Prediction, Parameter Estimation, and Sampling Distributions

February 23, 2018

Setup

- obtain data $(X_1, \dots, X_n)$
- hypothesize some model $f(x|\theta)$ for data
- estimate parameters $\widehat{\theta}$ in a model, e.g. maximum likelihood

Critical Questions:

- How do we assess model fit $\widehat{\theta}$?
- Is the model "correct"?

Model Goal: Prediction

For some problems, the main goal for our model is to predict future / unlabeled observations.

Some Examples:

- spam classifier: classify email as spam / not spam
- face recognition pipeline: detect faces in images

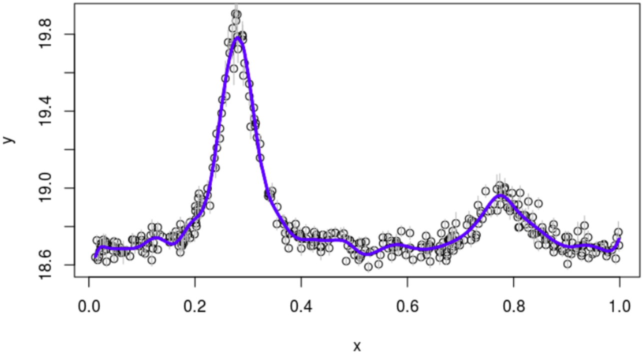
  https://jakevdp.github.io/PythonDataScienceHandbook/05.14-image-features.html
- NBA Point Spread: predict point spread for game so we can place bets

Model Assessment

- training / test data: estimate parameters using the training set. estimate performance on test set.
- cross–validation: divide data set into K sets and sequentially treat 1 set as the test set and the other $K - 1$ as training (useful when data sets not particularly large)

Important Point: The parameters in the proposed probabilistic model (even if it exists) could be wrong or uninterpretable, even when the model has good predictive performance.

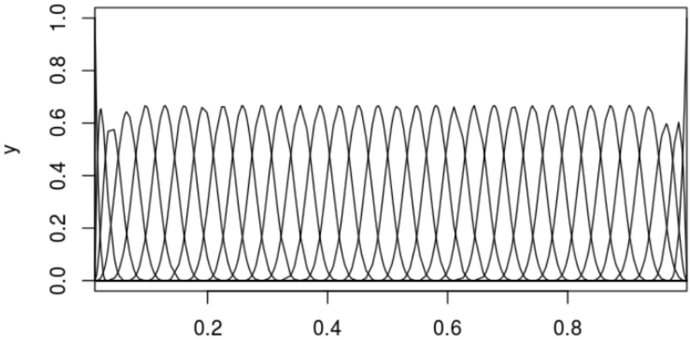
Example: Splines for Light Curve



Example: Splines Parameters and Basis

```
In [27]: ## regression splines
         N <- 30
         X <- bs(x,knots=(1:N)/(N+1),intercept=TRUE)
         coeffs <- matrix(lm.fit(X,y)$coefficients,ncol=1)
         preds <- X%*%coeffs
         print(t(coeffs))

            [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]
[1,] 18.64454 18.70615 18.68133 18.68579 18.68143 18.77694 18.65843 18.85829
           [,9]    [,10]    [,11]    [,12]   [,13]    [,14]    [,15]    [,16]
[1,] 18.8509 19.66277 19.93496 19.02237 18.8697 18.72123 18.73745 18.72066
           [,17]    [,18]    [,19]    [,20]    [,21]    [,22] [,23]   [,24]
[1,] 18.74385 18.63056 18.68058 18.71834 18.67976 18.68043 18.71 18.7545
           [,25]    [,26]    [,27]    [,28]    [,29]    [,30]   [,31]    [,32]
[1,] 18.85646 19.01132 18.86971 18.80584 18.72999 18.68257 18.7129 18.68144
           [,33]    [,34]
[1,] 18.65714 18.73399
```



Example: Splines for Light Curve

- the spline model may do a good job of predicting future observations
- but model parameters are not very interpretable, do not easily connect with any science

Model Goal: Parameter Estimation

Some Examples:

- Does drug increase survival? Data are death indicator (response either 0 or 1) and predictors (age, severity of disease, etc.) as well as received drug or placebo. Might use logistic model. One of the θ parameters corresponds to increase in survival rate for individuals taking drug.
- Fraction of voters who support policy? Proportion is some number $p \in [0, 1]$. Surveys are conducted to estimate p.
- What are masses of binary pair of stars in light curve? The masses are parameters in some physical model which is fit to the data, possibly using MLE. Even if the spline model fits better (in terms of predicting future observations) it will probably be less useful than the physical model for estimating pair masses.

