Single-Fit Bootstrapping

David Fletcher

University of Otago

Single-Fit Bootstrapping

- Bootstrapping
  1. Parametric vs Non-parametric
  2. Percentile vs Studentised
  3. Many variations (e.g. conditional, stratified, balanced,...)
- All involve refitting the model (or models) many times
- Can we sometimes avoid refitting the model?
- Modelling zero-inflated skewed data (Fletcher et al 2005)

Single-Fit Bootstrapping

- Bootstrapping
  1. Parametric vs Non-parametric
  2. Percentile vs Studentised
  3. Many variations (e.g. conditional, stratified, balanced,...)
- All involve refitting the model (or models) many times
- Can we sometimes avoid refitting the model?
- Modelling zero-inflated skewed data (Fletcher et al 2005)

Single-Fit Bootstrapping

Two settings of interest

1. Likelihood-based model $f_Y(Y;\beta_0)$ where
   - n is large enough for asymptotic likelihood theory to work
   - parameter of interest is $\theta_0 = g(\beta_0)$, with $g(\cdot)$ non-linear
2. Generalised linear model where
   - we have overdispersion
   - use quasi-likelihood (vs random effects or correlated errors)
   - want a confidence interval for the amount of overdispersion
   - example of a scale parameter

Single-Fit Bootstrapping

- Likelihood-based model $f_Y(Y;\beta_0)$ where
  - n is large enough for asymptotic likelihood theory to work
  - parameter of interest is $\theta_0 = g(\beta_0)$, with $g(\cdot)$ non-linear
- Asymptotic theory gives

$$\hat{\beta} \sim N(\beta_0, \Sigma_{\hat{\beta}})$$

where

$$\Sigma_{\hat{\beta}}^{-1} = \mathbb{E}\{-\nabla^2 \ell(\beta)\}\Big|_{\beta=\beta_0}$$

$\ell(\beta) =$ log-likelihood $\qquad \beta_0 =$ true value of β

Single-Fit Bootstrapping

- Likelihood-based model $f_Y(Y;\beta_0)$ where
  - n is large enough for asymptotic likelihood theory to work
  - parameter of interest is $\theta_0 = g(\beta_0)$, with $g(\cdot)$ non-linear
- Asymptotic theory gives

$$\hat{\beta} \sim N(\beta_0, \Sigma_{\hat{\beta}})$$

where

$$\Sigma_{\hat{\beta}}^{-1} = \mathbb{E}\{-\nabla^2 \ell(\beta)\}\Big|_{\beta=\beta_0}$$

$\ell(\beta) =$ log-likelihood $\qquad \beta_0 =$ true value of β

- Why bootstrap?

Single-Fit Bootstrapping

- Likelihood-based model $f_Y(Y; \beta_0)$ where
  - n is large enough for asymptotic likelihood theory to work
  - parameter of interest is $\theta_0 = g(\beta_0)$, with $g(\cdot)$ non-linear
- Asymptotic theory gives

$$\hat{\beta} \sim N(\beta_0, \Sigma_{\hat{\beta}})$$

where

$$\Sigma_{\hat{\beta}}^{-1} = \mathbb{E}\{-\nabla^2 \ell(\beta)\}\big|_{\beta=\beta_0}$$

$\ell(\beta) =$ log-likelihood $\qquad \beta_0 =$ true value of β

- Why bootstrap? Avoid problems with delta method...

Single-Fit Bootstrapping

- Parametric bootstrapping pretends that $\beta_0 = \hat{\beta}$ and uses

$$y^* \sim f_Y(Y; \hat{\beta})$$

which leads to

$$\hat{\beta}^* \sim N(\hat{\beta}, \Sigma_{\hat{\beta}^*})$$

where

$$\Sigma_{\hat{\beta}^*}^{-1} = \mathbb{E}\{-\nabla^2 \ell(\beta)\}\big|_{\beta=\hat{\beta}}$$

Single-Fit Bootstrapping

- Parametric bootstrapping uses

$$\hat{\beta}^* \sim N(\hat{\beta}, \Sigma_{\hat{\beta}*}) \quad \text{where} \quad \Sigma_{\hat{\beta}^*}^{-1} = \mathbb{E}\{-\nabla^2 \ell(\beta)\}\big|_{\beta=\hat{\beta}}$$

