NONPARAMETRIC RANDOM EFFECTS MODELS AND LIKELIHOOD RATIO TESTS

Oct 11, 2002

David Ruppert

Cornell University

www.orie.cornell.edu/˜davidr

(These transparencies and preprints available — link to “Recent Talks” and “Recent Papers”)

Work done jointly with

Ciprian Crainiceanu, Cornell University

OUTLINE

- Smoothing can be done using standard mixed models software because
    - Splines can be viewed as BLUPs in mixed models
- This random-effects spline model extends to:
    - Semiparametric models (allows parametric submodels)
    - Longitudinal data
    - nested families of curves

EXAMPLE

- Rick Canfield and Chuck Henderson, Jr. at Cornell are working on effects of low-level lead exposure on IQ of children.
- They have a mixed model but the dose-response curve should be modeled nonparametrically.

EXAMPLE — CONT

- They asked SAS is a ”PROC GAMMIXED” would be available someday.
  - short answer was ”no”
- Then, they found Matt Wand’s work and then contacted me.
- Now they know that GAMMIXED $\subset$ GLMMIXED.
  - SAS has GAMMIXED and does not know it!

TESTING IN THIS FRAMEWORK

- In principle, likelihood ratio tests (LRTs) could be used to test for effects of interest
  - E.g., hypothesis that a curve is linear or that an effect is zero $\Longleftrightarrow$ a variance component (and possibly a fixed effect) is zero
  - allows an elegant, unified theory

TESTING — CONT

- However, the distribution theory of LRTs is complex:
  - the null hypothesis is on the boundary of the parameter space, so “standard theory” suggests chi-squared mixtures as the asympototic distribution.
  - but standard asymptotics do not apply because of correlation
  - for the case of one variance component, we now have asymptotics that do apply

UNIVARIATE NONPARAMETRIC REGRESSION

- model

$$y_i = f(x_i) + \epsilon_i$$

- letting f be a spline

$$f(x) = \sum_{k=0}^{p} \beta_k x^k + \sum_{k=1}^{K} b_k (x - \kappa_k)_+^p$$

  - $b_1, \dots, b_K$ will be treated as “random effects”
  - assume they are iid $N(0, \sigma_b^2)$
  - size of σ_b^2 controls the amount of shrinkage or smoothing.

NONPARAMETRIC MODELS FOR LONGITUDINAL DATA

- y_{ij} is jth observation on ith subject
- consider the nonparametric model

$$y_{ij} = f(x_{ij}) + f_i(x_{ij}) + \epsilon_{ij}$$

- model the “population” curve f as a spline:

$$f(x) = \sum_{k=0}^{p} \beta_k x^k + \sum_{k=1}^{K} b_k (x - \kappa_k)_+^p$$

- model the “ith subject” curve f_i as another spline:

$$f_i(x) = \sum_{k=0}^{p} u_k^{(i)} x^k + \sum_{k=1}^{K} b_k^{(i)} (x - \kappa_k)_+^p$$

POPULATION CURVE

- Recall:

$$f(x) = \sum_{k=0}^{p} \beta_k x^k + \sum_{k=1}^{K} b_k (x - \kappa_k)_+^p$$

- $\beta_0, \ldots, \beta_p$ will be treated as “fixed effects”
- $b_1, \ldots, b_K$ will be treated as “random effects”
  - assume they are iid $N(0, \sigma^2_{b,P})$ (P = “population”)
  - this assumption can be viewed as a Bayesian model
  - somewhat different that usual interpretation of random effects

SUBJECT CURVES

- Recall:

$$f_i(x) = \sum_{k=0}^{p} u_k^{(i)} x^k + \sum_{k=1}^{K} b_k^{(i)} (x - \kappa_k)_+^p$$

- $u_0^{(i)}, \ldots, u_p^{(i)}$ will be treated as “random effects”
  - assume they are iid $N(0, \sigma_u^2)$
  - this is a typical “random effects” assumption
- $b_1^{(i)}, \ldots, b_K^{(i)}$ will also be treated as “random effects”
  - assume they are iid $N(0, \sigma_{b,S}^2)$ (S = “subject”)

NULL HYPOTHESES OF INTEREST

- Recall:

$$f_i(x) = \sum_{k=0}^{p} u_k^{(i)} x^k + \sum_{k=1}^{K} b_k^{(i)} (x - \kappa_k)_+^p$$

- $\sigma_u^2 = \sigma_{b,S}^2 = 0 \iff$ no subject effects

- $\sigma_{b,S}^2 = 0 \iff$ subject effects are pth degree polynomials

RELATED WORK

- Brumback and Rice (1998)
- Zhang, Lin, Raz, and Sowers (1998)
- Lin and Zhang (1999)
- Rice and Wu (2001)

See references at end.

