Applied Asymptotics

Case studies in higher order inference

Nancy Reid

University of Windsor

March 31, 2006

A.C. Davison, A. R. Brazzale

- higher order approximations in practical examples
- parametric models
- *p*-values and confidence intervals for a scalar parameter of interest
- likelihood based approximations
  - r - (generalized) likelihood ratio statistic $\dot{\sim} N(0,1)$
  - q - another likelihood based quantity $\dot{\sim} N(0,1)$
  - $r^* = r + \frac{1}{r}\log\frac{q}{r} \;\dot{\sim}\; N(0,1)$
- Illustration: top quark $Y \sim Poisson(\mu + b)$: is $\mu > 0$?
- Illustration: cost of medical treatment is $\mu_1 = \mu_2$?

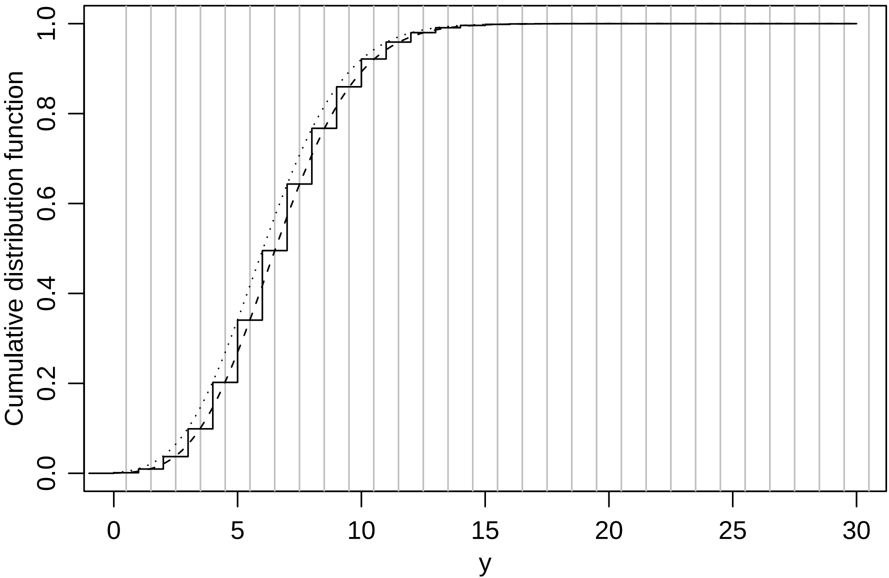
Top quark[1]

- model $Y \sim \text{Poisson}\ (\mu + b)$, b known
- data $y = 27$ $b = 6.7$
- mid p-value $\Pr(Y > 27) + \frac{1}{2}\Pr(Y = 27) = 18.45 \times 10^{-10}$
- approximation: $\Phi(r^*) = \Phi\left(r + \frac{1}{r}\log\frac{q}{r}\right) = 15.85 \times 10^{-10}$
- r = likelihood ratio statistic
  q = standardized maximum likelihood estimate for canonical parameter
- continuity correction: $\Phi\{r^*(y + \frac{1}{2})\} = 32.36 \times 10^{-10}$
  $\Pr(Y \geq 27) = 29.83 \times 10^{-10}$

b unknown; F R Wong (2004)

[1] Abe et al. 1995

	$\Pr(Y \geq y^0; \mu = 0)$	$\Pr(Y > y^0; \mu = 0)$
Exact	29.83×10^{-10}	7.06×10^{-10}
Continuity correction	32.36×10^{-10}	7.69×10^{-10}
Mid-p	18.45×10^{-10}	
r^*	15.85×10^{-10}	
$N(\theta, \theta)$	4.45×10^{-4}	2.21×10^{-5}
$N(\theta, \hat{\theta})$	1.02×10^{-4}	4.68×10^{-5}


Cumulative distribution function
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
y

Two-sample comparison

data on cost of treatment: standard vs new treatment[2]

Group 1	30	172	210	212	335	489	651	1263
	1875	2213	2998	4935				
Group 2	121	172	201	214	228	261	278	279
	561	622	694	848	853	1086	1110	1243

model $Y_{1i} \sim \text{Exp}(\mu_1), \quad Y_{2i} \sim \text{Exp}(\mu_2) \quad \psi = \mu_1/\mu_2$

