Prediction of Random Effects and Effects of Misspecification of Their Distribution

Charles E. McCulloch and John Neuhaus
Division of Biostatistics,
Department of Epidemiology and Biostatistics
University of California, San Francisco

CAPS, March 2008

Outline

1) Introduction and motivating examples
2) Prediction of random effects
   a) Are parametric assumptions important?
3) Brief review of effects of misspecification (more generally) in mixed models
4) Theoretical calculations (Linear Mixed Model)
5) Theoretical calculations (Binary Matched Pairs)
6) Simulations (Linear Mixed Model)
7) Example: Hormone and Estrogen Replacement Study
8) Summary

1. Introduction: Examples

Example 1: Measuring cognitive decline in elderly women (Women Who Maintain Optimal Cognitive Function into Old Age. Barnes DE, Cauley JA, Lui L-Y, Fink HA, McCulloch CE, Stone KL, Yaffe K. *J Amer Geriatics Soc*, 2007). A modified Mini-Mental status examination was given at baseline and years 6, 8, 10 and 15 in a prospective cohort study (Study of Osteoporotic Fractures). Which participants are thought to be in mental decline and what predicts that decline?

Example 2: Effect of pre-hypertension at an early age in the CARDIA study. (Prehypertension During Young Adulthood and Presence of Coronary Calcium Later in Life: The Coronary Artery Risk Development In Young Adults (CARDIA) Study. MJ Pletcher, K Bibbins-Domingo, CE Lewis, G Wei, S Sidney, JJ Carr, E Vittinghoff, CE McCulloch, SB Hulley, submitted). Blood pressure measured every five years since 1986. How to approximate previous and cumulative blood pressure exposure?

Example 3: Predicting those at risk for developing high blood pressure in HERS (The Heart and Estrogen Replacement Study - Hulley, et al, J. American Medical Association, 1998). HERS was a randomized, blinded, placebo controlled trial for women with previous coronary disease. We will use it as a prospective cohort study for prediction of high blood pressure. 2,763 women were enrolled and followed yearly for 5 subsequent visits. We will consider only the subset that were not diabetic and with systolic blood pressure less than 140 at the beginning of the study.

2. Mixed models and prediction of random effects

One way to address the questions above is to utilize mixed models and derive predicted values of the random effects.

Example 1: (cognitive decline):

$$MMSE_{it} = \text{cognitive measure for participant } i \text{ at time } t$$

$$= b_{0i} + b_{1i}t + \text{covariates} + \varepsilon_{it},$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \text{indep. } N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{pmatrix} \right)$$

$$\text{calculate } \tilde{b}_{1i} = \text{predicted decline for participant } i.$$

Some realistic but made up data:

```
. table visit, c(mean mmse n mmse sd mmse)

----------------------------------------
    visit | mean(mmse)  N(mmse)  sd(mmse)
----------+-----------------------------
        0 |   27.08       2,031    2.2
        1 |   27.17       1,931    2.3
        2 |   27.10       1,850    2.3
        3 |   27.08       1,750    2.3
        4 |   27.04       1,361    2.3
        5 |   27.10         269    2.2
----------------------------------------
```

So little change in average MMSE over time.

```
xi: xtmixed mmse visit exercise avgdrpwk || pptid: visit, cov(uns)

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -11662.158
Iteration 1:   log restricted-likelihood =  -11662.14
Iteration 2:   log restricted-likelihood =  -11662.14

Computing standard errors:

Mixed-effects REML regression       Number of obs      =      9110
Group variable: pptid               Number of groups   =      2032

                                    Obs per group: min =         1
                                                   avg =       4.5
                                                   max =         6

                                    Wald chi2(3)       =     27.24
Log restr-likelihood =  -11662.14   Prob > chi2        =    0.0000
```

```
------------------------------------------------------------------------------
        mmse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       visit |   -.0060353   .0059123    -1.02   0.307    -.0176231    .0055526
    exercise |    .0773954   .0179999     4.30   0.000     .0421162    .1126746
    avgdrpwk |   -.0097331   .0037005    -2.63   0.009    -.0169859   -.0024803
       _cons |    27.11017   .0495455   547.18   0.000     27.01307    27.20728
------------------------------------------------------------------------------

------------------------------------------------------------------------------
Random-effects Parameters  |   Estimate   Std.Err.     [95% Conf. Interval]
---------------------------+--------------------------------------------------
pptid: Unstructured        |
                 sd(visit) |    .1942305   .005721      .183333    .2057752
                 sd(_cons) |    2.158639   .034978      2.09115    2.228296
          corr(visit,_cons) |   -.0426722   .031013     -.103225    .0181959
---------------------------+--------------------------------------------------
               sd(Residual) |     .481975   .004743      .472767    .4913616
------------------------------------------------------------------------------
LR test vs. lin regression: chi2(3) = 17033.5  Prob > chi2 = 0.000
```

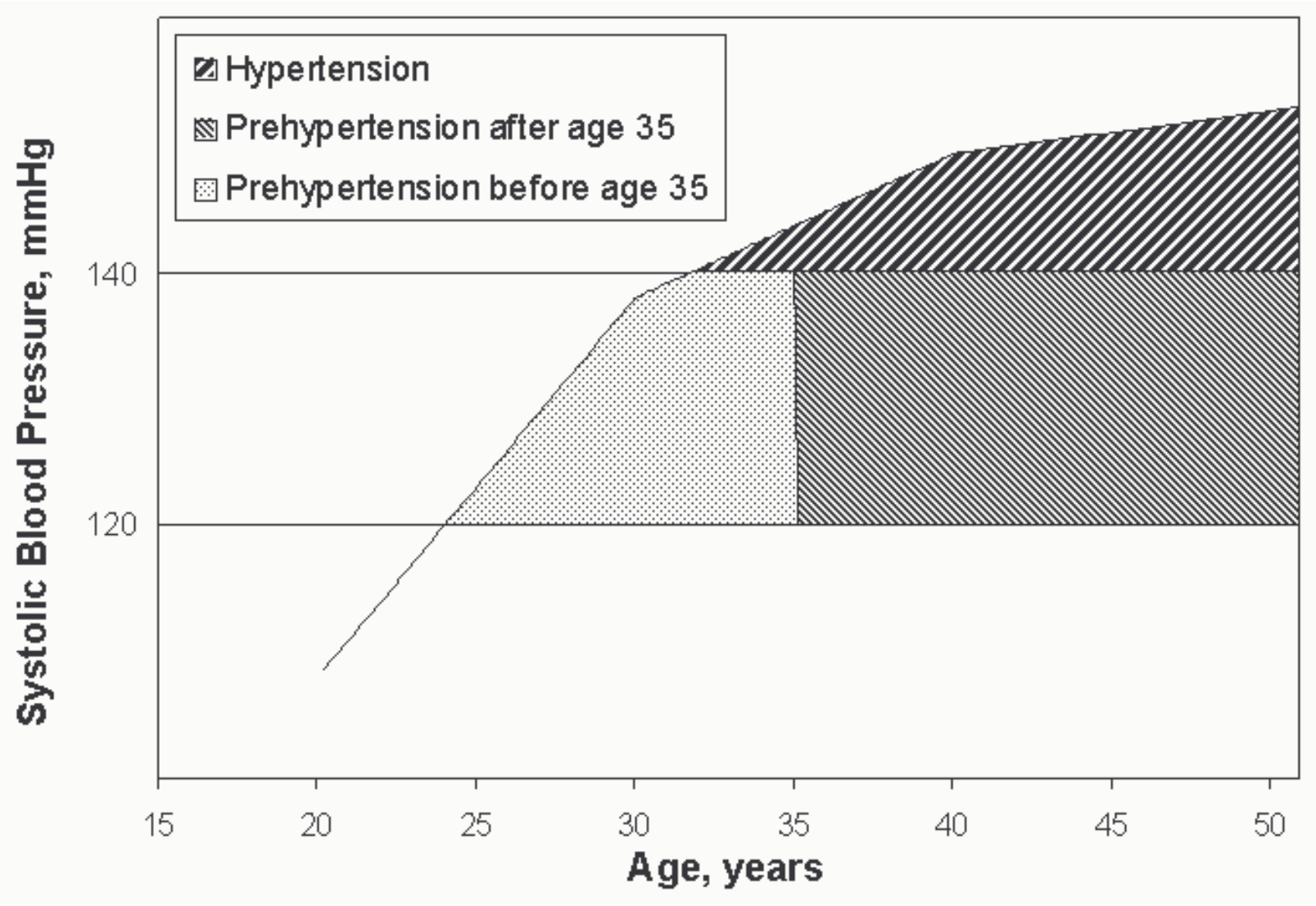
```
. predict rslopedev rintdev, reffects
. gen predslope=_b[visit]+rslopedev
. collapse rslopedev rintdev predslope, by(pptid)
. gen deltammse=6*predslope

. summarize

 Variable |   Obs     Mean    Std. Dev.       Min      Max
----------+-----------------------------------------------
    pptid |    2032   1394.65   794.41          1      2761
rslopedev |    2032   1.3e-10   .1421     -.7615     .9239
  rintdev |    2032   3.9e-10   2.133    -9.8275    3.0384
predslope |    2032  -.006035   .1421     -.7676     .9178
deltammse |    2032  -.036211   .8530    -4.6057    5.5071

. summarize deltammse predslope if deltammse<-2

 Variable |   Obs    Mean   Std. Dev.       Min        Max
----------+-----------------------------------------------
deltammse |    40  -2.634   .4937       -4.6057    -2.0331
predslope |    40  -.4390   .0822        -.7676     -.3388
```

Example 2: (pre-hypertension):

$$BP_{it} = \text{blood pressure for participant } i \text{ at time } t$$
$$= spline_i(t) + \text{covariates} + \varepsilon_{it},$$
$$(spline\ terms) \sim \text{indep. } N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

calculate predicted spline for participant i.

