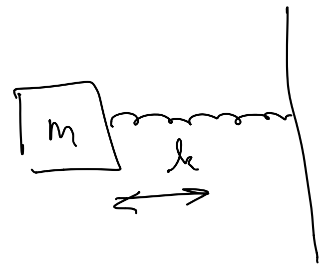
Models and Likelihood

Statistical Models

- Mathematical models
- Statistical models
  - Parameters
  - Likelihood

mathematical model that incorporates random behavior



motion governed by Newton laws

"fit model" to the situation.

First example: coin flipping

- Model a coin flipping experiment as a Bernoulli random variable with parameter p.
- Flip the coin 100 times and get 55 heads and 45 tails.

what is p?

$$L = \binom{100}{55} p^{55}(1-p)^{45}$$

what choice of p makes our result most likely.

- **Maximum Likelihood -** forget the constant as it doesn't effect the result.

$$\frac{dL}{dp} = 55\cancel{p^{54}}(1-p)^{45} - 45p^{55}(1-p)^{44} = 0$$

1 − p p

yields

$$\boxed{55(1-p) = 45p}$$

$$\frac{p}{1-p} = \frac{55}{45}$$

or

$$p = 55/100 = .55$$

$$p = 55/100 = .55$$

Independent normally distributed errors

- Back to our temperature model. We assume that the errors in our measurements are normally distributed around zero. There is one parameter: the variance σ^2 in our density function for a single measurement

$$\boxed{p_\sigma(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-x^2/(2\sigma^2)}}$$

- We make n independent measurements of temperature

controlled : true temp is known

$$x_1, \ldots, x_n$$

What does this tell us about σ^2? The likelihood for independent measurements is the density

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\|x\|^2/(2\sigma^2)}$$

- Maximize the density at this point. Use the *log-likelihood* as it is easier.

$$\log P(x) = -n\log\sigma - \frac{\|x\|^2}{2\sigma^2} + C$$

$$\frac{1}{2}\sigma^{-2} = -\sigma^{-3}$$

- Take the derivative and set it equal to zero.

$$\frac{dP}{d\sigma} = -\frac{n}{\sigma} + \frac{\|x\|^2}{\sigma^3} = 0$$

$$\Rightarrow \boxed{\sigma^2 = \frac{\|x\|^2}{n}}$$

MSE

$$\|x\|^2 = \sum_{i=1}^{n} (t - t_0)^2$$

M

- The *maximum likelihood estimate of the variance is the mean squared error!*

Linear Regression and likelihood

- Model says that our N data points (x_i, y_i) arose from a process

$$y = mx + b + \epsilon$$

Linear Model.

where ϵ is a normally distributed error term with variance σ^2.

- How should we set m, b, σ to make the observed data most likely?
- The density function is

$$P(m, b, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N e^{-\sum_{i=1}^N (y_i - mx_i - b)^2/(2\sigma^2)}$$

$\varepsilon = y - mx - b$

measured (x_i, y_i)

$\varepsilon_i = y_i - mx_i - b.$

- The *log likelihood* is

$$\log P = -N \log \sigma - \frac{1}{2\sigma^2} \underbrace{\sum_i (y_i - mx_i - b)^2}_{\|x\|^2} + C$$

- The derivatives with respect to m and b give the least squares estimates.

$$\frac{\partial \log P}{\partial m} = \frac{\partial}{\partial m} \sum (y_i - mx_i - b)^2 \quad \text{up to constants, etc}$$

$$\frac{\partial \log P}{\partial b} = \frac{\partial}{\partial b} \sum (y_i - mx_i - b)^2 \quad \text{up to constant...}$$

m, b are given the OLS solution.

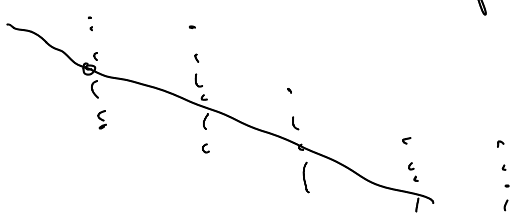
- The derivative with respect to σ gives the best estimate when σ^2 is the mean squared error.

$$\frac{\partial \log P}{\partial \sigma} = -\frac{N}{\sigma} + \frac{\sum (y_i - mx_i - b)^2}{\sigma^3} = 0$$

$$\sigma^2 = \frac{1}{N} \sum (y_i - mx_i - b)^2 = MSE(m, b).$$

- **Ordinary Least Squares is the maximum likelihood solution assuming independent normally distributed errors**

OLS: variance doesn't depend on x.



if errors don't have this property:
variance changes with x
heteroscedasticity