Outline

MLE under logistic model

- Binary regression model:

$$
\begin{aligned}
Y_i &\sim \mathrm{Ber}(\mu_i), \quad \text{independently} \\
g(\mu_i) &= \boldsymbol{X}_i'\beta,
\end{aligned}
$$

  where $g(\mu) = \log \frac{\mu}{1-\mu}$ is the logistic link function.
- Log-likelihood under logistic link:

$$
\text{log-L}(\beta) = \sum_{i=1}^{n} \left[Y_i(\boldsymbol{X}_i'\beta) - \log\{1 + \exp(\boldsymbol{X}_i'\beta)\} \right].
$$

- Let $\hat{\beta}$ denote the MLE of β. Then $\hat{\beta}$ satisfies the score equation

$$
\mathbf{0} \equiv \frac{d\text{Log-L}(\beta)}{d\beta} = \sum_{i=1}^{n} \left(Y_i - \text{logit}(\boldsymbol{X}_i'\beta) \right) \boldsymbol{X}_i
$$

- No closed form for $\hat{\beta}$.

Fitting a binary regression GLM: IRLS (Optional)

Algorithm:

1. Initialize: set $\hat{\mu}_i = 0.999$ or 0.001 depending on whether $Y_i = 1$ or 0.
2. Compute $Z_i \leftarrow g(\hat{\mu}_i) + g'(\hat{\mu}_i)(Y_i - \hat{\mu}_i)$.
3. Obtain $\hat{\beta}$ by regressing $\boldsymbol{Z}$ onto $\boldsymbol{X}$ using WLS with weights $W_i^{-1} = g'(\hat{\mu}_i)^2 V(\hat{\mu}_i)$ to
4. Compute $\hat{\mu}_i = g^{-1}(\boldsymbol{X}_i'\hat{\beta})$.
5. Repeat steps 2–4 until convergence.

Rough idea

- Recall that the link function $g(\cdot) : [0,1] \mapsto \mathbb{R}$ connects $\mathbb{E}(Y_i)$ to $\boldsymbol{X\beta}$.
- Taylor expansion of link function at $\hat{\mu}_i$: $g(\mathbb{E}(Y_i)) \approx g(\hat{\mu}_i) + g'(\hat{\mu}_i)(Y_i - \hat{\mu}_i)$.
- Introduce a "working" response $Z_i \leftarrow g(\hat{\mu}_i) + g'(\hat{\mu}_i)(Y_i - \hat{\mu}_i)$.
- By Bernoulli model assumption:

$$\mathbb{E}(Z_i|X_i) \approx \boldsymbol{X}_i\boldsymbol{\beta},$$
$$\text{Var}(Z_i|X_i) = g'(\hat{\mu}_i)^2\text{Var}(Y_i), \text{ where } \text{Var}(Y_i) \approx V(\hat{\mu}_i).$$

- Roughly equivalent to a weighted least square problem where $\boldsymbol{Z}$ is a (continuous) response, $\boldsymbol{X}$ is the predictor, and $g'(\hat{\mu}_i)^2 V(\hat{\mu}_i)$ is the error variance.

Goodness of Fit Statistics

Deviation of Model

We define the deviation of a model:

$$\text{Dev}(\hat{\boldsymbol{\mu}}, \boldsymbol{Y}) \overset{\text{def}}{=} -2 \times \text{log-L}(\hat{\boldsymbol{\mu}}, \boldsymbol{Y}) - 2 \times \text{log-L}(\boldsymbol{Y}, \boldsymbol{Y})$$

where $\hat{\boldsymbol{\mu}}$ denotes the fitted mean based on the specified model and $\boldsymbol{Y}$ the observations.