Exact inference $\left(\frac{\bar{Y}_{1.}}{\mu_1}\right) / \left(\frac{\bar{Y}_{2.}}{\mu_2}\right) \sim F_{2n,2m}$

exact 95% confidence interval for ψ: $(0.98, 4.19)$
approx 95% confidence interval for ψ: $(0.98, 4.185)$

[2] Evans et al. 1999

- $Y \sim f(y;\theta), \quad y \in R^n \quad \theta \in R^d \quad \theta = (\psi, \lambda) \quad \psi \in R$
- $\ell(\theta) = \ell(\theta; y) = \log f(y;\theta)$
- what values of ψ are supported by the data?
- ideal: a density on R : $f_m(t(y);\psi)$ or $f_c(t(y) \mid s(y);\psi)$
- approximate pivot, e.g. $(\hat\psi - \psi)/\widehat{\mathrm{s.e.}}(\hat\psi)$
- approximate pivot $r^*(\psi)$, similar, but better approximated by $N(0,1)$
- $r^* = r + (1/r)\log(q/r)$
- $r = \pm\surd[2\{\ell(\hat\psi,\hat\lambda) - \ell(\psi,\hat\lambda_\psi)\}]$
- $q =??$
- top quark: $q = (\hat\varphi - \varphi)/\widehat{\mathrm{s.e.}}(\hat\varphi) = \log(\hat\mu/\mu)\hat\mu^{1/2}$
- cost data: $q = \dfrac{\hat\psi - \psi}{\widehat{\mathrm{s.e.}}(\hat\psi)}\dfrac{\hat\psi\hat\lambda}{\psi\hat\lambda_\psi}$
- in other examples q is a standardized score statistic

Logistic regression

The first ten out of 79 sets of observations on the physical characteristics of urine. Presence/absence of calcium oxalate crystals is indicated by 1/0. Two cases with missing values.[3]

Case	Crystals	Specific gravity	pH	Osmolarity	Conductivity	Urea	Calcium
1	0	1.021	4.91	725	—	443	2.45
2	0	1.017	5.74	577	20.0	296	4.49
3	0	1.008	7.20	321	14.9	101	2.36
4	0	1.011	5.51	408	12.6	224	2.15
5	0	1.005	6.52	187	7.5	91	1.16
6	0	1.020	5.27	668	25.3	252	3.34
7	0	1.012	5.62	461	17.4	195	1.40
8	0	1.029	5.67	1107	35.9	550	8.48
9	0	1.015	5.41	543	21.9	170	1.16
10	0	1.021	6.13	779	25.7	382	2.21
⋮							⋮

[3] Andrews & Herzberg, 1985

Model: Independent binary responses $Y_1, \ldots, Y_n$ with

$$\Pr(Y_i = 1) = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}$$

Fitting generalized linear model in R:

```
data(urine)
fit <- glm(r~gravity+ph+osmo+cond+urea+calc,
family = binomial, data=urine)
summary(fit)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -355.33771  222.76696  -1.595  0.11069
gravity      355.94379  222.11004   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 **
```

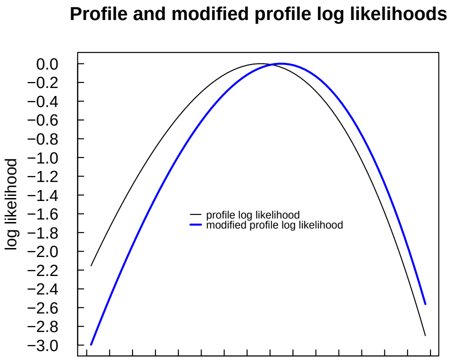
A closer look at coefficient of `urea`

method	lower bound	upper bound	p-value for 0
$\Phi(q)$	-0.063	-0.0006	0.047
$\Phi(r)$	-0.067	-0.0025	0.033
$\Phi(r^*)$	-0.058	0.0002	0.052

```
library(cond) # part of package 'hoa' on cran-r

urine.cond.urea <- cond.glm(urine.glm,offset=urea)
> summary(urine.cond.urea,test=0)

...

Test statistics
---------------
 hypothesis : coef( urea ) = 0
                                              statistic   tail prob.
Wald pivot                                       -1.986      0.02351
Wald pivot (cond. MLE)                           -1.852      0.03202
Likelihood root                                  -2.133      0.01648
Modified likelihood root                         -1.925      0.02713
Modified likelihood root (cont. corr.)           -1.917      0.02760
```

