Observed Confidence Levels: Theory and Practice

Northern Illinois Chapter
American Statistical Association
Fall Meeting
October 19, 2006

Dr. Alan M. Polansky
Division of Statistics
Northern Illinois University
De Kalb, IL 60115

Statistical Framework

- Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} F \in \mathfrak{F}$, a collection of d-dimensional random vectors.

- $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$

- $\theta = T(F)$ where $T : \mathfrak{F} \to \mathbb{R}^p$.

- Parameter Space: $\Theta = \{T(F) : F \in \mathfrak{F}\}$

- Let $\widehat{F}$ be an estimator of F based on $\mathbf{X}$ such that $\widehat{F} \in \mathfrak{F}$ with probability 1.

- Let $\widehat{\theta} = T(\widehat{F})$ be a plug-in estimator of θ.

The Problem of Regions

- Suppose there exist subsets $\{\Theta_i\}_{i=1}^{\infty} \subset \Theta$ such that

$$\Theta = \bigcup_{i=1}^{\infty} \Theta_i.$$

- The problem of regions consists of determining if $\theta \in \Theta_i$ for each i based on $\mathbf{X}$.

Example Applications

- Selecting the best treatment in a designed experiment.
- Determining the number of modes of an unknown density.
- Determining whether a mean profile in longitudinal data analysis is monotonically increasing, decreasing, or neither.
- Determining the ranking of the means of k populations.

Methods

- Multiple comparison techniques
- Best subset selection
- Posterior probabilities
- Confidence measures
  - Attained confidence levels (Efron and Tibshirani, 1998) based on one-sided p-values (paradoxes).
  - Observed confidence levels (Polansky, 2003) based on confidence regions.

Observed Confidence Levels

- Let $\mathbf{C}(\alpha, \omega; \mathbf{X}) \subset \Theta$ be a $100\alpha\%$ confidence region for θ based on $\mathbf{X}$.

- $\omega \in \Omega_\alpha \subset \mathbb{R}^q$ is a vector parameter that controls the shape of the confidence region for a given α.

- Suppose there exists sequences $\{\alpha_i\}_{i=1}^{\infty} \in [0, 1]$ with associated $\omega_i \in \Omega_{\alpha_i}$ such that

$$\mathbf{C}(\alpha_i, \omega_i; \mathbf{X}) = \Theta_i$$

  for $i = 1, 2, \ldots$.

- The confidence coefficient α_i is the observed confidence level of Θ_i for $i = 1, 2, \ldots$.

Example: Normal Mean

- Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$.

- Let $\Theta_i = [t_i, t_{i+1}]$ where $-\infty < t_i < t_{i+1} < \infty$.

- Confidence Region: For $\alpha = \omega_U - \omega_L$,

$$C(\alpha, \omega; X) = [\hat{\theta} - t_{n-1;1-\omega_L} n^{-1/2} \hat{\sigma}, \hat{\theta} - t_{n-1;1-\omega_U} n^{-1/2} \hat{\sigma}]$$

- Observed Confidence Level for $\Theta_i = [t_i, t_{i+1}]$:

$$\alpha(\Theta_i) = T_{n-1}\left[\frac{n^{1/2}(\hat{\theta} - t_i)}{\hat{\sigma}}\right] - T_{n-1}\left[\frac{n^{1/2}(\hat{\theta} - t_{i+1})}{\hat{\sigma}}\right]$$

Smooth Function Model

- Suppose $\mu = E\{\mathbf{X}_i\}$ with $\theta = g(\mu)$ for some smooth function $g : \mathbb{R}^d \to \mathbb{R}^p$.

- $\widehat{\theta} = g(\bar{\mathbf{X}})$.

- F satisfies Cramér's continuity condition, which states that ψ is the characteristic function of F, then

$$\limsup_{||t|| \to \infty} |\psi(\mathbf{t})| < 1.$$

- The asymptotic covariance matrix of $\widehat{\theta}$ is

$$\Sigma = \lim_{n \to \infty} V\{n^{1/2}\widehat{\theta}\} = h(\mu)$$

  for a smooth function $h : \mathbb{R}^d \to \mathbb{R}^p \times \mathbb{R}^p$.

