

FISH 604

Module 4:

Statistical estimation II

Instructor: Franz Muetter
Lena Point, Rm 315
796-5448
fmueter@alaska.edu

Review/Preview: Estimation methods

- Ordinary least-squares (OLS)
 - Weighted least-squares (WLS)
 - Generalized least-squares (GLS)
-
- Today { ■ Maximum likelihood estimation (MLE)

Objectives & Outcomes

- You should understand ...
 - ... the concepts of "likelihood", maximum likelihood estimation, and likelihood profiles
- You should be able to...
 - ... compute simple likelihoods and a likelihood profile or likelihood surface

Maximum likelihood estimation

- Fit data to a model by maximizing the likelihood (probability) $L(\theta)$ of the data
 - Requires a model (with a set of parameters θ) that describe the process generating the observations
 - Requires specification of the probability distribution of the data
- Assumptions:
 - Independent, identically distributed observations (assumption can be relaxed if we can describe dependence using suitable parameterization)
 - Data follow a known distribution

Maximum likelihood estimation

- Probability of the data given parameter(s) θ :

$$\Pr(x_1, x_2, \dots, x_n) = f_D(x_1, \dots, x_n | \theta)$$

- where:

- x_1, x_2, \dots, x_n is an iid sample

- f_D is a **probability density function** which depends on parameter(s) θ (e.g. normal, binomial, gamma, etc)

- Once we collect data, the data are given and we find the set of parameters θ that maximizes the likelihood, i.e. that make the observed data "most probable"

$$L(\theta | x_1, x_2, \dots, x_n)$$

Same function as f_D above, but looked at from 'perspective' of parameters

Example: Poisson likelihood

- Data: Number of sharks caught on halibut longline per 24 hour set. We have data for a single set with $k_1 = 4$ sharks and would like to estimate the catch rate λ (#/24 hr)

$$\text{Likelihood}(\text{parameter} \mid \text{data}) = L(\lambda \mid k_i)$$

- This is the likelihood that λ is the correct parameter (catch rate), given that we caught k sharks. (We observe k and would like to estimate λ).
- Our best estimate (given one data point k_1) is of course a catch rate of $\hat{\lambda} = k_1 = 4$
- But, can we say anything about the uncertainty in this estimate?
 - We can, if we assume the catch rate can be approximated by a Poisson distribution (i.e. a rare, random event)

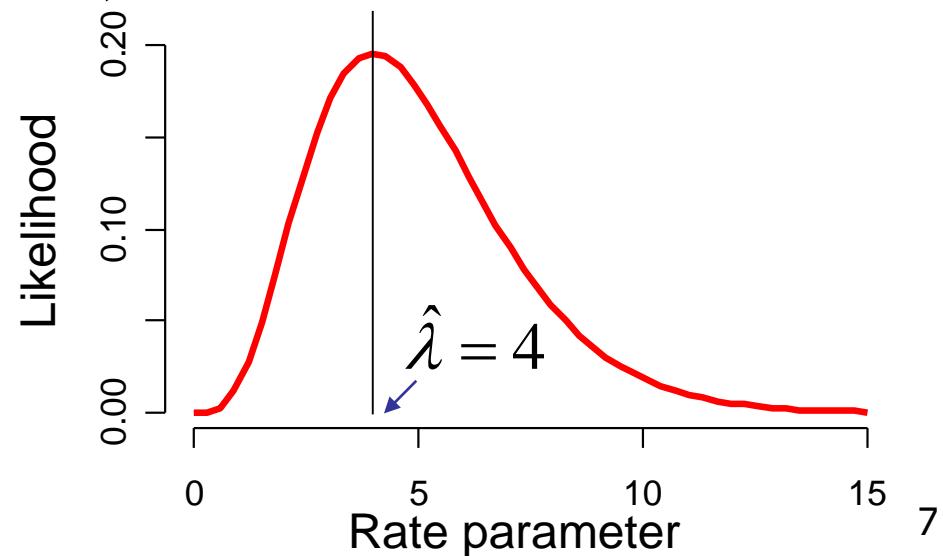
Example: Poisson likelihood

- Data: Number of sharks caught on halibut longline per 24 hour set. We have data for a single set with $k_1 = 4$ sharks and would like to estimate the catch rate λ (#/24 hr)

$$\text{Likelihood}(\text{parameter} \mid \text{data}) = L(\lambda \mid k_i)$$

- This is the likelihood that λ is the correct parameter (catch rate), given that we caught k sharks. (We observe k and would like to estimate λ).

The maximum likelihood estimate is the value of λ that maximizes the Poisson likelihood. Plotting the likelihood as a function of λ , assuming a Poisson distribution for k , gives us a sense of the range of rate parameters that are probable.