The area under the predicted blood pressure trajectory between 120 and 140 mmHg was integrated over time as a cumulative pre-hypertension exposure (in years of mmHg). This was then used as a predictor of coronary calcification.


Hypertension
Prehypertension after age 35
Prehypertension before age 35
Systolic Blood Pressure, mmHg
140
120
15
20
25
30
35
40
45
50
Age, years

Example 3: (high blood pressure):

$BP_{it} = 1$ if blood pressure is high for subject i at time t, and 0 o/w

$\text{logit}(P\{BP_{it} = 1\}) = b_{0i} + \text{covariates}$

$b_{0i} \sim \text{i.i.d. } N\left(\beta_0, \sigma_b^2\right)$

calculate $\tilde{b}_{0i}$ = predicted intercept for participant i.

Predicted values of random effects available from `gllamm` or the new (Ver 10) multilevel logit command `xtmelogit`

Standard software (e.g., SAS Proc MIXED or NLMIXED; Stata `xtmixed` or `xtlogit`) fit the models using regular or restricted maximum likelihood. So they use a parametric assumption for both the distribution of the outcome and the distribution of the random effects, the latter typically that the distributions are normal.

Key question: Is the parametric assumption of the random effects distribution important?

This is especially crucial since we don't get to directly observe the random effects. Unfortunately, the predicted random effects may not reflect the shape of the underlying distribution. (More on this point later).

3. Review of impact of misspecification in mixed models

A number of investigations have focused on the effect of misspecifications in parametric mixed models. They can be grouped as:

1. Getting the distributional shape wrong.
2. Falsely assuming the random effect is independent of the cluster size.
3. Falsely assuming the random effect is independent of covariates, e.g.,
    a. Mean of random effects distribution could be associated with a covariate.
    b. Variance of random effects distribution could be associated with a covariate.

Most investigations have concentrated on the impact on estimation of the fixed effects portion of the model.

General assessment:

1) Getting the distributional shape wrong has little impact on inferences about the fixed effects.

2) Incorrectly assuming the random effects distribution is independent of the cluster size may affect inferences about the intercept, but does not seriously impact inferences about the regression parameters.

3) However, assuming the random effects distribution is independent of the covariates when it is not is potentially serious. (Related to mean: Neuhaus and McCulloch, JRSSB, 2006; related to the variance: Heagerty and Kurland, Biometrika, 2001).

What about inference about the predictions of the random effects?

We'll concentrate on the issue of wrong distributional shape, where fixed effects inferences seem largely unaffected.

Intuition: the assumed form of the random effects distribution may be a more crucial assumption in this case.

4. Theoretical calculations (Linear Mixed Model)

First consider an easy situation. Assumed model is a one-way random effects model with known intercept and variance components and normally distributed random effects:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \ldots, n_i; i = 1, \ldots, q$$

$$b_i \sim i.i.d.\ N\left(0, \sigma_b^2\right)$$

$$\varepsilon_{it} \sim i.i.d.\ N\left(0, \sigma_\varepsilon^2\right)$$

$$\varepsilon_{it} \perp b_i, \mu, \sigma_\varepsilon^2, and\ \sigma_b^2\ known$$

The Best Linear Unbiased Predictor is defined as the prediction, $\tilde{b}_i$, that minimizes $E[(\tilde{b}_i - b_i)^2]$ among linear functions of Y that are unbiased: $E[\tilde{b}_i - b_i] = 0$.

Easy calculation shows this is given by $E[b_i \mid Y]$.

$$
\begin{aligned}
E[(\tilde{b}_i - b_i)^2] &= E[\{(\tilde{b}_i - E[b_i \mid y]) + (E[b_i \mid y] - b_i)\}^2] \\
&= E[(\tilde{b}_i - E[b_i \mid y])^2] + E[(E[b_i \mid y] - b_i)^2] + 0 \\
&= \text{nonnegative quantity} + \text{quantity indep of } \tilde{b}_i
\end{aligned}
$$

which is minimized when the nonnegative quantity is chosen to be zero, i.e., $\tilde{b}_i = E[b_i \mid y]$.

For our model, the Best Linear Unbiased Predictor is given by

$$\tilde{b}_i = E[b_i \mid Y] = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i}\left(\bar{Y}_{i\cdot} - \mu\right)$$

Writing this out:

$$\begin{aligned}
\tilde{b}_i &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i}\left(\bar{Y}_{i\cdot} - \mu\right) \\
&= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i}\left(\mu + b_i + \bar{\varepsilon}_{i\cdot} - \mu\right) \\
&= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i}\left(b_i + \bar{\varepsilon}_{i\cdot}\right)
\end{aligned}$$

Conditional on b_i, the Y_{it} are independent $N(\mu + b_i, \sigma_\varepsilon^2)$. So

$$\tilde{b}_i \mid b_i \sim N\left(\mu_{\tilde{b}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} b_i, \left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i} \right)^2 \frac{\sigma_\varepsilon^2}{n_i} \right)$$

and $\tilde{b}_i$ is conditionally biased. Since the calculations are conditional on b_i, results do not depend on the distribution of the b_i and so the conditional bias does *not* depend on the distribution.

$$\tilde{b}_i = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2 / n_i}\left(b_i + \bar{\varepsilon}_{i\cdot}\right)$$

$$\rightarrow \frac{\sigma_b^2}{\sigma_b^2}\left(b_i + 0\right) = b_i \text{ as } n_i \rightarrow \infty$$

So $\tilde{b}_i$ converges in probability to the true value as $n_i \rightarrow \infty$. But asymptotic calculations as $n_i \rightarrow \infty$ are not usually of interest for a random effects model.

What does the distribution of the $\tilde{b}_i$ look like?

And what if the assumption of normality for the b_i is incorrect, i.e., not normal?

If n_i is large then each $\tilde{b}_i$ is close to b_i and hence the distribution is approximately correct.

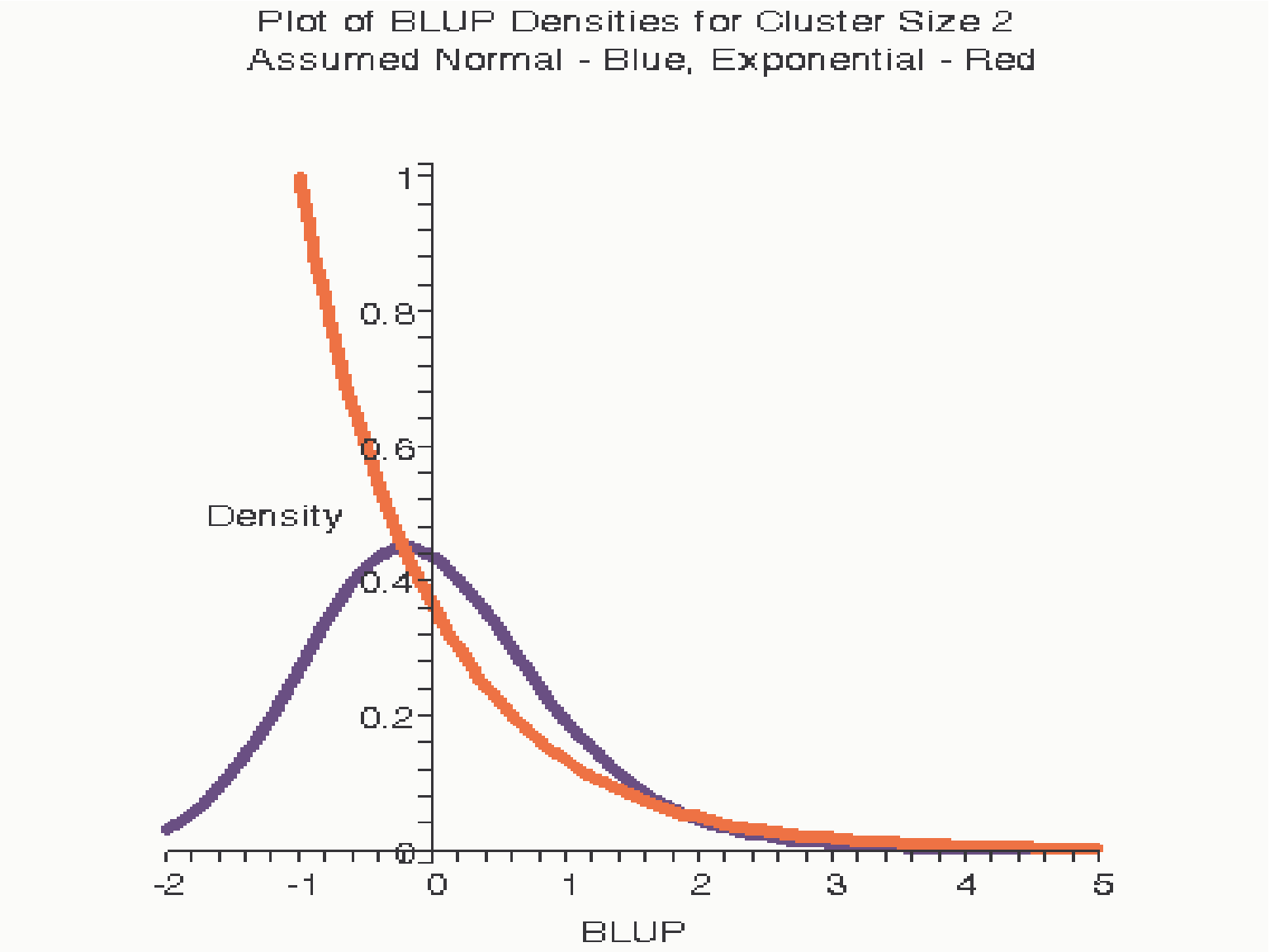
But what about the case when n_i is not large, the usual case of interest?