BALANCE 1-WAY ANOVA

- model:

$$Y_{ij} = \mu + b_i + \epsilon_{ij},\ i = 1, \ldots, I \text{ and } j = 1, \ldots, J.$$

and

$$b_i \sim N(0, \sigma_b^2)$$

- null hypothesis:

$$H_0 : \sigma_b^2 = 0.$$

- If $I \to \infty$ with J fixed, then

$$-2\log(LR) \to \tfrac{1}{2}\chi_0^2 + \tfrac{1}{2}\chi_1^2.$$

(Self and Liang, 1987; Stram and Lee, 1994)

  - This is the iid case if we take the subjects as "observations"

Note: The equivalent fixed effects hypothesis is $b_1 = \cdots = b_I = 0$.

- Then the LR test is equivalent to the F-test
- $-2\log(LR) \to \chi^2_{I-1}$ under H_0

- If $J \to \infty$ with I fixed, then

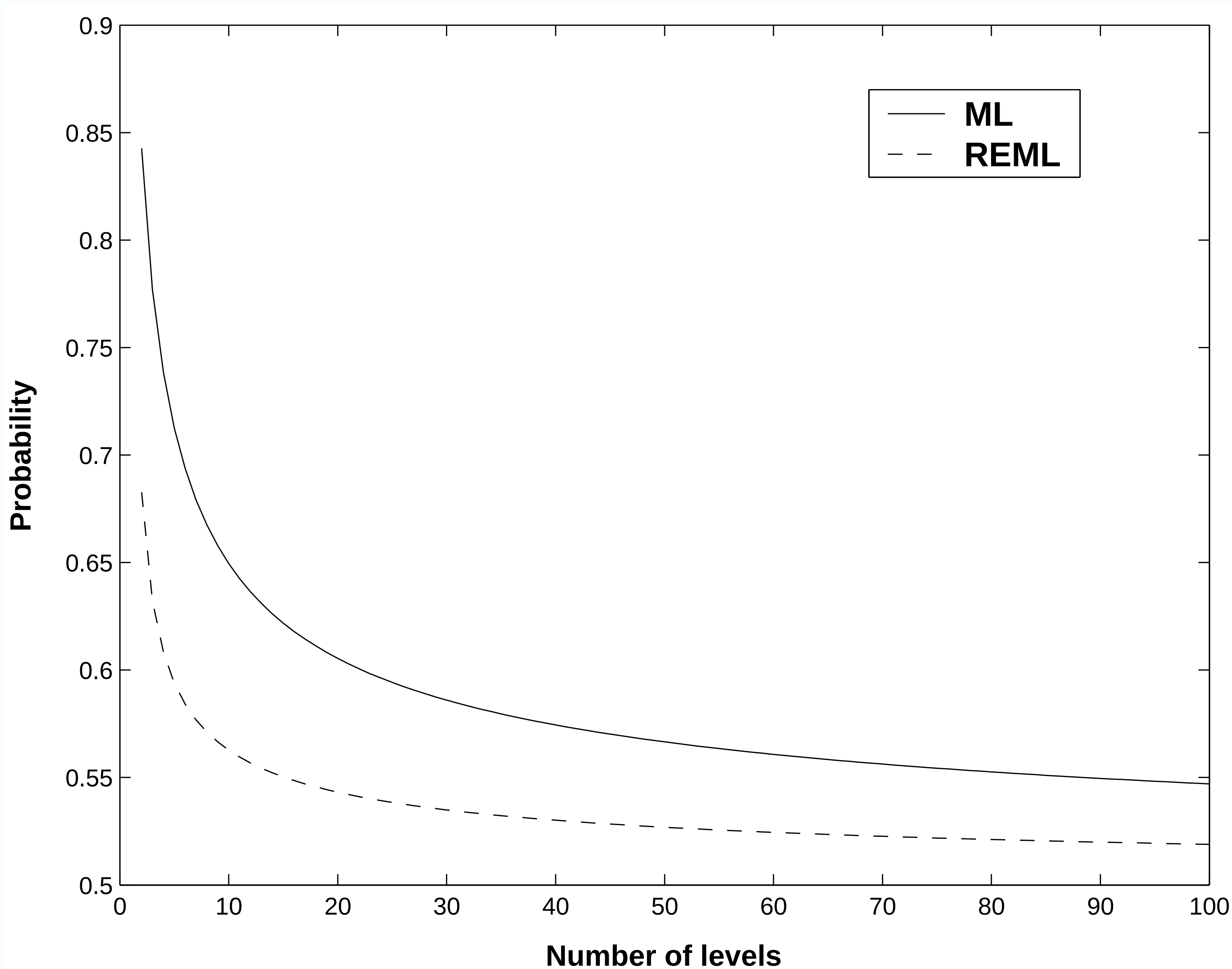
$$-2\log(LR) \Rightarrow I\left\{X^*_{I-1} - 1 - \log(X^*_{I-1})\right\}\mathcal{I}_{\{X^*_{I-1}>1\}},$$

