Finish lik. inf. (testing)

\[ f(y; \varphi) = \exp \left\{ \varphi' t(y) - c(\varphi) - d(y) \right\} \]

\[ f(t; \varphi) = \int f(y; \varphi) dy \quad y = (y_1, \ldots, y_n) \]
\[ \{y; t(y) = t\} \]
\[ = \exp \left\{ \varphi' t - nc(\varphi) \right\} h(t) \]

\[ f(t; t_0; \varphi_j) = \exp \left\{ \varphi_j t - c_{t_0}(\varphi_j) \right\} h_{t_0}(t) \]

1 par. exp'ld family

true for any \( \varphi \) that's linear in \( y \)

\[ f(t; t_2; \varphi) = \exp \left\{ \varphi t - c_{t_2}(\varphi) \right\} h_{t_2}(t) \]

\[ t(y) = \left( \sum_{i=1}^n t_i(y_i), \ldots, \sum t_p(y) \right) \]
In our asymptotic, we used

\[ \sup_{\mathcal{F}_2} l (\varphi_1, \varphi_2) = l_p (\varphi_1) \]

found \( \hat{\varphi}_1 - \varphi_1 \sim N(0, \ldots) \)

We can instead use

\[ l_c (\varphi_1) = t_1 \varphi_1 - c_{t_2} (\varphi_1) \]

and get \( \mathcal{W}_c, \mathcal{W}_{c,e}, \mathcal{W}_c, \mathcal{W}_c, \ldots \)

same asymptotic applies

\[ \mathbb{E} l_c (\varphi_1) = 0 \quad \text{but} \quad \mathbb{E} l_p (\varphi_1) \neq 0 \]

\( O(n^{-1}) \)

genuine log-lik.

Example: logistic regression \( Y_i \sim \text{Ber} (p_i) \)

\[ \log \left( \frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 x_i \]

\[ L(\beta_0, \beta_1) = \exp \left( \beta_0 \sum y_i + \beta_1 \sum x_i y_i - \ln \left( 1 + e^{\beta_0 + \beta_1 x_i} \right) \right) \]
\( \psi = \beta_1 \) inference

\[
f(s_1 = \sum x_i y_i ; s_0 = \sum y_i ; \beta_1) \quad s_0 \# \text{success, } s_1 = \chi_1
\]

\[
f(s_1, \beta_1) = c(s_0, s_1) \frac{\beta s_1}{\sum \frac{c(s_0, u) e^{\beta_i u}}{u}}
\]

\[c(s_0, s_1) \# \text{subsets of } \{x_1, \ldots, x_n\} \text{ of size } s_0\]

\[\text{that add to } s_1\]

\[\chi_i = \begin{cases} 1 & i = 1, \ldots, n_1 \\ 0 & i \geq n_1 + 1, \ldots, n_1 + n_2 \end{cases} \text{ that control}\]

\[\sum \chi_i y_i = \# \text{successes in that group}\]

\[
\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_i
\]

\[
= \beta_0 + \beta_1 \chi_i = 0, \quad \chi_i = 0
\]

\[
= \beta_0 + \beta_1 \quad \chi_i = 1
\]
\[ f(s_i | s_0; \beta) = \frac{(n_1)(n_2)}{(s_i)(s_0-s_i)} e^{\beta s_i} \sum_u (\text{disjoint}) e^{\beta u} \]

\[ c(s_0, s_i) = \# \text{ of ways of choosing a set of size } s_0 \text{ with } s_i, s_0-s_i, 0's \]

\[ f(s_i | s_0; \beta = 0) = \frac{(n_1)(n_2)}{(s_i)(s_0-s_i)} / (s_i) \]

basis of Fisher's exact test of independence in 2x2 table

\[
\begin{array}{c|ccc}
& 1 & 0 & \text{df} \\
\hline
1 & n_1 - \beta & n_2 - \beta & n_1 + n_2 \\
0 & s_0 - s_1 & n_2 - s_0 + s_1 & n_0 \\
\hline
s_0 & n - s_0 & n - s_0 & n
\end{array}
\]