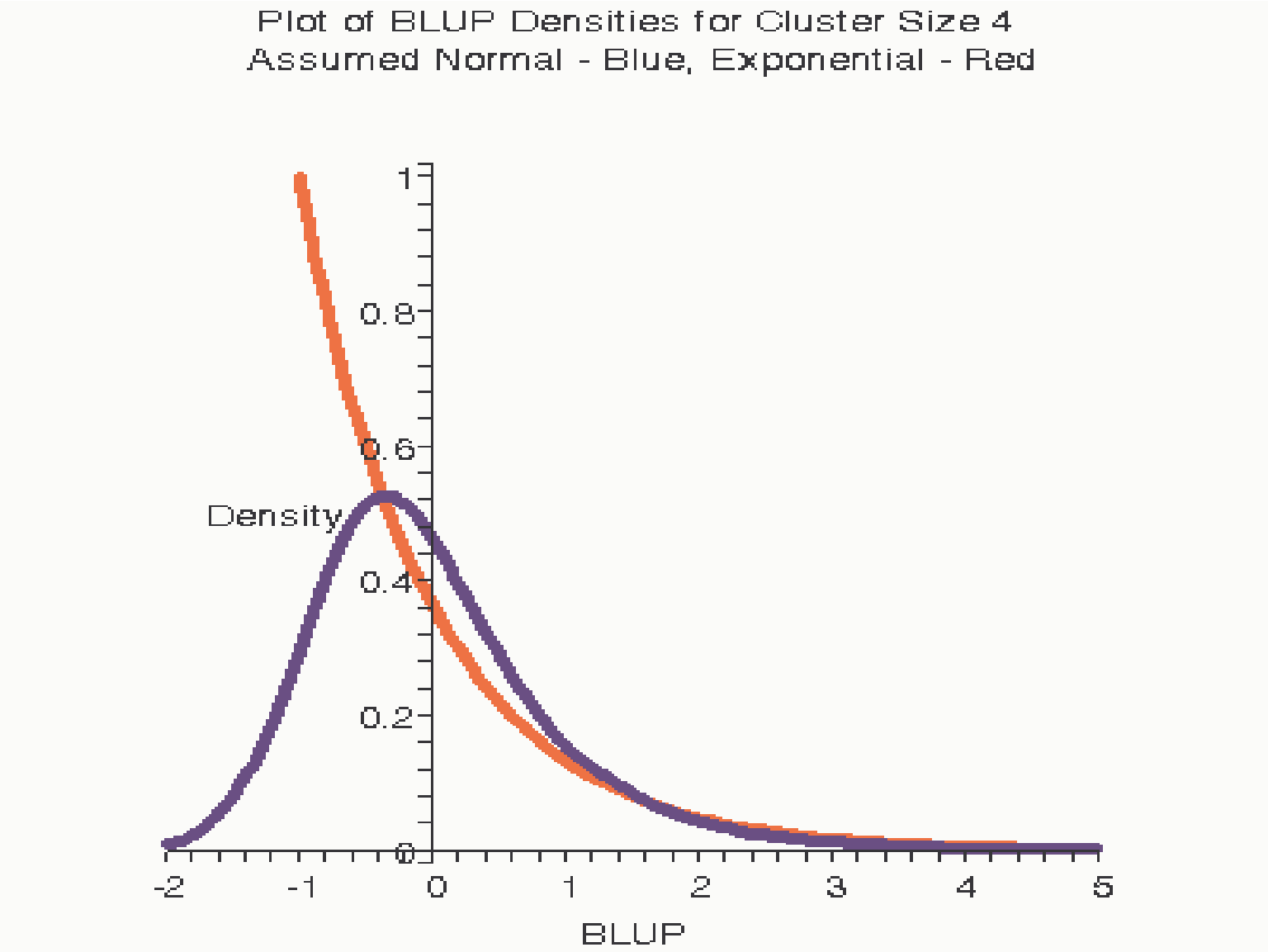
Then the distribution of $\tilde{b}_i$ is the convolution of the true density with the conditional density of $\tilde{b}_i$ given b_i.

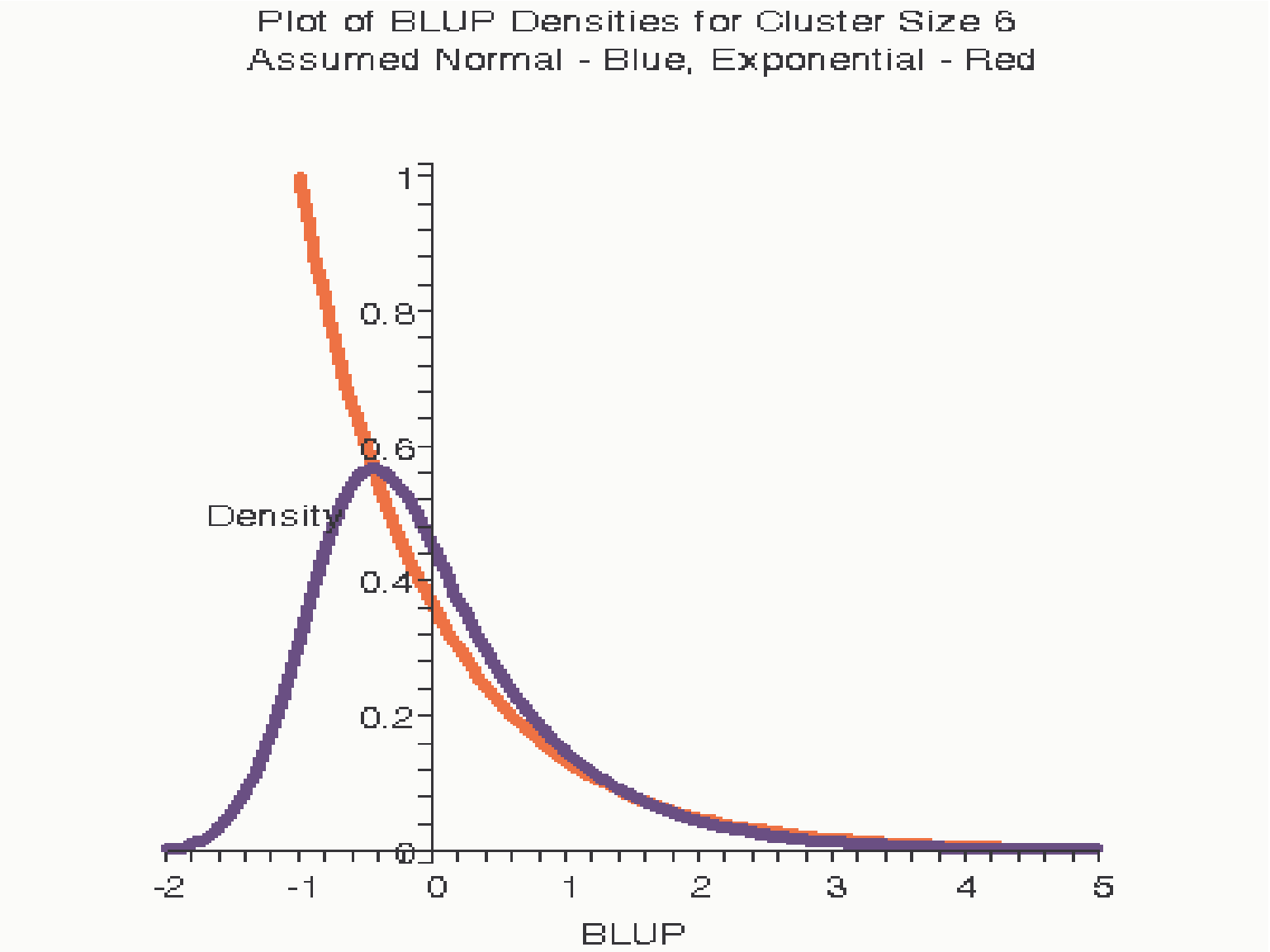
For example, suppose the true density is exponential(1), shifted to have mean 0. Then the density of $\tilde{b}_i$ is given by