Sampling Distributions

- Standardized Distribution:

$$G_n(\mathbf{t}) = P\{n^{1/2}\mathbf{\Sigma}^{-1/2}(\widehat{\theta} - \theta) \leq \mathbf{t} | \mathbf{X}_1, \ldots \mathbf{X}_n \overset{\text{iid}}{\sim} F\}$$

- Studentized Distribution:

$$H_n(\mathbf{t}) = P\{n^{1/2}\widehat{\mathbf{\Sigma}}^{-1/2}(\widehat{\theta} - \theta) \leq \mathbf{t} | \mathbf{X}_1, \ldots \mathbf{X}_n \overset{\text{iid}}{\sim} F\}$$

Confidence Regions

- Studentized Region

$$\mathbf{C}(\alpha,\omega;\mathbf{X}) = \{\widehat{\theta} - n^{-1/2}\widehat{\Sigma}^{1/2}\mathbf{t} : \mathbf{t} \in \mathcal{H}_\alpha\}$$

  where $\mathcal{H}_\alpha$ is any region such that

$$\int_{\mathcal{H}_\alpha} dH_n(\mathbf{t}) = \alpha.$$

- Hybrid Region: Uses the approximation $\mathcal{H}_\alpha \simeq \mathcal{G}_\alpha$,

$$\mathbf{C}(\alpha,\omega;\mathbf{X}) = \{\widehat{\theta} - n^{-1/2}\widehat{\Sigma}^{1/2}\mathbf{t} : \mathbf{t} \in \mathcal{G}_\alpha\}$$

  where $\mathcal{G}_\alpha$ is any region such that

$$\int_{\mathcal{G}_\alpha} dG_n(\mathbf{t}) = \alpha.$$

Confidence Regions

- Normal Approximation Uses the approximation $\mathcal{H}_\alpha \simeq \mathcal{N}_\alpha$,

$$\mathbf{C}(\alpha, \omega; \mathbf{X}) = \{\widehat{\theta} - n^{-1/2}\widehat{\Sigma}^{1/2}\mathbf{t} : \mathbf{t} \in \mathcal{N}_\alpha\}$$

where $\mathcal{N}_\alpha$ is any region such that

$$\int_{\mathcal{N}_\alpha} d\Phi(\mathbf{t}) = \alpha.$$

Observed Confidence Levels

- Let $\Psi \subset \Theta$ be an arbitrary region in the parameter space.

- Studentized

$$\alpha_{\text{stud}}(\Psi) = \int_{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)} dH_n(\mathbf{t})$$

where

$$n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta} - \Psi) = \{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta} - \psi) : \psi \in \Psi\}$$

Observed Confidence Levels

- Hybrid

$$\alpha_{\mathsf{hyb}}(\Psi) = \int_{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)} dG_n(\mathbf{t})$$

- Normal Approximation

$$\widehat{\alpha}_{\mathsf{stud}}(\Psi) = \int_{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)} d\Phi(\mathbf{t})$$

Bootstrap Approximation

- If G_n and H_n are unknown, and the normal approximation is not accurate enough, then G_n and H_n can be estimated using the bootstrap:

$$\widehat{G}_n(\mathbf{t}) = P^*\{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}^* - \widehat{\theta}) \leq \mathbf{t} | \mathbf{X}_1^*, \ldots \mathbf{X}_n^* \overset{\text{iid}}{\sim} \widehat{F}_n\}$$

$$\widehat{H}_n(\mathbf{t}) = P^*\{n^{1/2}\widehat{\Sigma}^{*-1/2}(\widehat{\theta}^* - \widehat{\theta}) \leq \mathbf{t} | \mathbf{X}_1^*, \ldots \mathbf{X}_n^* \overset{\text{iid}}{\sim} \widehat{F}_n\}$$

where

  - $P^*\{A\} = P\{A|\mathbf{X}\}$
  - $\widehat{F}_n$ is the empirical distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
  - $\widehat{\theta}^* = g(\bar{\mathbf{X}}^*)$, $\widehat{\Sigma}^* = h(\bar{\mathbf{X}}^*)$.
  - $\bar{\mathbf{X}}^* = n^{-1} \sum_{i=1}^n \mathbf{X}_i^*$

Bootstrap Estimates

- Studentized

$$\widehat{\alpha}^*_{\mathsf{stud}}(\Psi) = \int_{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)} d\widehat{H}_n(\mathbf{t})$$

- Hybrid

$$\widehat{\alpha}^*_{\mathsf{hyb}}(\Psi) = \int_{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)} d\widehat{G}_n(\mathbf{t})$$

Calculating the Bootstrap Estimates

- $\widehat{G}_n(\mathbf{t})$ and $\widehat{H}_n(\mathbf{t})$ rarely exist in a closed form.