A closer look at coefficient of `urea`

method	lower bound	upper bound	p-value for 0
$\Phi(q)$	-0.063	-0.0006	0.047
$\Phi(r)$	-0.067	-0.0025	0.033
$\Phi(r^*)$	-0.058	0.0002	0.052



A little more theory...

- linear exponential families: $f(y;\theta) = \exp\{\psi s + \lambda^T t - c(\theta) - d(y)\}$
- $f_c(s \mid t;\psi) = \exp\{\psi s - c_t(\psi) - d_t(s)\}$
- $\Phi(r^*)$ approximates $F_c(s \mid t;\psi)$ to $O(n^{-3/2})$
- with $q = (\hat\psi - \psi) j_p^{1/2}(\hat\psi)\rho(\psi,\hat\psi)$
- $\rho(\psi,\hat\psi) = \dfrac{|j_{\lambda\lambda}(\hat\psi,\hat\lambda)|}{|j_{\lambda\lambda}(\psi,\hat\lambda_\psi)|}$
- conditional log-likelihood approximated by $\ell_p(\psi) + \frac{1}{2}\log j_{\lambda\lambda}(\psi,\hat\lambda_\psi)$

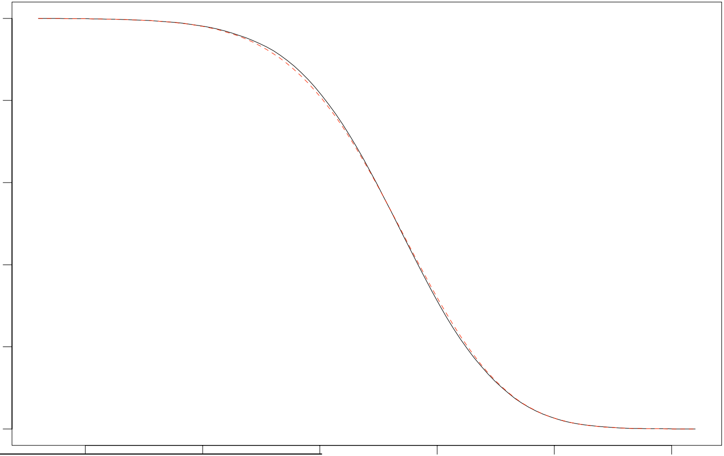
- $n = 77, d = 7$, but the information in 77 binary observations seems to be comparable to the information in just 10 continuous observations
- Bayesian inference (3rd order) also straightforward in linear exponential families
- there is a unique non-informative (matching) prior

method	lower bound	upper bound
$\Phi(r)$	-0.067	-0.0025
$\Phi(r^*)$	-0.058	0.00029
$\Phi(r_B^*)$	-0.058	0.00028

A matching prior ensures that to $O(n^{-3/2})$, the posterior upper α limit has frequentist coverage $1 - \alpha$:

$$\begin{aligned} \mathrm{Pr_m}(\psi > \psi^{(1-\alpha/2)}(y) \mid y) &= \alpha \\ \mathrm{Pr}(\psi^{(1-\alpha/2)}(Y) < \psi \mid \theta) &\doteq \alpha \end{aligned}$$

Comparison of posterior limits for ψ with frequentist p-value, $0 < \alpha < 1$[4]



[4] Computations by Ana-Maria Staicu, U of T

Several 2×2 tables[5]

Institution	y_1	m_1	y_2	m_2	Institution	y_1	m_1	y_2	m_2
1	3	4	1	3	12	2	2	0	2
2	3	4	8	11	13	1	4	1	5
3	2	2	2	3	14	2	3	2	4
4	2	2	2	2	15	2	4	4	6
5	2	2	0	3	16	4	12	3	9
6	1	3	2	3	17	1	2	2	3
7	2	2	2	3	18	3	3	1	4
8	1	5	4	4	19	1	4	2	3
9	2	2	2	3	20	0	3	0	2
10	0	2	2	3	21	2	4	1	5
11	3	3	3	3					

[5] Lipsitz et al. 1988: Tests of homogeneity for the risk difference when the data are sparse. *Biometrics*.