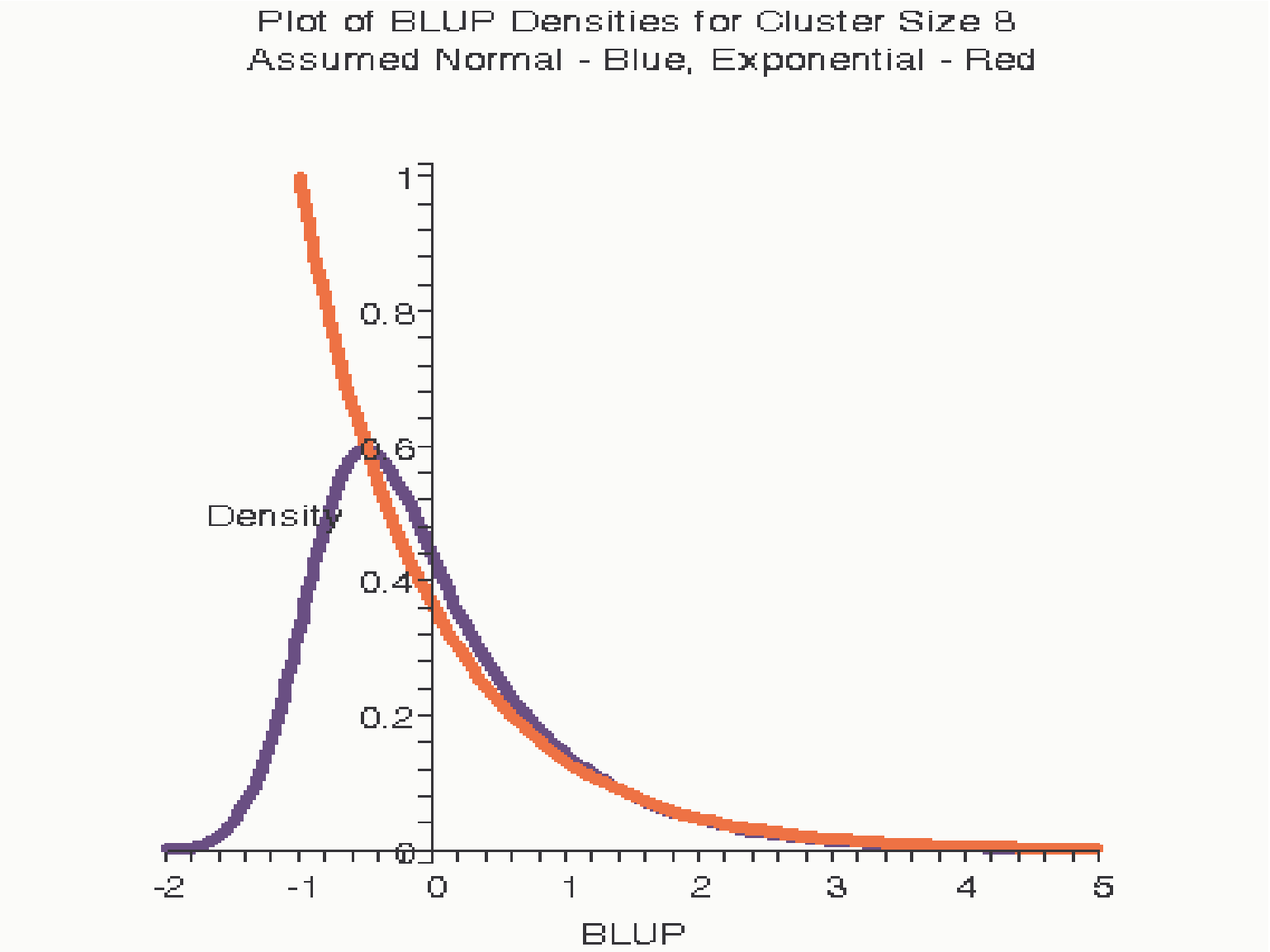
$$\int_0^\infty \exp\left\{-(\tilde{b}-\mu_{\tilde{b}})^2 n_i/(2\sigma_\varepsilon^2)\right\}\exp(-\tilde{b}-1)d\tilde{b},$$

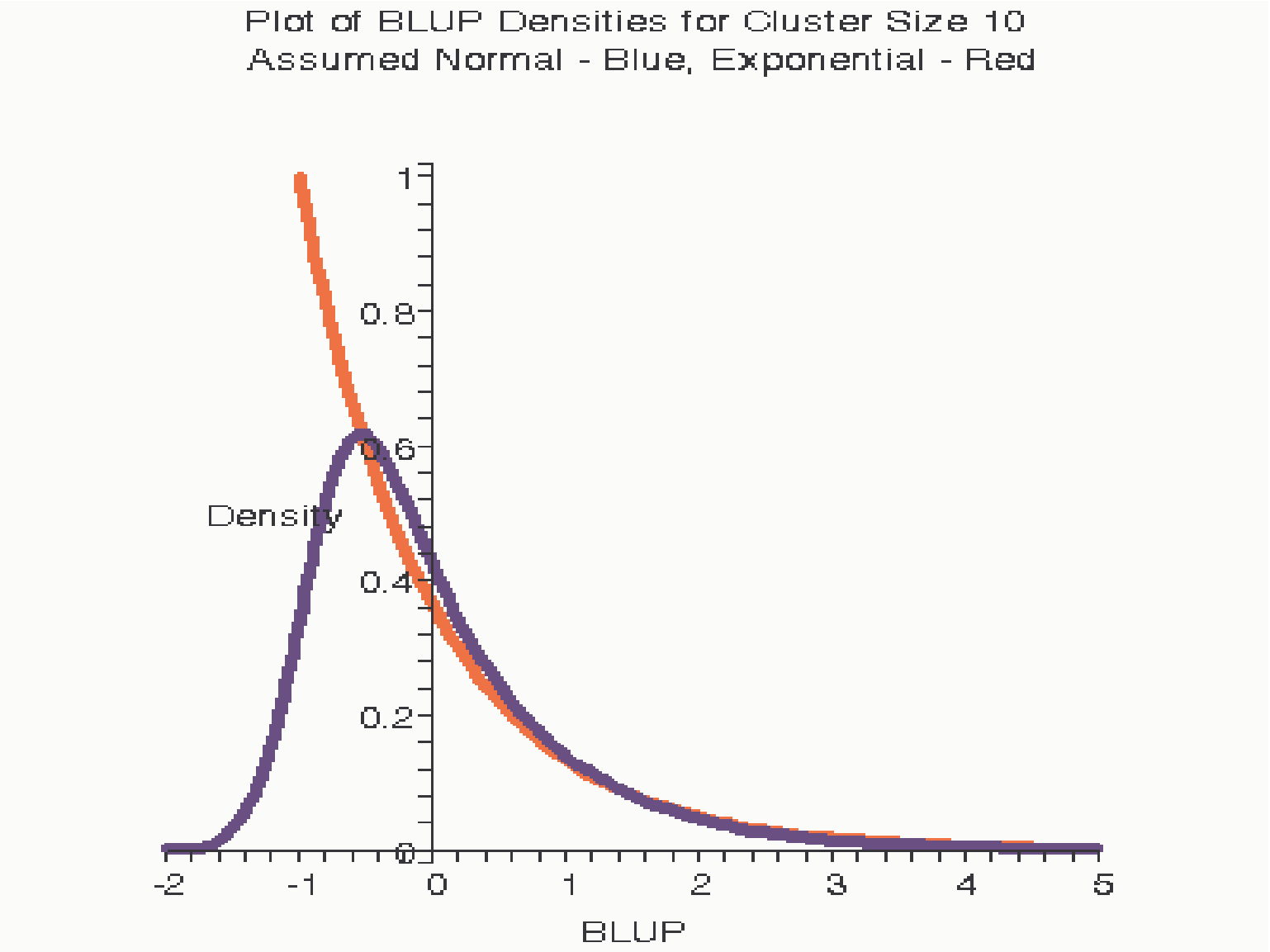
which is straightforward to evaluate numerically:

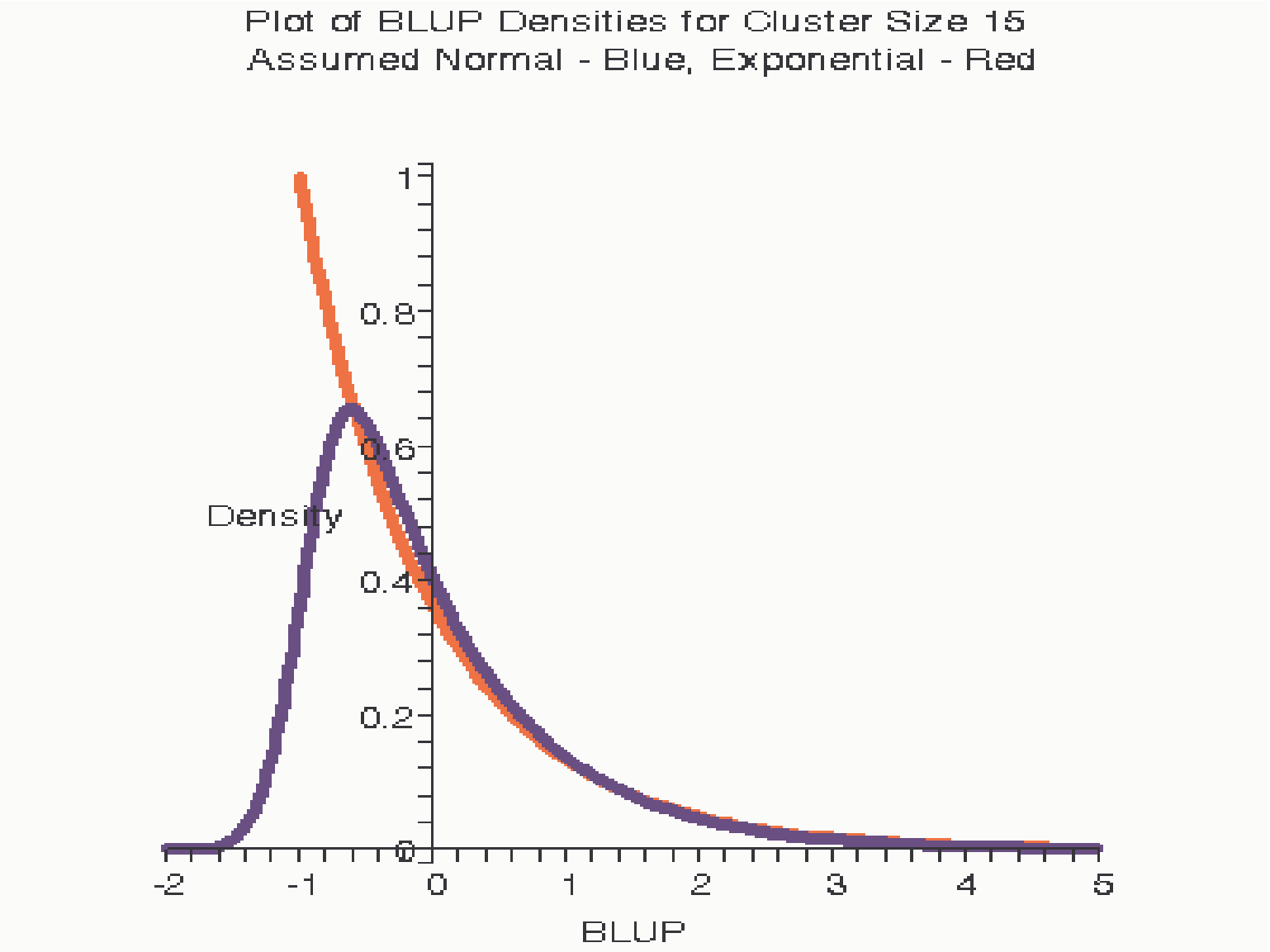

Plot of BLUP Densities for Cluster Size 2
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP

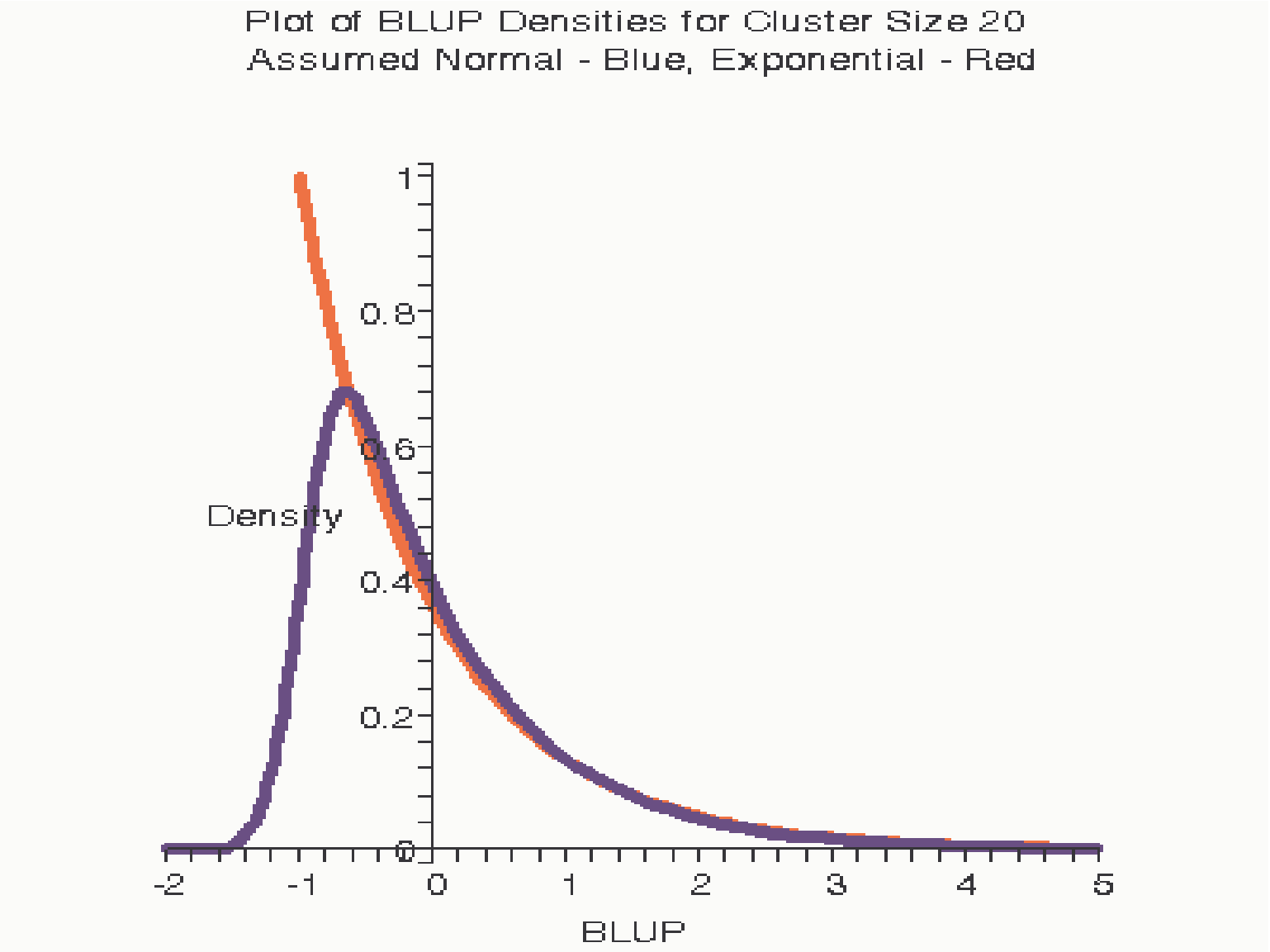

Plot of BLUP Densities for Cluster Size 4
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP


Plot of BLUP Densities for Cluster Size 6
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP


Plot of BLUP Densities for Cluster Size 8
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP


Plot of BLUP Densities for Cluster Size 10
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP


Plot of BLUP Densities for Cluster Size 15
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP


Plot of BLUP Densities for Cluster Size 20
Assumed Normal - Blue, Exponential - Red
Density
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
5
BLUP

What is the BLUP under the exponential assumption?

