Further, I make confidence limits for \( \beta_j \) as
\[
\text{lower limit}: \hat{\beta}_j - t_{\alpha/2, n-k} \sqrt{\text{MSE} \left( G(D' D)^{-1} \right)} \\
\text{upper limit}: \hat{\beta}_j + t_{\alpha, n-k} \sqrt{\text{MSE} \left( G(D' D)^{-1} \right)}
\]

Inference for a simple mean response
Suppose I wish to estimate \( f(\xi, \beta) \). Fact 4 suggests
\[
f(\xi, \kappa_0) \sim N(f(\xi, \beta), \sigma^2 G(D' D)^{-1} G')
\]
and a \( t_{n-k} \) reference distribution

\[
\hat{G} = \left( \begin{array}{c} \frac{\partial f(\xi, \beta)}{\partial \beta_j} \\ \vdots \\ \frac{\partial f(\xi, \beta)}{\partial \beta_j} \end{array} \right)_{1 \times k}
\]

and approximate confidence limits for \( f(\xi, \beta) \) become
\[
f(\xi, \kappa_0) \pm t_{\alpha/2, n-k} \sqrt{\text{MSE} \left( G(D' D)^{-1} \right)}
\]

where
\[
G = \left( \begin{array}{c} \frac{\partial f(\xi, \beta)}{\partial \beta_j} \\ \vdots \\ \frac{\partial f(\xi, \beta)}{\partial \beta_j} \end{array} \right)_{1 \times k}
\]

Reasoning exactly as for inference for a simple \( \beta_j \), I can test \( H_0: f(\xi, \beta) = \beta \) using
\[
T = \frac{f(\xi, \kappa_0) - \beta}{\sqrt{\text{MSE} \left( G(D' D)^{-1} \right)}}
\]

Prediction
Suppose that in the future I will observe \( y^* \) that is normal with mean \( h(\beta) \) and variance \( \sigma^2 \) independent of \( y_1, y_2, \ldots, y_k \). For example, \( y^* \) might be a new observation at \( \xi \) in this case
\[
h(\beta) = f(\xi, \beta) \quad \text{and} \quad n = 1
\]
or \( y^* \) might be a difference in \( z \) new observations at \( \xi^1, \xi^2 \) in this case
\[
h(\beta) = f(\xi^1, \beta) - f(\xi^2, \beta) \quad \text{and} \quad n = 2
The standard reasoning leads to prediction limits for $y^*$

$$h(b_{ols}) = t \sqrt{\text{MSE}} \sqrt{\hat{y} + \hat{\Sigma} (\hat{\beta}' \hat{\Sigma})^{-1} \hat{\beta}}$$

Methods Based on the Large n Behavior of the Likelihood Function

The MNN likelihood function for the nonlinear model is

$$L(\beta, \sigma^2 | Y) = (2\pi)^{-r} \frac{1}{J^n} \exp \left( -\frac{1}{2\sigma^2} \sum (y_i - f(x_i, \beta))^2 \right)$$

$$l(\theta) = l(\theta_1, \theta_2)$$

Suppose that for every $\theta_1$, $\hat{\theta}_2(\theta_1)$ maximizes

$$l(\theta_1, \hat{\theta}_2)$$

For choices of $\theta_1$. As it turns out, a large sample confidence set for $\theta_1$ is

$$\{ \theta_1 | l(\theta_1, \hat{\theta}_2(\theta_1)) > (1-\alpha) \text{ level} \}$$

$$l(\hat{\theta}_{MLE}) - \frac{1}{2} \chi^2_p$$

upper $\alpha$ pt of $\chi^2_p$ and the "log-likelihood function" is

$$\lambda(\beta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2$$

$$- \frac{1}{2 \sigma^2} \sum (y_i - f(x_i, \beta))^2$$

Temporarily think of $\theta$ as a general parameter vector of dimension $r$ and a general log-likelihood (see Section 7.3.2 of class outline)

$$\theta = (\theta_1)_{p \times 1} \quad (\theta_2)_{(r-p) \times 1}$$

This is the set of all $\theta_1$'s for which there is a $\theta_2$ so that $l(\theta_1, \theta_2)$ is within $\frac{1}{2} \chi^2_p$ of the maximum of the log-likelihood.

