Likelihood Inference in Exponential Families and Generic Directions of Recession

Charles J. Geyer
School of Statistics
University of Minnesota

Elizabeth A. Thompson
Department of Statistics
University of Washington

http://www.stat.umn.edu/geyer/gdor/

What this Talk is About

In exponential families for discrete data — logistic and Poisson regression, log-linear models for contingency tables — MLE does not always exist in the conventional sense.

Then available software produces nonsense, often with no error or warning. Hypothesis tests and confidence intervals based on "usual" asymptotics of MLE do not work.

We now have the solution! Old theory (Barndorff-Nielsen, 1978). New software (R package `rcdd`, Geyer and Meeden, 2008, which uses `cddlib` computational geometry package, Fukuda, 2008).

Binomial Example

x is Binomial(n, p). MLE is $\hat{p} = x/n$.

Natural parameter is

$$\theta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

When $\hat{p} = 0$ or $\hat{p} = 1$, MLE of the natural parameter $\hat{\theta} = \text{logit}(\hat{p})$ does not exist.

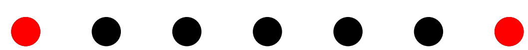
When $\hat{p} = 0$ or $\hat{p} = 1$, distribution for MLE is degenerate.

When $\hat{p} = 0$ or $\hat{p} = 1$, "usual" confidence interval

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

does not work.

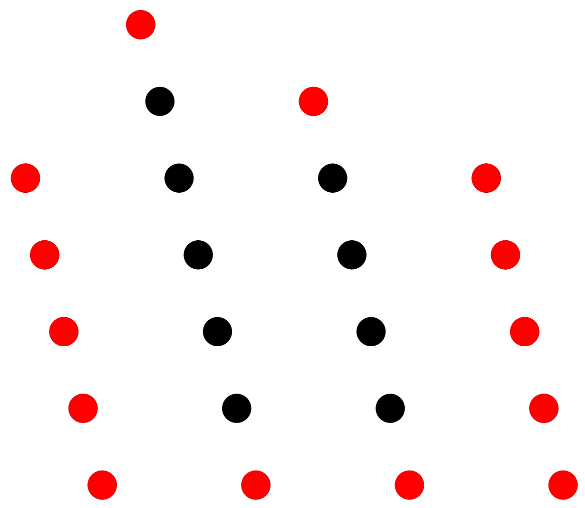
One-Dimensional Exponential Family Example



Black points: MLE exists. Usual asymptotics o. k.

Red points: MLE for natural parameter does not exist. MLE distribution degenerate. Usual asymptotics no good.

Two-Dimensional Exponential Family Example



Black points: MLE exists. Usual asymptotics o. k.

Red points: MLE for natural parameter does not exist. MLE distribution degenerate. Usual asymptotics no good.

Curse of Dimensionality

For dimensions higher than two, no nice pictures, same geometry.

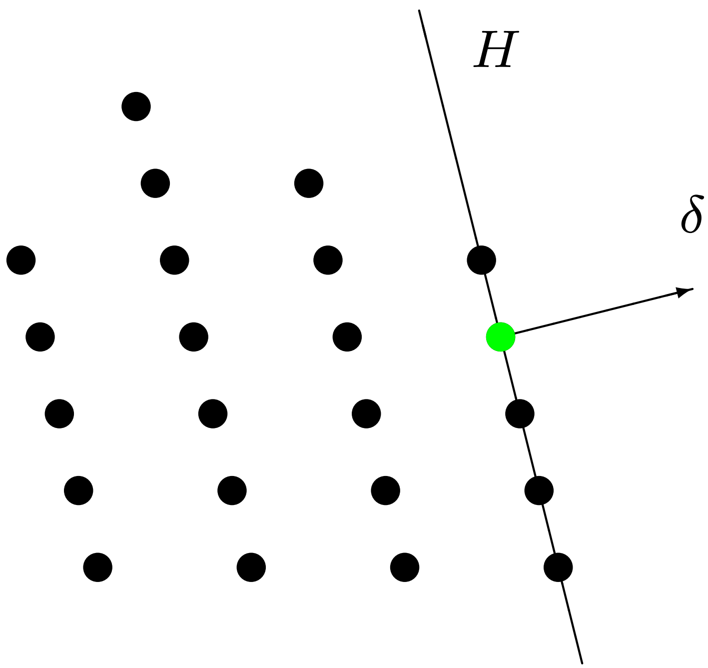
The higher the dimension, the more boundary points (red) where MLE does not exist.

