2 Profile Likelihood

The term "profile likelihood" can mean multiple things.

In all cases, this is a way to get a point estimate and assess uncertainty in a proportion of the parameter vector while essentially ignoring the other parameters.

2.1 Analytical Methods via Profile Likelihoods

In certain problems it is possible to maximize the log likelihood for part of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ without actually knowing the value of the other part.

$\underline{\theta}_1$

Say $\tilde{\underline{\theta}}_2(\underline{\theta}_1)$ maximizes the log likelihood for any value of $\underline{\theta}_1$.

For any $\underline{\theta}_1$, $\tilde{\underline{\theta}}_2(\underline{\theta}_1)$ maximizes the likelihood wrt $\underline{\theta}_2$.

$\tilde{\underline{\theta}}_2(\underline{\theta}_1)$

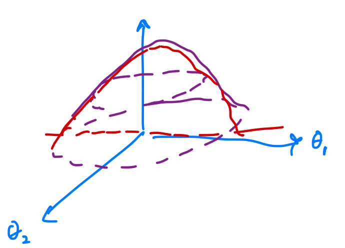
that maximizes part of the likelihood

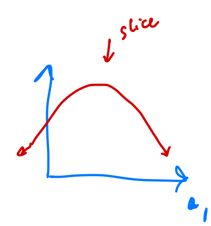
The **profile likelihood** is the usual likelihood with the known function of part of the parameter vector inserted for that parameter, making the likelihood only a function of one part of the vector.

$\underline{\theta}_2$

$\underline{\theta}_1$

i.e. the profile likelihood is $L(\underline{\theta}_1, \tilde{\underline{\theta}}_2(\underline{\theta}_1))$ is a function of only $\underline{\theta}_1$ (lower dimension).

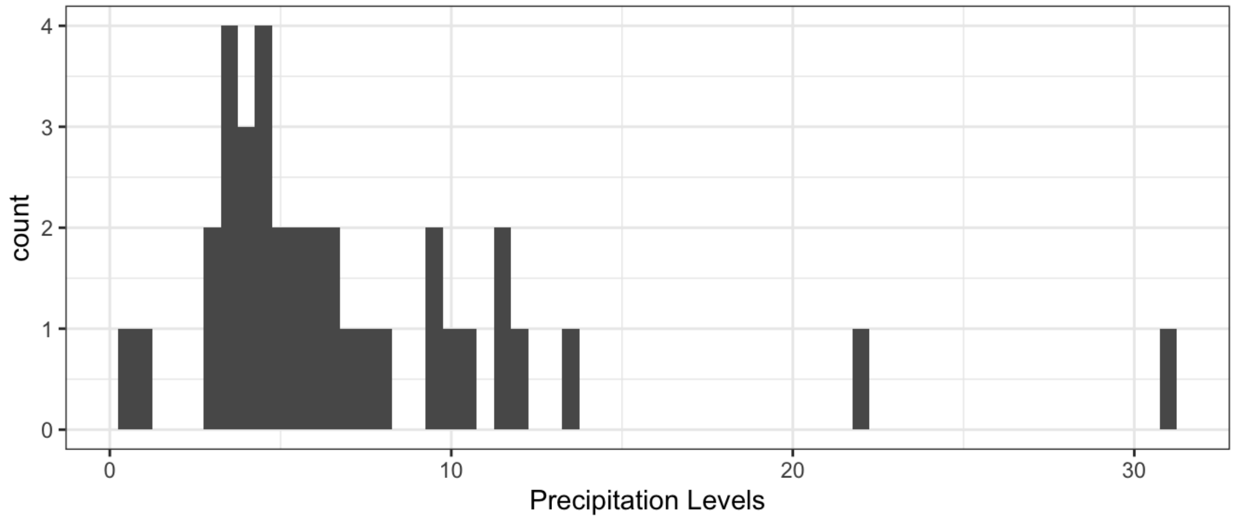




Then we need only maximize $L(\underline{\theta}_1, \tilde{\underline{\theta}}_2(\underline{\theta}_1))$ wrt $\underline{\theta}_1$ to get $\hat{\underline{\theta}}_1 \Rightarrow \hat{\underline{\theta}}_2 = \tilde{\underline{\theta}}_2(\hat{\underline{\theta}}_1)$

In this case, we are assuming we can write $\tilde{\underline{\theta}}_2(\underline{\theta}_1)$ as an analytical function.

