Generalized Linear Models (GLMs/GLIMs)

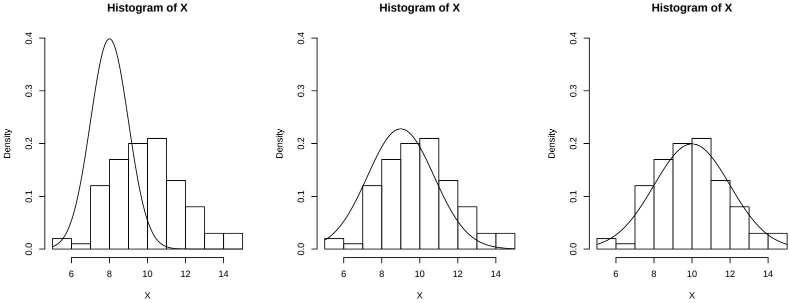
MATH 420/620

Friday, 23 Feb, 2018

Maximum Likelihood Estimators

If $\hat{\theta}$ maximizes the likelihood, we call it a Maximum Likelihood Estimator (MLE). The Cramer-Rao Inequality tells us MLEs are *minimum variance estimators* but MLEs are often *biased*.

Intuition: Suppose you estimate the mean μ and variance σ^2 of a Normal distribution by plotting a histogram, choose an initial μ and σ and overlay the corresponding density curve, then iteratively adjust μ and σ until it looks like a good match.

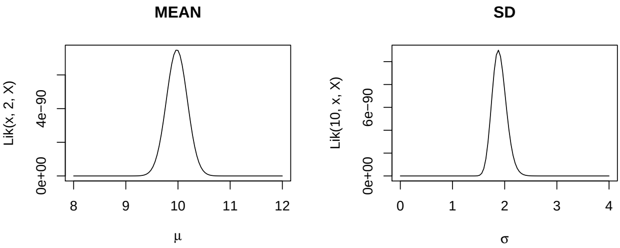


MLE Example

For iid Normal data ($f(x_i; \mu, \sigma)$) the joint density $f_X(\mathbf{x}; \mu, \sigma) = \prod_{i=1}^{n} f(x_i, \mu, \sigma)$ defines the likelihood function for parmeters μ and σ,

$$\mathcal{L}(\mu, \sigma; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma).$$

Plotting likelihood values over a range of possible parameter values (here holding one parameter constant while varying the other) in **R** yields...



The MLEs for μ and σ are the pair of values that yield the maximum likelihood value. In this case, using the `optim()` function yields $\mu = 9.98$ and $\sigma = 1.88$.

Likelihood Examples

Q: What is a plausible likelihood function for the following data set where an unfair coin was tossed 14 times and the number of outcomes that were heads was counted: $y = (4, 6, 6, 7, 7, 9, 10, 8, 2, 6)$?

Q: What is the likelihood function for a data set of tree heights?

Q: How would you determine a reasonable likelihood function for the distances from UNR to all of the Starbucks stores in Nevada?

Other statistical frameworks?

1. (Ordinary) Least Squares and Weighted Least Squares
2. Regression Models: SLR, MLR, GLM, GAM,
3. Bayesian Statistics