Curse of Dimension Reduction

GLM with data vector y and model matrix M: natural sufficient statistic is $M^T y$.

Relevant geometry is for $M^T y$ not y. Need linear programming to determine points on boundary.

Generic Directions of Recession



Black dots: possible values of $M^T y$. Green dot: the observed value of $M^T y$.

Direction of recession δ is vector orthogonal to hyperplane H containing green dot with all black dots in H or on side opposite to δ.

Vector δ is *generic* direction of recession (GDOR) if fewest dots in H of any such construction and some dots not in H.

MLE does not exist in the conventional sense if and only if a GDOR exists.

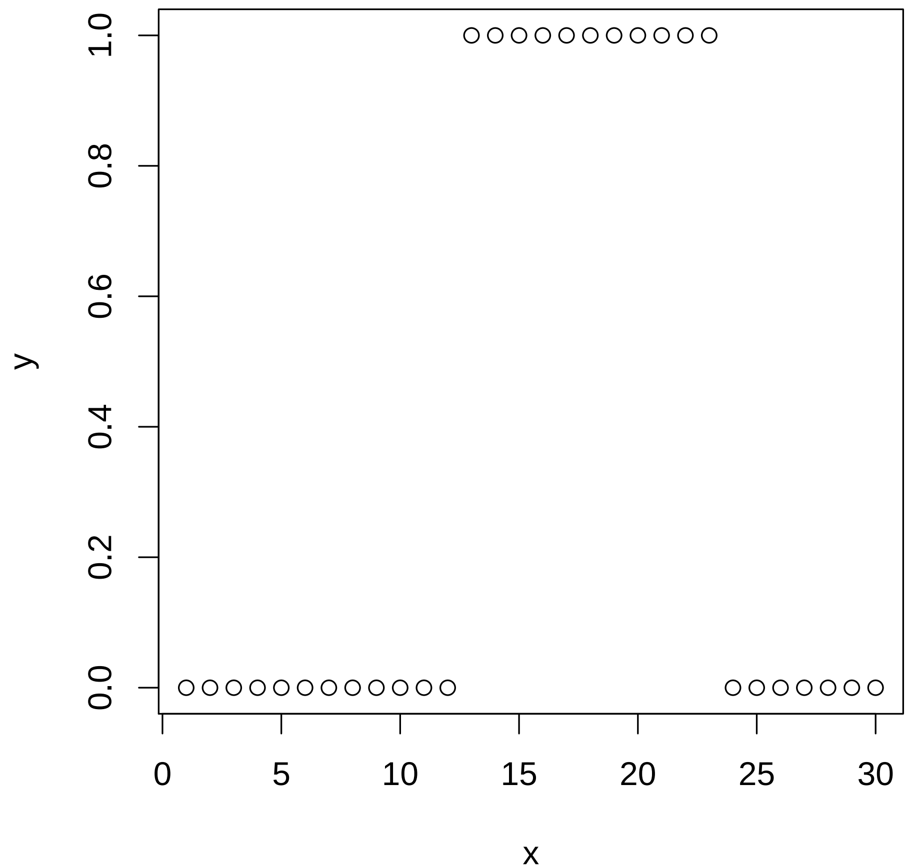
Logistic Regression Example

Response vector y, predictor vector x, and y_i is Bernoulli(p_i) with

$$\theta_i = \text{logit}(p_i) = \beta_1 + x_i\beta_2 + x_i^2\beta_3$$

(quadratic logistic regression). How hard can it be?

Logistic Regression Example (cont.)



If data as shown, MLE does not exist!

Logistic Regression Example (cont.)

```
>  x <- 1:30
>  y <- c(rep(0, 12), rep(1, 11), rep(0, 7))
>  out <- glm(y ~ x + I(x^2), family = binomial, x = TRUE)
Warning messages:
1: In glm.fit(x = X, y = Y, weights = weights, start = start, etasta
  algorithm did not converge
2: In glm.fit(x = X, y = Y, weights = weights, start = start, etasta
  fitted probabilities numerically 0 or 1 occurred
```

The `glm` function suggests — somewhat indirectly — that the MLE may not exist.