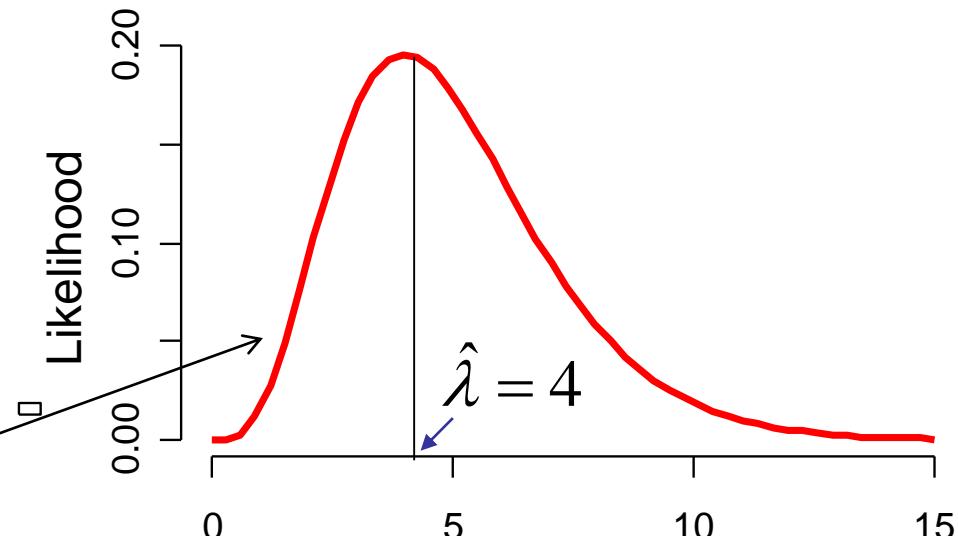


Example: Poisson likelihood

Likelihood

$$L(\lambda | 4) = \frac{e^{-\lambda} \lambda^4}{4!}$$

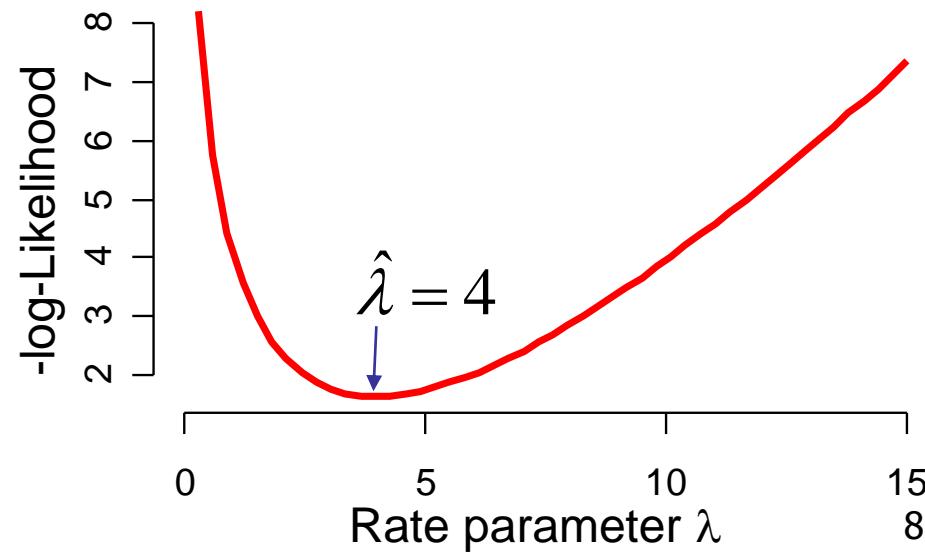
(evaluated over a range of values of λ)



Negative log-likelihood

$$\begin{aligned} -\log L(\lambda | 4) &= -\log \left(\frac{e^{-\lambda} \lambda^4}{4!} \right) \\ &= \lambda - 4 \cdot \log(\lambda) + \log(4!) \\ &\approx \lambda - 4 \cdot \log(\lambda) \end{aligned}$$

(Constants may be dropped. Caution when comparing likelihoods!)



Example: Poisson likelihood

Multiple samples: Assume we did 3 longline sets that caught $k_1 = 4$, $k_2 = 7$, and $k_3 = 2$ sharks. What is our best estimate of the rate parameter and what's a likely range of parameter values?

We assume each set was done independently (i.e. the probability of catching k sharks in one set doesn't hold any information about the probability of catching sharks in the second set), then the probability of the data for a given rate parameter is:

Likelihood:

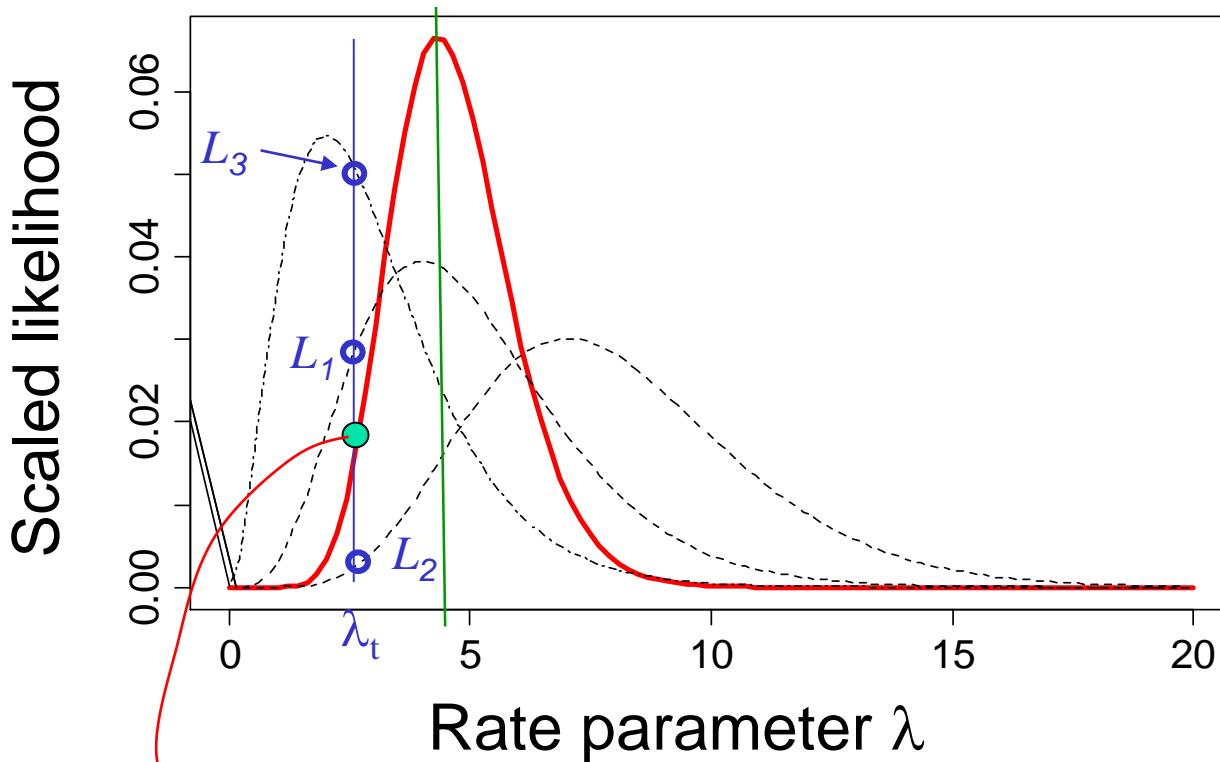
$$\begin{aligned} L(\lambda | k_1, k_2, k_3) &= L(\lambda | k_1) \cdot L(\lambda | k_2) \cdot L(\lambda | k_3) \\ &= L_1 \cdot L_2 \cdot L_3 \end{aligned}$$

Negative log-likelihood:

$$-\log L(\lambda | k_1, k_2, k_3) = -(\log L_1 + \log L_2 + \log L_3)$$

Example: Poisson likelihood

$$\text{Likelihood}(\lambda | k_1, k_2, k_3)$$



Maximum
likelihood
Estimate
(MLE):