Model: $Y_{1i} \sim \mathrm{Binomial}(m_{1i}, p_{1i}) \quad Y_{2i} \sim \mathrm{Binomial}(m_{2i}, p_{2i})$

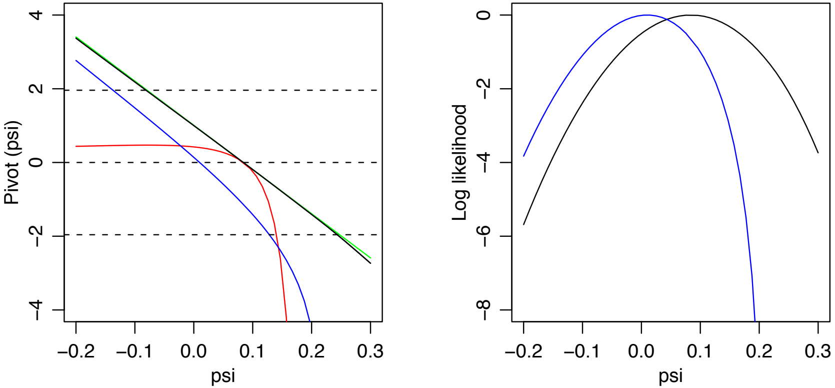
parameter of interest $\psi = p_{2i} - p_{1i}$
nuisance parameters $p_{1i}, i = 1, \ldots 21$

inference for ψ :

	lower	upper	point estimate	p-value for $\psi = 0$
$\Phi(r)$	-0.081	0.243	0.10	0.16
$\Phi(r^*)$	-0.137	0.126	0.01	0.45

weighted least squares estimates[6] $0.06, 0.02, 0.07$

[6] Kuhnert et al 2004



likelihood root, r^*, q

profile log likelihood,
modified profile

- in this model ψ, the parameter of interest, is not a component of the canonical parameter (it would be on the log-odds scale)
- there is no exact elimination of nuisance parameters by conditioning
- the construction of q incorporates approximate conditioning
- q is neither a standardized mle/Wald nor a standardized score/Rao statistic, but something 'in between'
- using $\Phi(r^*)$ also automatically incorporates smoothing of discrete distributions, via mid p-values

$$q = \frac{|\varphi(\hat{\theta}) - \varphi(\hat{\theta}_\psi) \quad \varphi_{\lambda'}(\hat{\theta}_\psi)|}{|\varphi_{\theta'}(\hat{\theta})|} \left\{ \frac{|j(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{1/2}$$

Example G: Cox & Snell, 1980

	cost	date	T1	T2	cap	PR	NE	CT	BW	N	PT
1	460.05	68.58	14	46	687	0	1	0	0	14	0
2	452.99	67.33	10	73	1065	0	0	1	0	1	0
3	443.22	67.33	10	85	1065	1	0	1	0	1	0
4	652.32	68.00	11	67	1065	0	1	1	0	12	0
5	642.23	68.00	11	78	1065	1	1	1	0	12	0
6	345.39	67.92	13	51	514	0	1	1	0	3	0
7	272.37	68.17	12	50	822	0	0	0	0	5	0
8	317.21	68.42	14	59	457	0	0	0	0	1	0
9	457.12	68.42	15	55	822	1	0	0	0	5	0
10	690.19	68.33	12	71	792	0	1	1	1	2	0
11	350.63	68.58	12	64	560	0	0	0	0	3	0
12	402.59	68.75	13	47	790	0	1	0	0	6	0
13	412.18	68.42	15	62	530	0	0	1	0	2	0
14	495.58	68.92	17	52	1050	0	0	0	0	7	0
15	394.36	68.92	13	65	850	0	0	0	1	16	0