and

$$-2\log(RLR) \Rightarrow (I-1)\left\{X_{I-1} - 1 - \log(X_{I-1})\right\}\mathcal{I}_{\{X_{I-1}>1\}},$$

where $X_{I-1} \sim \frac{\chi^2_{I-1}}{I-1}$ and $X^*_{I-1} \sim \frac{\chi^2_{I-1}}{I}$.

(Crainiceanu and Ruppert, 2002)

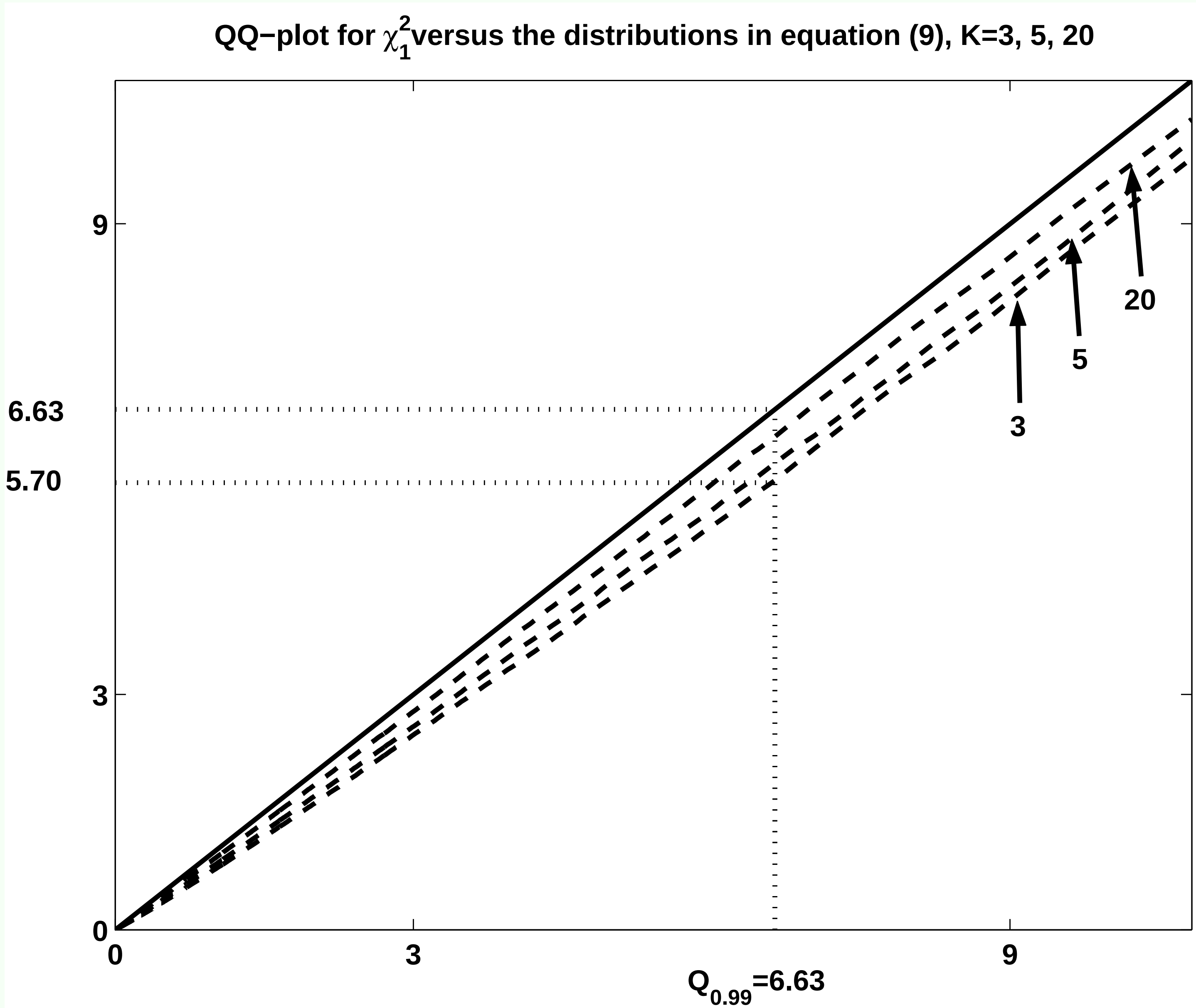