Model Validity

- Is the model a good fit for the data?
- If not, purported interpretation of parameter may be invalid.
- This is more of a problem when the goal is parameter estimation, than for prediction.
- Methods for assessing goodness–of–fit are often fairly model specific. (often hypothesis testing)
- We do not discuss these issues in depth in this course.

Parameter Uncertainty

- Estimate parameters with some method, such as maximum likelihood.
- The estimates are not the exact true values.
- **Frequentist** statistics assesses uncertainty via the sampling distribution of the estimator.
- **Rough Idea:** If we repeat the data collection process over and over again, how much spread is there in the estimator around the truth.

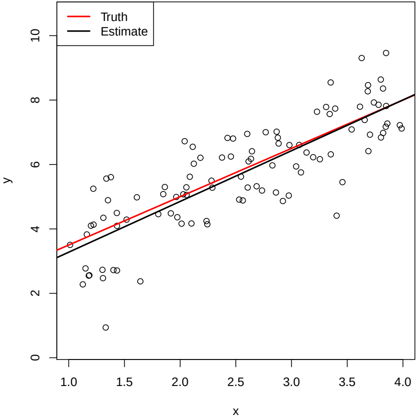
Example:

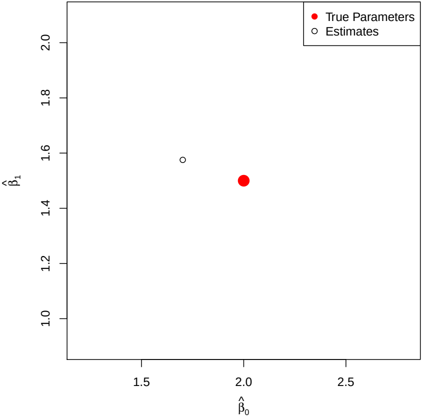
$$Y = X\beta + \epsilon$$

where $\epsilon \sim N(0, \sigma^2 I)$ then the maximum likelihood estimator for β is

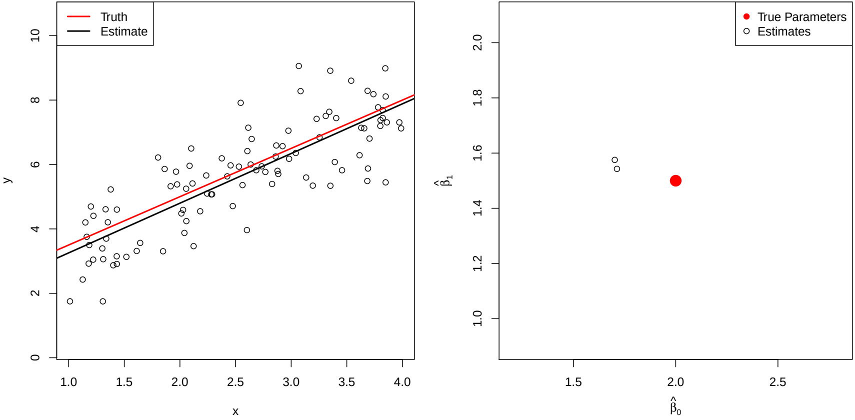
$$\widehat{\beta} = (X^T X)^{-1} X^T Y.$$

Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

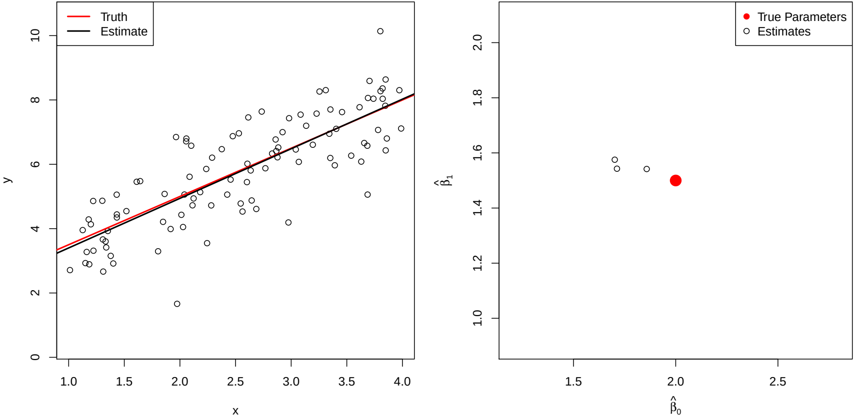




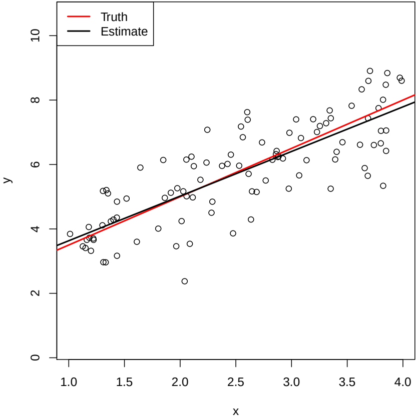
Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