- Simulation methods are used to approximate the bootstrap estimates:

  - Simulate b samples of size n from $\widehat{F}_n$ conditional on $\mathbf{X}$ (resampling).

  - For each sample, calculate $\mathbf{T}_i^* = n^{1/2}\widehat{\mathbf{\Sigma}}^{*-1/2}(\widehat{\theta}^* - \widehat{\theta})$.

  - The bootstrap estimate $\widehat{H}_n(\mathbf{t})$ can then be approximated by the empirical distribution of $\mathbf{T}_1^*, \ldots, \mathbf{T}_b^*$:

$$\widehat{H}_n(\mathbf{t}) \simeq \tilde{H}_n(\mathbf{t}) = b^{-1} \sum_{i=1}^{b} \delta(\mathbf{T}_i^* \leq \mathbf{t})$$

Calculating the Bootstrap Estimates

- The bootstrap estimate of α_{stud} can then be approximated by

$$
\begin{aligned}
\widehat{\alpha}^*_{\text{stud}}(\Psi) &\simeq \int_{n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)} d\tilde{H}_n(\mathbf{t}) \\
&= b^{-1}\sum_{i=1}^{b}\delta(\mathbf{T}_i^* \in n^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\Psi)) \\
&= b^{-1}\sum_{i=1}^{b}\delta(\widehat{\theta}-\widehat{\Sigma}^{1/2}\widehat{\Sigma}_i^{*-1/2}(\widehat{\theta}_i^*-\widehat{\theta}) \in \Psi)
\end{aligned}
$$

- Similar methods can be used to approximate $\widehat{\alpha}^*_{\text{hyb}}$.

Efron's Percentile Method

- An additional observed confidence level can be computed based on the percentile method confidence region of Efron (1979).

- A $100\alpha\%$ percentile method confidence region for θ is any region $\mathcal{V}_\alpha$ such that

$$\int_{\mathcal{V}_\alpha} d\widehat{V}_n(\mathbf{t}) = \alpha,$$

where $V_n(\mathbf{t}) = P\{\widehat{\theta} \leq \mathbf{t} | \mathbf{X}_1, \ldots \mathbf{X}_n \overset{\text{iid}}{\sim} F\}$ and the bootstrap estimate of $V_n(\mathbf{t})$ is given by

$$\widehat{V}_n(\mathbf{t}) = P^*\{\widehat{\theta}_i^* \leq \mathbf{t} | \mathbf{X}_1^*, \ldots \mathbf{X}_n^* \overset{\text{iid}}{\sim} \widehat{F}_n\}$$

Percentile Observed Confidence Level

- The corresponding observed confidence level based on the percentile method is

$$\widehat{\alpha}^*_{\text{perc}}(\Psi) = \int_{\Psi} d\widehat{V}_n(\mathbf{t})$$

Simulation Approximation

- Simulate b samples of size n from $\widehat{F}_n$.

- For each sample compute $\widehat{\theta}^*$.

- The distribution $\widehat{V}_n$ is approximated with the empirical distribution of $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_b^*$,

$$\widehat{V}_n(\mathbf{t}) \simeq \tilde{V}_n(\mathbf{t}) = b^{-1} \sum_{i=1}^{b} \delta\{\widehat{\theta}_i^* \leq \mathbf{t}\}$$

- The corresponding approximation of $\widehat{\alpha}^*_{\text{perc}}$ is

$$\widehat{\alpha}^*_{\text{perc}}(\boldsymbol{\Psi}) \simeq \int_{\boldsymbol{\Psi}} d\tilde{V}_n(\mathbf{t}) = b^{-1} \sum_{i=1}^{b} \delta\{\widehat{\theta}_i^* \in \boldsymbol{\Psi}\}$$

Asymptotic Accuracy

- A method for computing an observed confidence level is accurate if $\alpha(\Psi) = \alpha$ whenever Ψ is a $100\alpha\%$ confidence region for θ.

- The studentized measure α_{stud} is accurate by this definition.

- A method for computing an observed confidence level is k^{th}-order accurate if

$$\alpha(\Psi) = \alpha + O(n^{-k/2})$$

as $n \to \infty$ whenever Ψ is a $100\alpha\%$ confidence region for θ.