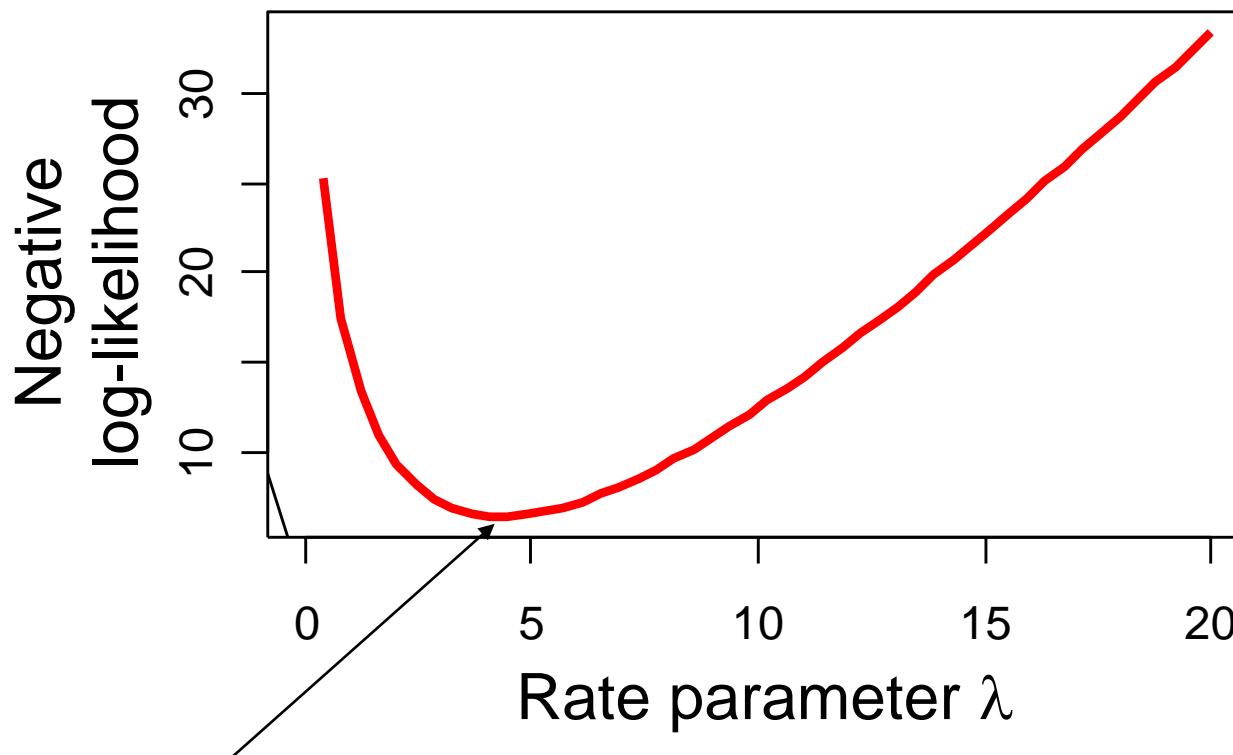
$$\hat{\lambda} = 4.33$$

Total likelihood: $L(\lambda_t) = L_1(\lambda_t) * L_2(\lambda_t) * L_3(\lambda_t)$

Example: Poisson likelihood

Negative log-likelihood:

$$-\log L(\lambda | k_1, k_2, k_3) = -(\log L_1 + \log L_2 + \log L_3)$$



(Find minimum of negative log-likelihood)

Maximum likelihood estimate:

$$\hat{\lambda} = 4.33$$

Solution in this case is rather trivial:
 $= (4+7+2)/3$
 $= 4.33$

Maximum likelihood estimation

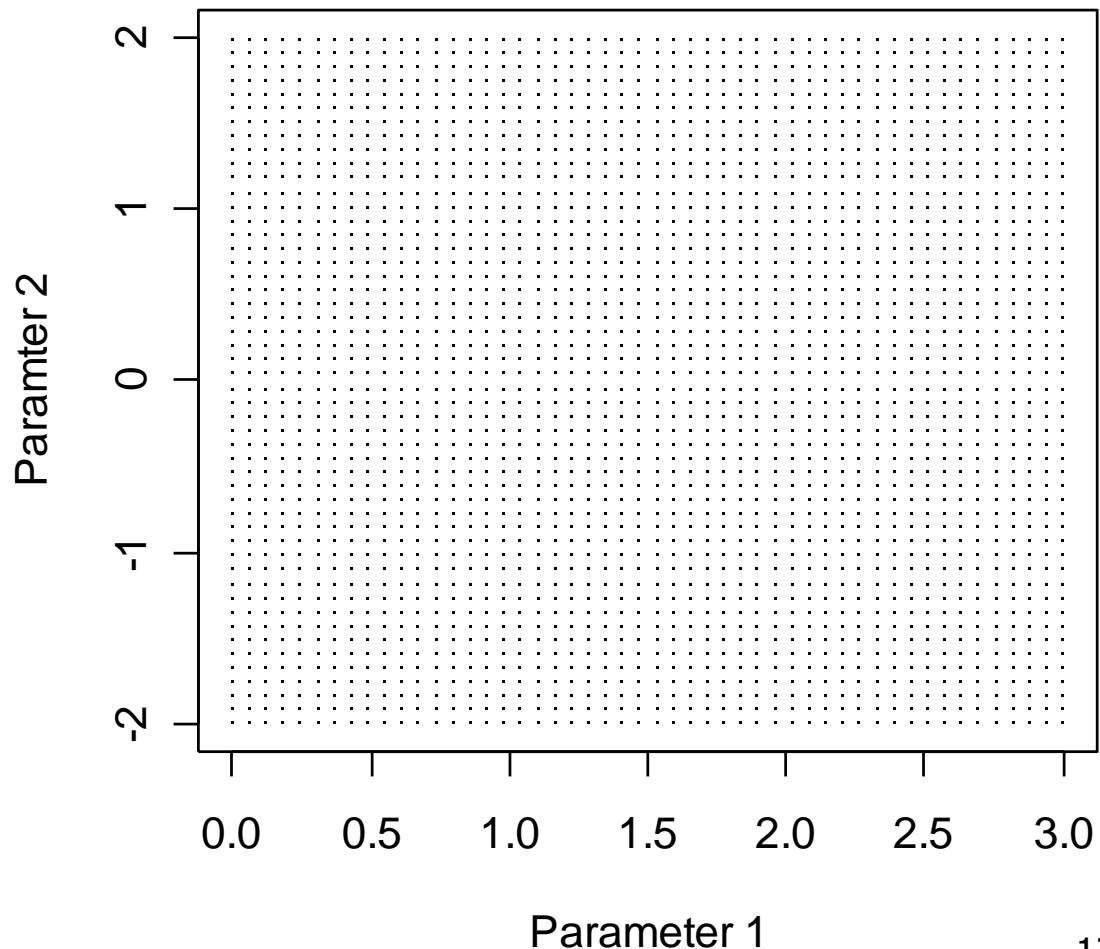
- In simple cases (such as Poisson case), the maximum likelihood estimate can be obtained by simple algebra
- Typically, we wish to maximize likelihood over several (or numerous) parameters at once.
- In most cases, numerical methods are required
 - Most numerical algorithms are designed to find the minimum of a function (also called 'objective function') over the "parameter space", i.e. over all possible combinations of parameters
 - Therefore, maximum likelihood routines search for the minimum of the negative log-likelihood
 - Grid searches
 - Gradient methods

Maximum likelihood estimation

■ Grid search:

Compute likelihood as a function of parameters for each combination of parameter values and find the value that maximizes likelihood!