- Deviation is used to assess the goodness of fit of the model.
- For Gaussian linear model, deviation is proportional to SSE:

$$\text{Dev}(\hat{\boldsymbol{\mu}}, \boldsymbol{Y}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2.$$

- For Bernoulli logistic model:

$$\text{Dev}(\hat{\boldsymbol{\mu}}, \boldsymbol{Y}) = -2 \left(\sum^{n} Y_i \log \mu_i + (1 - Y_i) \log(1 - \hat{\mu}_i) \right).$$

Outline

1. Estimation of logistic regression coefficients
2. Binomial regression
3. Poisson Regression

Binomial regression

- Suppose that we have k independent observations $y_1, \ldots, y_K$, each of which is independent drawn of a Binomial random variable Y_i:

$$Y_i \sim \mathrm{B}(n_i, \pi_i), \quad \text{independently for all } i = 1, \ldots, K.$$

- Here n_i is the (known) binomial denominator and $\pi_i \in [0, 1]$ is the (unknown) success probability.
- Logistic models on $\{\pi_i\}$:

$$\mathrm{logit}(\pi_i) = \boldsymbol{X}_i'\boldsymbol{\beta}, \quad \text{for all } i = 1, \ldots, K.$$

  where $\boldsymbol{X}_i' = (1, X_{i1}, \ldots, X_{i,p-1})' \in \mathbb{R}^{1\times p}$ is the predictor for i-th observation, and $\boldsymbol{\beta} \in \mathbb{R}^{p\times 1}$ is the parameter of interest.
- R command: `myweight=`$c(n_1, \ldots, n_K)$
  `glm(Y/myweight~X,family="binomial",`
  `weights=myweight)`

Example: one-factor logistic model

A study surveys 1,607 individuals in Madison. Their answers to "whether travel during Thanksgiving" are tabulated below.

Age (i)	Travel (Y_i)	Not travel ($n_i - Y_i$)	Total (n_i)
< 25	72	325	397
25–29	105	299	404
30–39	237	375	612
40–49	93	101	194
Total	507	1100	1607

Questions:

- Does the data supports a common probability of traveling for the four age groups?
- What is the estimated probability of traveling for people under age 25? The 95% confidence interval?

Example: one-factor logistic model

- Consider a one-factor model where we allow each age group to have its own "success probability" π_{ij}.
- The model can be expressed as

$$\text{logit}(\pi_i) = \eta + \alpha_i, \quad \text{where } i = 1, 2, 3.4$$

  where we impose $\alpha_1 = 0$.
- Why impose $\alpha_1 = 0$? Note that age is a categorical factor that has four levels.
- Interpretation: η is the logit of the reference group (i.e. $i = 1$), and α_i measures the difference in logits between level i of the factor and the reference group.
- Parameter estimation:

Parameter	Symbol	Estimate	Std. Error	z-ratio
Constant	η	-1.507	0.130	-11.57
Age 25-29	α_2	0.461	0.173	2.67
30-39	α_3	1.048	0.154	6.79
40-49	α_4	1.425	0.194	7.35

Example: Two-factor logistic model

Actual Travel vs. Plan and Age:

Age (i)	Plan (j)	Travel (Y_{ij})	Not travel ($n_{ij} - Y_{ij}$)	All (n_{ij})
< 25	No	58	265	323
	Yes	14	60	74
25-29	No	68	215	283
	Yes	37	84	121
30-39	No	79	230	309
	Yes	158	145	303
40-49	No	14	43	57
	Yes	79	58	137
Total		507	1100	1607

- Let Y_{ij} denotes the number of individuals who actually travel, where $i = 1, \ldots, 4$ refers the four age groups and $j = 1, 2$ denotes the two categories of plan.

Example: Two-way logistic model

- Model assumption:

  $$Y_{ij} \sim B(n_{ij}, \pi_{ij}), \quad \text{independently for } i = 1, \ldots, 4 \text{ and } j = 1, 2.$$

- Additive model on $\{\pi_{ij}\}$

  $$\text{logit}(\pi_{ij}) = \eta + \alpha_i + \beta_j.$$

  For parameter identifiability, we impose $\alpha_1 = 0$ and $\beta_1 = 0$.
- Interpretation:
  - η is the logit of traveling probability for individuals under 25 who do not plan for traveling (reference level).
  - $\alpha_i : i = 2, 3, 4$ represents the main effect of age 25-29, 30-39, 40-49, compared to individuals under age 25 in the same plan group.
  - β_2 represents the main effect of planning, compared to individuals who do not plan for traveling in the same age group.