$n = 32, d = 8$

- Model $Y_i = \beta_0 + x_i^T\beta + \sigma\epsilon_i$
- $\epsilon \sim N(0,1)$ or $\epsilon \sim t_\nu$

	Normal		t_4, first order		t_4, third order	
	Est (SE)	z	Est (SE)	z	Est (SE)	z
Constant	−13.26 (3.140)	−4.22	−11.30 (3.67)	−3.01	−11.86 (3.70)	−3.21
date	0.212 (0.043)	4.91	0.191 (0.048)	3.97	0.196 (0.049)	4.02
log(cap)	0.723 (0.119)	6.09	0.648 (0.113)	5.71	0.682 (0.129)	5.31
NE	0.249 (0.074)	3.36	0.242 (0.077)	3.12	0.239 (0.080)	2.97
CT	0.140 (0.060)	2.32	0.144 (0.054)	2.68	0.143 (0.063)	2.26
log(N)	−0.088 (0.042)	−2.11	−0.060 (0.043)	−1.40	−0.072 (0.048)	−1.51
PT	−0.226 (0.114)	−1.99	−0.282 (0.101)	−2.80	−0.265 (0.110)	−2.42

ν	log(N)		PT	
	First order	Third order	First order	Third order
4	0.162	0.151	0.005	0.024
6	0.110	0.116	0.007	0.032
8	0.081	0.098	0.009	0.036
10	0.064	0.086	0.011	0.038
20	0.036	0.064	0.016	0.045
40	0.025	0.053	0.029	0.050
100	0.020	0.047	0.022	0.053
∞	0.035	0.045	0.046	0.057

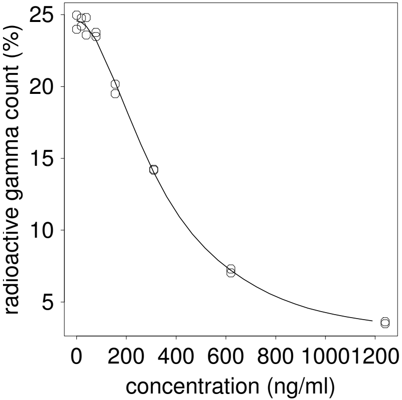
```
library(marg)
# part of package 'hoa' on cran-r
data(nuclear)

# Fit normal-theory linear model and examine its contents:

nuc.norm <- lm( log(cost) ~ date + log(cap) + NE + CT + log(N) + PT,
+                  data = nuclear )
summary(nuc.norm)

# Fit linear model with t errors and 4 df and examine its contents:

nuc.t4 <- rsm( log(cost) ~ date + log(cap) + NE + CT + log(N) + PT,
+                 data = nuclear, family = student(4) )
summary(nuc.t4)
plot(nuc.t4)

# Conditional analysis for partial turnkey guarantee:

nuc.t4.pt <- cond( nuc.t4, offset = PT )
summary(nuc.t4.pt)
plot(nuc.t4.pt)

# For conditional analysis for other covariates, replace pt by
# log(N), ...
```

A little more theory...

- linear regression: $y_i = x_i^T \beta + \sigma \epsilon_i$
- $\epsilon_i \sim f(\cdot)$ (known)
- marginal density of $t_1 = (\hat{\beta}_1 - \beta_1)/s$, conditional on configuration, is free of $\beta_{(2)}, \sigma$
- $\Phi(r^*)$ approximates $F_m(t_1; \beta_1 \mid \hat{\epsilon})$ to $O(n^{-3/2})$
- with $q = \ell_p'(\psi) j_p^{-1/2}(\hat{\psi}) \rho^{-1}(\psi, \hat{\psi})$
- $\rho(\psi, \hat{\psi}) = \dfrac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|}$
- marginal[7] log-likelihood approximated by $\ell_p(\psi) - \frac{1}{2} \log j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)$

[7] conditioned on configuration

Radio-immunoassay[8]