1-Way ANOVA: $\lim_{n\to\infty} P_{H_0}\{\log(LR) = 0\}$

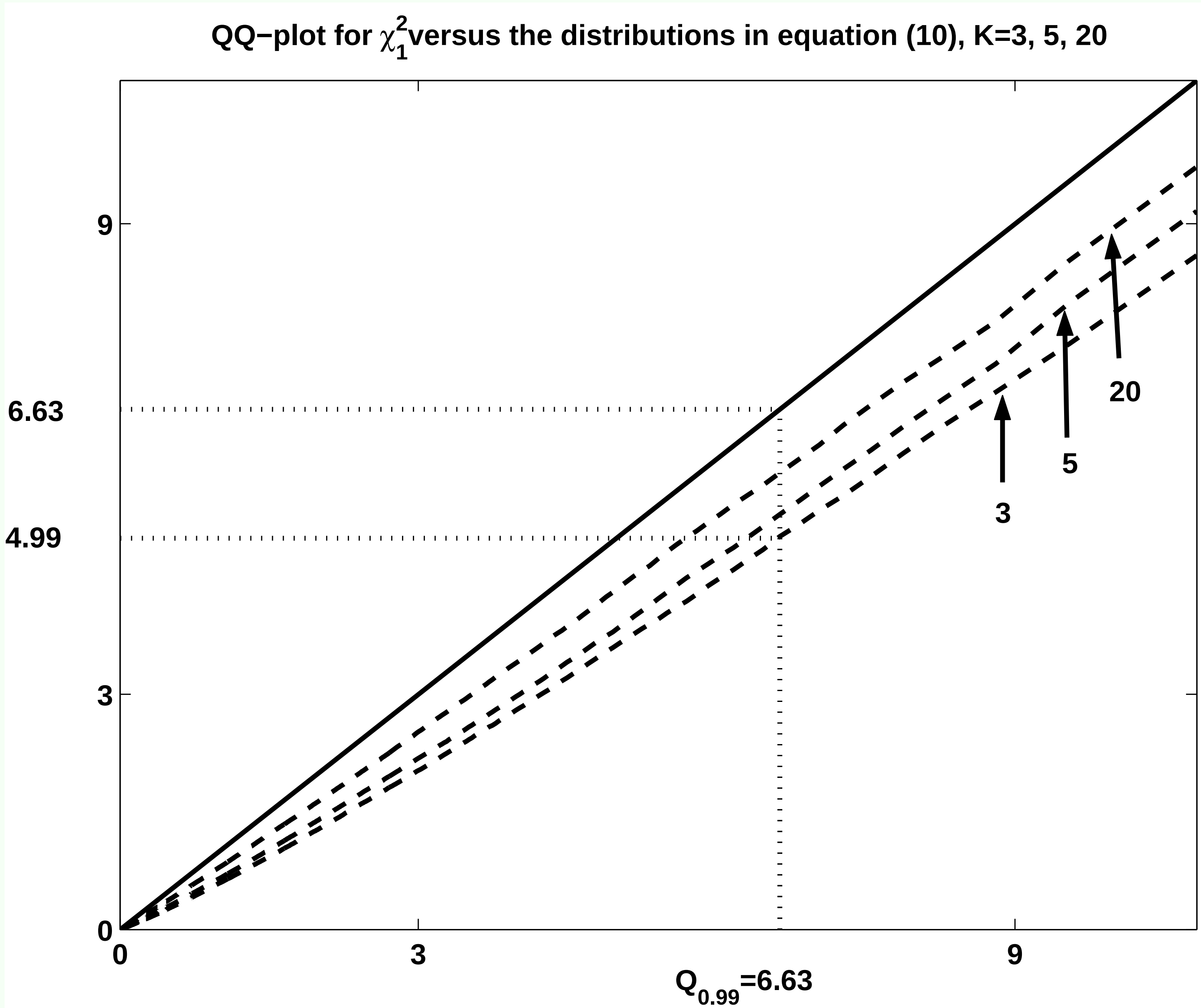
Pinheiro and Bates (2000, p. 87)

- simulated the LRT
- found some empirical evidence that the $.5\chi_0^2 + .5\chi_1^2$ mixture is better replaced by $p_0\chi_0^2 + (1 - p_0)\chi_1^2$ for $p_0 > .5$.