Example: Two-way logistic model

Parameter estimates for additive logistic model.

Parameter	Symbol	Estimate	Std. Error	z-ratio
Constant	η	-1.694	0.130	-12.53
Age 25-29	α_2	0.368	0.175	2.10
30-39	α_3	0.808	0.160	5.06
40-49	α_4	1.023	0.204	5.01
Plan Yes	β_2	0.824	0.117	7.04

- The estimates of the $\alpha_j' s$ show a strong monotonic age effect.
- The estimate of β_2 shows a strong effect for planning.

Example: Two-way logistic model

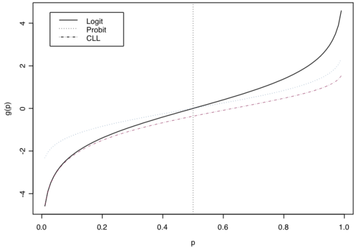
ANOVA table:

Model	logit(π_{ij})	Deviance	df
Null	η	145.7	7
Age only	$\eta + \alpha_i$	66.5	4
Plan only	$\eta + \beta_j$	54.0	6
Additive	$\eta + \alpha_i + \beta_j$	16.8	3
Full	η_{ij}	0	0

- Age main effect $H_0 : \alpha_2 = \alpha_3 = \alpha_4 = 0$. Highly significant.
- Plan main effect $H_0 : \beta_2 = 0$. Highly significant.

Remarks on Binomial regression

Link	$\eta = g(\pi)$	$\pi = g^{-1}(\eta)$
identity	π	η
logarithmic	$\log \pi$	e^{η}
logistic	$\log\left(\frac{\pi}{1-\pi}\right)$	$\frac{e^{\eta}}{1+e^{\eta}}$
probit	$\Phi^{-1}(\pi)$	$\Phi(\eta)$
log-log	$\log(-\log \pi)$	$\exp(-e^{\eta})$
complementary log-log	$\log(-\log(1-\pi))$	$1-\exp(-e^{\eta})$



- For Binomial mode, the identity or logarithmic link may not be the best choice. ($\pi = g^{-1}(\eta)$ may lie outside [0,1])

Outline

Poisson Regression

- Poisson regression is useful when the outcome is a count, with large-count outcomes being rare events.
- The outcomes are counts with $Y_i = 0, 1, 2, \ldots, i = 1, \ldots, n$.
- Let Y_i follow an indepdent Poisson distribution with mean μ_i.

$$Y_i \sim \text{Poi}(\mu_i), \quad \text{independently.}$$

- The probability distribution of Y_i is

$$f(Y_i) = \frac{\mu_i^{Y_i} \exp(-\mu_i)}{Y_i!}$$

- Note $\mathbb{E}(Y_i) = Var(Y_i) = \mu_i > 0$.
- The joint probability distribution is

$$f(Y_1, \ldots, Y_n) = \prod_{i=1}^{n} f_i(Y_i) = \prod_{i=1}^{n} \frac{\mu_i^{Y_i} \exp(-\mu_i)}{Y_i!}$$

Poisson Regression

- Poisson regression model:

  Y_i are independent Poisson random variables with

$$\mathbb{E}(Y_i) = \mu_i, \text{ where } \mu_i = g(\boldsymbol{X}_i'\boldsymbol{\beta}).$$

- $g(\cdot) : \mathbb{R} \mapsto \mathbb{R}_+$ maps the linear predictors $\boldsymbol{X}_i'\boldsymbol{\beta} \in \mathbb{R}$ to the Poisson mean $\in \mathbb{R}_+$.
- A common “link" function is

$$\mu_i = g(\boldsymbol{X}_i'\boldsymbol{\beta}) = \exp(\boldsymbol{X}_i'\boldsymbol{\beta})$$