Model $y_i = \mu(x_i, \beta) + \sigma\epsilon_i$

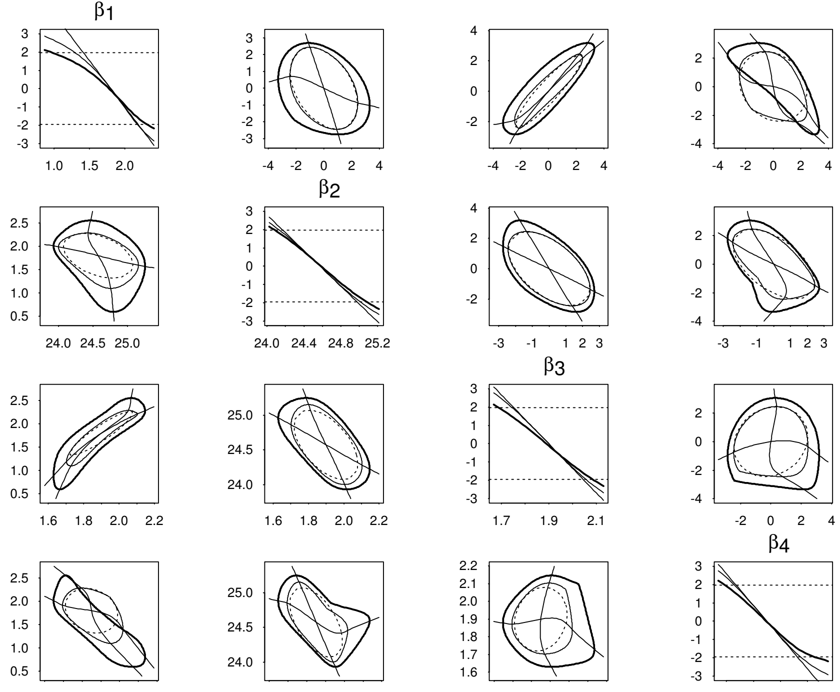
[8] Belanger et al., 1996

mean function: $\mu(x, \beta) = \beta_1 + \dfrac{\beta_2 - \beta_1}{1 + (x/\beta_4)^{\beta_3}}$,

normal errors: $\epsilon_i \sim N(0, \sigma^2 V^2(x_i, \beta))$

variance function: $V^2(x, \beta) = \mu^\gamma(x, \beta)$

library `nlreg` gives maximum likelihood estimates and second order inference using a method due to Skovgaard

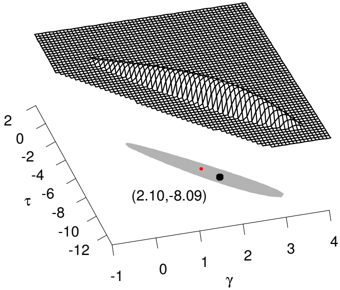


β_1
β_2
β_3
β_4

Inference for variance parameters $\log \sigma$ and γ

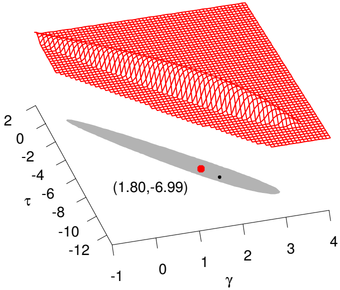
PROFILE LIKELIHOOD

(MLE and 95% confidence region added)



MODIFIED PROFILE LIKELIHOOD

(MLE and 95% confidence region added)



Coverage of 95% confidence intervals
estimated by parametric bootstrap

	β_1	β_2	β_3	β_4	γ	$\log\sigma^2$
Wald	79.3	89.5	83.7	85.7	72.7	66.8
r	82.7	92.1	85.9	87.9	78.3	76.0
r^*	92.9	94.5	93.2	93.7	93.3	93.6
studentized	100	99.3	100	100	98.7	99.3
percentile	100	99.3	100	99.3	65.1	51.7