- But this is equivalent to

$$\hat{\beta}^* \sim N(\hat{\beta}, \hat{\Sigma}_{\hat{\beta}}) \qquad \text{since} \quad \hat{\Sigma}_{\hat{\beta}}^{-1} = \mathbb{E}\{-\nabla^2 \ell(\beta)\}\big|_{\beta=\hat{\beta}}$$

Single-Fit Bootstrapping

- Parametric bootstrapping equivalent to using

$$\hat{\beta}^* \sim N(\hat{\beta}, \hat{\Sigma}_{\hat{\beta}}) \qquad \hat{\theta}^* = g(\hat{\beta}^*)$$

- Requires no refitting of the model
- Percentiles of $\hat{\theta}^*$ provide a confidence interval

Single-Fit Bootstrapping

- Parametric bootstrapping equivalent to using

$$\hat{\beta}^* \sim N(\hat{\beta}, \hat{\Sigma}_{\hat{\beta}}) \qquad \hat{\theta}^* = g(\hat{\beta}^*)$$

- Requires no refitting of the model
- Percentiles of $\hat{\theta}^*$ provide a confidence interval
- c.f. asymptotic Bayes
  - $p(\beta|y) \approx N(\hat{\beta}, \hat{\Sigma}_{\hat{\beta}})$
  - Posterior distribution vs Bootstrap distribution
  - Hastie, Tibshirani and Friedman (2009)

Single-Fit Bootstrapping

Example 1: Zero-inflated count data

- Hurdle (or mixture) model
  1. Binomial model for probability of a zero
  2. (Zero-truncated) negative binomial model for non-zeros
- $\mathbb{E}(Y) =$ non-linear function of the two sets of β's

Example 2: Probability of surviving to adulthood

- Mark-recapture model (e.g. chicks banded at fledging)

$$S_{0A} = S_J^k \qquad (k = \text{age at reaching adulthood})$$

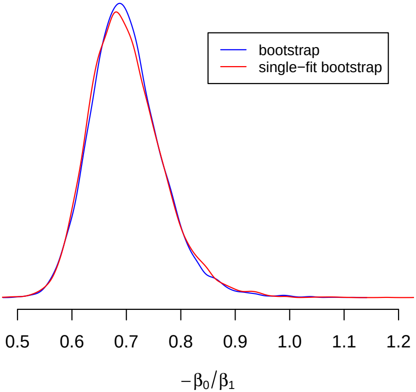
$$\log\left(\frac{S_J}{1-S_J}\right) = \beta_0 + \beta_1 x_1 + \ldots$$

Example 3: Tim's maxima...

Single-Fit Bootstrapping

Example 4: LD50

- Binomial model: $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x$ (LD50 $= -\beta_0/\beta_1$)
- Example: $n = 30$, $m = 10$, $\beta_0 = -1$, $\beta_1 = 1.6$, $x \in [0, 1]$



Single-Fit Bootstrapping

Confidence Interval for Overdispersion (with Michel de Lange)

- Overdispersion
  - Arises when fitting a generalised linear model to count data
  - Greater variation in Y than we would expect from the model
  - Occurs with binomial, Poisson and multinomial models
  - Leads to
    1. Confidence intervals being too narrow
    2. Model selection/averaging: too much weight to larger models
- Motivation: mark-recapture data (multinomial models)
  - Multi-state mark-recapture models can take a while to fit
  - Standard bootstrapping not straightforward to apply

Single-Fit Bootstrapping

- Reasons for overdispersion
  1. Heterogeneity in model parameters (e.g. probability of survival)
  2. Dependence in observations (e.g. breeding pair of birds)

- Possible remedies
  1. Add a random effect (e.g. individual probabilities of survival)
  2. Model the dependence structure (less common, more difficult)

- Semi-parametric approach: Quasi-likelihood
  1. Robust to model-misspecification
     - special case of robust likelihood (Royall and Tsou 2003)
  2. Simple to use (adjust intervals, model selection/averaging)

Single-Fit Bootstrapping

Quasi-likelihood

- Specify only the mean and variance of Y
- var$(Y_i) = \phi V_i$ $(i = 1, \dots, n)$
- V_i = variance of Y predicted by the assumed model
- Estimate of ϕ used to adjust
  1. Confidence intervals (wider)
  2. Model selection/averaging (less weight to larger models)

Single-Fit Bootstrapping

- Mark-recapture setting (multinomial models)
  1. Farzana Afroz PhD: better point estimator of ϕ
  2. How reliable is the estimator?
- Confidence interval for ϕ
  1. Useful in its own right (this talk)

     e.g. $\hat{\phi} = 4.6$; CI = [0.8, 10.3] vs [4.1, 5.4]
  2. Effect on (future work)
     - Confidence intervals for model parameters
     - Model selection/averaging