Asymptotic Accuracy

- Accurate: α_{stud}

- Multivariate Edgeworth expansion theory can be used to prove the following asymptotic accuracy results:
  - First-Order Accurate: $\widehat{\alpha}_{\text{stud}}$, α_{hyb}, $\widehat{\alpha}^*_{\text{hyb}}$, $\widehat{\alpha}^*_{\text{perc}}$
  - Second-order Accurate: $\widehat{\alpha}^*_{\text{stud}}$

Empirical Study

- $\mathbf{X}_1, \ldots, \mathbf{X}_n \stackrel{\text{iid}}{\sim} F$, where F is a bivariate normal mixture and $n = 10$, 25 and 50.

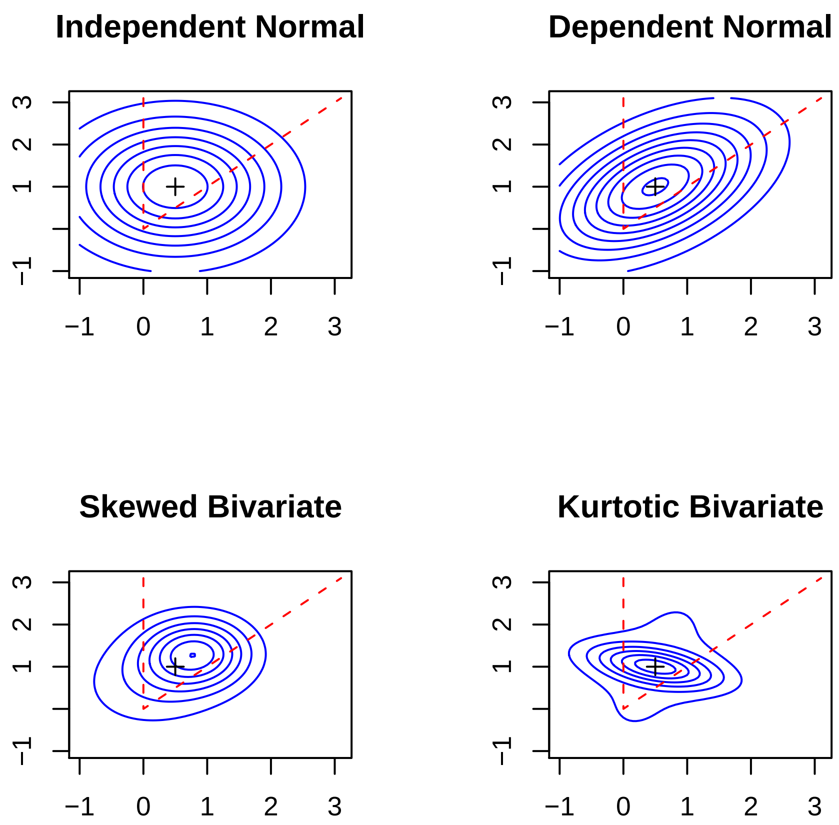
- $\theta = E\{\mathbf{X}_i\} = (0.5, 1.0)'$ with $\Theta = \mathbb{R}^2$

- $\Psi = \{\theta' = (\theta_1 \ \theta_2) : 0 \leq \theta_1 \leq \theta_2 < \infty\}$