Finding a GDOR

```
> library(rcdd)
> tanv <- out$x
> tanv[y == 1, ] <- (-tanv[y == 1, ])
> vrep <- cbind(0, 0, tanv)
> lout <- linearity(vrep, rep = "V")
> lout
integer(0)
> p <- ncol(tanv)
> hrep <- cbind(-vrep, -1)
> hrep <- rbind(hrep, c(0, 1, rep(0, p), -1))
> objv <- c(rep(0, p), 1)
> pout <- lpcdd(hrep, objv, minimize = FALSE)
> gdor <- pout$primal.solution[1:p]
> gdor
[1] -53.3636364   6.5454545  -0.1818182
```

Finding a GDOR (cont.)

In the immortal words of a comment in the UNIX source code, "you are not expected to understand this," but the output tells us two things:

```
> gdor
[1] -53.3636364   6.5454545  -0.1818182
```

gives the GDOR and

```
> lout
integer(0)
```

says that hyperplane H contains no $M^T y$ except for observed value.

Limiting Conditional Model

Take limits of probability distributions

$$\lim_{s \to \infty} P_{\beta + s\delta}(\,\cdot\,) = P_{\beta}(\,\cdot \mid M^T Y \in H)$$

set of all such limits is called *limiting conditional model* (LCM).

LCM is exponential family; same natural parameter and statistic.

If l is log likelihood for original model and $\hat{\beta}$ is MLE for LCM, then

$$\lim_{s \to \infty} l(\hat{\beta} + s\delta) = \sup_{\beta} l(\beta)$$

so can consider MLE for original model to be $\hat{\beta}$ sent to infinity in direction δ.

Logistic Regression Example (cont.)

So what? The MLE distribution in the LCM is degenerate, concentrated at one point. It says we could never observe data different from what we did observe. Nobody believes that.

The sample is not the population. Estimates are not parameters.

We need confidence intervals, necessarily one-sided, saying how close s is to infinity in $\hat{\beta} + s\delta$ and how close the corresponding mean value parameters $\mu_i = E_\beta(Y_i)$ are to their observed values.

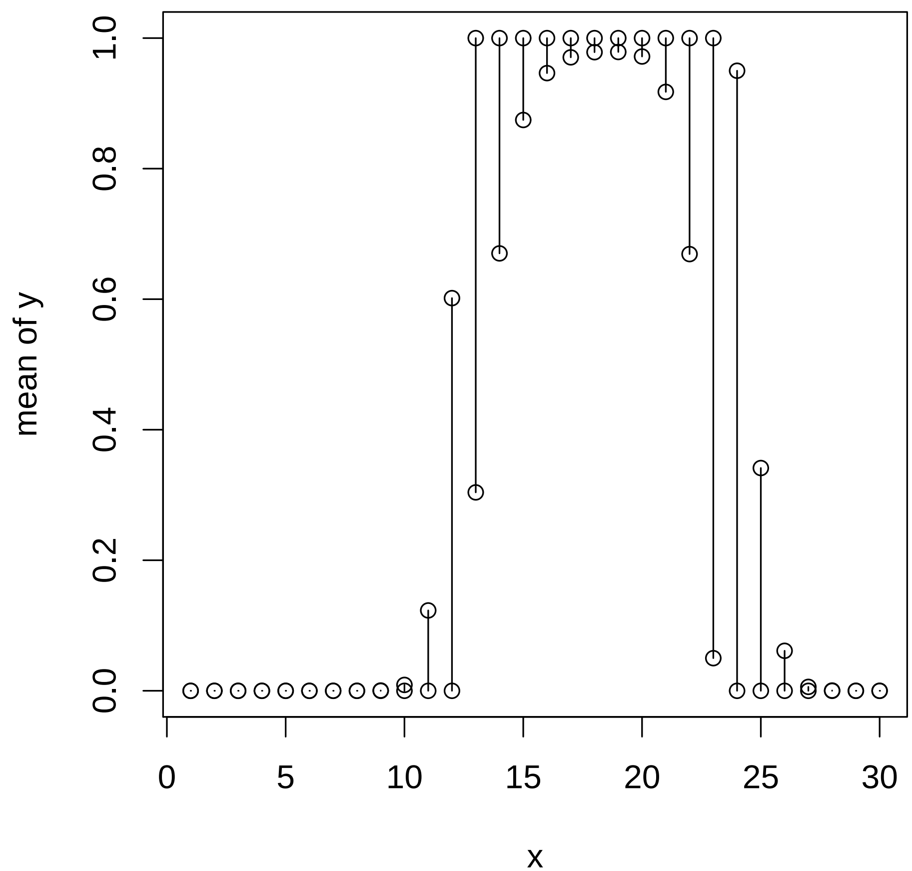
One-Sided Confidence Intervals

Let B denote set of all MLE for the LCM — affine subspace, all points corresponding to same distribution — then

$$\{\, \beta \in B : P_\beta(M^T Y \in H) \geq \alpha \,\}$$

is $1 - \alpha$ confidence region saying how close natural parameter is to infinity.

