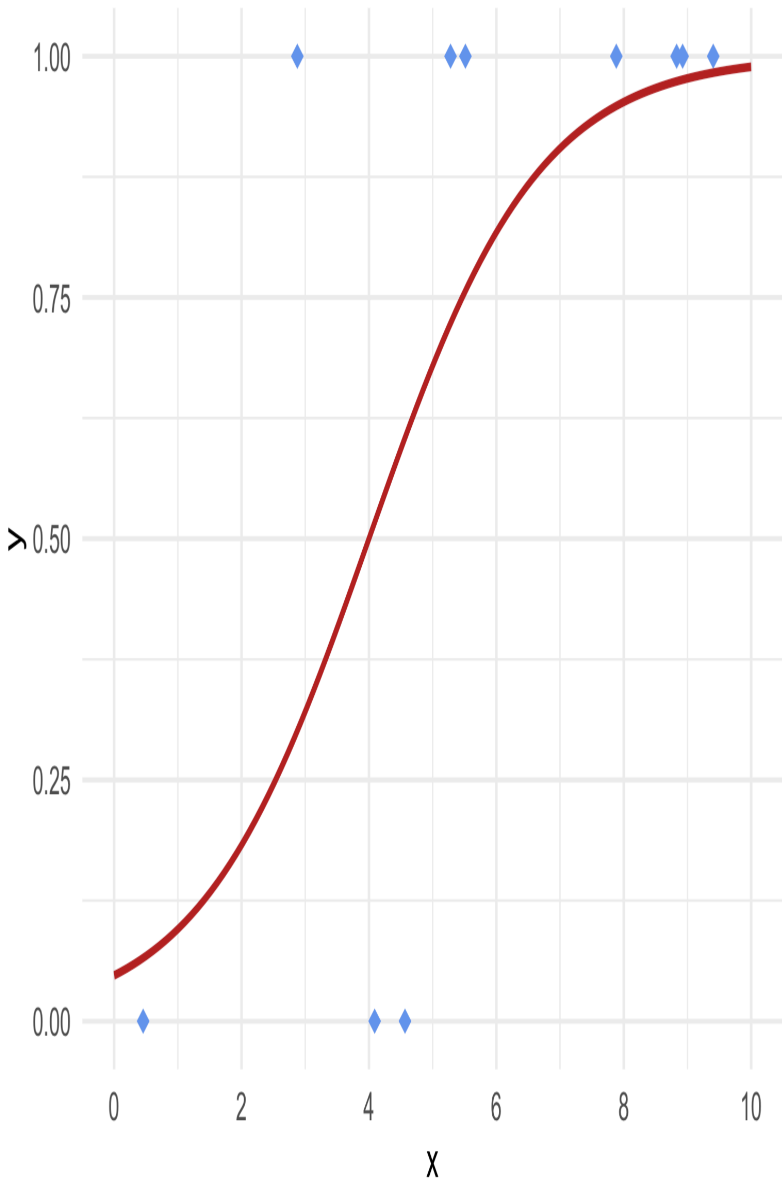


Likelihood



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)
logit(p) = $\beta_0 + \beta_1 X$
logit^{-1} = logistic
p = logistic($\beta_0 + \beta_1 x$)

Coefficient estimation



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