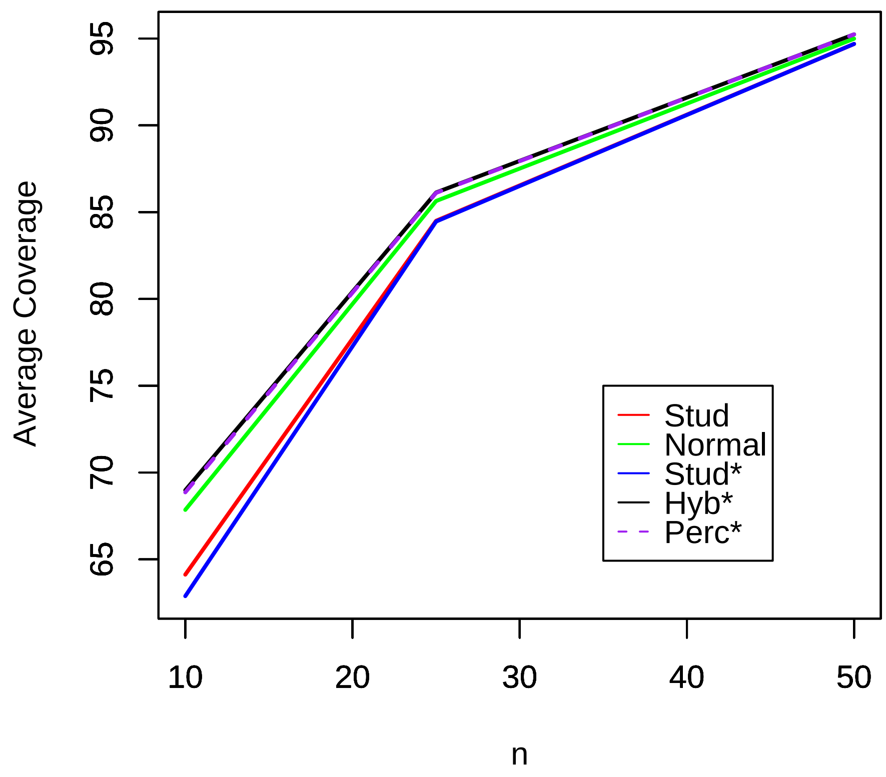
- 100 samples for each size and distribution were generated using R.

- The average observed confidence level over the 100 simulated samples for each of the methods α_{stud}, $\widehat{\alpha}^*_{\text{stud}}$, $\widehat{\alpha}_{\text{stud}}$, α_{hyb}, $\widehat{\alpha}^*_{\text{hyb}}$ and $\widehat{\alpha}^*_{\text{perc}}$ was computed.

