Various ‘types’ of likelihood

1. likelihood, marginal likelihood, conditional likelihood, profile likelihood, adjusted profile likelihood
2. semi-parametric likelihood, partial likelihood
3. empirical likelihood, penalized likelihood
4. quasi-likelihood, composite likelihood
5. simulated likelihood, indirect inference
6. bootstrap likelihood, *h*-likelihood, weighted likelihood, pseudo-likelihood, local likelihood, sieve likelihood

October 23

- consistency of mle (sketch and refs, use draft)
- careful derivation of limit theorem
- limit theorem for Bayes posterior; Laplace approx
- reminder: profile and limit theorems (handout)
- marginal and conditional likelihood (slides)
- adjusted profile likelihood (slides)

Asymptotics for Bayesian inference

- $\pi(\theta \mid y) = \dfrac{\exp\{\ell(\theta; y)\}\pi(\theta)}{\int \exp\{\ell(\theta; y)\}\pi(\theta) d\theta}$

- expand numerator and denominator about $\hat{\theta}$, assuming $\ell'(\hat{\theta}) = 0$

- $\pi(\theta \mid y) \doteq N\{\hat{\theta}, j^{-1}(\hat{\theta})\}$ "data swamps the prior"

- a similar argument would give $\pi(\theta \mid y) \doteq N\{\hat{\theta}_\pi, j_\pi^{-1}(\hat{\theta}_\pi)\}$,

$\hat{\theta}_\pi$ solves $h'(\theta) = 0$; $h(\theta) = \ell(\theta) + \log \pi(\theta)$

$\hat{\theta} = \hat{\theta}_\pi + O_p(n^{-1})$

Posterior is asymptotically normal

careful statement 1: Berger, Ch.4; Walker, 1969

For any $a, b \in \mathbb{R}, a < b$, let $a_n = a_n(y) = \hat{\theta}_n + aj^{-1/2}(\hat{\theta}_n)$, $b_n = b_n(y) = \hat{\theta}_n + bj^{-1/2}(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the solution of $\ell'(\theta; y) = 0$, assumed unique, and $j(\theta) = -\ell''(\theta; y)$. Then

$$\int_{a_n}^{b_n} \pi(\theta \mid y) \longrightarrow \Phi(b) - \Phi(a), \quad n \to \infty.$$

careful statement 2: Johnson, 1970

If $\pi(\theta_0) > 0$ and $\pi'(\theta)$ is continuous in a neighbourhood of θ_0, there exist constants D and n_y s.t.

$$|F_n(\xi) - \Phi(\xi)| < Dn^{-1/2}, \quad \text{for all } n > n_y,$$

on an almost-sure set with respect to $\pi(\theta_0)f(y; \theta_0)$

$y = (y_1, \ldots, y_n)$ is i.i.d. from $f(y; \theta_0)$

Laplace approximation

- $\pi(\theta \mid y) = \dfrac{\exp\{\ell(\theta)\}\pi(\theta)}{\int \exp\{\ell(\theta)\}\pi(\theta)d\theta}$
- expand denominator only about $\hat{\theta}$

$$\pi(\theta \mid y) \doteq \frac{1}{(2\pi)^{d/2}}|j(\hat{\theta})|^{+1/2}\exp\{\ell(\theta; y) - \ell(\hat{\theta}; y)\}\frac{\pi(\theta)}{\pi(\hat{\theta})}$$

- more precisely

$$\pi(\theta \mid y) = \frac{1}{(2\pi)^{d/2}}|j(\hat{\theta})|^{+1/2}\exp\{\ell(\theta; y) - \ell(\hat{\theta}; y)\}\frac{\pi(\theta)}{\pi(\hat{\theta})}\{1 + O(n^{-1})\}$$