"exact test"

\[ \chi^2 \]

obs'd table \[ \sum_{\text{cells}} \frac{(O-E)^2}{E} \]

p-value \[ P_n(S_i > s_i | s_0; \beta = 0) \]

profile test
\[
\mu_1 - \mu_2 = 0 \quad \text{in} \quad \text{control group}
\]

\[
Y_{11}, \ldots, Y_{1n}, \quad Y_{21}, \ldots, Y_{2n}
\]

\[
t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p / \sqrt{n}}
\]

\[
P_r (T \geq t_{obs} ; \mu_1 = \mu_2) = 1 - pt (n-1, t_{obs})
\]

\[
\begin{align*}
Y_i & = 1 & \chi_i & = 1 & (i = 1, \ldots, n) \\
0 & & 0 & &
\end{align*}
\]

\[
\text{Matched pairs} \quad (Y_{j1}, \ Y_{j2}) \text{ ind't } \text{Ber} \quad j = 1, \ldots, m
\]

\[
P_r (Y_{j1} = 1) = \frac{e^{\lambda_j}}{1 + e^{\lambda_j}} \quad P_r (Y_{j2} = 1) = \frac{e^{4 + \lambda_j}}{1 + e^{4 + \lambda_j}}
\]

\[
\text{logit } (P_{j1} = 1) = \chi_j \quad \text{logit } (P_{j2} = 1) = 4 + \lambda_j
\]

\[
\begin{align*}
\text{"} \chi_j = 0 \text{"} & \quad \text{"} \lambda_j = 1 \text{"} \\
\text{cases} & \quad \text{controls}
\end{align*}
\]

\[
l(\Psi, \chi) = \psi \sum_{j=1}^{m} Y_{j2} + \sum_{j=1}^{m} \lambda_j (Y_{j1} + Y_{j2}) - c (\chi, \Psi)
\]
\[ \begin{pmatrix} \chi_{i,j} \end{pmatrix} \begin{cases} -\infty & 0,0 \\ \frac{1}{2} \psi & y_{i,j}, y'_{i,j} = 0,1 \\ +\infty & 1,1 \end{cases} \]

\[ l_P(\psi) = l(\psi; \hat{\psi}) \quad l_P(\hat{\psi}) = 0 \]

\[ \sum_{j=1}^{m} y_{j,2} - n_0 + \frac{r_1}{1 + e^{\psi/2}} = 0 \]

\[ n_0 = \#(0,0) \quad n_1 = \#(0,1) \lor (1,0) \]

\[ 1 + e^{\psi/2} = \frac{r_1/m}{(\sum y_{j,2} - n_0)/m} \quad \xrightarrow{p} 1 + e^\psi \]

\[ \hat{\psi} \xrightarrow{p} 2\psi \quad m \to \infty \quad \lambda_1, \ldots, \lambda_m \]

\[ f \left( \sum_{j=1}^{m} y_{j,2} \mid y_{i,j} + y'_{i,j} ; j = 1, \ldots, m \right) \]

\[ n_{01} \sim \text{Bin} \left( r_{01} ; \frac{e^\psi}{1 + e^\psi} \right) \quad \text{unqwe distr} \]

\[ n_{01} = \sum_{j=1}^{m} y_{j,2} \quad n_1 = \#(1,0) \lor (0,1) \]

\[ S_{12} \sim \text{fix} \quad \text{margins in table.} \]
\[ f(t; \nu) = f_c(t_1 \mid t_2; \eta_1) f_m(t_2; \eta) \]

Ignore info on \( \eta_1 \), here

\( t \) is discrete

Binomial

Bernoulli

\( f(t_1 \mid t_2) \) "more discrete" fewer pt. of the prob.

\[
\begin{array}{ccc}
3 & 1 & \\
2 & 4 & \\
\end{array}
\]

\[ P_n(T > t_1) = .027 \quad P_n(T > t_1) = .001 \]

\[ P_n(T > t_1, -1) = .053 \]

- Rasch model in psychometrics
- Choice based sampling in econometrics
- conditional in exponential families (after marginal)
  \[ y \rightarrow t \rightarrow t_i \]
  \[ \int \quad \text{cond'g} \]

