Inference for Generalized Linear Models

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

April 1, 2008

Complicated Example

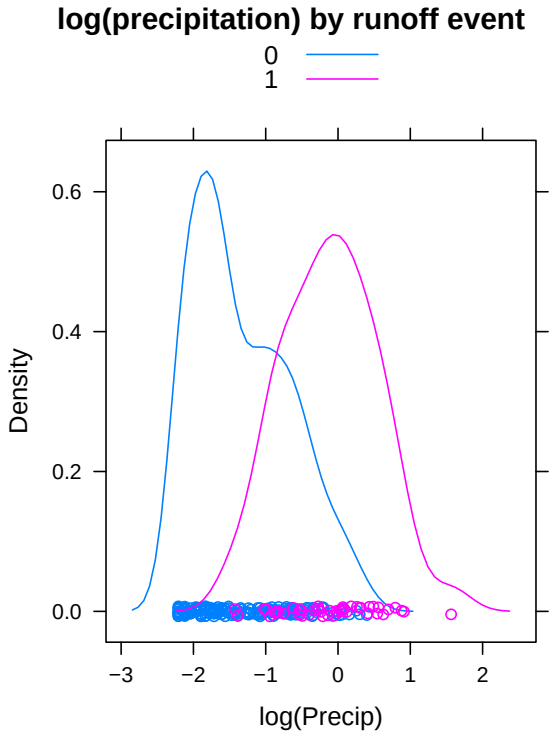
- Recall the runoff example from earlier in the semester.
- We had a model to predict runoff events.
- Say we want to predict the amount of runoff for a storm.
- We can have a logistic regression model to predict if the storm produces any measurable runoff, and a classical linear model to predict the amount of runoff when it does.
- We can use simulation to estimate the error in the predictions.

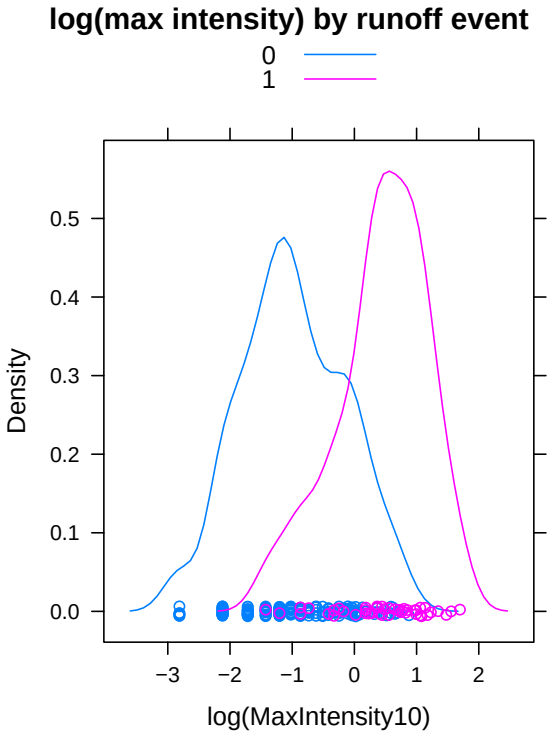
Models

- We fit a logistic regression model for runoff events based on the logarithm of precipitation and the logarithm of maximum storm intensity (over 10 minutes).
- Given a runoff event, we model the logarithm of the amount of runoff on the basis of the logarithms of precipitation, maximum ten minute storm intensity, and the time since the last storm.

```
> runoff = read.table("RunoffClean.txt", header = T)
> fit1 = glm(RunoffEvent ~ log(Precip) + log(MaxIntensity10),
+     data = runoff, family = binomial)
> fit2 = lm(log(Runoff) ~ log(Precip) + log(MaxIntensity10) +
+     log(LastStorm), data = runoff, subset = (Runoff > 0))
```


Graphs of Transformed Variables





Fitted Logistic Regression Model

```
> display(fit1, digits = 3)

glm(formula = RunoffEvent ~ log(Precip) + log(MaxIntensity10),
    family = binomial, data = runoff)
                   coef.est coef.se
(Intercept)        0.215    0.320
log(Precip)        1.813    0.429
log(MaxIntensity10) 1.593    0.358
---
  n = 231, k = 3
  residual deviance = 115.0, null deviance = 227.8 (difference = 112.8)
```


Fitted Regression Model

```
> display(fit2, digits = 3)

lm(formula = log(Runoff) ~ log(Precip) + log(MaxIntensity10) +
    log(LastStorm), data = runoff, subset = (Runoff > 0))
                   coef.est coef.se
(Intercept)         2.694    0.830
log(Precip)         1.370    0.363
log(MaxIntensity10)  0.763    0.326
log(LastStorm)      -0.266    0.117
---
n = 45, k = 4
residual sd = 1.451, R-Squared = 0.47
```

Simulations for Prediction

```
> sim1 = sim(fit1, 1000)
> apply(sim1$beta, 2, mean)
```

```
       (Intercept)         log(Precip) log(MaxIntensity10)
         0.2215979           1.8028966           1.6048461
```

```
> apply(sim1$beta, 2, sd)
```

```
       (Intercept)         log(Precip) log(MaxIntensity10)
         0.3214574           0.4357967           0.3610295
```

```
> mean(sim1$sigma)
```

```
[1] 1
```


Simulations for Prediction

```
> sim2 = sim(fit2, 1000)
> apply(sim2$beta, 2, mean)
```

```
       (Intercept)         log(Precip) log(MaxIntensity10)
         2.7139484           1.3753978           0.7657414
    log(LastStorm)
        -0.2682446
```

```
> apply(sim2$beta, 2, sd)
```

```
       (Intercept)         log(Precip) log(MaxIntensity10)
         0.8570230           0.3770254           0.3255932
    log(LastStorm)
         0.1194336
```

```
> mean(sim2$sigma)
```

```
[1] 1.480807
```

Prediction of Runoff

- Storm with 0.5 inches of rain, maximum ten minute intensity of 0.75 inches, 14,000 minutes since the previous storm.

```
> x1.new = c(1, log(0.5), log(0.75))
> y1.new = c(1, log(0.5), log(0.75), log(14000))
> prob1 = invlogit(sim1$beta %*% x1.new)
> invlogit(predict(fit1, data.frame(Precip = 0.5, MaxIntensity10 = 0.75)))
[1] 0.1823968
> range(prob1)
[1] 0.08522118 0.33643572
> event1 = rbinom(1000, 1, prob1)
> run1 = exp(rnorm(1000, sim2$beta %*% y1.new, sim2$sigma))
> run1[event1 == 0] = 0
> sum(run1 == 0)/1000
[1] 0.827
> quantile(run1, c(0.025, 0.975))
    2.5%    97.5%
0.000000 2.113863
```


Prediction of Runoff

- storm with 1.0 inches of rain, maximum ten minute intensity of 4.25 inches, 100 minutes since the previous storm.

```
> x2.new = c(1, log(1), log(4.25))
> y2.new = c(1, log(1), log(4.25), log(100))
> prob2 = invlogit(sim1$beta %*% x2.new)
> invlogit(predict(fit1, data.frame(Precip = 1, MaxIntensity10 = 4.25)))
[1] 0.9255018
> range(prob2)
[1] 0.5974186 0.9904506
> event2 = rbinom(1000, 1, prob2)
> run2 = exp(rnorm(1000, sim2$beta %*% y1.new, sim2$sigma))
> run2[event2 == 0] = 0
> sum(run2 == 0)/1000
[1] 0.065
> quantile(run2, c(0.025, 0.975))
    2.5%    97.5%
0.000000 8.589538
```