Model:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \ldots, n_i; i = 1, \ldots, q$$
$$b_i \sim \text{i.i.d.}\, \sigma_b(\mathcal{E}(1) - 1)$$
$$\varepsilon_{it} \sim \text{i.i.d.}\, N\left(0, \sigma_\varepsilon^2\right)$$
$$\varepsilon_{it} \perp b_i;\ \mu, \sigma_\varepsilon^2, \text{and}\ \sigma_b^2 \text{ known}$$

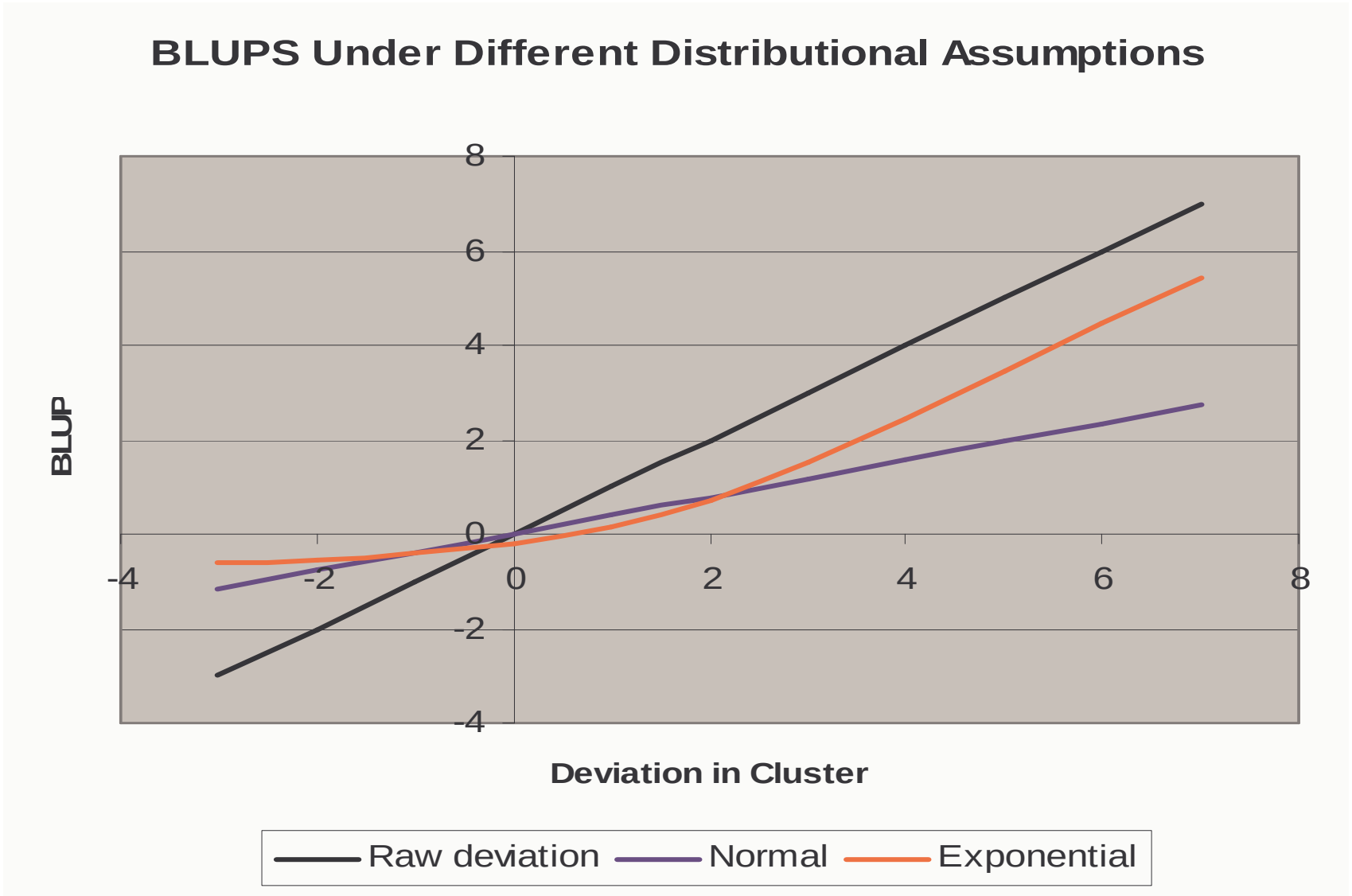
Define $\Delta = \dfrac{\sqrt{n}}{\sigma_\varepsilon}\left(\bar{Y} - \mu + \sigma_b\right) - \dfrac{\sigma_\varepsilon}{\sqrt{n}\sigma_b}$.

Then, with $\tilde{b}_i = E[b_i \mid Y]$,

$$\tilde{b}_i = \bar{Y} - \mu - \frac{\sigma_\varepsilon^2}{n\sigma_b} + \frac{\phi(\Delta)\sigma_\varepsilon}{\Phi(\Delta)\sqrt{n}} \ ,$$

where $\phi(t)$ and $\Phi(t)$ are the standard normal p.d.f. and c.d.f.

How do the assumed normal and assumed exponential BLUPs compare?


BLUPS Under Different Distributional Assumptions
BLUP
8
6
4
2
0
-2
-4
-4
-2
0
2
4
6
8
Deviation in Cluster
Raw deviation
Normal
Exponential

5. Theoretical calculations (Binary matched pairs)

Assumed model

$$Y_{it} \mid b_i \sim \text{Binomial}(p_{it}), i = 1,\ldots,q;\ t = 1,2$$

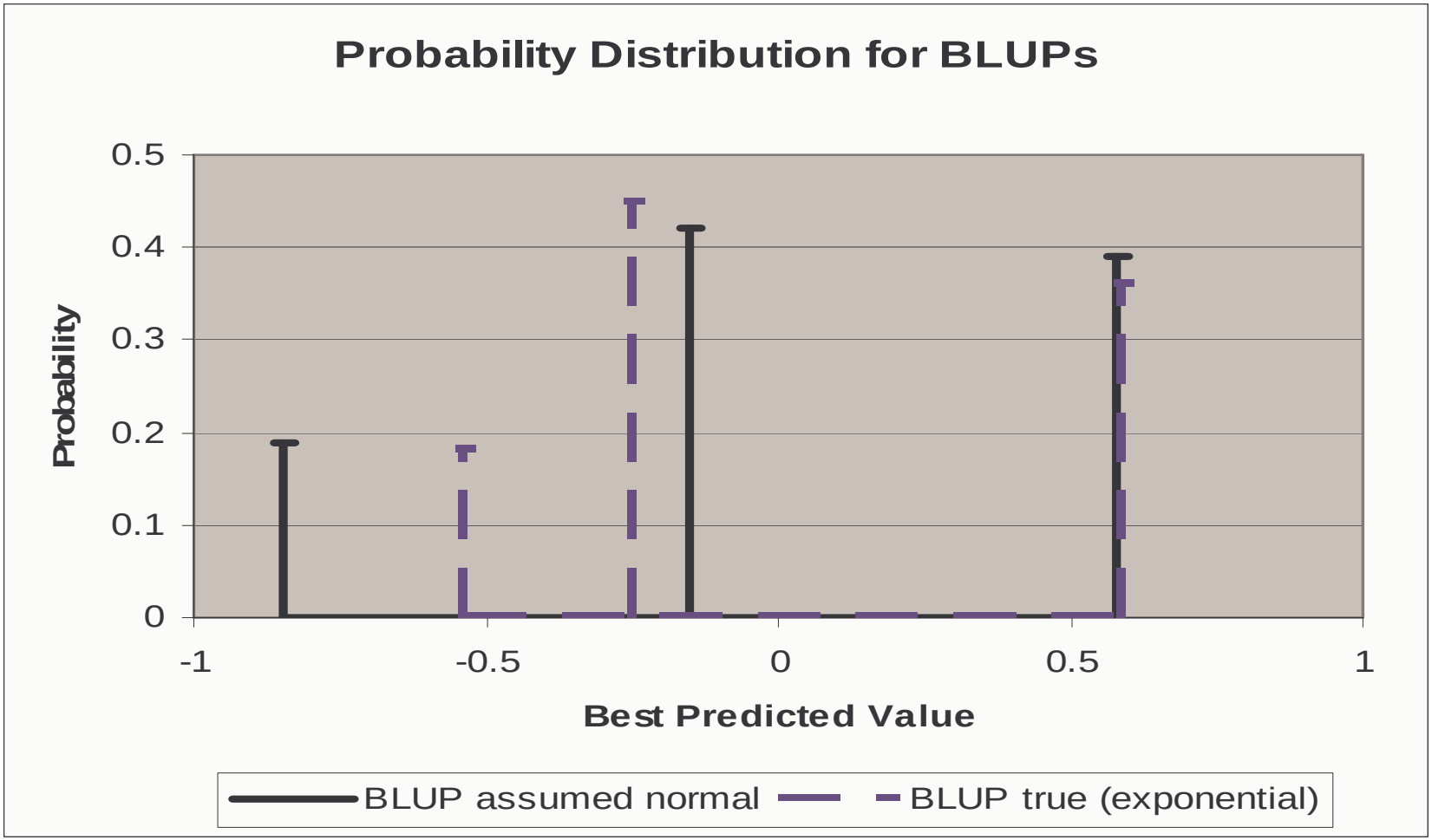
$$\text{logit}(p_{it}) = \mu + b_i + \beta I_{\{t=2\}}$$

$$b_i \sim \text{i.i.d. } N\left(\beta_0, \sigma_b^2\right)$$

Since there are only 4 data configurations per cluster there are only four possible values for $\tilde{b}_i$, for a given set of parameter values. For example, when $y_{i1} = y_{i2} = 1$, $\tilde{b}_i$ is given by (with $p(t) = 1/(1+e^{-t})$)

$$\tilde{b}_i = \frac{\int_{-\infty}^{\infty} b\phi(b)\, p(\mu + \sigma_b b)\, p(\mu + \sigma_b b + \beta) db}{\int_{-\infty}^{\infty} \phi(b)\, p(\mu + \sigma_b b)\, p(\mu + \sigma_b b + \beta) db}$$

These depend on the assumed distribution. The probabilities of the four (actually three) values depends on the true distribution.