- marginal inference in transform families

Example: Log-normal, \( y \sim \text{log-normal} \( y; \mu, \sigma \)

\[ f(\mu, \sigma | \alpha; \mu, \sigma) \]

\[ f(\hat{\mu}, \hat{\sigma} | \alpha; \mu, \sigma) = \frac{L(\mu, \sigma; \hat{\mu}, \hat{\sigma}, \alpha)}{\int L(\mu, \sigma) \, d\mu \, d\sigma} \]

\[ a_i = (y_i - \mu) / \hat{\sigma} \]

\[ y \rightarrow \hat{\mu}, \hat{\sigma} \quad \text{by cond'g inference} \]

\[ \mu, \sigma \quad \text{matches in dimension} \]

\[ \Psi = \mu \quad \text{let} \]

\[ t_1 = (\hat{\mu} - \mu) / \hat{\sigma} \]

\[ t_2 = \hat{\sigma} / \sigma \quad 1-1 \text{ transf.} \]

\[ \mu \in \mathbb{R}, \sigma \in \mathbb{R^+} \]
\[ f(t_1, t_2 | \alpha, \mu, \sigma) \leq f(\hat{\mu}, \hat{\sigma} | \alpha, \mu, \sigma) \]

\( t_1 \) is a unique pivotal quantity for inference about \( \mu \) (\( \sigma \))

\[ f_m(t_1) = f_m(t_1 | \mu, y_1) \quad \text{for fixed } \mu = \mu_0 \]

\( f_m(t_1) \) is the only marginal density free of \( \sigma \), based on \( y_1 \) a

any quantity dep. on \( y_1 | \alpha, \mu \) is dist-free of \( \sigma \), is a \( F \) of \( t_1 \).

"\( t_1 \) is a maximal invariant under the
locin scale group \( \text{dist-free of } \sigma \)"

\[ \frac{f(y_1 | \mu, \sigma)}{f(\hat{\mu}, \hat{\sigma} | \alpha, \mu, \sigma)} \]

\( TSH \) - needs more abstract argument than
If \( y_i = \mathbf{x}_i^T \beta + \sigma e_i \)

\[
\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}} = t_j, \quad t_{p+m} = \frac{\hat{\sigma}}{\sigma}
\]

\( f_m(t_j(\beta_j)) \) free of \( \beta(-j), \sigma \)

\( f_n(\hat{\sigma}) \) free of \( \beta \) essentially

\[
\frac{L(\beta, \sigma; \hat{\beta}, \hat{\sigma}, \alpha)}{\int L(\beta, \sigma) d\beta d\sigma} = f(\hat{\beta}, \hat{\sigma} \mid \alpha)
\]

\( f_m(\beta_j) \) require \( \int_{\mathbb{R}^{p-1} \times \mathbb{R}^+} d\beta_j \sigma \hat{\sigma} \)

Example \( N(\mu, \sigma^2) \)

\[\hat{\mu} = \bar{y}, \quad \hat{\sigma}^2 = \frac{\sum (y_i - \bar{y})^2}{n} \]

\[\uparrow = \frac{s_y^2}{n}\]

m.l.e. from \( L(\mu, \sigma) \)
\[ f(\bar{y}, s_y^2 | \mu, \sigma) = f(\bar{y}, s_y^2 \mid \mu, \sigma) = N(\mu, \frac{s_y^2}{n}) \times \sigma^2 \chi^2_{n-1} \]

\[ (\bar{y}, s_y^2) \rightarrow (t_1, t_2) \quad t_1 = \frac{\bar{y} - \mu}{s_y} \quad t_2 = \frac{s_y}{\sigma^2} \]

\[ f_m(t_1) \times \chi^2_{n-1} \text{ for } \frac{s_y^2}{\sigma^2} \text{ and } t_1 \text{ (N)} \]

\[ P_n(t_1 > t_{obs} \mid \mu_0) = P_n(\chi^2_{n-1} > \chi^2_{obs}) \]

- can be computed from \( \chi^2_{n-1} \)

\[ f_m(s_y^2) = \frac{(s_y^2)^{n/2 - 1}}{\Gamma(n/2)} e^{-s_y^2/2\sigma^2} \left( \frac{1}{2^{n/2} \Gamma(n/2)} \right) \]

\[ l_m(\sigma^2) = -\frac{n-1}{2} \log \sigma^2 - \frac{s_y^2}{2\sigma^2} \]

\[ l_m(\sigma^2) = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{s^2}{n-1} \]
"rule" from marginal is better

\[ y \rightarrow \mu \mid \sigma^2 \sim \sigma^2 \mid \sigma \]

\[ \text{Cond} = \text{marginal} \]

Marginal log-likelihood for \( \sigma^2 \) gives better inference than the full log-likelihood.