- Equivalently, $\log(\mu_i) = \boldsymbol{X}_i'\boldsymbol{\beta}$.
- The log-likelihood function is

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} Y_i \log \mu_i - \sum_{i=1}^{n} \mu_i + C.$$

Poisson Regression

- Iteratively reweighted least squares can again be used to obtain MLEs of $\boldsymbol{\beta}$.
- Given $\hat{\boldsymbol{\beta}}$, the fitted response function is

$$\hat{\mu}_i = \exp(\boldsymbol{X}_i'\hat{\boldsymbol{\beta}})$$

- Model inference for a Poisson regression model is carried out in a similar fashion to that for logistic regression:
  - Testing for individual coefficients based on Wald test statistics.
  - Testing for groups of coefficients based on the likelihood ratio test statistic.
- Deviance for fitted Poisson regression:

$$D(\boldsymbol{Y}, \boldsymbol{\mu}) = 2\sum_{i=1}^{n}\left[-Y_i \log\frac{Y_i}{\hat{\mu}_i} - (Y_i - \hat{\mu}_i)\right]$$

.

Miller Lumber Company Example

- The Miller Lumber Company is a large retailer of lumber and paint. During a two-week period, in-store surveys were conducted and addresses of customers were obtained.
- The total number of customers who visited the store from each center within a 10-mile radius was determined
- Relevant demographic information for each center (average income, number of housing units, etc.) was obtained.
- Several other variables expected to be related to customer counts were constructed from maps, including distance from center to nearest competitor and distance to store.

Miller Lumber Company Example

- Initial screening of the potential predictor variables was conducted which led to the retention of five predictor variables:
  X_1: Number of housing units
  X_2: Average income, in dollars
  X_3: Average housing unit age, in years
  X_4 : Distance to nearest competitor, in miles
  X_5: Distance to store, in miles
- Response Y : Number of customers who visited store from census tract

```
> mydata = read.table("MillerLumber.txt", header=T); attach(mydata)
> head(mydata)
   Y  X1    X2 X3   X4   X5
1  9 606 41393  3 3.04 6.32
2  6 641 23635 18 1.95 8.89
3 28 505 55475 27 6.54 2.05
4 11 866 64646 31 1.67 5.81
5  4 599 31972  7 0.72 8.11
6  4 520 41755 23 2.24 6.81
```

Miller Lumber Company Example

```
> glm5 = glm(Y~., data=mydata, family=poisson("log"))
> summary(glm5)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.942e+00  2.072e-01  14.198  < 2e-16 ***
X1           6.058e-04  1.421e-04   4.262 2.02e-05 ***
X2          -1.169e-05  2.112e-06  -5.534 3.13e-08 ***
X3          -3.726e-03  1.782e-03  -2.091   0.0365 *
X4           1.684e-01  2.577e-02   6.534 6.39e-11 ***
X5          -1.288e-01  1.620e-02  -7.948 1.89e-15 ***
---

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 422.22  on 109  degrees of freedom
Residual deviance: 114.99  on 104  degrees of freedom
AIC: 571.02

Number of Fisher Scoring iterations: 4
```

- The fitted Poisson response function is

$$\hat{\mu} = \exp(2.942 + .00061X_1 - .000012X_2 - .0037X_3 + .17X4 - .13X_5)$$

Miller Lumber Company Example

```
> library(car)
> Anova(glm5, type="III")
Analysis of Deviance Table (Type III tests)

Response: Y
   LR Chisq Df Pr(>Chisq)
X1   18.203  1  1.986e-05 ***
X2   31.794  1  1.714e-08 ***
X3    4.379  1    0.03638 *
X4   41.660  1  1.086e-10 ***
X5   67.500  1  < 2.2e-16 ***
```