One-Sided Intervals for Logistic Regression



One-sided exact simultaneous 95% confidence intervals for mean value parameters $\mu_i = E_\beta(Y_i)$.

Lessons Learned from Logistic Regression Example

GDOR notion. General.

LCM construction. General.

LCM concentrated at one point so any β is an MLE for the LCM. Not general.

Usually LCM fixes only some, not all components of response vector at observed values. Hence need to find MLE for LCM.

Loglinear Model Example

$2 \times 2 \times \cdots \times 2$ contingency table with seven dimensions hence $2^7 = 128$ cells.

```
> dat <- read.table(
+     url("http://www.stat.umn.edu/geyer/gdor/catrec.txt"),
+     header = TRUE)
```

gets the data.

```
> out3 <- glm(y ~ (v1 + v2 + v3 + v4 + v5 + v6 + v7)^3,
+     family = poisson, data = dat, x = TRUE)
```

fits the model. R gives no error or warning, but MLE does not exist.

Loglinear Model Example (cont.)

```
> tanv <- out3$x
> vrep <- cbind(0, 0, tanv)
> vrep[dat$y > 0, 1] <- 1
> lout <- linearity(vrep, rep = "V")
> linear <- dat$y > 0
> linear[lout] <- TRUE
> sum(linear)
[1] 112
> length(linear) - sum(linear)
[1] 16
```

The LCM fixes 16 components of the response vector at their observed value zero and leaves 112 components random.

Loglinear Model Example (cont.)

Fit LCM

```
> dat.cond <- dat[linear, ]
> out3.cond <- glm(y ~ (v1 + v2 + v3 + v4 + v5 + v6 + v7)^3,
+     family = poisson, data = dat.cond)
> summary(out3.cond)
```

Voluminous output not shown (64 regression coefficients!). This fit `out3.cond` can be used to produce valid hypothesis tests and confidence intervals about the 112 components of the response not fixed in the LCM.

Loglinear Model Example (cont.)

For the 16 components of the response fixed at zero in LCM, proceed as before. Find GDOR.

```
> p <- ncol(tanv)
> hrep <- cbind(0, 0, -tanv, 0)
> hrep[!linear, ncol(hrep)] <- (-1)
> hrep[linear, 1] <- 1
> hrep <- rbind(hrep, c(0, 1, rep(0, p), -1))
> objv <- c(rep(0, p), 1)
> pout <- lpcdd(hrep, objv, minimize = FALSE)
> gdor <- pout$primal.solution[1:p]
```

and find one-sided confidence intervals.

One-Sided Intervals for Loglinear Model

One-sided exact simultaneous 95% confidence intervals for mean value parameters $\mu_i = E_\beta(Y_i)$ based on multinomial sampling (not Poisson). Sample size `sum(dat$y)` is 544.

v_1	v_2	v_3	v_4	v_5	v_6	v_7	lower	upper
0	0	0	0	0	0	0	0	0.2855
0	0	0	1	0	0	0	0	0.1404
1	1	0	0	1	0	0	0	0.2194
1	1	0	1	1	0	0	0	0.4198
0	0	0	0	0	1	0	0	0.0892
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	0	1	1	0	1	0	0.2639
0	0	0	0	0	1	1	0	0.0665
0	0	0	1	0	1	1	0	0.1543
1	1	0	0	1	1	1	0	0.1406
1	1	0	1	1	1	1	0	0.3230

Hypothesis Tests

"Usual" hypothesis tests valid if MLE exists in the conventional sense for null hypothesis.

If not, then base test on LCM for null hypothesis (S. Fienberg, personal communication).

More about RCDD

Exact, infinite-precision, rational arithmetic. Can be used for computational geometry operations, including the `linearity` and `lpcdd` functions, and for ordinary arithmetic and comparison.

When exact arithmetic is used, computer proofs are as rigorous as pencil-and-paper proofs.

See package vignette for everything `rcdd` can do.

Web Pages

Paper.

Electronic Journal of Statistics **3**, (2009), 259–289 (electronic)

Two technical reports done with `Sweave` so every number in the paper and this talk is reproducible by anyone who has R. Also slides for this talk.

`http://www.stat.umn.edu/geyer/gdor/`