Extended to mixed & random effects.

\( \text{(4)} \) "REML" = marginal for variance estimates.

2 model classes

1. marg. \( \rightarrow \) cond \( \rightarrow \) examples...

1. cond \( \rightarrow \) marg.

Both difficult than profiling.
You can show that to some order of approx, the marginal log-lik & conditional log-lik are approximated by

\[ \ell_{\text{approx}}(\psi) = \ell_{p}(\psi) - \frac{1}{2} \log | \hat{j}^{\psi}(4, \hat{\lambda}) | \]

\[ = \ell(4, \hat{\psi}) - \frac{1}{2} \log | \hat{j}^{\lambda}(\psi, \hat{\psi}) | \]

\[ \hat{j}(\theta) = - \frac{1}{\theta} \frac{\partial^{2} \ell}{\partial \theta \partial \theta^{\top}} = \left[ \begin{array}{c} \hat{j}^{\psi} \\ \hat{j}^{\lambda} \end{array} \right] \]

\[ \text{Example: } N(\mu, \sigma^{2}) - \frac{n}{2} \log \sigma^{2} + \frac{1}{2} \log \sigma^{2} \]

\[ (\psi, \lambda) \]

\[ \text{Inference on } \psi, \text{ eliminate } \lambda \]

- exact calc = 5

- approx calc = 5 (hint)
Summary of Likelihood

\[ \theta = (\psi, \lambda) \]

\[ y = (y_1, \ldots, y_n) \]

\[ f(y; \psi, \lambda) \]

\[ \psi \rightarrow (\psi, \phi(\lambda, \psi)) \]  

\[ \text{“interest-respecting”} \]

\[ \frac{dw(\psi)}{\text{inv}} = 2 \{ l_\theta(\hat{\psi}) - l_\theta(\psi) \} \]

\[ \text{und.} \]

\[ \mathcal{W}(\psi) = \left( \hat{\psi} - \psi \right)^T j_\theta^{-1}(\hat{\psi}) \left( \hat{\psi} - \psi \right) \sim \chi^2 \]

\[ \text{not inv.} \]

\[ \mathcal{W}_n(\psi) = \left[ l_\theta'(\psi) \right]^T j_\theta^{-1}(\hat{\psi}) [l_\theta'(\psi)] \]

\[ 2ll(\psi, \hat{\psi}) = 0 \]

\[ \begin{align*}
\text{if } \psi \text{ scalar} & \quad \chi^2 \\
\tau &= \text{sign}(\hat{\psi} - \psi) \sqrt{\mathcal{W}(\psi)} \sim \mathcal{N}(0,1) \\
\tau_e &= (\hat{\psi} - \psi)^\frac{1}{2} j_\theta(\hat{\psi}) \sim \mathcal{N}(0,1) \\
\tau_u &= l_\theta'(\psi)^\frac{1}{2} j_\theta^{-1}(\hat{\psi}) \sim \mathcal{N}(0,1)
\end{align*} \]

plot \[ \tau (\psi; \psi^0) \text{ vs } \psi \]

plot \[ \mathcal{W}(\psi) \text{ vs } \psi \]

\[ l_\theta(\psi) \]

\[ p(\psi) \]
\[ W_e = (\hat{\psi} - \psi)^T \tilde{j}_p(\hat{\psi}) (\hat{\psi} - \psi) \]

\[ W_u = \ell_p(\hat{\psi}) \tilde{j}_p(\hat{\psi}) \ell_p(\hat{\psi}) \]

\[ j_p(\psi) = (\tilde{j}_{\psi \psi} - \tilde{j}_{\psi})^{-1} \]

\[ \psi = 0 \text{ of interest} \]

\[ \hat{\lambda}_0, -\frac{\partial^2 l}{\partial \theta^2} \text{ at } (0, \hat{\lambda}_0) \]

Score test, using 'null' pt. for information.

Require just one model fit.