Make sure to cover the range of possible (reasonable) parameter values



Maximum likelihood estimation

- Poisson example showed how to find Maximum Likelihood Estimate of a simple distributional parameter (λ)
- What about fitting a model that relates the rate λ to potential explanatory variables?
 - Principle is exactly the same, but the distributional parameter itself (i.e. the rate parameter λ) now becomes a function of other model parameters
 - Problem becomes finding the combination of model parameters that maximizes likelihood

Example: Poisson regression

- Problem: Model shark catch rate (# of sharks per set) as a function of depth, salinity, time of year, etc.
- Assume Poisson distribution for catch, but actual catch rate (λ) may vary with depth and other covariates:

Linear model: catch rate = $\alpha + \beta_1 * (\text{depth}) + \beta_2 * (\text{depth})^2$

or: $E(k) = \lambda$

$$= \alpha + \beta_1 \text{depth} + \beta_2 \text{depth}^2$$

Example: Poisson regression

Model:

$$\lambda = \alpha + \beta_1 x + \beta_2 x^2$$

Data: n longline sets with $k_1, k_2, k_3, \dots, k_n$ sharks

Distribution of data: $\text{Poisson}(\lambda)$

Likelihood for one data point i with observed catch k_i and catch rate λ_i :

$$L(\lambda_i | k_i) = \frac{e^{\lambda_i} \lambda_i^{k_i}}{k_i!} = \frac{e^{\alpha + \beta_1 x_i + \beta_2 x_i^2} (\alpha + \beta_1 x_i + \beta_2 x_i^2)^{k_i}}{k_i!}$$

Example: Poisson regression

Likelihood of all the data:

$$L(\lambda | k_1, k_2, \dots, k_n) = \frac{e^{\lambda_1} \lambda_1^{k_1}}{k_1!} \times \frac{e^{\lambda_2} \lambda_2^{k_2}}{k_2!} \times \cdots \times \frac{e^{\lambda_n} \lambda_n^{k_n}}{k_n!}$$

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

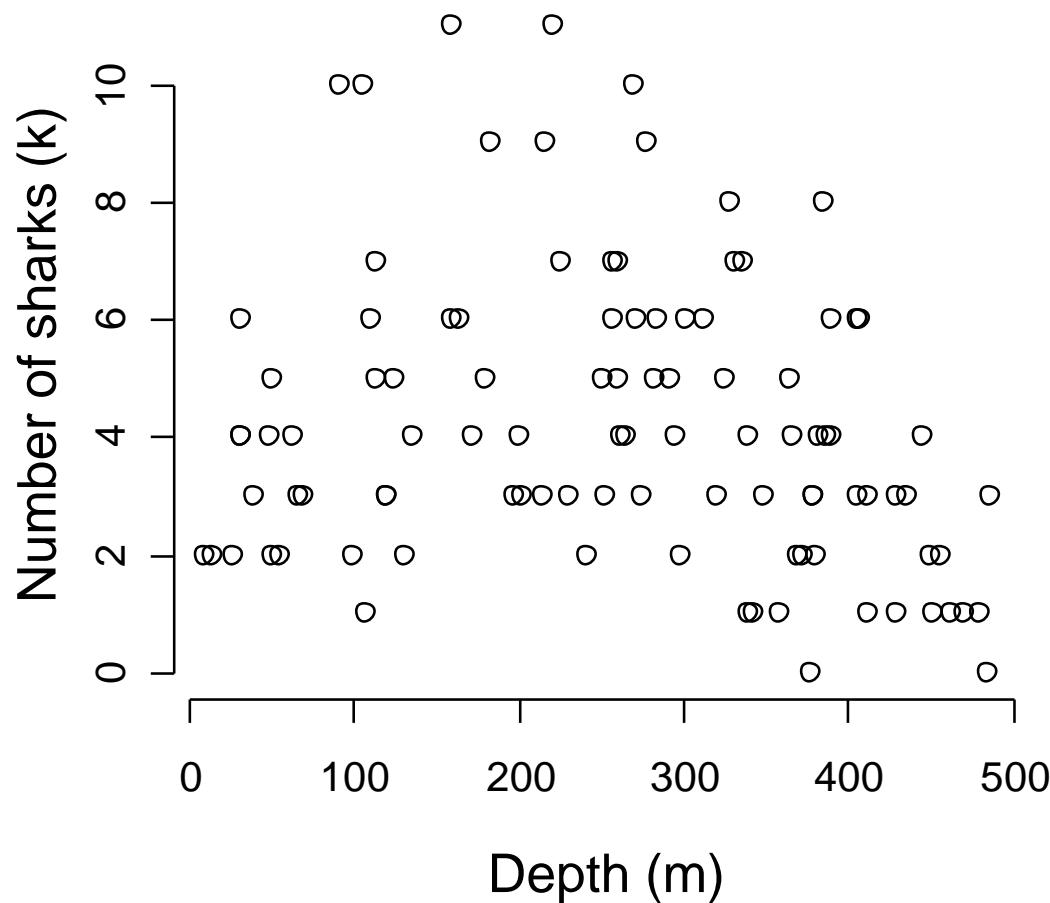
$$= \prod_{i=1}^n \frac{e^{\lambda_i} \lambda_i^{k_i}}{k_i!}$$

$$= \prod_{i=1}^n \frac{e^{\alpha + \beta_1 x_i + \beta_2 x_i^2} (\alpha + \beta_1 x_i + \beta_2 x_i^2)^{k_i}}{k_i!}$$