These theoretical results help explain their findings.



1-Way ANOVA: asympt. null dist of RLR, given RLR > 0



1-Way ANOVA: asympt. null dist of LR, given LR $>$ 0

PENALIZED SPLINES

- model:

$$y_i = m\left(x_i\right) + \epsilon_i \; ,$$

- null hypothesis:

$$H_0 : \; m\left(x\right) = \beta_0 + \beta_1 x + ... + \beta_{p+1-q} x^{p-q}, \; q \geq 0.$$

- alternative hypothesis:

$$H_A : \; m(x) = \beta_0 + \beta_1 x + ... + \beta_p x^p + \sum_{k=1}^{K} b_k \left(x - \kappa_k\right)_+^p \; ,$$

- notation:

$$\boldsymbol{\theta} = \left(\beta_0, \ldots, \beta_p, b_1, \ldots, b_K\right)^T$$

- penalized least squares:

minimize

$$\sum_{i=1}^{n} \{y_i - m\left(x_i; \boldsymbol{\theta}\right)\}^2 + \lambda \boldsymbol{\theta}^T \boldsymbol{L} \boldsymbol{\theta} \, ,$$

with

$$\boldsymbol{L} = \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{\Sigma}^{-1} \end{bmatrix} ,$$

- same as BLUP in a linear mixed model with

$$\text{Cov}(\boldsymbol{b}) = \sigma_b^2 \boldsymbol{\Sigma}$$

and

$$\lambda = \frac{\sigma_\epsilon^2}{\sigma_b^2}$$

(Brumback, Ruppert, and Wand, 1999)

- new form of null:

if $q = 0$

$$\sigma_b^2 = 0$$

or, if $q > 0$,

$$\beta_{p-q+1} = \cdots = \beta_p = 0 \quad \text{and} \quad \sigma_b^2 = 0.$$

Example: (Crainiceanu and Ruppert, 2002)

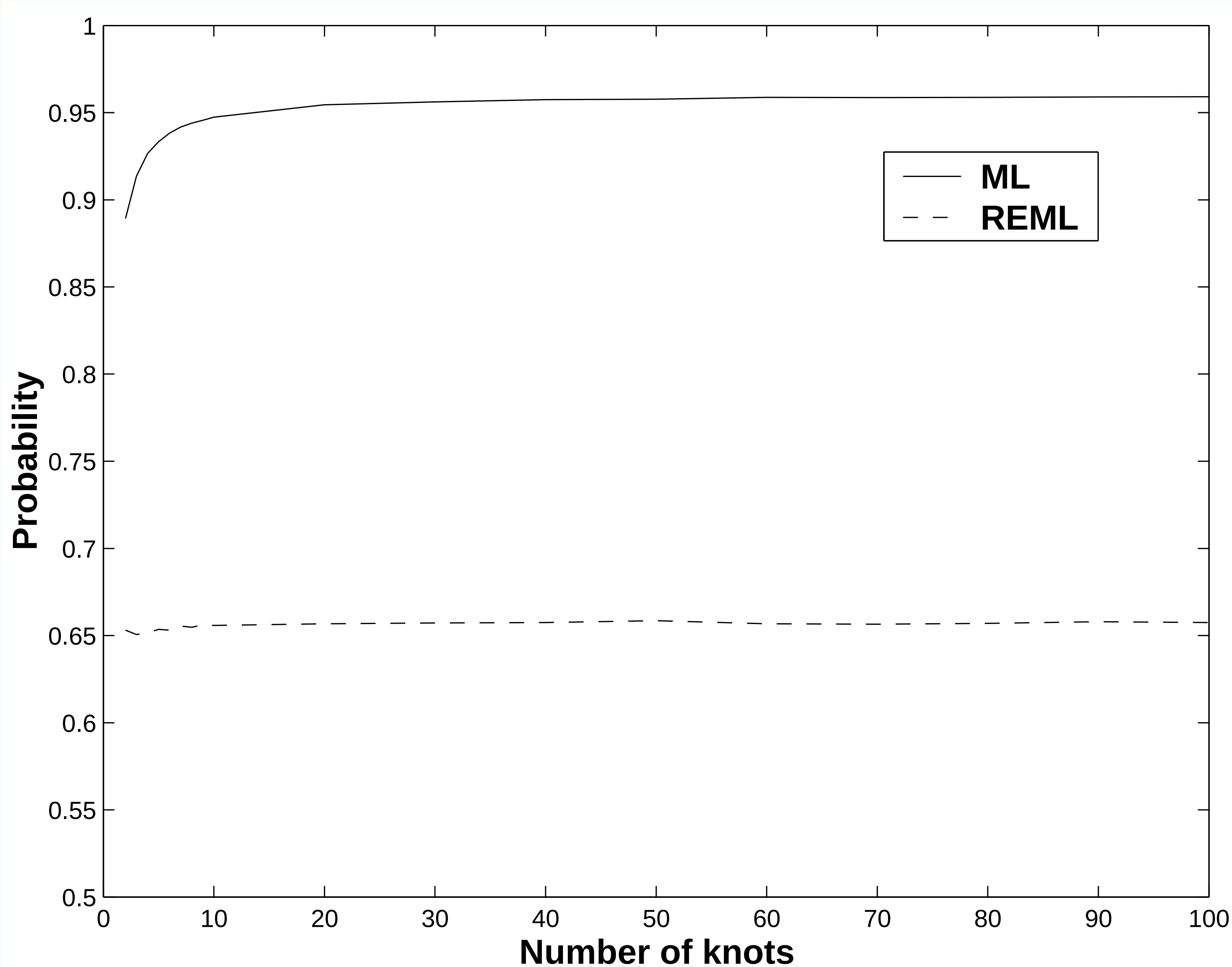
- x_i's equally spaced
- 20 equally spaced knots
- $p = q = 0$ (constant mean versus piecewise constant mean)

