

# Lab 1: Generalized linear models

March 29, 2018

# Generalized linear models

---

- Specify distribution for response variable
- Specify linear predictor
- Specify link function
  - Calculates expected response given linear predictor
- Example
  - Counts for local densities

$$c_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \mathbf{x}_i \boldsymbol{\beta}$$

---

# Common distributions for data

## – Discrete

Name	Notation	Domain	Range
Bernoulli	$B \sim \text{Bernoulli}(p)$	$0 \leq p \leq 1$	$B = \{0, 1\}$
Binomial	$N \sim \text{Binomial}(p, n)$	$0 \leq p \leq 1$	$N = \{0, 1, \dots, n\}$
Poisson	$N \sim \text{Poisson}(\lambda)$	$\lambda > 0$	$N = \{0, 1, \dots, \infty\}$
Negative binomial	$N \sim \text{NegativeBinomial}(\lambda, \theta)$	$\lambda > 0$ $\theta > 0$	$N = \{0, 1, \dots, \infty\}$
Conway-Maxwell-Poisson	$N \sim \text{CMP}(\mu, \nu)$	$\mu > 0$ $\nu > 0$	$N = \{0, 1, \dots, \infty\}$

---

# Common distributions for data

## – Continuous

Name	Notation	Domain	Range
Normal	$Y \sim \text{Normal}(\mu, \sigma^2)$	$\sigma^2 > 0$	Unrestricted
Lognormal	$\log(Y) \sim \text{Normal}(\mu, \sigma^2)$	$\sigma^2 > 0$	$Y > 0$
Gamma	$Y \sim \text{Gamma}(\mu, CV)$	$\mu > 0$ $CV > 0$	$Y > 0$
Beta	$p \sim \text{Beta}(\alpha, \beta)$	$\alpha > 0, \beta > 0$	$0 < p < 1$

---

## How to choose a distribution for data?

- Choice 1 – is it *continuous* or *discrete*?
  - Continuous: normal, lognormal, beta, gamma
  - Discrete: Bernoulli, binomial, poisson, negative binomial
- Choice 2 – what is the range of possible values?
  - E.g., if discrete:
    - If it is 0 or 1, then it's Bernoulli
    - If it's between 0 and N, where N is the number of trials, then it's Binomial
- Choice 3 – How flexible do you want it?

---

## How to choose a distribution for data?

- Frequent null models:

1. Binomial
2. Poisson
3. Normal

---

## How to choose a distribution for data?

- Binomial

- If you have one or more binary events:

$$B_i \sim \text{Bernoulli}(p)$$

- Then the sum of successes...

$$N = \sum_{i=1}^{n_i} B_i$$

... follows a binomial distribution

$$N \sim \text{Binomial}(p, n)$$

- Characteristics:

$$\mathbb{E}(N) = pn$$

$$\mathbb{V}(N) = np(1 - p)$$

---

## How to choose a distribution for data?

- Poisson

- If you have a lot of independent events, each with low probability:

$$N \sim \text{Binomial}(p, n)$$

where  $np \gg 0$  and  $p \ll 1$

- Then the number of successes follows a Poisson distribution

$$N \sim \text{Poisson}(np)$$

- Characteristics:

$$\mathbb{E}(N) = np$$

$$\mathbb{V}(N) = np$$



## How to choose a distribution for data?

---

- Normal

- If you have one or more events:

$$B_i \sim g(\boldsymbol{\theta})$$

where  $g(\boldsymbol{\theta})$  is some unknown density function

- Then the sum of outcomes ...