(In practice, we take the log and minimize the sum of negative log-likelihoods)

Example: Poisson regression

- Data ($n = 100$ observations for k):



Example: Poisson regression

Evaluate negative log-likelihood at different parameter combinations β_1, β_2

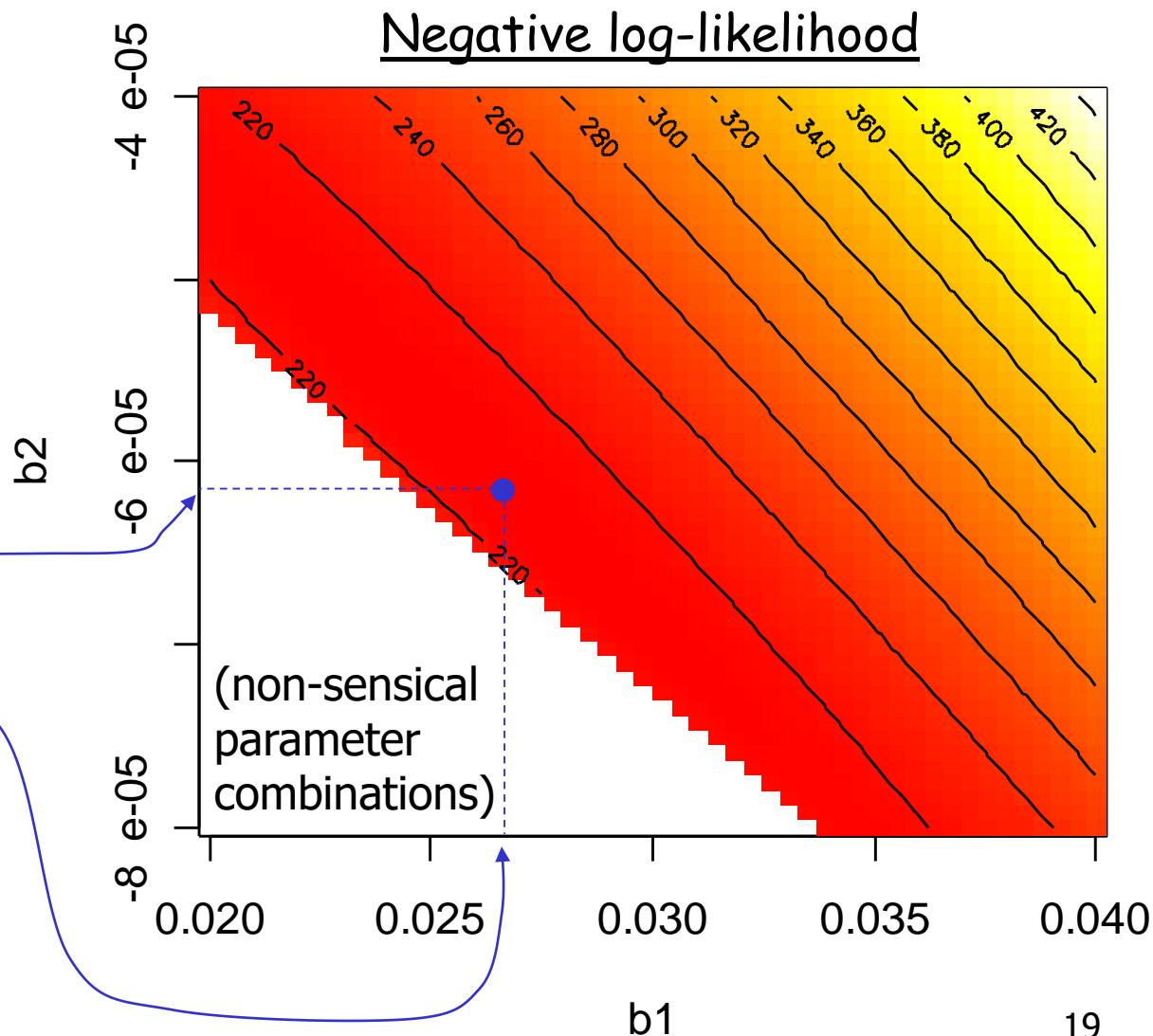
Find minimum:

$$\hat{\beta}_2 = -0.0000616$$

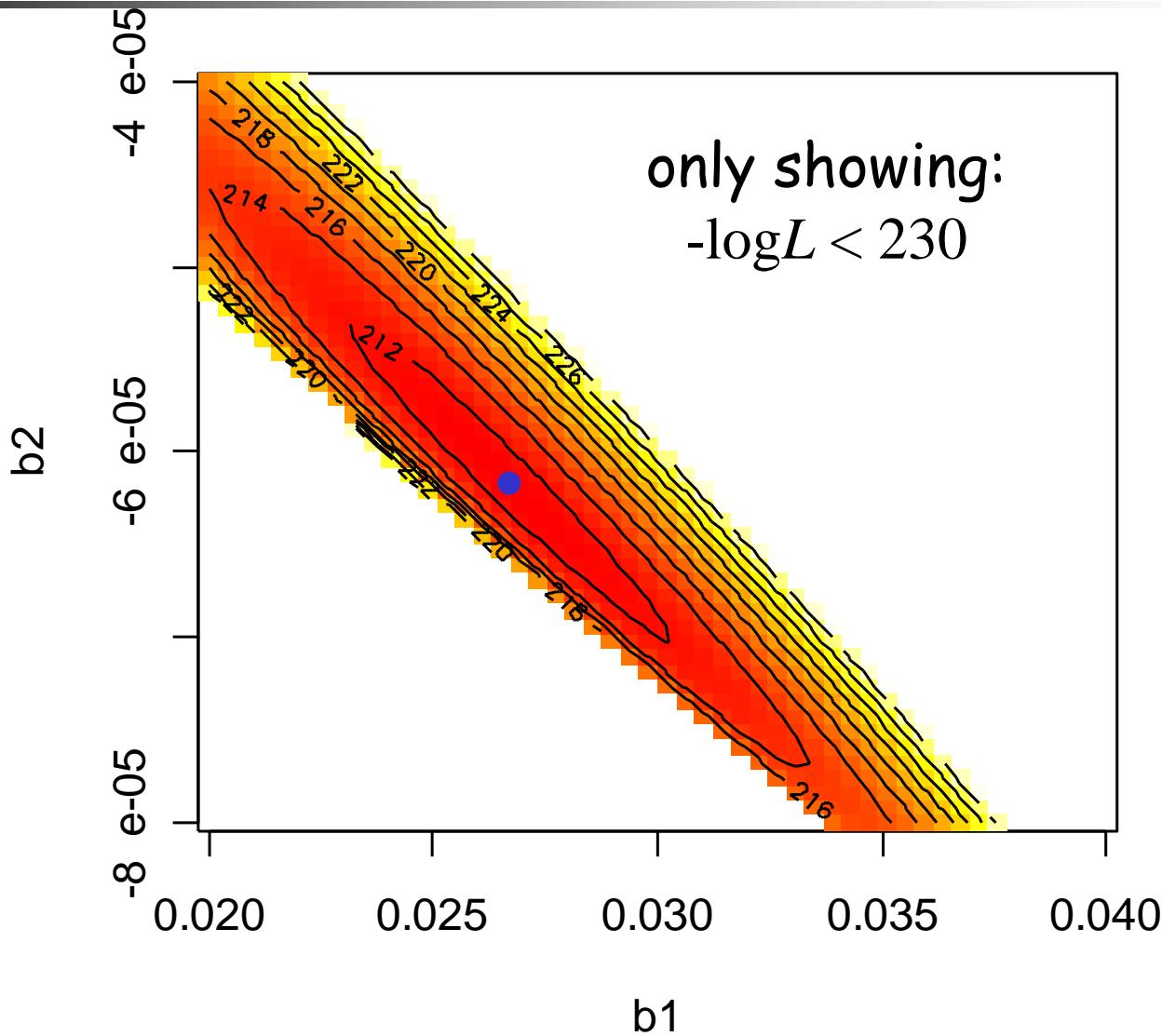
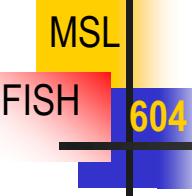
$$\hat{\beta}_1 = 0.0267$$

$$\hat{\alpha} = 2.557$$

(α fixed for this plot)



Example: Poisson regression



Example: Poisson regression

- Fitted model: (Estimated mean number of sharks at depth: λ)