Then,

$$\lim_{n \to \infty} P_{H_0}\{\log(RLR) = 0\} = .6567, \quad \text{not .5}$$

and

$$\lim_{n \to \infty} P_{H_0}\{\log(LR) = 0\} = .9545, \quad \text{not .5}$$

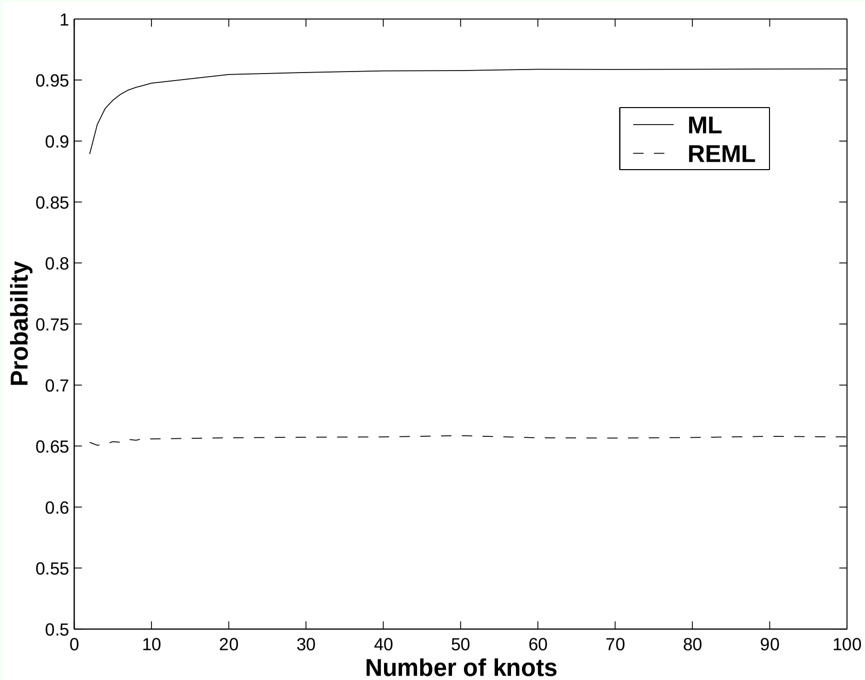


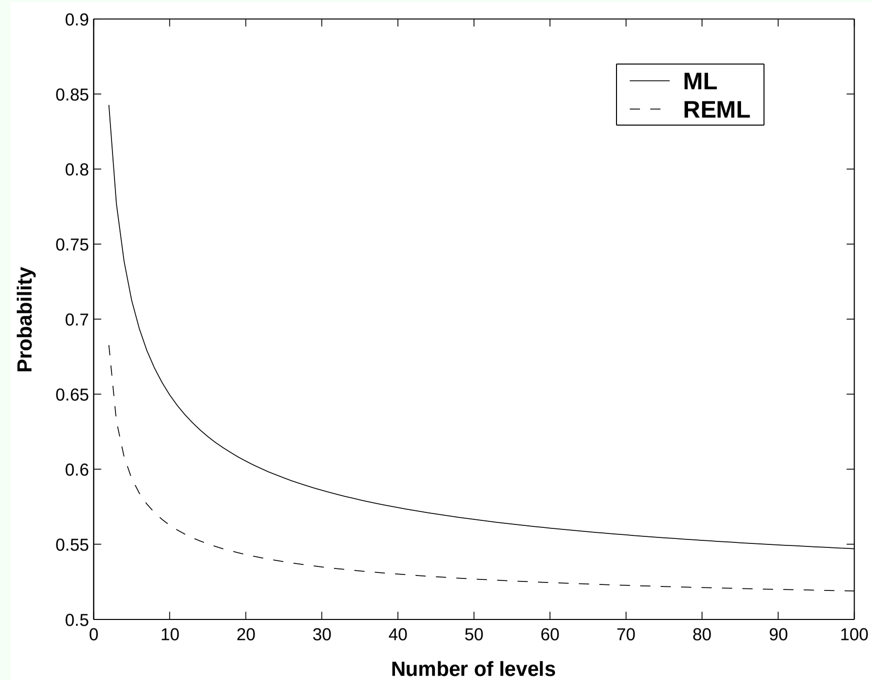
P-splines: $\lim_{n\to\infty} P_{H_0}\{\log(LR) = 0\}$

ORTHOGONALIZATION

- one can apply Gram-Schmidt to the “design matrix”
  - power functions are replaced by orthogonal polynomials
  - “Plus functions” are replaced by spline basis functions that are orthogonal to polynomials
- The asymptotics of the LRT are changed by this reparametrization

- Asymptotics are essentially the same as for 1-way ANOVA with
  - I ($=$ # levels) = K ($=$ # knots) +1
    - E.g., 5 levels is like 4 knots