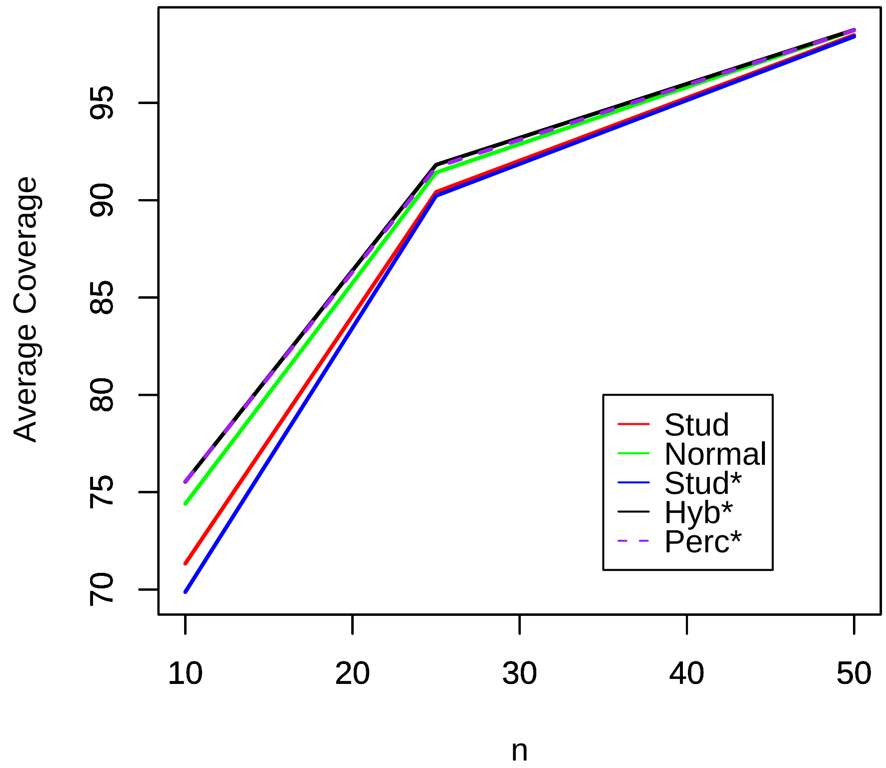
Normal Mixtures



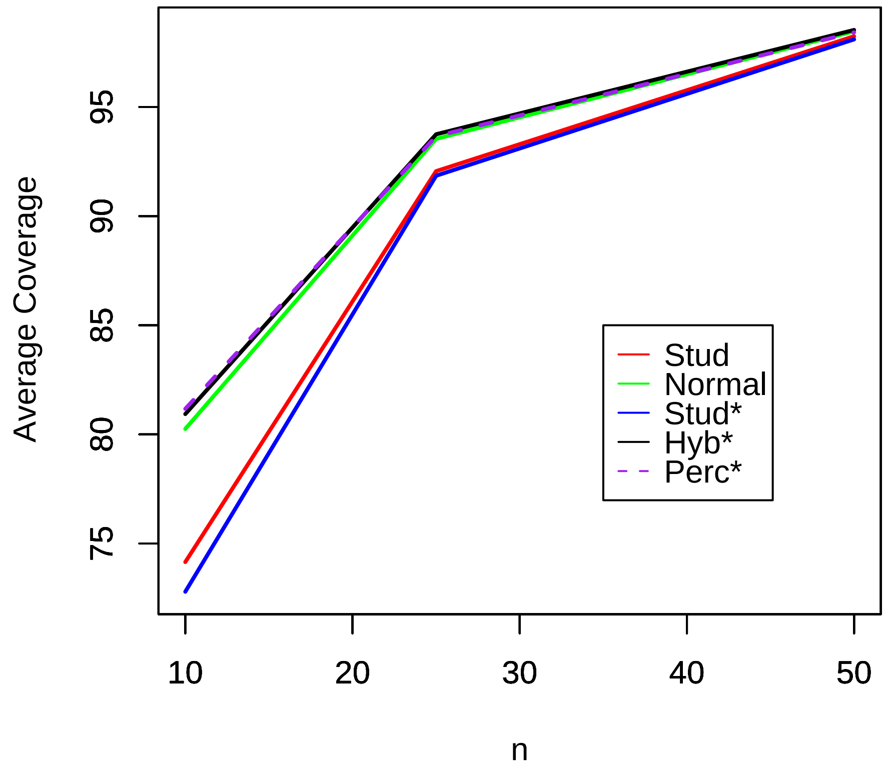
Simulation Results: Independent Normal



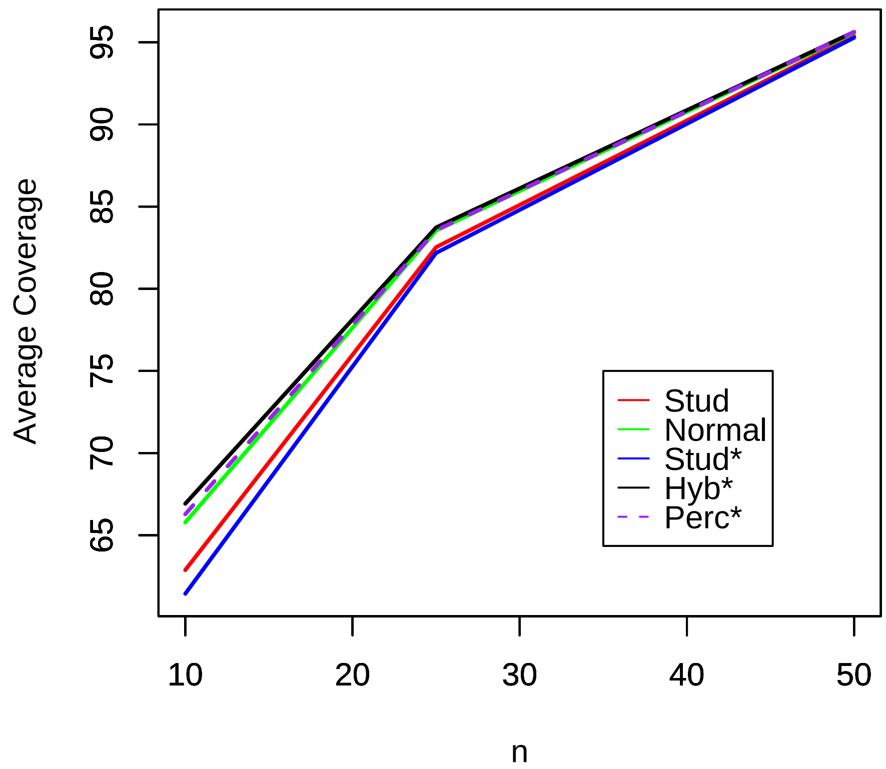
Simulation Results: Dependent Normal



Simulation Results: Skewed Bivariate



Simulation Results: Kurtotic Bivariate



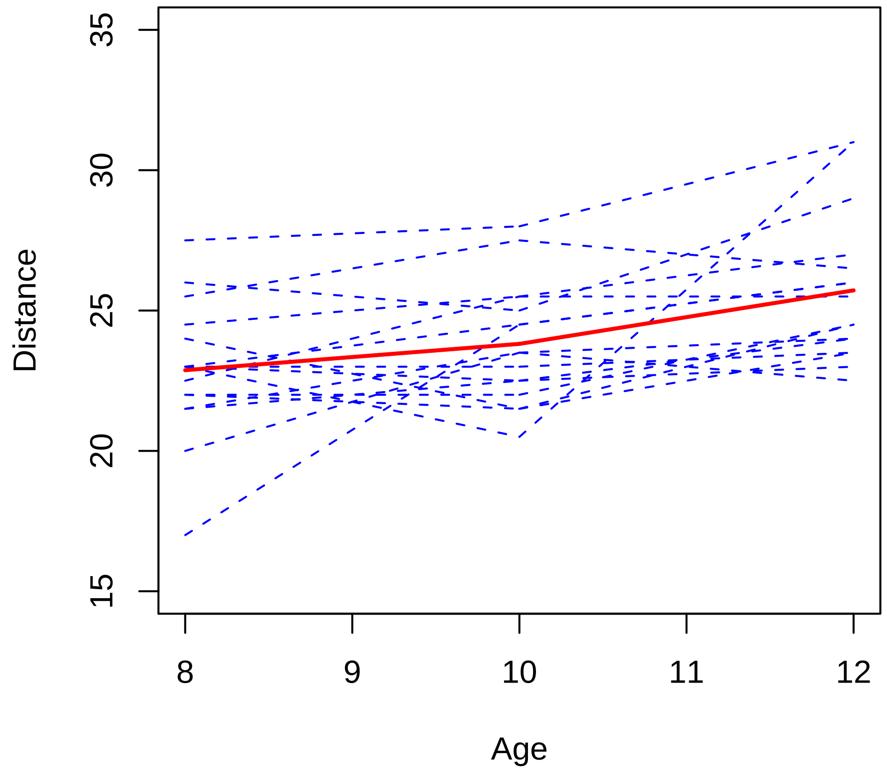
Example: Growth Curves

- Data from Potthoff and Roy (1964).

- Growth curves of the distance (mm) from the center of the pituitary gland to the pterygomaxillary fissure in young children.

- First three observations from the male subjects.

Example: Growth Curves



Example: Growth Curves

- Let $\mathbf{X}_i$ be a random vector containing the three measurements from subject $i = 1, \ldots, 16$.

- Assume $\mathbf{X}_1, \ldots, \mathbf{X}_{16} \overset{\text{iid}}{\sim} F$ where F is unknown.

- Let $\theta = E\{\mathbf{X}_i\}$.