Example (Hurricane Data, Cont'd): For 36 hurricanes that had moved far inland on the East Coast of the US in 1900-1969, maximum 24-hour precipitation levels during the time they were over mountains.



We modeled the precipitation levels with a gamma distribution, which has log likelihood

$$\ell(\alpha, \beta) = -n \log \Gamma(\alpha) - n\alpha \log \beta + (\alpha - 1) \sum \log Y_i - \frac{\sum Y_i}{\beta}$$

Take a partial derivative wrt β:

$$\frac{\partial \ell(\alpha,\beta)}{\partial \beta} = S_2(\alpha,\beta) = -\frac{n\alpha}{\beta} + \frac{\sum Y_i}{\beta^2} \overset{\text{set}}{=} 0$$

$$\Rightarrow \tilde{\beta}(\alpha) = \frac{\bar{Y}}{\alpha}$$

We can substitute this back into $\ell(\alpha,\beta)$:

$$\ell(\alpha, \tilde{\beta}(\alpha)) = -n \log \Gamma(\alpha) - n\alpha\left(\log \bar{Y} - \log \alpha\right) + (\alpha-1)\sum \log Y_i - n\alpha.$$

"profile log-likelihood"

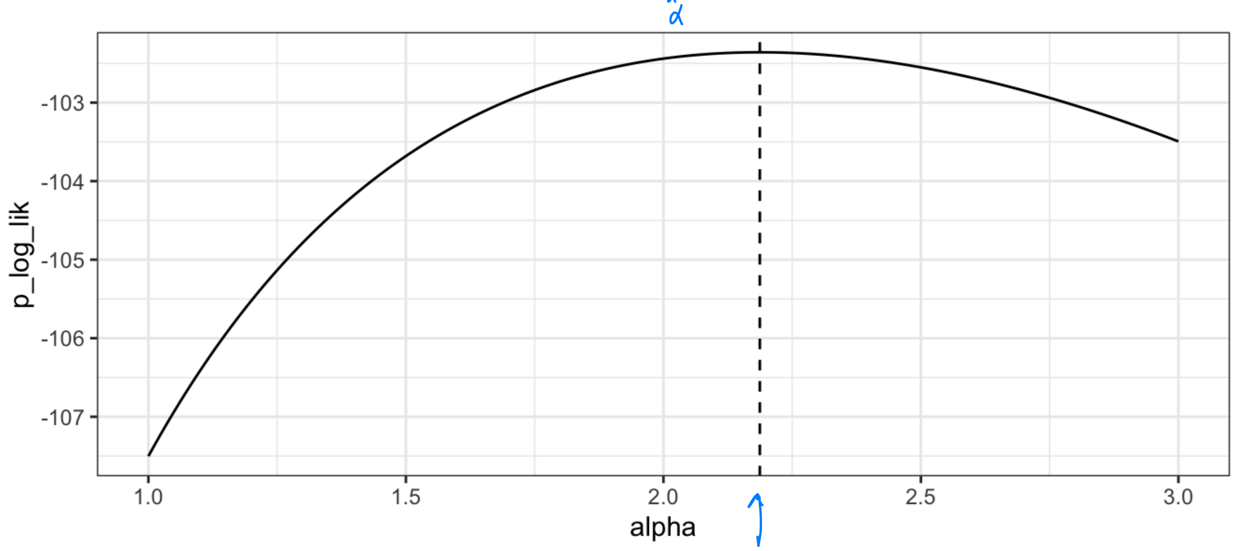
only a function of α (not β).

```
gamma_prof_loglik <- function(alpha, data) {
  beta <- mean(data) / alpha
  sum(dgamma(data, alpha, scale = beta, log = TRUE))
}

## get maximum profile likelihood estimate
alpha_mple <- optim(1, gamma_prof_loglik, data = hurr_rain, method =
        "BFGS", control = list(fnscale = -1))

## plot profile likelihood
data.frame(alpha = seq(1, 3, length.out = 200)) |>
  rowwise() |>
  mutate(p_log_lik = gamma_prof_loglik(alpha, hurr_rain)) |>
  ggplot() +
  geom_line(aes(alpha, p_log_lik)) +
  geom_vline(aes(xintercept = alpha_mple$par), lty = 2)
```

$\hat{\alpha}$



$\hat{\alpha} = 2.19$

$\Rightarrow \hat{\beta} = \frac{7.29}{2.19} = 3.33$

Same values we found before by maximizing $\ell(\alpha, \beta)$ in 2 dimensions.

but we only needed to optimize in 1 dimension!