$$N = \sum_{i=1}^{n_i} b_i$$

... will converge on a normal distribution

$$N \sim \text{Normal}(\mu, \sigma_b^2)$$

... as the number of events gets large  $n_i \rightarrow \infty$

$$\mathbb{E}(N) = \mu = n_i \mathbb{E}(g(\boldsymbol{\theta}))$$

$$\mathbb{V}(N) = \sigma_b^2 = n_i^2 \mathbb{V}(g(\boldsymbol{\theta}))$$

---

## Review:

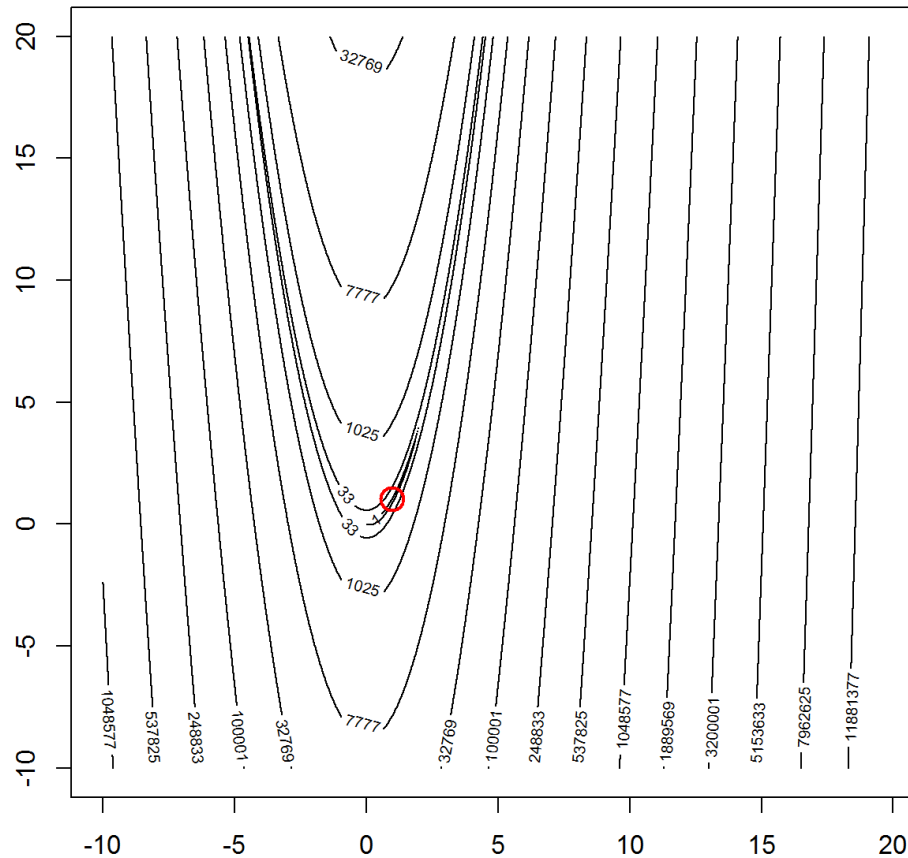
- Maximum likelihood estimation (MLE)

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}; \mathbf{y}))$$

- Where  $\hat{\boldsymbol{\theta}}$  is the MLE estimate of parameters
- Where  $\operatorname{argmax}_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}; \mathbf{y}))$  is the maximum value for  $L(\boldsymbol{\theta}; \mathbf{y})$  that can be achieved for any value of  $\boldsymbol{\theta}$
- *argmax* is done using maximization algorithms

# How to maximize the likelihood function

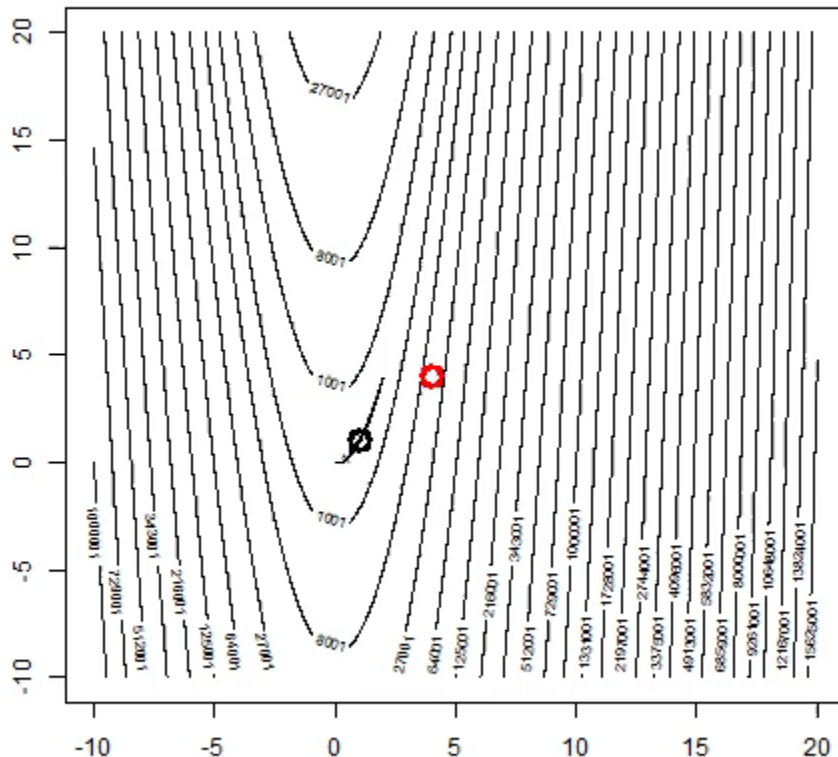
- Nonlinear minimizers
- Test using Rosenbrook “Banana” function