BTW $l(\theta_1, \hat{\theta}_2(\theta_1)) = \lambda^*(\theta_1)$ is sometimes called the profile log-likelihood for $\theta_1$ and this prescription is then "the set of all $\theta_1$'s with profile log-likelihood within $\frac{1}{2} \chi^2_p$ of the maximum."
Profile log-likelihood for $\sigma^2$

$$
\frac{-n}{2} \log \sigma^2 - \frac{1}{2} \log \frac{SSE}{n} - \frac{1}{2} \frac{SSE}{\sigma^2}
$$

Application #1

For any given $r^2$, $x(\theta, \sigma^2)$ is maximized as function of $\sigma^2$ by $\bar{y}$, so

$$
\hat{\theta}_1 = \bar{y} - \frac{1}{2} \log \frac{SSE}{n} - \frac{1}{2} \frac{SSE}{\sigma^2}
$$

Now make applications of this to the

nonlinear model

\[ l(\theta_1, \theta_2) \]

Contours:

\[ \frac{-n}{2} \log \sigma^2 - \frac{n}{2} \log \frac{SSE}{n} - \frac{1}{2} \frac{SSE}{\sigma^2} = C \]

contour plot for $l(\theta_1, \theta_2)$

\[ \hat{\theta}_2 = \bar{y} \]

maximum $l(\theta_2) = l(\hat{\theta}_2)$
This is an alternative to approximating

\[ \frac{SSE}{\hat{\sigma}^2} \sim \chi^2_{n-k} \]

as in the LM case and using

\[
\left( \begin{array}{c}
\frac{SSE}{\text{upper } \chi^2} \\
\frac{SSE}{\text{lower } \chi^2}
\end{array} \right)
\]

Application #2 \[ \theta_1 = \beta \]

For any \( \beta \), \( \lambda(\beta, \hat{\sigma}^2) \) is maximized as a function of \( \hat{\sigma}^2 \) by

\[ \hat{\sigma}^2(\beta) = \frac{1}{n} \sum (y_i - f(\Xi_i, \beta))^2 \]

So we have

\[ \lambda(\theta_1, \hat{\sigma}^2(\theta_1)) = \lambda(\beta, \hat{\sigma}^2(\beta)) \]

and

\[ \lambda(\theta_{MLE}) = \lambda(\hat{\theta}_1, \hat{\sigma}^2(\theta_1)) \]

= \{ \beta | \Sigma (y_i - f(\Xi_i, \beta))^2 < \text{SSE} e^{\frac{1}{2} \chi^2_k} \}

So an approximate confidence region for \( \beta \) is

\[ \{ \beta | \log \Sigma (y_i - f(\Xi_i, \beta))^2 - \log \text{SSE} < \frac{1}{n} \chi^2_{k+\alpha} \} \]

\( k = 2 \) example

\[ \beta_2 \]

contour plot of \[ n \hat{\sigma}^2(\beta) = \Sigma (y_i - f(\Xi_i, \beta))^2 \]

contour plot of \( \lambda_{MLE} \) in more detail.
As it turns out in the linear model, an exact confidence set is
\[ \{ \beta \in \mathbb{R}^k \mid \Sigma (y_i - f(\xi_i, \beta))^2 < \text{SSE} \left( 1 + \frac{k}{n-k} F_{k, n-k} \right) \} \]
upper \( \alpha \) pt for \((1-\alpha)\) level confidence

The "Beale" region — this is usually just carried over to the non-linear model

\[ (1 + \frac{1}{n-k} F_{k, n-k}) = (1 + \frac{\hat{e}_0^2}{n-k}) \]
upper \( \alpha \) pt for \((1-\alpha)\) level confidence

Application 43: \( \beta_1 = \beta_j \)
(\text{confidence regions for single entries of } \beta) — the same reasoning as before leads to
\[ \{ \beta_j \mid \min_{\beta_j} \Sigma (y_i - f(\xi_i, \beta_j))^2 < \text{SSE} \} \]
Appealing to exact theory for the LM case it's common to replace \( e_t^2 \) with

Cartoon \( k=2 \) (contour plot \( \Sigma (y_i - f(\xi_i, \beta))^2 \))

\begin{align*}
\text{CI for } \beta_1 & \quad \text{CI for } \beta_2 \\
\text{CI for } \beta_1 & \quad \text{CI for } \beta_2 \\
\text{contour when } \Sigma (y_i - f(\xi_i, \beta))^2 = \text{SSE} (1 + \frac{\hat{e}_0^2}{n-2}) & \quad \text{contour when } \Sigma (y_i - f(\xi_i, \beta))^2 = \text{SSE} (1 + \frac{\hat{e}_0^2}{n-2})
\end{align*}
BTW... folklore is that these regions do better (in terms of holding nominal coverage levels) than inferences made on basis of maximum likelihood theory for (BOls, $\frac{SSR}{n}$) under

Second Generalization of the LM... Mixed Linear Models. These are models that can be written in the form

(usually covariance matrices for $u, \varepsilon$ are assumed to depend upon a few parameters... usually thought of as "variance components")

It is standard to assume that $u$ and $\varepsilon$ are uncorrelated, so if I call

\[ Var(\varepsilon) = R \]
\[ Var(u) = G \]

\[ E\left(\begin{pmatrix} u \\ \varepsilon \end{pmatrix}\right) = 0 \]
\[ Var\left(\begin{pmatrix} u \\ \varepsilon \end{pmatrix}\right) = \left(\begin{array}{cc} G & 0 \\ 0 & R \end{array}\right) \]

\[ Var\left(\begin{pmatrix} Z'GZ + \varepsilon \\ \varepsilon \end{pmatrix}\right) \]

\[ EY = E(X\beta + Zu + \varepsilon) = X\beta \]
\[ VarY = Var(Zu + \varepsilon) \]
\[ = Var(zu) + Var(\varepsilon) \]
\[ = ZGZ' + R \]