Mixed linear models (normal error)

$$y_i = X_i\beta + Z_i b_i + \epsilon_i$$

$$b_i \sim N(0, \Omega), \quad \epsilon_i \sim N(0, \Sigma)$$

$$\ell(\beta, \rho) = -\frac{1}{2}\log|V_i(\rho)| - \frac{1}{2}\sum(y_i - X_i\beta)V_i^{-1}(\rho)(y_i - X_i\beta)$$

where $V_i(\rho) = X_i\Omega Z_i + \Sigma$, ρ are the parameters in Ω and Σ

residual log likelihood for inference on ρ (reml)

$$\ell_R(\rho) = -\frac{1}{2}\log|V_i(\rho)| - \frac{1}{2}r_i^T V_i^{-1}(\rho)r_i - \frac{1}{2}\sum X_i^T V_i(\rho)X_i$$

where $r_i = y_i - X_i\hat{\beta}_\rho$

Mixed linear models (normal error): Guolo & Brazzale (2005)

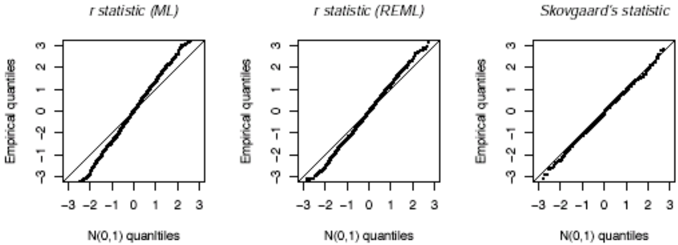


Figure2. Normal Q-Q plots of r (left) and $\tilde{r}$ (right) based upon $R = 1{,}000$ replicates in a simulation study where we tested whether $\sigma_{01} = 0$ in model (4). The middle panel reports the normal Q-Q plot for the r statistic based upon the restricted likelihood function.

`lme-hoa` package; extension to nonlinear regression uses approximate linearization of Lindstrom and Bates

Type II censored data[9]

40 units on test, 28 failures at (log) times

0.0507	0.0579	0.0784	0.0954	0.1376	0.2249	0.2362	0.2481
0.2501	0.2811	0.3027	0.3091	0.4296	0.5379	0.5621	0.5781
0.7811	0.8228	0.9455	0.9871	1.0060	1.0335	1.0377	1.0471
1.0876	1.2473	1.2776	1.3445				

Weibull model: $f(y;\mu,\sigma) = e^{(y-\mu)/\sigma}\exp\{-e^{(y-\mu)/\sigma}\}$

90% confidence intervals

	μ	σ
$\Phi(r)$	$(-0.116, 0.476)$	$(0.700, 1.217)$
$\Phi(r^*)$	$(-0.107, 0.510)$	$(0.743, 1.320)$
Exact (num. int.)	$(-0.11, 0.51)$	$(0.724, 1.277)$

[9] Lawless 2003 Ch.5; Wong & Wu 2003

Air-conditioning units[10]

aircraft number	sample size n_i	failure times y_{ij}
1	23	413 14 58 37 100 65 9 169 447 184 36 201 118 34 31 18 18 67 57 62 7 22 34
2	29	90 10 60 186 61 49 14 24 56 20 79 84 44 59 29 118 25 156 310 76 26 44 23 62 130 208 70 101 208
3	15	74 57 48 29 502 12 70 21 29 386 59 27 153 26 326
4	14	55 320 56 104 220 239 47 246 176 182 33 15 104 35
5	30	23 261 87 7 120 14 62 47 225 71 246 21 42 20 5 12 120 11 3 14 71 11 14 11 16 90 1 16 52 95
6	27	97 51 11 4 141 18 142 68 77 80 1 16 106 206 82 54 31 216 46 111 39 63 18 191 18 163 24
7	24	50 44 102 72 22 39 3 15 197 188 79 88 46 5 5 36 22 139 210 97 30 23 13 14
8	9	359 9 12 270 603 3 104 2 438
9	12	487 18 100 7 98 5 85 91 43 230 3 130
10	16	102 209 14 57 54 32 67 59 134 152 27 14 230 66 61 34