# How to maximize the likelihood function

- Methods without gradients are slow

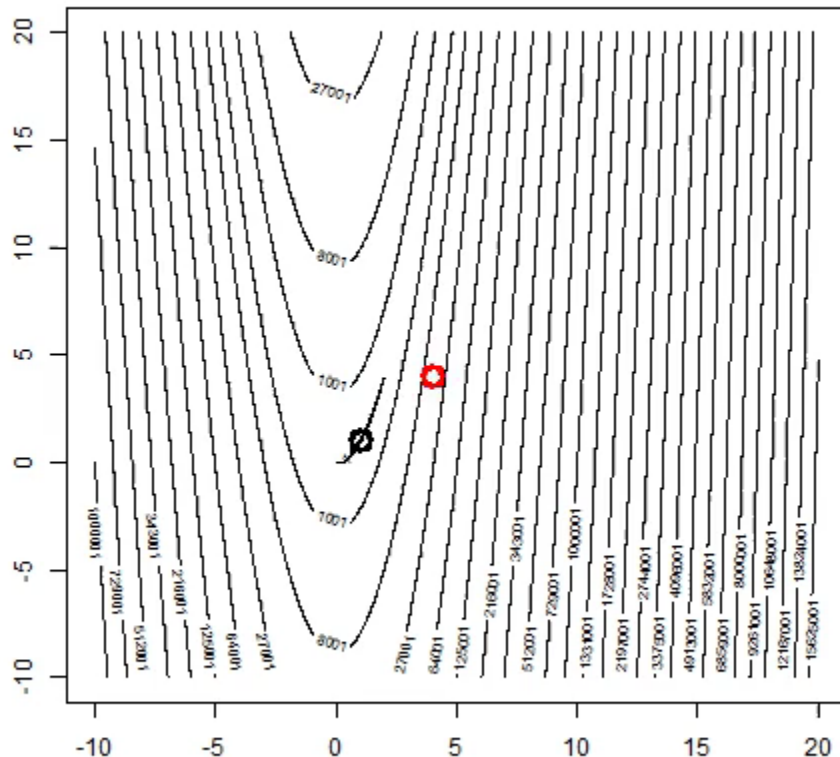
Quasi-Newton



# How to maximize the likelihood function

- Methods with gradients are much faster!

TMB using Nelder-Mead



---

Example #1 – What is the mean density of canary rockfish in the California Current?