Posterior cdf, $d = 1$

$$
\begin{aligned}
\int_{-\infty}^{\theta} \pi(\vartheta \mid y) d\vartheta &\doteq \int_{-\infty}^{\theta} \frac{1}{(2\pi)^{1/2}} |j(\hat\theta)|^{+1/2} \exp\{\ell(\theta; y) - \ell(\hat\theta; y)\} \frac{\pi(\theta)}{\pi(\hat\theta)} d\theta \\
&= \int_{\infty}^{r} \frac{1}{(2\pi)^{1/2}} e^{-r^2/2} j^{1/2}(\hat\theta) \frac{\pi(\theta)}{\pi(\hat\theta)} \frac{r}{-\ell'(\theta)} dr \\
&= \int_{\infty}^{r} \phi(r) \left(\frac{r}{q} + 1 - 1 \right) dr \\
&= \Phi(r) + \int_{\infty}^{r} r\phi(r) \left(\frac{1}{q} - \frac{1}{r} \right) dr \\
&= \Phi(r) + \phi(r) \left(\frac{1}{r} - \frac{1}{q} \right) + \int_{\infty}^{r} \phi(r) d \left(\frac{1}{r} - \frac{1}{q} \right)
\end{aligned}
$$

show that $r = q + aq^2/\sqrt{n} + bq^3/n$ implies $(1/r - 1/q)$ is linear in r; see for example BDR, §8.6 and A.2

Nuisance parameters

- $\theta = (\psi, \lambda) = (\psi_1, \ldots, \psi_q, \lambda_1, \ldots, \lambda_{d-q})$
- $U(\theta) = \begin{pmatrix} U_\psi(\theta) \\ U_\lambda(\theta) \end{pmatrix}, \qquad U_\lambda(\psi, \hat\lambda_\psi) = 0$
- $i(\theta) = \begin{pmatrix} i_{\psi\psi} & i_{\psi\lambda} \\ i_{\lambda\psi} & i_{\lambda\lambda} \end{pmatrix} \quad j(\theta) = \begin{pmatrix} j_{\psi\psi} & j_{\psi\lambda} \\ j_{\lambda\psi} & j_{\lambda\lambda} \end{pmatrix}$
- $i^{-1}(\theta) = \begin{pmatrix} i^{\psi\psi} & i^{\psi\lambda} \\ i^{\lambda\psi} & i^{\lambda\lambda} \end{pmatrix} \quad j^{-1}(\theta) = \begin{pmatrix} j^{\psi\psi} & j^{\psi\lambda} \\ j^{\lambda\psi} & j^{\lambda\lambda} \end{pmatrix}.$
- $i^{\psi\psi}(\theta) = \{i_{\psi\psi}(\theta) - i_{\psi\lambda}(\theta) i^{-1}_{\lambda\lambda}(\theta) i_{\lambda\psi}(\theta)\}^{-1},$
- $\ell_{\mathrm{P}}(\psi) = \ell(\psi, \hat\lambda_\psi), \qquad j_{\mathrm{P}}(\psi) = -\ell''_{\mathrm{P}}(\psi)$

Inference from limiting distributions, nuisance parameters

$$\begin{aligned} w_u(\psi) = U_\psi(\psi,\hat\lambda_\psi)^T\{i^{\psi\psi}(\psi,\hat\lambda_\psi)\}U_\psi(\psi,\hat\lambda_\psi) &\quad \dot\sim \quad \chi^2_q \\ w_e(\psi) = (\hat\psi-\psi)\{i^{\psi\psi}(\hat\psi,\hat\lambda)\}^{-1}(\hat\psi-\psi) &\quad \dot\sim \quad \chi^2_q \\ w(\psi) = 2\{\ell(\hat\psi,\hat\lambda)-\ell(\psi,\hat\lambda_\psi)\} = 2\{\ell_{\rm P}(\hat\psi)-\ell_{\rm P}(\psi)\} &\quad \dot\sim \quad \chi^2_q; \end{aligned}$$