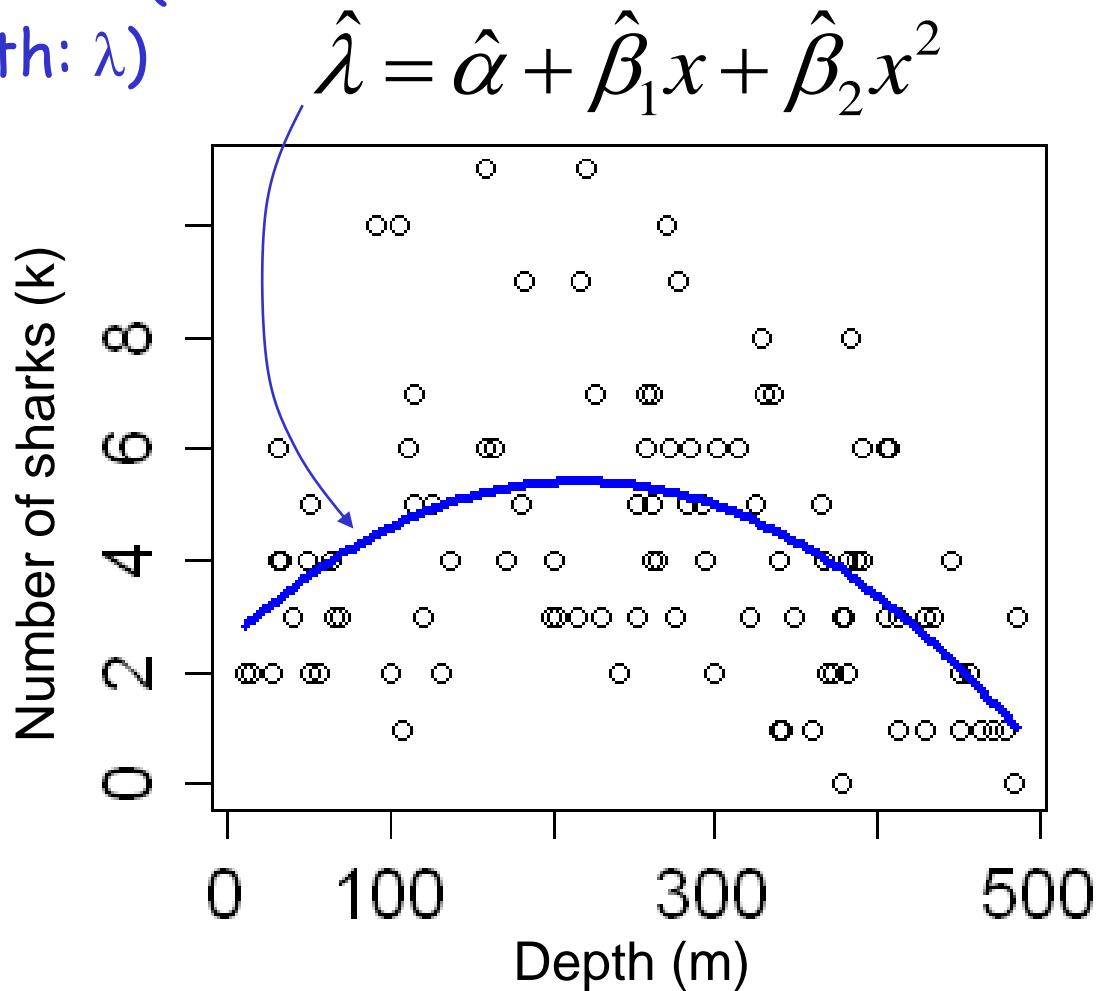
$$\hat{\lambda} = \hat{\alpha} + \hat{\beta}_1 x + \hat{\beta}_2 x^2$$

Log-likelihood was minimized for:

$$\hat{\alpha} = 2.557$$

$$\hat{\beta}_1 = 0.0267$$

$$\hat{\beta}_2 = -0.00000616$$



Example: Poisson regression

- In R, Poisson regression can be fit via maximum likelihood using function:
`glm()` (for Generalized Linear Models)
- We will work through the complete R code for fitting the Poisson regression example in lab (*Lab 4.R*)

Normal likelihood

- In the case of normally distributed data, **least-squares regression** and **maximum likelihood estimation** are (nearly) equivalent!
- Recall the pdf of a normal distribution, which describes the probability of a random variable having a particular value y , given parameters μ (mean) and σ^2 (variance) of the normal distribution:

$$\Pr(y | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

➡ Again, looking at this from a likelihood perspective, we make this a function of the parameters...

Normal likelihood

- In the case of normally distributed data, **least-squares regression** and **maximum likelihood estimation** are (nearly) equivalent!
- Recall the pdf of a normal distribution, which describes the probability of a random variable having a particular value y , given parameters μ (mean) and σ^2 (variance) of the normal distribution:

$$L(\mu, \sigma^2 | y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

Normal likelihood

Likelihood given a single observation:

$$L(\mu, \sigma^2 | y_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

Likelihood given multiple observation:

$$L(\mu_i, \sigma^2 | \mathbf{y}) = \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right)$$

Data vector

In regression, the mean is modeled as a function of independent variables x_i : $\mu_i = f(x_i / \text{parameters})$ and L is maximized over the "parameter space" (all possible combinations of the parameters)

Normal likelihood

Negative log-likelihood, normal distribution:

$$\begin{aligned}
 -\log L(\mu, \sigma^2 | \mathbf{y}) &= -\log \left(\prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{(y_i - \mu)^2}{2\sigma^2} \right) \right) \\
 &= n \log(\sigma) + \frac{n}{2} \log(2\pi) + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \\
 &= \text{constant} + \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^n (y_i - \mu)^2}_{=\text{RSS}}
 \end{aligned}$$

→ For normal distribution, maximizing the likelihood is equivalent to minimizing the residual sum of squares

Suggested Reading

- Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* 47: 90-100