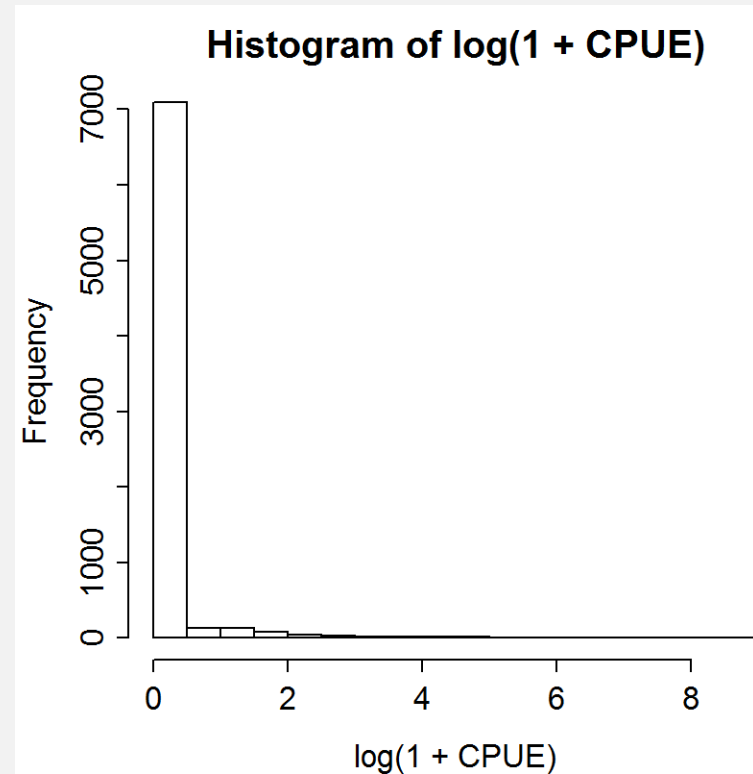
- Define linear predictor matrix

$$x_i = 1$$

- i.e.,

$$\mathbf{X} = \mathbf{1}$$

- We call  $\mathbf{X}$  an intercept matrix



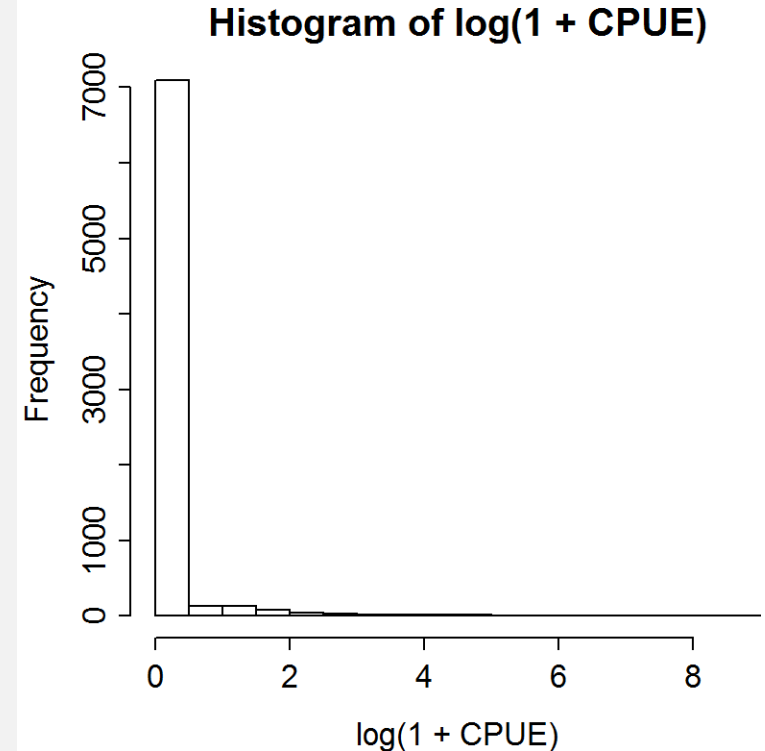
- Generalized linear models
  - Specify distribution for response variable
  - Specify function for expected value

- Canary catch rates

$$\log(\lambda_i) = \mathbf{x}_i \boldsymbol{\beta}$$

$$\Pr(C = c_i)$$

$$= \begin{cases} \theta_1 & \text{if } c_i = 0 \\ (1 - \theta_1) \text{Lognormal}(\lambda_i, \theta_2) & \text{if } c_i > 0 \end{cases}$$



Hint

---

If:

$$\Pr(c_i) = \begin{cases} \theta_1 & \text{if } c_i = 0 \\ (1 - \theta_1) \text{Lognormal}(\lambda_i, \theta_2) & \text{if } c_i > 0 \end{cases}$$

Then:

$$\begin{aligned} & \log(\Pr(c_i)) \\ &= \begin{cases} \log(\theta_1) & \text{if } c_i = 0 \\ \log(1 - \theta_1) + \log(\text{Lognormal}(\lambda_i, \theta_2)) & \text{if } c_i > 0 \end{cases} \end{aligned}$$