↗ maybe we can't write $\tilde{\theta}_2(\underline{\theta}_1)$ as an analytical function.

2.2 Numerical Methods via Profile Likelihoods

The log likelihood can be maximized over one portion of the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ for any fixed value of the other, even if that maximization cannot be expressed as an explicit function.

↳ in other words, this can still be useful if both optimizations are done numerically.

This is the most commonly found situation for profile likelihood methods.

We can more formally define a profile likelihood as

$$L^p(\boldsymbol{\theta}_1) = \max_{\boldsymbol{\theta}_2} L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \quad \text{for any } \underline{\theta}_1 \in \Theta_1$$

then the log profile likelihood is

$$\ell^p(\underline{\theta}_1) = \max_{\underline{\theta}_2} \log L(\underline{\theta}_1, \underline{\theta}_2)$$

The profile likelihood and log profile likelihood behave in many ways like true likelihood functions:

1. The estimate of $\boldsymbol{\theta}_1$ found by maximizing $L^p(\boldsymbol{\theta}_1)$ is the MLE of $\boldsymbol{\theta}_1$.

MPLE → $$\max_{\theta_1} L^p(\underline{\theta}_1) = \max_{\theta_1} \max_{\theta_2} L(\underline{\theta}_1, \underline{\theta}_2) = \max_{\theta_1, \theta_2} L(\underline{\theta}_1, \underline{\theta}_2)$$

simultaneous MLE

2. A likelihood ratio test statistics formed with the profile likelihood has a limiting χ^2 distribution. For $\dim(\underline{\theta}_2) = p-r$, $\dim(\underline{\theta}_1) = r$

$$T(\underline{\theta}_1) = -2\left(\ell^p(\underline{\theta}_1) - \ell^p(\hat{\underline{\theta}}_{1MPLE})\right) \xrightarrow{d} \chi^2_r \quad \text{for any fixed } \underline{\theta}_1 \in \Theta_1.$$

3. A profile likelihood confidence region is a valid approximate confidence region for $\boldsymbol{\theta}_1$.

gives approximately the right coverage

$$CI: \left\{ \underline{\theta}_1^0 : -2\left[\ell^p(\underline{\theta}_1^0) - \ell^p(\hat{\underline{\theta}}_{1MPLE})\right] \le \chi^2_{r, 1-\alpha} \right\}$$

Where does this confidence region come from?

This is an inverted profile likelihood ratio test.

let $r=1$, then look at profile likelihood ratio test $H_0: \theta_1^0$ is the true parameter.

Then $\lambda = -2\left[\ell^P(\theta_1^0) - \ell^P(\hat{\theta}_{1,MPLE})\right] \overset{\circ}{\sim} \chi_1^2$ (asymptotically based on properties of LRT).

quantile of $\chi_1^2 = 3.84$.

$$\Longrightarrow P\left(-2\left[\ell^P(\theta_1^0) - \ell^P(\hat{\theta}_{1,MPLE})\right] > q_{0.95}\right) \approx 0.05$$

$$\Longrightarrow -2\left[\ell^P(\theta_1^0) - \ell^P(\hat{\theta}_{1,MPLE})\right] > 3.84$$

$$\vdots$$

$$\ell^P(\theta_1^0) < \ell^P(\hat{\theta}_{1,MPLE}) - 1.92$$

solve to get an confidence interval uses more of the likelihood surface than based on Fisher Information (which will be symmetric).

However, these are *not* full likelihood functions.

The derivatives of profile likelihood don't behave like the derivatives of full likelihoods:

$$E\,\frac{\partial \ell^P(\underline{\theta}_1)}{\partial \underline{\theta}_1} \neq 0 \text{ necessarily!}$$

When we hold $\underline{\theta}_2$ fixed, the uncertainty in estimator of $\underline{\theta}_2$ is ignored in the uncertainty of estimation of $\underline{\theta}_1$.

$\Longrightarrow$ There is not a "Wald-type" theory for profile-likelihood estimates.

asymptotically Normal