It is also straightforward to calculate the mean square error of prediction using the assumed and true models under the true model. For example, if the assumed model is normal, but the true is exponential here are some values of the mean square error of prediction:

Mean squared error of prediction MSEP = $E[(\tilde{b}_i - b_i)^2]$ with $\mu = 0, \sigma = 1$:

β	Normal (assumed)	Exponential (true)	Percent increase
0	0.77	0.75	3.5%
1	0.82	0.79	3.0%
2	0.85	0.83	2.1%
3	0.87	0.85	1.4%

6. Simulation

We simulated data from the one-way random model:

$$Y_{it} = \mu + b_i + \varepsilon_{it}, t = 1, \ldots, n_i; i = 1, \ldots, q$$
$$b_i \sim \text{i.i.d.} N\left(0, \sigma_b^2\right) \text{or } b_i \sim \text{i.i.d.} \sigma_b \{\mathcal{E}(1) - 1\}$$
$$\varepsilon_{it} \sim \text{i.i.d.} N\left(0, \sigma_\varepsilon^2\right), \varepsilon_{it} \perp b_i,$$

with $q = 10 = n_i$ and using the same random numbers for both the normal and exponential random effects (and the same error terms). 10,000 replications. An assumed normal model was fit.

Simulation results

Estimates of the parameters

Normal	True	Ave	SD	Ave SE
μ	1	1.00	0.33	0.32
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
Exponential				
μ	1	1.00	0.33	0.31
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47	0.29

*Excludes one outlier

Estimates of fixed effects parameters are little affected.

Estimates of the parameters

Normal	True	Ave	SD	Ave SE
μ	1	1.00	0.33	0.32
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
Exponential				
μ	1	1.00	0.33	0.31
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47*	0.29*

*Excludes one outlier

As is the estimate of log of the residual variance.

Estimates of the parameters

Normal	True	Ave	SD	Ave SE
μ	1	1.00	0.33	0.32
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
Exponential				
μ	1	1.00	0.33	0.31
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47*	0.29*

*Excludes one outlier

But the estimate of the random effects variance is off.

Estimates of the parameters

Normal	True	Ave	SD	Ave SE
μ	1	1.00	0.33	0.32
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.07	0.29*	0.27*
Exponential				
μ	1	1.00	0.33	0.31
$\ln(\sigma_\varepsilon^2)$	0	-0.01	0.075	0.075
$\ln(\sigma_b^2)$	0	-0.18*	0.47*	0.29*

*Excludes one outlier

Confidence interval coverage for μ was slightly lower than nominal for the normal (92%), and low for the exponential (88%).

Mean square error of prediction for the BLUPs was 1.87 for the normal model and 1.84 for the exponential.

Do the BLUPs calculated under the assumption of normality reflect the true underlying shape (exponential)?

For data simulated with normally distributed random effects the average skewness was -0.01 and the average kurtosis was 2.50 (with a normal having values 0 and 3).

For data simulated with exponentially distributed random effects the average skewness was 0.85 and the average kurtosis was 3.14 (with an exponential(1) having values 2 and 9).

7. Example (HERS)

Recall the HERS example: We will consider the 1,378 women who did not have high blood pressure and were not diabetic at the baseline visit. We will use the baseline and visits 1 through 3 to predict the blood pressure at visits 4 and 5 and whether or not the woman had developed high blood pressure on either visit 4 or 5.

Brief descriptive statistics:

Variable	Mean/Percentage	SD
Age	66.3	6.9
BMI	27.3	4.9
Weight	70.3 kg	13.4 kg
On BP meds	79%	

Predictive model (for baseline and visits 1, 2 and 3):

$$BP_{it} = \beta_0 + b_{0i} + \beta_1 BMI + \beta_2 EXER + \beta_3 AGE + \beta_4 MEDS + \beta_5 DM + \varepsilon_{it},$$

$$b_{0i} \sim \text{i.i.d. } N\left(0, \sigma_b^2\right) \text{ or } b_{0i} \sim \text{i.i.d. } \sigma_b \{\mathcal{E}(1) - 1\}$$

$$\text{calculate } \tilde{BP}_{it} = \hat{\beta}_0 + \tilde{b}_{0i} + \hat{\beta}_1 BMI + \hat{\beta}_2 EXER + \hat{\beta}_3 AGE + \hat{\beta}_4 MEDS + \hat{\beta}_5 DM \text{ (mixed model pred)}$$

$$\text{or } \hat{BP}_{it} = \hat{\beta}_0 + \hat{\beta}_1 BMI + \hat{\beta}_2 EXER + \hat{\beta}_3 AGE + \hat{\beta}_4 MEDS + \hat{\beta}_5 DM \text{ (fixed effects only)}$$

How well do the predictions work?

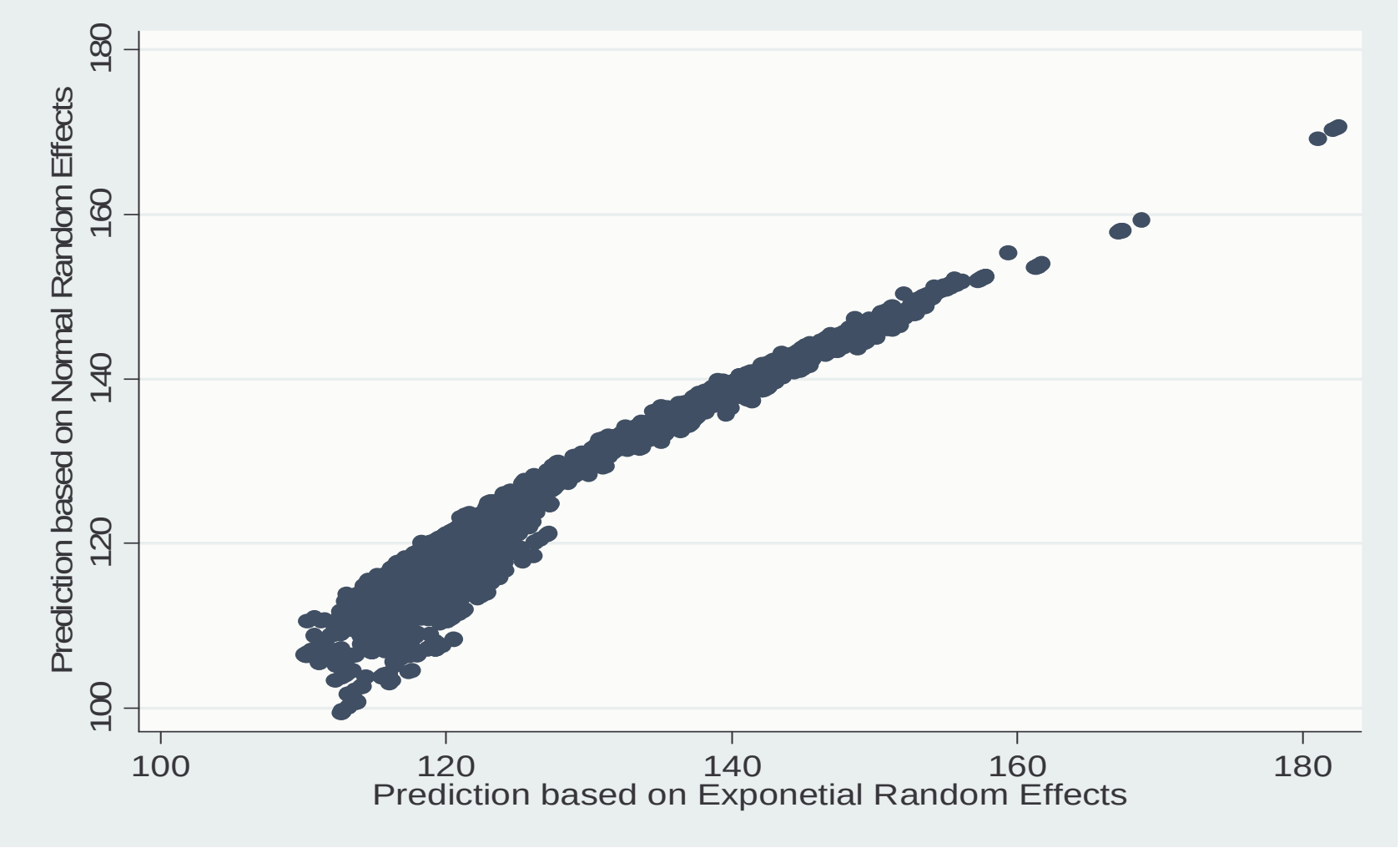
For predicting the actual systolic blood pressure:

	Prediction Errors		
Method	Ave	Ave abs	RMSE
Fixed effects only	3.4	13.8	18.1
Mixed model (normal)	3.9	11.0	14.9
Mixed model (exponential)	3.1	11.1	14.9

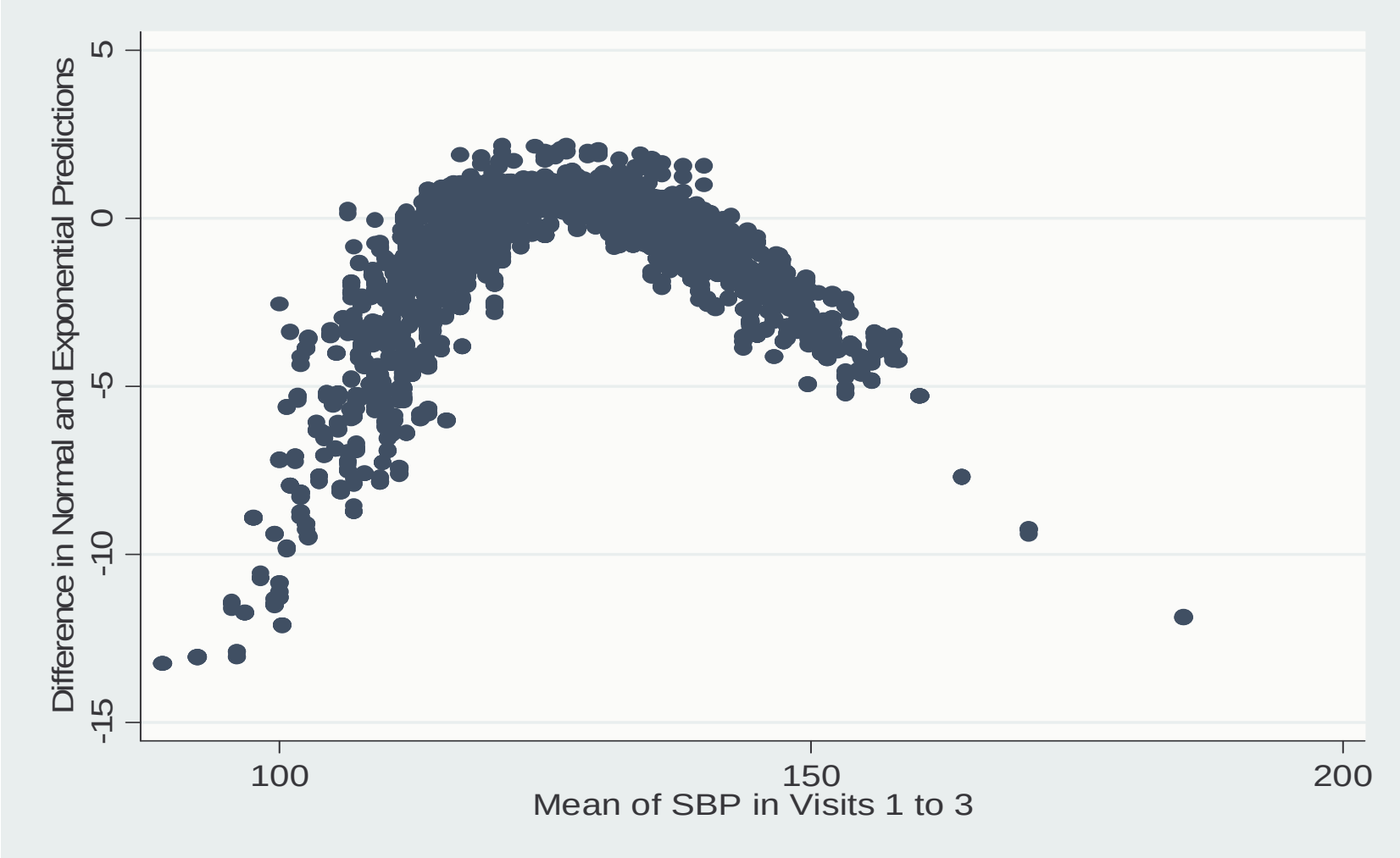
For predicting high BP or not:

Area under the ROC curve: Fixed effects – 0.55, Normal – 0.80, Exponential – 0.80.

Do they give the same predicted values? No, but close:



Here is a plot of the difference between the predicted values:



8. How to fit non-normal random effects distributions

Well known: If a variable is skewed right then taking the log of that variable often makes its distribution closer to normal. (Distribution is approximately lognormal).

Used in reverse: If a variable is normal then exponentiating it creates a skewed-right, lognormal variable.

Schematically:

Normal = ln (Skewed)

Exp(Normal) = Skewed

Basic idea: NLMIXED can only accommodate normally distributed random effects, but arbitrary nonlinear models. So instead of directly including the random effect, include exp(random effect).

By comparing inclusion of (random effect) versus exp(random effect) we can see if a lognormally distributed variate fits better.

Illustration (repeated measures on patients with two predictors, x_1 and x_2)

```
proc nlmixed data=work.temp;
  parms b0=1 b1=0 b2=0 ls2u=-1 ls2e=-1;
  nu=b0+b1*x1+b2*x2;
  mu=nu+pat;
  model y ~ normal(mu, exp(ls2e));
  random pat ~ normal(0, exp(ls2u)) subject=patntID;
run;
```

VERSUS

```
proc nlmixed data=work.temp;
  parms b0=1 b1=0 b2=0 ls2u=-1 ls2e=-1
    mupat=0;
  nu=b0+b1*x1+b2*x2;
  mu=nu+exp(mupat+pat);
  model y ~ normal(mu, exp(ls2e));
  random pat ~ normal(0, exp(ls2u)) subject=patntID;
run;
```

How to compare models?

Models are fit by maximizing the log of the likelihood. But this doesn't account for the fact that the lognormal model has an additional parameter and would be expected to fit better just because of that extra flexibility.

Also, not the usual (nested) model comparison.

A typical method of choosing between non-nested models is the Akaike Information Criterion (AIC).

$$\begin{aligned} \text{AIC} &= -2 \times \log \text{lik} + 2 \times (\text{no. of parameters}) \\ &= -2 \times \text{model fit} + \text{complexity penalty} \end{aligned}$$

So smaller is better.

Illustration: The normal model has 5 parameters (intercept, two betas, two variances) for a penalty of $2 \times 5 = 10$.

Lognormal model additionally has a mean parameter for 6 parameters and a penalty of 12.

Log likelihood of normal = -41.6
Log likelihood of lognormal = -37.4

AIC for normal = $-2 \times -41.6 + 10 = 83.8 + 10 = 93.8$
AIC for lognormal = $-2 \times -37.4 + 12 = 74.8 + 12 = 86.8$

So lognormal is preferred.

Johnson family of distributions

But what if the data are armed with a pointed stick? That is, what if the random effects are not normally distributed but also not lognormally distributed?

This trick can be extended using the "Johnson" family of transformations (NL Johnson, "Systems of frequency curves generated by methods of translation" *Biometrika*, 36: 149-176, 1949).

Johnson family

By considering three different transformations, distributions with any skewness or kurtosis can be modeled. Let u represent a random effect.

Normal (no transformation)

u

Transformation 1 (lognormal)

$\exp(\mu+u)$

Transformation 2

$\phi\sinh(\mu+u)$

Transformation 3

$\phi\exp(\mu + u)/[1+\exp(\mu + u)]$

Here is code for the sinh(.) one:

```
proc nlmixed data=work.tmp maxiter=500 maxfunc=10000;
   parms b0=1.1 bi1=0 bi2=0 muu=-2 phi=1 lsig2u=-1
         lsig2e=-1;
   nu=b0+bi1*infect1+bi2*infect2;
   title "Avidity sinh random effect, time 1";
   mu=nu+phi*sinh(muu+u);
   model type14 ~ normal(mu, exp(lsig2e));
   random u ~ normal(0, exp(lsig2u)) subject=matchid;
run;
```

Note: Only one will fit the data best and rest will have trouble converging.

9. Summary

- Predicted values of random effects show modest sensitivity to the assumed distributional shape.
- Distribution shape of BLUPs often not reflective of true random effects distribution.
- The ranking of predicted values is little affected.
- Fitting flexible distributional shapes is an easy way to check sensitivity of the results to the assumed shape.

I can be contacted at: chuck@biostat.ucsf.edu

and Estie has put the slides on the seminar web site if you wish to download them.