(last resort)
Example \( Y \sim \text{Mult} (n; \pi) \)

\[ \pi = (\pi_1, \ldots, \pi_m) \quad Y = (Y_1, \ldots, Y_m) \quad \sum Y_i = n \quad \sum \pi_i = 1 \]

\[ f(y; \pi) = \frac{n!}{y_1! \cdots y_m!} \pi_1^{y_1} \cdots \pi_m^{y_m} \]

\[ \pi_m = 1 - \pi_1 - \cdots - \pi_{m-1} \]
\[ y_m = n - y_1 - \cdots - y_{m-1} \]

\( W_1, W_2, W_m \) of \( \pi \)

\[ \begin{aligned} \frac{d \ell (\pi; y)}{d \pi_i} \bigg|_{\pi} &= 0 \quad \text{set} \quad m-1 \text{ eq's in } m-1 \text{ unk}. \\ \\
- \frac{d^2 \ell}{d \pi_i d \pi_k} \bigg|_{\pi} &= j(\hat{\pi}) \quad (m-1) \times (m-1) \end{aligned} \]

\[ \hat{\pi}_j = \frac{y_j}{n} \quad j(\pi) = i(\pi) \quad \text{bec exp'ld} \quad \text{freq} \]

SM \S 4.5, CH Ex 2.17
\[ i_{jk}(\pi) = \mathbb{1}_{\{j=k\}} \cdot \frac{1}{\pi_j} - \frac{1}{\pi_m} \]

\[ n \left[ \begin{array}{ccc}
\frac{1}{\pi_1} - \frac{1}{\pi_m} & \cdots & \frac{1}{\pi_m} \\
\frac{1}{\pi_m} & \cdots & \frac{1}{\pi_n-1} - \frac{1}{\pi_n} \\
\end{array} \right] \]

\[ g_j[in(I)]_{jk} = \frac{1}{n} \left\{ \begin{array}{ll}
\pi_j (1 - \pi_j) & \hat{y}_j = k \\
-\pi_j \pi_{k} & \hat{y} \neq k \\
\end{array} \right. \]

\[ \hat{\pi}_j = \frac{y_j}{n} \sim \text{Beta} \left( \frac{1}{n} \sum \text{Bin}(n, \pi_j) \right) \]

\[ \omega(\pi) = 2 \left\{ \ell(\hat{\pi}) - \ell(\pi) \right\} \]

\[ = 2 \sum_{j=1}^{m} (y_j \log \hat{\pi}_j - y_j \log \pi_j) \]

\[ = 2 \sum_{j=1}^{m} y_j \log \left( \frac{y_j}{n \pi_j} \right) \leq \log \left( \frac{O}{E} \right) \]

\[ \omega_e(\pi) = (\hat{\pi} - \pi)^T g (\hat{\pi} - \pi) \]

\[ = \sum_{j=1}^{m} (y_j - n \pi_j)^2 \]

\[ \frac{O}{E} = \frac{O - \epsilon}{\epsilon} + 1 \]
\[ \frac{\sum_{j=1}^{n} \frac{y_j}{y_j}}{\sum_{j=1}^{n} \frac{y_j}{y_j}} \leq \frac{\sum_{j=1}^{n} (y_j - n \pi_j)^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2 / \epsilon} \]

\[ W_u(\pi) = \left( U(\pi) = \frac{dP}{d\pi} \right)^T \pi^{-1}(\pi) U(\pi) \]

\[ = \cdots \sum_{j=i}^{m} \frac{(y_j - n \pi_j)^2}{n \pi_j} \sum_{j=1}^{m} (y_j - \bar{y})^2 / \epsilon \]

Typically \( \pi = \pi(\beta) \) \[ \text{dim} \beta = q < m-1 \]

e.g. if we had \[ \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
\hline
\end{array} \]

\[ \pi_i^* = \pi_i \cdot \pi_j^* \]

\[ 1 + 5 = 2 \]

\[ W(\beta) = 2 \sum_{i=1}^{m} y_i \log \left( \frac{y_i}{n \pi_i(\beta)} \right) \]

\[ W_e(\beta) = \sum_{i=1}^{m} \frac{(y_i - n \pi_i(\beta))^2}{n \pi_i(\beta)} \]