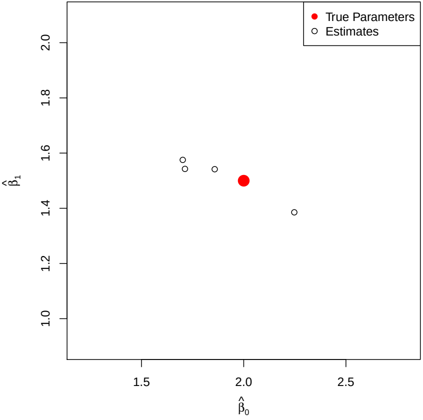


Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

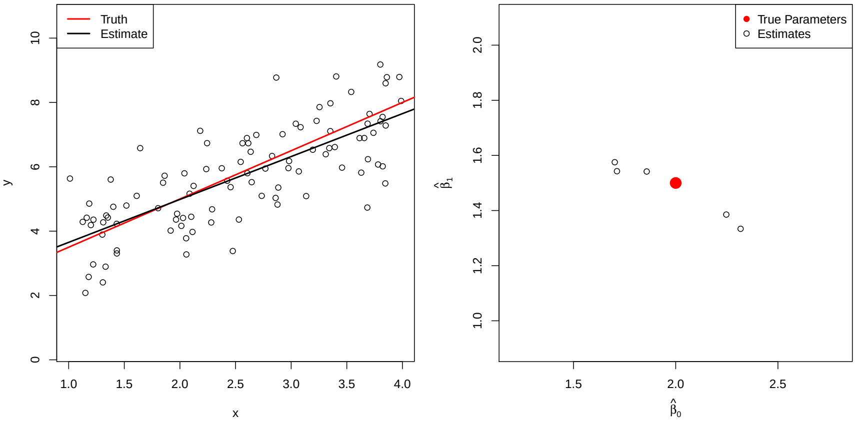


Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

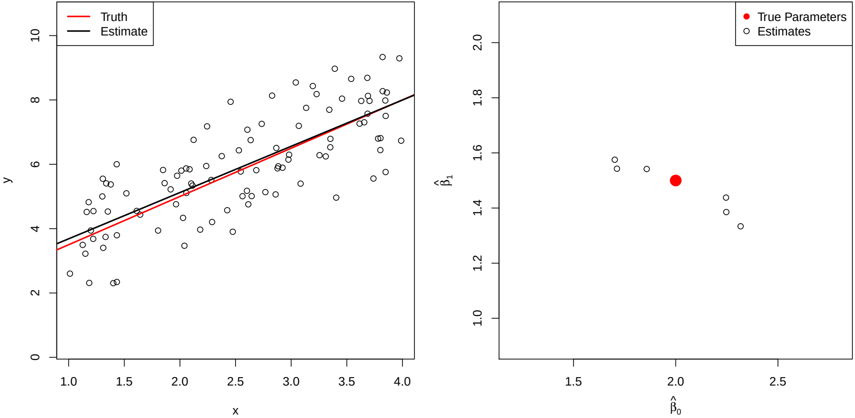




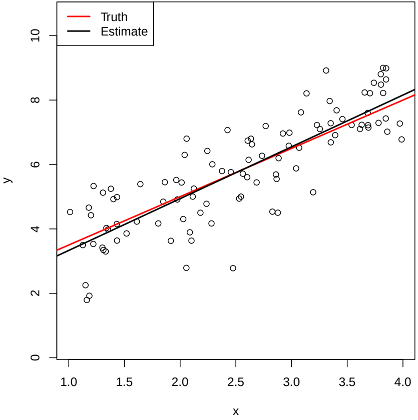
Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