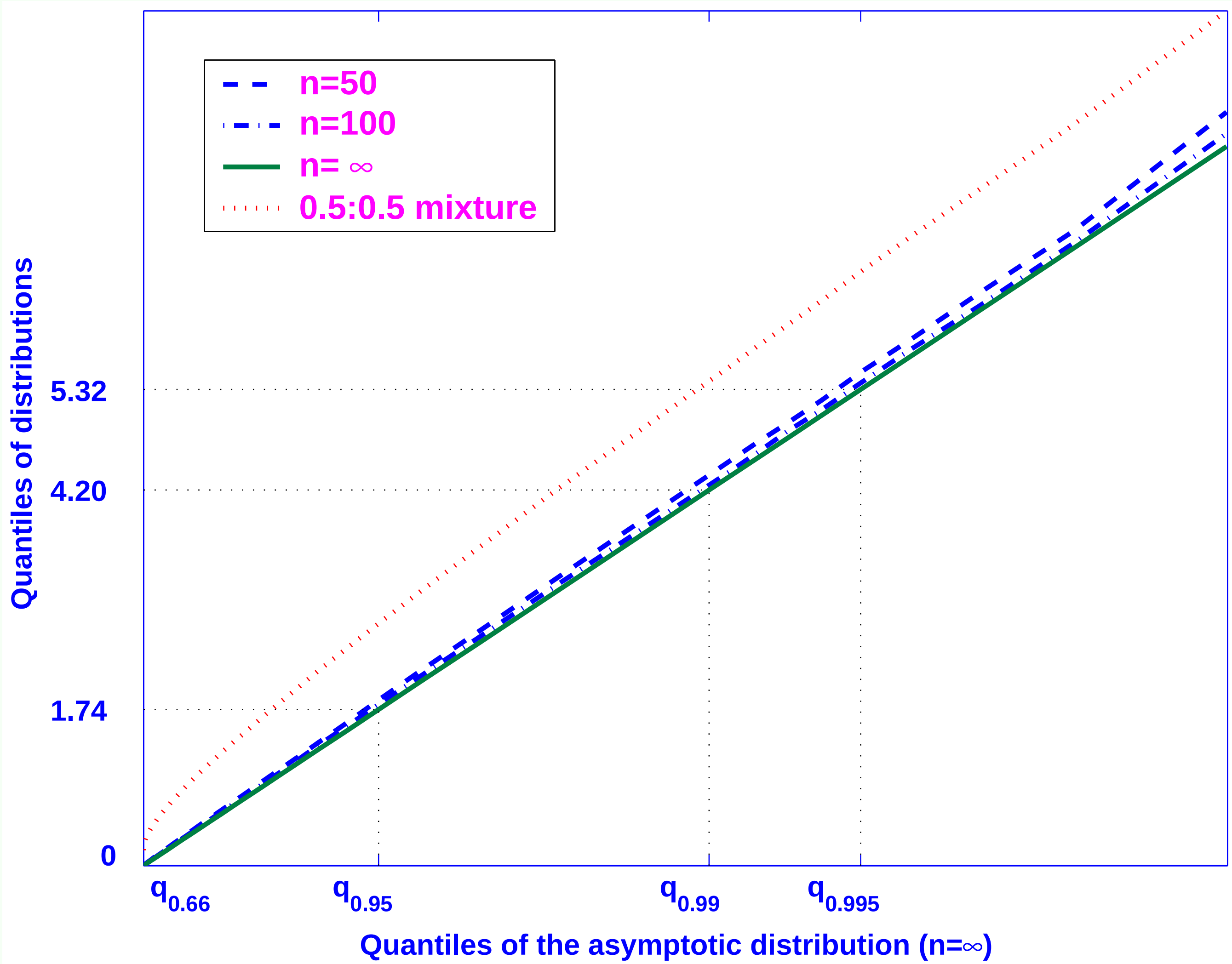




P-splines

Orthogonalized = 1-way ANOVA

Asymptotic null probabilities that log-LR is zero



Comparison of finite-sample and asymptotic quantiles

Hypotheses: linear trend versus 20-knot linear spline

Comparison of LRT with other tests

Reference: Crainiceanu, Ruppert, Aerts, Claeskens, and Wand (2002, in preparation)

- Results in next table are for testing
  - constant mean

  versus

  - general alternative
    - piecewise constant spline, or
    - linear spline

- The comparisons are made with an
  - increasing,
  - concave, and
  - periodic

  mean function, chosen so that good tests had power approximately 0.8

- R-test is from Cantoni and Hastie (2002)
- F-test is as in Hastie and Tibshirani (1990)
- “C” means alternative is a piecewise constant function
- “L” means alternative is a linear spline

- “1” means estimate under alternative has DF one greater than under null
- “ML” means smoothing parameter under alternative is chosen by ML
- “GCV” means smoothing parameter under alternative is chosen by GCV

Test	Average power	Maximum power	Minimum Power
RLRT-C	0.8885	0.9660	0.8166
R-GCV-L	0.8737	0.9910	0.7188
R-ML-C	0.8615	0.9916	0.7022
F-ML-L	0.8569	0.8796	0.8328
R-ML-L	0.8569	0.8796	0.8328
F-ML-C	0.8534	0.9928	0.6708
F-GCV-L	0.8482	0.9946	0.6634
LRT-L	0.7561	0.8466	0.6832
F-1-C	0.7087	0.8442	0.4816
F-1-L	0.6775	0.9414	0.3012
R-1-L	0.6239	0.9126	0.1462
R-GCV-C	0.6144	0.9284	0.3392

Conclusions

- Standard asymptotics are, in general, not suitable
- Better asymptotics for one variance component are feasible
- For more than one variance component, one might need to use simulation to get p-values

References