Approximate Pivots, $q = 1$

$$\begin{aligned} r_u(\psi) &= \ell_{\rm P}'(\psi)j_{\rm P}(\hat\psi)^{-1/2} \;\dot\sim\; N(0,1), \\ r_e(\psi) &= (\hat\psi-\psi)j_{\rm P}(\hat\psi)^{1/2} \;\dot\sim\; N(0,1), \\ r(\psi) &= \text{sign}(\hat\psi-\psi)[2\{\ell_{\rm P}(\hat\psi)-\ell_{\rm P}(\psi)\}]^{1/2} \;\dot\sim\; N(0,1) \end{aligned}$$

all based on treating profile log-likelihood as a one-parameter log-likelihood

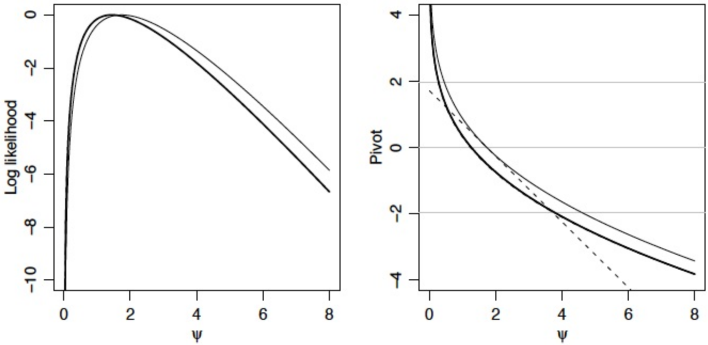


Figure 2.3: Inference for shape parameter ψ of gamma sample of size $n = 5$. Left: profile log likelihood ℓ_{p} (solid) and the log likelihood from the conditional density of u given v (heavy). Right: likelihood root $r(\psi)$ (solid), Wald pivot $t(\psi)$ (dashes), modified likelihood root $r^*(\psi)$ (heavy), and exact pivot overlying $r^*(\psi)$. The horizontal lines are at $0, \pm 1.96$.

Profile likelihood: examples

- regression

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2), \quad \psi = \sigma^2$$

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})^T(y - X\hat{\beta})$$

- Neyman-Scott

$$y_{ij} \sim N(\mu_i, \sigma^2), j = 1, \ldots, k; i = 1, \ldots, m$$

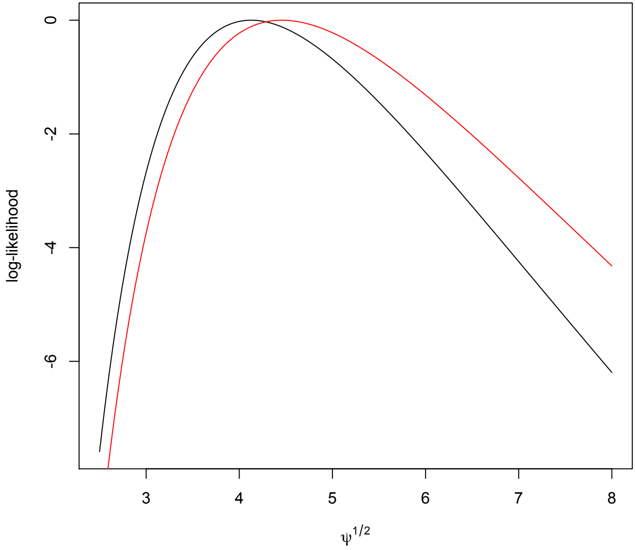
$$\hat{\sigma}^2 = \frac{1}{mk}\sum_{i=1}^{m}(y_{ij} - \bar{y}_{i.})^2$$

- 2×2 tables

$$y_{i1} \sim Bern(p_{i1}), y_{i2} \sim Bern(p_{i2}), i = 1, \ldots, n, \quad \log\{\frac{p_{i1}/(1-p_{i1})}{p_{i2}(1-p_{i2})}\} = \psi + \lambda_i$$

$$\hat{\psi} \xrightarrow{p} \psi/2$$

This is a plot of $-n \log \sigma - (y - X\hat{\beta})^T(y - X\hat{\beta})/2\sigma^2$ (black), and $-(n-p) \log \sigma - (y - X\hat{\beta})^T(y - X\hat{\beta})/2\sigma^2$ against σ (red) for given data



Eliminating nuisance parameters

- Profile likelihood poor if q large; fails if $q \to \infty$

- alternative: marginal likelihood:
  $f(\underline{y}_n; \psi, \lambda) \propto f_{\mathrm{m}}(t_1; \psi) f_{\mathrm{c}}(t_2 \mid t_1; \psi, \lambda)$ $\qquad t_j = t_j(\underline{y})$

