Models and Likelihood

Statistical Models

- Mathematical models
- Statistical models
  - Parameters
  - Likelihood

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.

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

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

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

yields

$$55(1-p) = 45p$$

or

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

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

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

- We make n independent measurements of temperature

$$x_1, \ldots, x_n$$

  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.

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

- Take the derivative and set it equal to zero.

- 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

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

  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

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

- The *log likelihood* is

  $$\log P = -N \log \sigma - \frac{1}{2\sigma^2} \sum_i (y_i - mx_i - b)^2$$

- The derivatives with respect to m and b give the least squares estimates.
- The derivative with respect to σ gives the best estimate when σ^2 is the mean squared error.
- **Ordinary Least Squares is the maximum likelihood solution assuming independent normally distributed errors**