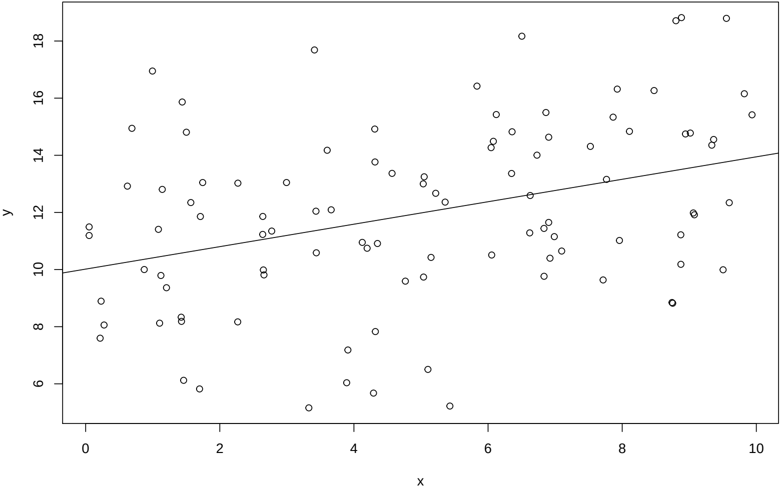
Simple and Multiple Linear Regression

```
x=runif(100,0,10); y=rnorm(100,x/2+9,3);
fit=lm(y~x)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9287 -2.1336  0.2398  1.9411  6.5422
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.0159     0.5858  17.098  < 2e-16 ***
## x             0.3931     0.1014   3.874 0.000193 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.965 on 98 degrees of freedom
## Multiple R-squared:  0.1328, Adjusted R-squared:  0.124
## F-statistic: 15.01 on 1 and 98 DF,  p-value: 0.0001931
```

Simple and Multiple Linear Regression

```
plot(x,y); abline(fit)
```



Simple and Multiple Linear Regression

```
x1=runif(100,0,10); x2=runif(100,100,200); y2=rnorm(100,x1/2+x2/10+17,2);
fit=lm(y2~x1+x2)
summary(fit)
```

```
##
## Call:
## lm(formula = y2 ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8992 -1.1548  0.1905  1.3867  4.3104
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.425932   1.132055  13.626  < 2e-16 ***
## x1           0.501901   0.068033   7.377 5.51e-11 ***
## x2           0.110914   0.006888  16.103  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.862 on 97 degrees of freedom
## Multiple R-squared:  0.7644, Adjusted R-squared:  0.7595
## F-statistic: 157.3 on 2 and 97 DF,  p-value: < 2.2e-16
```

GLM Model Framework

The GLM is described by three components:

1. The *random component* specifies the conditional distribution of $y_i|\vec{x}_i$ and is typically a member of the exponential family of distributions (Normal, binomial, Poisson, Negative-binomial, etc.) but other distributions are possible.
2. We call our linear sum of predictors the *linear predictor*, and denote it as $\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$
3. We call the transformation that links the expected response values $\mu_i = E(y_i|\vec{x})$ and the linear predictor η_i the *link function*: $g(\mu_i) = \eta_i$. This link function is assumed to be smooth (differentiable) and invertible. It's inverse g^{-1} is often called the *mean function* since $\mu_i = g^{-1}(\eta_i)$.

GLM vs MLR

Recall the MLR model with untransformed response values can be written as

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j \, x_{ij} + \epsilon_i .$$

In that case, we model $E(g(y_i))$ as a linear sum of x_i values, and further assume Normal errors with constant variance.

Consider, for now, the simple untransformed case (i.e., g is the identity function).

GLM vs MLR

One could pose the MLR model as a GLM (not to be confused with a General Linear Model) as follows:

1. The *random component* is Normally distributed.
2. The *linear predictor* is $\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$ (nothing new here!)
3. The *link function* is the identity function: $g(E(y_i)) = E(y_i) = \eta_i$

GLM vs MLR

Note that transforming Y values under MLR is different than specifying a non-identity link function!

GLMs model $g(E(y_i))$, the transformed expecation of the response, using the linear predictor. This gives more flexibility to apply linearizing transformations without affecting the distribution about that trend. For example, compare the two models by comparing y_i values and inverse-transforms:

$$\text{MLR: } y_i = g^{-1}(\eta_i + \epsilon_i)$$

$$\text{GLM: } y_i = g^{-1}(\eta_i) + \epsilon_i$$

This distinction often makes GLMs preferrable over MLR.

Parameter Estimation, etc.

Parameter estimation is done via Maximum Liklihood, and most of the diagnostics for multiple linear regresion carry over to GLMs.