- Example $N(X\beta, \sigma^2 I)$: $\quad f(y; \beta, \sigma^2) \propto f_{\mathrm{m}}(RSS; \sigma^2) f_{\mathrm{c}}(\hat{\beta} \mid RSS; \beta, \sigma^2)$

$$L_m(\sigma^2) \propto f_m(RSS; \sigma^2)$$

- alternative conditional likelihood:
  $f(\underline{y}; \psi, \lambda) \propto f_{\mathrm{c}}(t_1 \mid t_2; \psi) f_{\mathrm{m}}(t_2; \psi, \lambda)$

- Example 2×2 tables:
  $f(\underline{y}; \psi, \underline{\lambda}) \propto \prod_{i=1}^n f_c(y_{i1} \mid y_{i1} + y_{i2}; \psi) f_m(y_{i1} + y_{i2}; \psi, \lambda_i)$

$$L_c(\psi) = \prod f_c(y_{i1} \mid y_{i1} + y_{i2}; \psi)$$

Linear exponential families

- conditional density free of nuisance parameter
- $f(y_i; \psi, \lambda) = \exp\{\psi^T s(y_i) + \lambda^T t(y_i) - k(\psi, \lambda)\} h(y_i)$
- $f(y; \psi, \lambda) = \exp\{\psi^T \Sigma s(y_i) + \lambda^T \Sigma t(y_i) - nk(\psi, \lambda)\} \Pi h(y_i)$

Let $s = \Sigma s(y_i), t = \Sigma t(y_i)$

- $f(s, t; \psi, \lambda) = \exp\{\psi^T s + \lambda^T t - nk(\psi, \lambda)\} \tilde{h}(s)$

$$
\begin{aligned}
f(s \mid t; \psi) &= \frac{f(s, t; \psi, \lambda)}{\int f(s, t; \psi, \lambda) ds} \\
&= \frac{\exp\{\psi^T s + \lambda^T t - nk(\psi, \lambda)\} \tilde{h}(s)}{\int \exp\{\psi^T s + \lambda^T t - nk(\psi, \lambda)\} \tilde{h}(s) ds} \\
&= \frac{\exp\{\psi^T s\} \tilde{h}(s)}{\int \exp\{\psi^T s\} \tilde{h}(s) ds} \\
&= \exp\{\psi^T s - n\tilde{k}_t(\psi)\} \tilde{h}_t(s)
\end{aligned}
$$

$\tilde{k}_t$, $\tilde{h}_t$ just convenient notation for integral of denominator

Logistic regression

- $y_i \sim Binom(m_i, p_i), i = 1, \ldots, n$
- $\log\{p_i/(1 - p_i)\} = x_i^{\mathrm{T}}\beta$
- $f(y; \beta) = \exp\{\beta_1 \Sigma(x_{i1} y_i) + \cdots + \beta_p \Sigma(x_{ip} y_i) - \Sigma m_i \log(1 + e^{x_i^{\mathrm{T}}\beta})\}$
- $f_c(s_5 \mid s_{-(5)}; \beta_5) \propto \exp\{\beta_5 s_5 - \tilde{k}(\beta_5)\} h(s)$

... Logistic regression

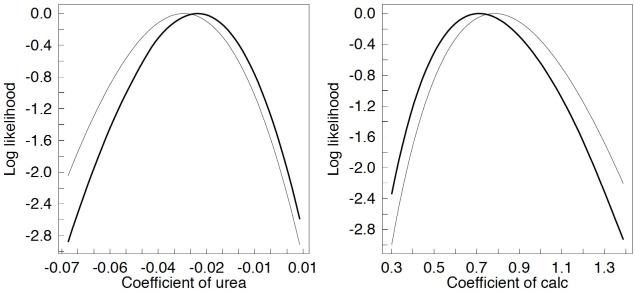


Figure 4.2: Comparison of log likelihoods for the urine data: profile log likelihood (solid line), approximate conditional log likelihood (bold line). The variables of interest are urea (left panel) and calcium concentration (right panel). The graphical output is obtained with the `plot` method of the `cond` package.