Example A "One way random effects model"
A "batch" process makes "widgets" -

Sample 2 widgets from each of 3 batches and measure hardness:

\[ y_{ij} = \text{measured hardness of } j \text{th widget from batch } i \]

I might model as a within-batch random effect:

\[ y_{ij} = M + \alpha_i + \epsilon_{ij} \]

Some "process average hardness" random effect for batch i

and with this notation I can write

\[
\begin{pmatrix}
  y_{11} \\
  y_{12} \\
  y_{21} \\
  y_{22} \\
  y_{31} \\
  y_{32}
\end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} M \end{pmatrix} + \begin{pmatrix}
  100 \\
  100 \\
  010 \\
  010 \\
  001 \\
  001
\end{pmatrix} \begin{pmatrix}
  \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\
  \epsilon_{12} \\
  \epsilon_{21} \\
  \epsilon_{22} \\
  \epsilon_{31} \\
  \epsilon_{32}
\end{pmatrix}
\]

\[ E(Y) = \begin{pmatrix} M \end{pmatrix} \text{ and } \ Var(Y) = G \ Opportune \ + \ R \]

\[ G = \begin{pmatrix}
  100 \\
  100 \\
  010 \\
  010 \\
  001 \\
  001
\end{pmatrix} \begin{pmatrix}
  100 & 100 & 010 & 010 & 001 & 001
\end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix}
  100 & 100 & 010 & 010 & 001 & 001
\end{pmatrix} \]

\[ \text{Var} \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & 0 \\
  0 & \sigma^2 & 0 \\
  0 & 0 & \sigma^2
\end{pmatrix} \]

\[ E(\varepsilon) = 0 \text{ and } \ Var(\varepsilon) = \sigma^2 I \]

\[ E(\varepsilon) = 0 \text{ and } \ Var(\varepsilon) = \sigma^2 I \]

\[ \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} \begin{pmatrix} \sigma^2 \end{pmatrix} = G \]

\[ \text{Var} \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} = \begin{pmatrix} \sigma^2 \end{pmatrix} \]

\[ \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} \begin{pmatrix} \sigma^2 \end{pmatrix} = G \]

\[ \text{Var} \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} = \begin{pmatrix} \sigma^2 \end{pmatrix} \]

\[ \text{Var} \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} = \begin{pmatrix} \sigma^2 \end{pmatrix} \]

\[ \text{Var} \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} = \begin{pmatrix} \sigma^2 \end{pmatrix} \]

\[ \text{Var} \begin{pmatrix} \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{pmatrix} = \begin{pmatrix} \sigma^2 \end{pmatrix} \]
\[
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
+ \gamma^2 I
\]

\[
= \begin{pmatrix}
\gamma \alpha^2 & \gamma \alpha^2 & \gamma \alpha^2 & \gamma \alpha^2 \\
\gamma \alpha^2 & \gamma \alpha^2 & \gamma \alpha^2 & \gamma \alpha^2 \\
0 & \text{same} & 0 & \text{same} \\
0 & \text{same} & 0 & \text{same}
\end{pmatrix}
\]

Kronecker product: For \( A = (a_{ij}) \) rxc and \( B = (b_{ij}) \) sx d, \( A \otimes B \) is an rxc matrix of sx d matrices where in the \( i \)th row and \( j \)th column is \( a_{ij} B \).

For the \( k \)th analysis of specimen \( i \) by chemist \( j \),

\[
y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}
\]

Example B: A "Two Way Mixed Effects model Without Interaction"

\( 2 \) analytical chemists each make \( 2 \) analyses on same two specimens (each is split into \( 4 \) parts)
How to write in "standard form"?

With

\[
E(x_1) = \bar{x}, \quad \text{Var}(x_2) = \sigma_x^2 I,
\]
\[
E(\varepsilon) = 0, \quad \text{Var}(\varepsilon) = \sigma^2 I.
\]

I might write this model (assuming \(x_1\)'s and \(\varepsilon\)'s are not correlated) as

\[
\begin{pmatrix}
Y \\
\end{pmatrix} = \begin{pmatrix}
X \\
\end{pmatrix} \beta + \begin{pmatrix}
\varepsilon \\
\end{pmatrix}
\]

\[
\text{Var}(Y) = \sigma_x^2 \beta \beta' + \sigma^2 I
\]

\[
= \sigma_x^2 \begin{pmatrix} I & J \\ J' & I \end{pmatrix} + \sigma^2 I
\]

Again, this is not the usual LM (not even an Aitken version).

Estimation in the Mixed Linear Model?
Consider normal maximum likelihood...

With

\[
V = ZGZ' + R
\]

\[
\text{usally a function of variance components}
\]

\[
\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2
\]

Write \(V(\sigma^2)\) and note that the normal likelihood function is...