- Regions of interest: $\Theta_1 = \{\theta : \theta_1 \leq \theta_2 \leq \theta_3\}$, $\Theta_2 = \{\theta : \theta_1 \leq \theta_3 \leq \theta_2\}$, $\Theta_3 = \{\theta : \theta_2 \leq \theta_1 \leq \theta_3\}$, $\Theta_4 = \{\theta : \theta_2 \leq \theta_3 \leq \theta_1\}$, $\Theta_5 = \{\theta : \theta_3 \leq \theta_1 \leq \theta_2\}$, $\Theta_6 = \{\theta : \theta_3 \leq \theta_2 \leq \theta_1\}$.

Observed Confidence Levels

Region	Ordering	$\widehat{\alpha}^*_{\text{stud}}$	$\widehat{\alpha}^*_{\text{hyb}}$	$\widehat{\alpha}^*_{\text{perc}}$
Θ_1	$\theta_1 \leq \theta_2 \leq \theta_3$	72.2	92.5	96.1
Θ_2	$\theta_1 \leq \theta_3 \leq \theta_2$	13.6	0.2	0.0
Θ_3	$\theta_2 \leq \theta_1 \leq \theta_3$	6.6	7.2	3.9
Θ_4	$\theta_2 \leq \theta_3 \leq \theta_1$	0.1	0.0	0.0
Θ_5	$\theta_3 \leq \theta_1 \leq \theta_2$	7.5	0.0	0.0
Θ_6	$\theta_3 \leq \theta_2 \leq \theta_1$	0.0	0.1	0.0

Further Research

- Regression and Linear Models [in progress and Polansky (2003)]

- Nonparametric regression and density estimation [Efron and Tibshirani (1998), Hall and Ooi (2004)].

- Generalized Linear Models

- Goodness-of-Fit