Brumback, B., and Rice, J., (1998), Smoothing spline models for the analsysi of nested and crossed samples of curves, *JASA*, **93**, 944–961

Brumback, B., Ruppert, D., and Wand, M.P., (1999). Comment on “Variable selection and function estimation in additive nonparametric regression using data-based prior” by Shively, Kohn, and Wood, *JASA*, **94**, 794–797.

Cantoni, E., and Hastie, T.J., 2002. Degrees of freedom tests for smoothing splines, *Biometrika*, **89**. 251–263.

Crainiceanu, C. M., and Ruppert, D., (2002), Asymptotic distribution of likelihood ratio tests in linear mixed models, submitted

Crainiceanu, C. M., Ruppert, D., and Vogelsang, T. J., (2002). Probability that the mle of a variance component is zero with applications to likelihood ratio Tests, manuscript

Crainiceanu, C. M., Ruppert, D., Aerts, M., Claeskens, G., and Wand, M., (2002), Tests of rolynomial regression against a general alternative, in preparation.

Hastie, T.J., Tibshirani, R., 1990. *Generalized Additive Models*, London: Chapman and Hall.

Lin, X., and Zhang, D. (1999), Inference in generalized additive mixed models by using smoothing splines, *JRSS-B*, 61, 381–400.

Pinheiro, J., and Bates, D., (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York.

Rice, J., and Wu, C., (2001), Nonparametric mixed effects models for unequally sampled noisy curves," *Biometrics*, **57**, 253–259.

Self, S., and Liang, K., (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions, *JASA*, **82**, 605–610.

Stram, D., Lee, J., (1994). "Variance Components Testing in the Longitudinal Mixed Effects Model," *Biometrics*, **50**, 1171–1177.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998), Semi-parametric stochastic mixed models for longitudinal data, *JASA*, **93**, 710–719