[10] Cox & Snell, Example T

D. R. Cox and E. J. Snell

APPLIED STATISTICS

Principles and Examples

				23					
	90			261					
	10			87					
	60			7	97				
	186			120	51				
	61			14	11				
	49			62	4	50			
413	14			47	141	44			
14	24			225	18	102			
58	56			71	142	72			
37	20			246	68	22			
100	79			21	77	39			
65	84			42	80	3			
9	44			20	1	15			
169	59			5	16	197			102
447	29	74		12	106	188			209
184	118	57	55	120	206	79			14
36	25	48	320	11	82	88			57
201	156	29	56	3	54	46		487	54
118	310	502	104	14	31	5		18	32
34	76	12	220	71	216	5		100	67
31	26	70	239	11	46	36	359	7	59
18	44	21	47	14	111	22	9	98	134
18	23	29	246	11	39	139	12	5	152
67	62	386	176	16	63	210	270	85	27
57	130	59	182	90	18	97	603	91	14
62	208	27	33	1	191	30	3	43	230
7	70	153	15	16	18	23	104	230	66
22	101	26	104	52	163	13	2	3	61
34	208	326	35	95	24	14	438	130	34

Model

$$f(y_{ij}; \beta_i, \mu_i) = \frac{1}{\Gamma(\beta_i)} \left(\frac{\beta_i}{\mu_i}\right)^{\beta_i} y_{ij}^{\beta_i - 1} e^{-\beta_i y_{ij}/\mu_i}$$

Are data consistent with a common value for β?
Vector parameter of interest.

First order theory uses generalized likelihood ratio test
$w = 2\{\ell(\underline{\widehat{\beta}}, \underline{\widehat{\mu}}) - \ell(\widehat{\beta}, \underline{\widehat{\mu}})\} \sim \chi^2_9 \quad O(n^{-1})$

Bartlett: $w^{**} = w/E(w); \quad O(n^{-2})$

Skovgaard $w^* = w(1 - (1/w)\log\gamma)^2$

First order p-value 0.073; using Bartlett correction (bootstrap) 0.107; using Skovgaard 0.130.

A little more theory...

- general (continuous) families: $f(y;\theta) \quad y \in R^n \quad \theta \in R^p$
- approximate conditioning to construct a density on R^p (location-type)
- approximate marginalization to find a density on R (exponential-type)
- the density is $\Phi(r^*)$
- with

$$q = \frac{|\varphi(\hat{\theta}) - \varphi(\hat{\theta}_\psi) \quad \varphi_{\lambda'}(\hat{\theta}_\psi)|}{|\varphi_{\theta'}(\hat{\theta})|} \left\{ \frac{|j(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{1/2}$$

- log-likelihood approximated by $\ell_p(\psi) - \frac{1}{2} \log j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) - a(\psi)$

A lot more theory...

- $\varphi(\theta) = \ell_{;V}(\theta; y)$: sample space differentiation
- V is a set of vectors tangent to an approximate ancillary
- $V_i = \left(\frac{\partial z_i}{\partial \theta}\right)^{-1} \frac{\partial z_i}{\partial y_i}\bigg|_{y^0}$
- z_i is a pivotal quantity for the ith observation, e.g. $z_i = (y_i - \mu)/\sigma$ in a location-scale model; $F(y;\theta)$ in general
- needs some adjustment for discrete data

Conclusions

- Easy to use hoa in generalized linear models and in non-normal linear regression
- Not too hard to use hoa for non-canonical parameters, non-linear regression
- Some problems can be re-expressed as exponential families (e.g. Butler et al. 1989, Wong & Wu SJS)
- Most numerical work is very convincing
- Exception: proportional hazards model with q taken as Wald statistic
- Relies on the log likelihood function (and in particular on differentiating log likelihood on the sample space): how to adapt to pseudo-likelihoods

...conclusions

- non-parametric or semi-parametric models?
- Example: Poisson (log-linear) regression in time series of air pollution[11]

$$\log \mu = \beta PM_{10} + S(time, df) + S(temperature, df) + ...$$

- $S(x, df) = \sum b_i f_i(x)$ where $f_i(\cdot)$ is a spline basis function
- scalar parameter of interest; very high dimensional nuisance parameter
- interest in choosing the number of basis functions

[11] Peng et al 2004