$$f_c(s_5 \mid s_{-(5)}; \beta_5) \propto \exp\{\beta_5 s_5 - \tilde{k}(\beta_5)\} h(s)$$

Summary 4.1 Approximate conditional inference for the urine data.

```
> urine.glm <- glm( formula=r~I(100*(gravity-1))+ph+osmo+conduct+urea+calc,
+                    family=binomial, data=urine )

> summary(urine.glm)

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)             0.60609    3.79582   0.160  0.87314
I(100 * (gravity - 1))  3.55944    2.22110   1.603  0.10903
ph                     -0.49570    0.56976  -0.870  0.38429
osmo                    0.01681    0.01782   0.944  0.34536
conduct                -0.43282    0.25123  -1.723  0.08493 .
urea                   -0.03201    0.01612  -1.986  0.04703 *
calc                    0.78369    0.24216   3.236  0.00121 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 105.17  on 76  degrees of freedom
Residual deviance:  57.56  on 70  degrees of freedom
AIC: 71.56

> urine.cond.urea <- cond( urine.glm, offset=urea )

> coef( urine.cond.urea )
            Estimate   Std. Error
uncond.  -0.03201315   0.01611884
cond.    -0.02759202   0.01489919

> summary( urine.cond.urea, coef=F )

Confidence intervals
--------------------
 level = 95 %
                                         lower two-sided      upper
Wald pivot                             -0.06361          -0.0004208
Wald pivot (cond. MLE)                 -0.05679           0.0016100
Likelihood root                        -0.06677          -0.0024570
Modified likelihood root               -0.05874           0.0004687
```

Summary 4.1 Approximate conditional inference for the urine data (cont.).

```
> urine.cond.calc <- cond( urine.glm, offset=calc )

> coef( urine.cond.calc )
         Estimate  Std. Error
uncond.  0.7836913   0.2421638
cond.    0.7110584   0.2282501

> summary( urine.cond.calc, coef=F )

Confidence intervals
 level = 95 %
                                       lower two-sided upper
Wald pivot                            0.3091           1.258
Wald pivot (cond. MLE)                0.2637           1.158
Likelihood root                       0.3815           1.342
Modified likelihood root              0.3193           1.213
Modified likelihood root (cont. corr.) 0.3044          1.254

Diagnostics:
-----------
    INF      NP
0.08451 0.32878
```

Marginal and conditional likelihoods

$$L_c(\psi) = \log f_c\{s(y) \mid t(y); \psi\},$$
$$L_m(\psi) = \log f_m\{s(y); \psi\}$$

- Inference based on usual asymptotics applies, under regularity conditions on $f(y; \psi, \lambda)$
- likelihoods based on observable random variables
- Bartlett identities apply directly
- use conditional or marginal Fisher information, etc.

- might lose information in other component

$$f(y; \psi, \lambda) \propto f_m(s; \psi) f_c(t \mid s; \psi, \lambda)$$

- marginal likelihoods associated with transformation models

REML

Approximate conditional inference

- $\ell_c(\psi) \doteq \ell_p(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|$ $\quad i_{\psi\lambda}(\theta) = 0$
- $\ell_m(\psi) \doteq \ell_p(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|$
- $\ell_c(\psi) \doteq \ell_p(\psi) + \frac{1}{2}\log|j_{\eta\eta}(\psi, \hat\eta_\psi)|$ $\quad \exp\{\psi^{\mathrm{T}}s + \eta^{\mathrm{T}}t - c(\psi,\eta)\}$
- adjusted profile log-likelihood

$$\ell_A(\psi) = \ell_p(\psi) + A(\psi)$$

$A(\psi)$ assumed to be $O_p(1)$

- generic form is $A_{FR}(\psi) = +\frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)| - \log|\frac{d(\lambda)}{d\hat\lambda_\psi}|$ Fraser 03
- closely related $A_{BN}(\psi) = -\frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)| + \log|\frac{d\hat\lambda}{d\hat\lambda_\psi}|$ SM §12.4.1

BN 83