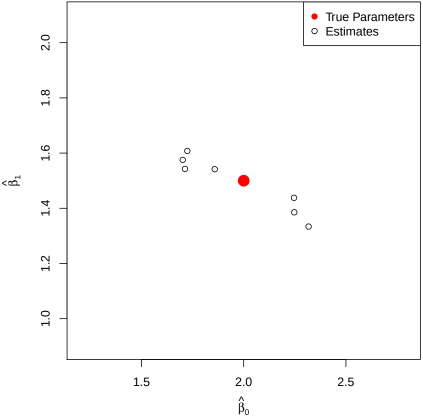


Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

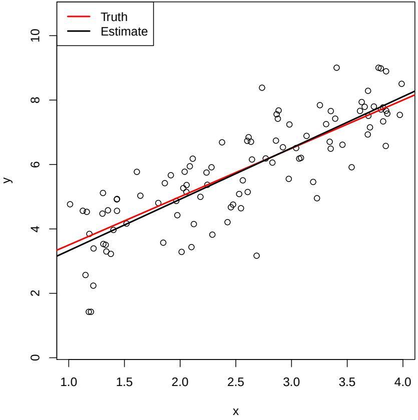


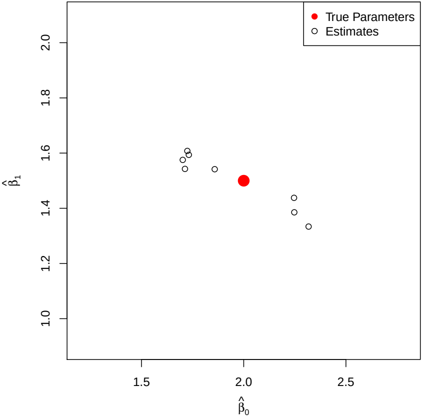
Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$



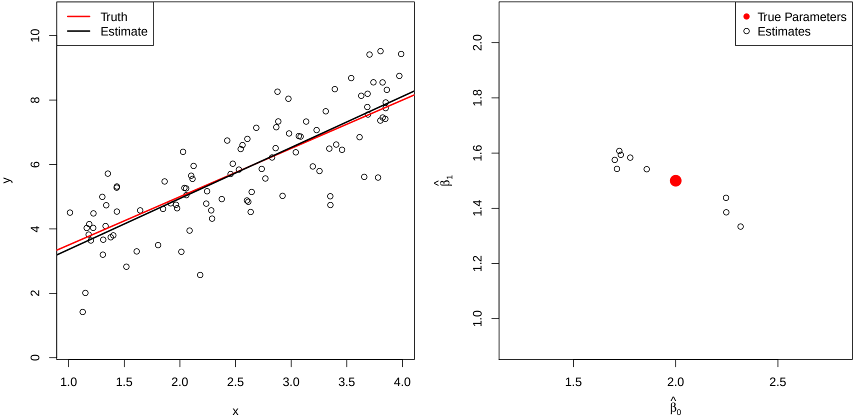


Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

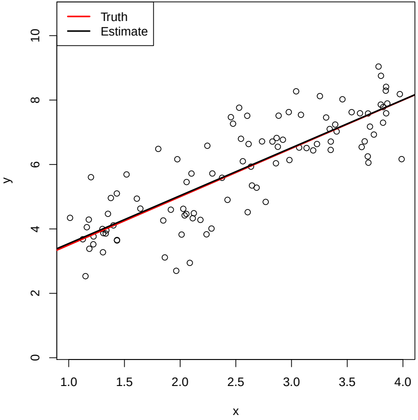


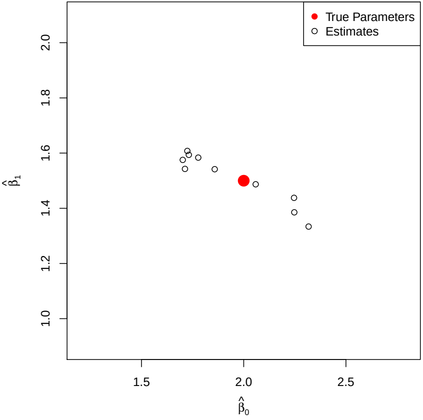


Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$



Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

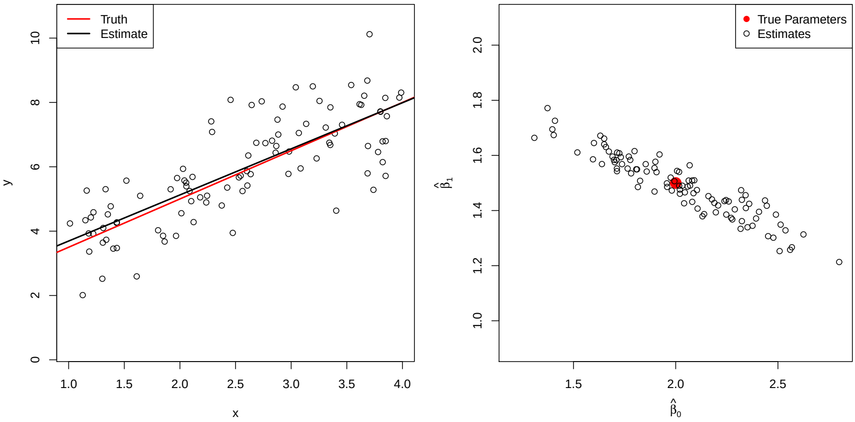




Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$

Repeat 89 more times.

Example: $\beta = (2, 1.5)^T$, $\sigma^2 = 1$



Sampling Distribution

- In practice we only see one estimate.
- The variance of the estimates around the truth is the variance of the MLE.
- Surprisingly we can estimate this variance from the likelihood function of a single sample.
- This estimate is (very conveniently) a by product of Newton's optimization method.