For more information, please see Ch. 15 of *Applied Regression Analysis & Generalized Linear Models* by John Fox. http://www.sagepub.com/sites/default/files/upm-binaries/21121_Chapter_15.pdf

Example: Logistic Regression (Sheather, Ch. 8)

A common form of response data are counts of a particular type of outcome among *m* trials. For example, the number of individuals in a sample with a specific genotype. In such cases, the data are best modeled using a binomial distribution, not a Normal distribution, using *logistic regression*.

$$E(Y|x) \sim binom(m, p)$$

Example: Logistic Regression

Here the parameter of interest is p – the probability of a *success* on each of our m trials. Since m is known and not a parameter that needs to be estimated, the goal is to estimate p as a function of our *linear predictor*. In logistic regression, this is done by assuming a logit link function,

$$E(Y|X) = m\,p = m\frac{\exp(\eta_i)}{1+\exp(\eta_i)} = \frac{m}{1+\exp(-\eta_i)}$$

Thus, a little algebra gives that

$$\eta_i = \log\left(\frac{p(x)}{1-p(x)}\right)$$

We call the right side of that equation the *logit* function, and $p(x)/(1-p(x))$ the *odds*.

Example: Logistic Regression

To see how this can be cast as a GLM, note that:

1. The distribution is binomial.
2. The relationship between the mean (let's use $\mathrm{E}(y_i/m = \theta_i)$) and linear predictor η_i is given by the logit function

$$\eta_i = g(\theta) = \log\left(\frac{\theta(x)}{1 - \theta(x)}\right)$$

Example: Logistic Regression

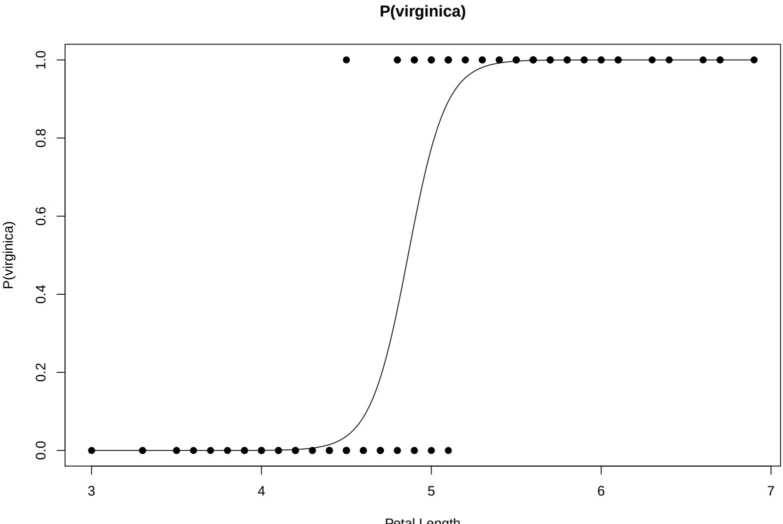
Using R's built in **iris** data, let's see if we can use petal length to differentiate between two species (setosa vs versicolor):

```
iris2 <- iris[iris$Species=="virginica" | iris$Species=="versicolor",]
iris2$Pvirginica=iris2$Species=="virginica"
#Species <- iris2$Species; Petal.Length <- iris2$Petal.Length;
fit <- glm(Species~Petal.Length, data=iris2, family=binomial(link="logit"))
fit
```

```
##
## Call:  glm(formula = Species ~ Petal.Length, family = binomial(link = "logit
##     data = iris2)
##
## Coefficients:
##  (Intercept)  Petal.Length
##      -43.781         9.002
##
## Degrees of Freedom: 99 Total (i.e. Null);  98 Residual
## Null Deviance:       138.6
## Residual Deviance: 33.43     AIC: 37.43
```

Example: Logistic Regression

```
plot(iris2$Petal.Length, iris2$Pvirginica, ylab="P(virginica)", pch=19, xlab="P
tmpdat=data.frame(Petal.Length=seq(min(iris2$Petal.Length),max(iris2$Petal.Leng
points(tmpdat[,1],predict(fit,tmpdat,type="response"),type="l")
```


P(virginica)
P(virginica)
Petal.Length