2 Profile Likelihood

The term "profile likelihood" can mean multiple things.

In all cases, this is a way to assess uncertainty (or get a point estimate) in a portion of the parameter while essentially ignoring the others.

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

i.e. for any value $\underline{\theta}_1$, $\underline{\tilde{\theta}}_2(\underline{\theta}_1)$ maximizes the likelihood wrt $\underline{\theta}_2$.

$\underline{\tilde{\theta}}_2(\underline{\theta}_1)$ that maximized part of the likelihood

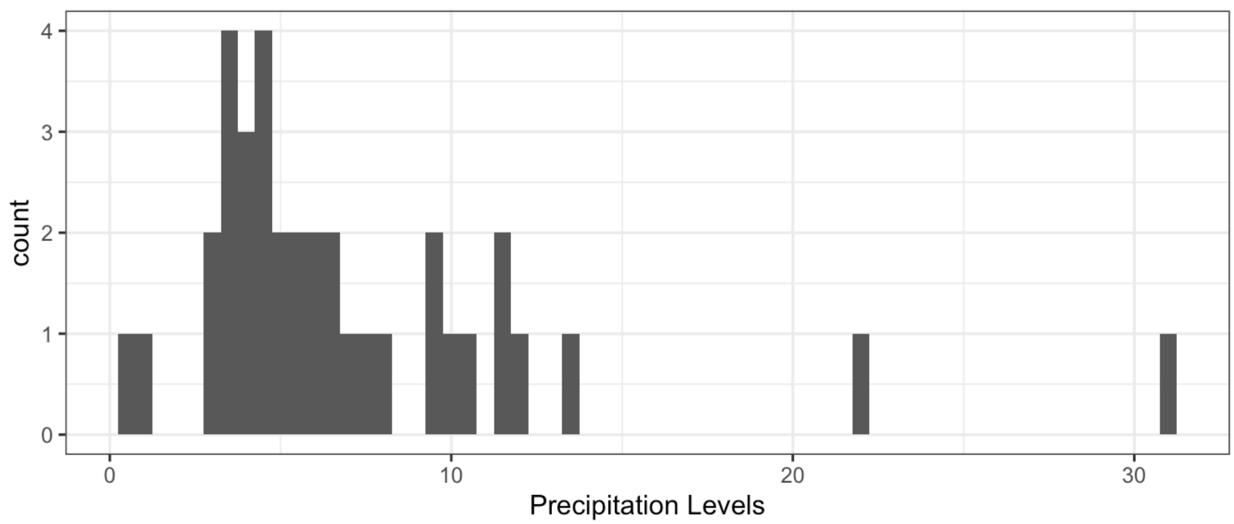
The **profile likelihood** is the usual likelihood with the known function of part of the parameter vector inserted for that parameter, making the likelihood only a function of one part of the vector.
$\underline{\theta}_2$
$\underline{\theta}_1$

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

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

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

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

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

Taking a partial derivative wrt β:

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

$$\Rightarrow \tilde{\beta}(\alpha) = \frac{\bar{Y}}{\alpha}$$ and we can substitute this back into $\ell(\alpha,\beta)$:

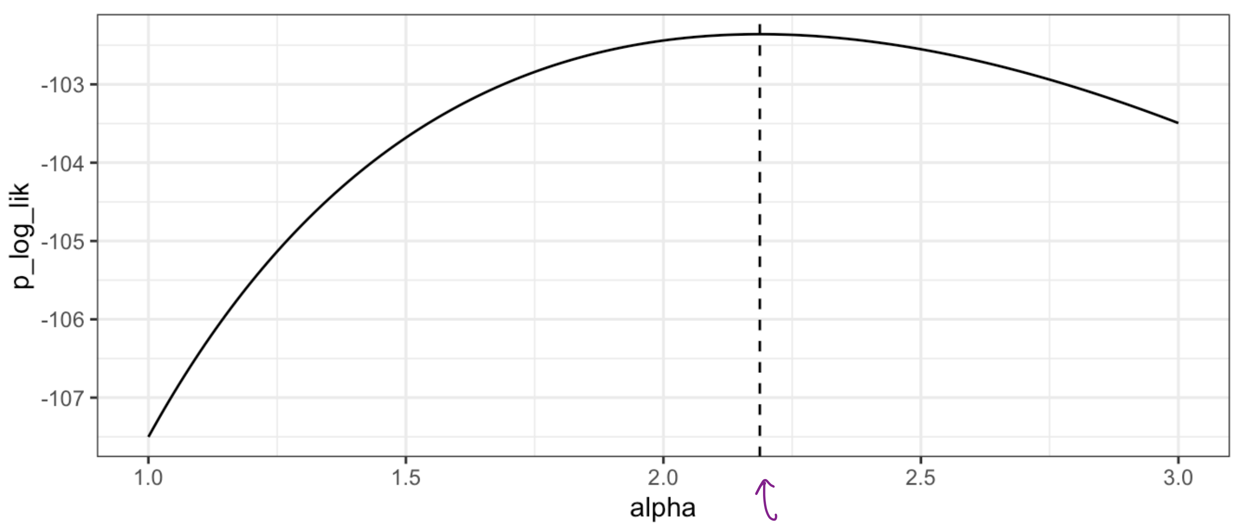
$$\ell(\alpha, \tilde{\beta}(\alpha)) = -n \log \Gamma(\alpha) - n\alpha(\log \bar{Y} - \log \alpha) + (\alpha - 1)\sum \log Y_i - n\alpha$$

"profile log-likelihood"

```
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)
```



$\hat{\alpha} = 2.19$

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

Same values we found before by maximizing $\ell(\alpha, \beta)$ in 2 dimensions, but we only needed to optimize a function in 1 dimension.

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

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

Then the log profile likelihood is

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

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 $\boldsymbol{\theta}_1$. ↙ simultaneous.

$$\max_{\underset{\sim}{\theta}_1} L^p(\underset{\sim}{\theta}_1) = \max_{\underset{\sim}{\theta}_1} \max_{\underset{\sim}{\theta}_2} L(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2)$$
$$= \max_{\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2} L(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2).$$

2. A likelihood ratio test statistics formed with the profile likelihood has a limiting χ^2 distribution.

For $\dim(\underset{\sim}{\theta}_2) = p-r$
$\dim(\underset{\sim}{\theta}_1) = r$

$$T(\underset{\sim}{\theta}_1) = -2\Big[\ell^p(\underset{\sim}{\theta}_1^\circ) - \underbrace{\ell^p(\hat{\underset{\sim}{\theta}}_{1,MLE})}_{= \ell(\hat{\underset{\sim}{\theta}}_{1MLE}, \hat{\underset{\sim}{\theta}}_{2MLE})}\Big] \xrightarrow{d} \chi^2_r \quad \text{for any fixed } \underset{\sim}{\theta}_1^\circ \in \Theta_1,$$

3. A profile likelihood confidence region is a $\underbrace{\text{valid approximate}}_{\text{gives approximately right coverage.}}$ confidence region for $\boldsymbol{\theta}_1$.

$$\text{CI}: \left\{ \underset{\sim}{\theta}_1^\circ : -2\left[\ell^p(\underset{\sim}{\theta}_1^\circ) - \ell^p(\hat{\underset{\sim}{\theta}}_{1,MLE})\right] \le \chi^2_{r,1-\alpha} \right\}$$

Where does this confidence region come from?

This is an inverted profile likelihood ratio test.

Let $r=1$, then look at the profile likelihood ratio test: $H_0: \underline{\theta}_1^0$ is the true parameter.

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

quantile of $\chi_1^2 = 3.84$

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

$$\Rightarrow \quad -2\left[\ell^p(\theta_1^0) - \ell^p(\hat{\theta}_{1,MLE})\right] > 3.84$$

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

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

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

However, these are *not* full likelihood functions.

The derivatives of profile likelihoods don't behave like the derivatives of full likelihoods, e.g.

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

When we hold $\underline{\theta}_2$ fixed, the uncertainty in estimator of θ_2 is ignored in the uncertainty of estimation of $\underline{\theta}_1$.

$\Rightarrow$ there is not a "Wald-type" (Normal) theory for profile likelihood estimates.