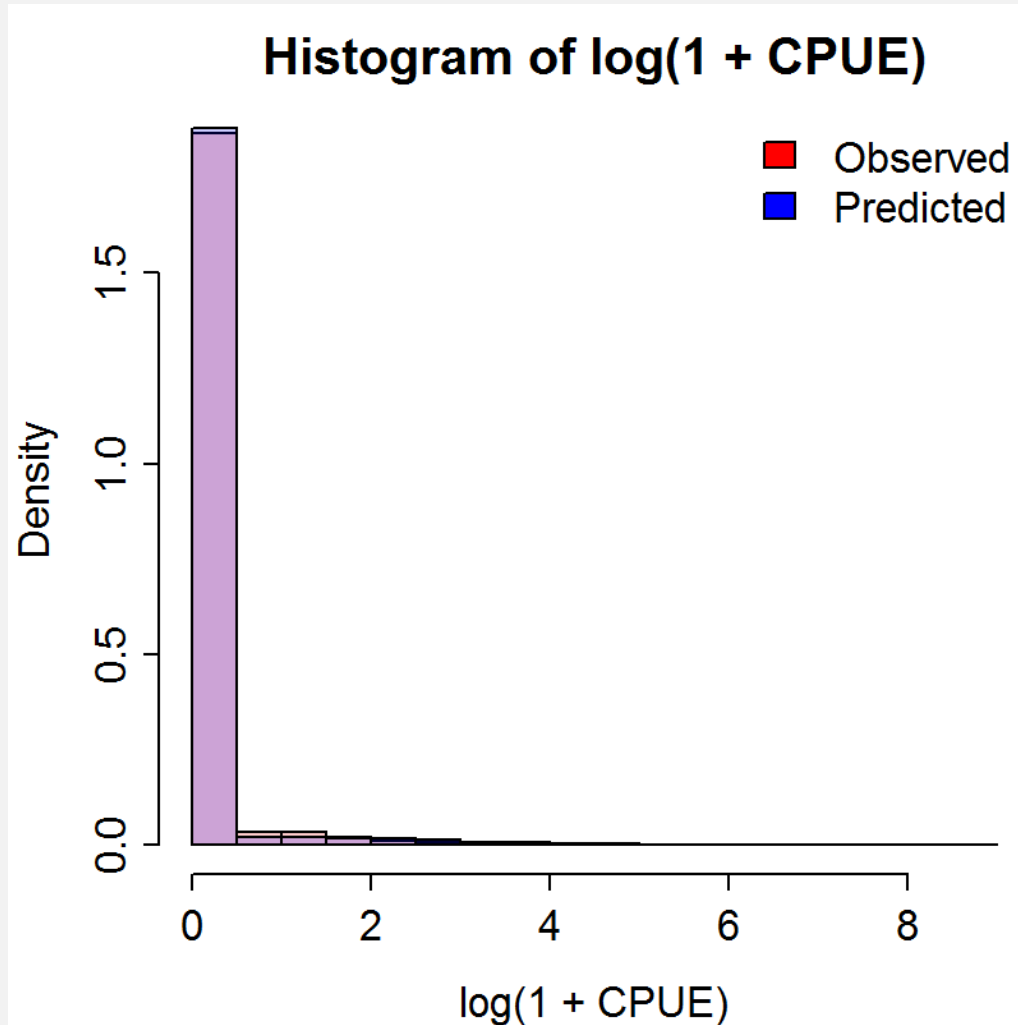


---

[Work on TMB code in groups of 2 for 20 minutes]

# Conclusion

- Decent fit...



# How do we assess fit?

---

- We want expected predictive loss
  - Assume there's a true “data-generating process” (DGP)

$$f(y_i)$$

- Where  $p(\mathbf{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{y})$  is your specified probability distribution

$$\text{predictive probability} = \int p(y^*|\hat{\boldsymbol{\theta}}) f(y^*) dy^*$$

- Where
  - $y^*$  is some future data

Then

$$\text{expected predictive log. probability} = \sum_{j=1}^J \log(p(y_j|\hat{\boldsymbol{\theta}}))$$

- Where
  - $y_j$  is some data that were “held out” when estimating parameters  $\hat{\boldsymbol{\theta}}$

More reading: Gelman, A., Hwang, J. & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat. Comput.*, 24, 997–1016.

---

## How do we assess fit?

- K-fold crossvalidation
  1. Divide data set into  $K$  even partitions
  2. Calculate predictive probability for 1<sup>st</sup> partition
    - For each piece  $K$ , fit the model to all data except data in that partition
    - Calculate the predictive probability of data in partition  $K$  using this model
    - Record predictive probability
  3. Repeat step 2 for all  $K$  partitions
  4. Chose the model with the highest predictive probability

---

## Confidence interval:

- Parameter estimates are normally distributed
- Computation

$$CI_{x\%}(\hat{\theta}) = \hat{\theta} \pm \widehat{SE}(\hat{\theta}) \times \Phi^{-1}\left(\frac{x}{2}\right)$$

- Where  $CI_{x\%}$  contains the true value  $x\%$  of the time if the model is correct
- $\Phi^{-1}$  is the inverse cumulative distribution for a normal distribution
- $\hat{\theta}$  is the estimate for parameter  $\theta$
- $\widehat{SE}(\hat{\theta})$  is the estimated standard error for parameter  $\theta$

## Confidence interval coverage

---

- *Coverage* – the expected proportion of times that an estimated  $x\%$  confidence interval contains the true value given an estimation model and true “data-generating process”

### *Estimation:*

1. Simulate data with a known value for parameter  $\theta$
2. Record true parameter values
3. Apply estimator
4. Record confidence interval  $CI_{x\%}(\hat{\theta})$  for parameter  $\theta$
5. Repeat steps 1-4 hundreds of times
6. Compute the proportion of times where  $CI_{x\%}(\hat{\theta})$  contains the true value for parameter  $\theta$

---

[Work on TMB code in groups of 2 for 20 more minutes]

## Homework assignment:

---

- Due at beginning of Lab #2
- Must turn in your own code
- Cannot cut-paste any code from other students
  - You can hand-write your own code while working with someone else, or looking at my example code