Single-Fit Bootstrapping

- Confidence interval for ϕ using the pivot

$$(n-p)\hat{\phi}/\phi \sim \chi^2_{n-p}$$

- Leads to a $100(1-2\alpha)\%$ interval given by

$$(n-p)\hat{\phi}/\phi_L = \chi^2_{n-p,1-\alpha} \qquad (n-p)\hat{\phi}/\phi_U = \chi^2_{n-p,\alpha}$$

- What if $(n-p)\hat{\phi}/\phi$ not χ^2_{n-p}?

  (e.g. $\phi > 1$ and/or n small)

Single-Fit Bootstrapping

- Confidence interval for ϕ using the approximate pivot

$$(n-p)\hat{\phi}/\phi \sim G_{n-p}$$

- Use bootstrap to estimate G_{n-p} (c.f. studentised bootstrap)
  1. B bootstrap estimates give $(n-p)\hat{\phi}^*_b/\hat{\phi} \quad (b = 1, \ldots, B)$
  2. Relevant percentiles of bootstrap distribution give

$$\phi_L = \hat{\phi}^2/\hat{\phi}^*_{1-\alpha} \qquad \phi_U = \hat{\phi}^2/\hat{\phi}^*_{\alpha}$$

- Resample Y and refit the model OR resample the residuals?

$$\hat{\phi}_1 = \sum_{i=1}^{n} e_i^2/(n-p) \quad \text{where} \quad e_i = (y_i - \hat{\mu}_i)/\hat{V}_i^{1/2}$$

$$\hat{\phi}_2 = \hat{\phi}_1/(1+\bar{s}) \quad \text{where} \quad s_i = (y_i - \hat{\mu}_i)/\hat{V}_i$$

Single-Fit Bootstrapping

- Resampling the residuals

$$\hat{\phi}_1 = \sum_{i=1}^{n} e_i^2/(n-p) \quad \text{where} \quad e_i = (y_i - \hat{\mu}_i)/\hat{V}_i^{1/2}$$

$$\hat{\phi}_2 = \hat{\phi}_1/(1+\bar{s}) \quad \text{where} \quad s_i = (y_i - \hat{\mu}_i)/\hat{V}_i$$

- $\hat{\phi}_2$ has smaller variance (especially when data are sparse)
- "hat-values" allowed for in the denominator $(n-p)$
- Avoids the need for conditioning upon $\hat{\mu}_i$ (McCullagh 1986)

Single-Fit Bootstrapping

- Resampling the residuals

$$\hat{\phi}_1 = \sum_{i=1}^n e_i^2/(n-p) \quad \text{where} \quad e_i = (y_i - \hat{\mu}_i)/\hat{V}_i^{1/2}$$

$$\hat{\phi}_2 = \hat{\phi}_1/(1+\bar{s}) \quad \text{where} \quad s_i = (y_i - \hat{\mu}_i)/\hat{V}_i$$

- $\mathbb{E}(e_i) \approx 0 \qquad \mathbb{E}(e_i^2) \approx \left(\frac{n-p}{n}\right)\phi \qquad \mathbb{E}(e_i e_j) \approx 0$

  e_i approximately exchangeable up to second moment

- $\mathbb{E}(e_i^3) \propto V_i'/V_i^{1/2} \implies$ not exhangeable on 3rd moment

  Stratify by $V_i'/V_i^{1/2}$ (Davison and Hinkley, 1997)

Single-Fit Bootstrapping

- Resampling the residuals

$$\hat{\phi}_1 = \sum_{i=1}^n e_i^2/(n-p) \quad \text{where} \quad e_i = (y_i - \hat{\mu}_i)/\hat{V}_i^{1/2}$$

$$\hat{\phi}_2 = \hat{\phi}_1/(1+\bar{s}) \quad \text{where} \quad s_i = (y_i - \hat{\mu}_i)/\hat{V}_i$$

- $\mathbb{E}(s_i) \approx 0 \qquad \mathbb{E}(s_i^2) \approx \left(\frac{n-p}{n}\right)\phi/V_i \qquad \mathbb{E}(s_i s_j) \approx 0$

  s_i approximately exchangeable on first moment only

  $\hat{\phi}_2$ better for point estimation

  what about interval estimation?

Single-Fit Bootstrapping

- Resampling the residuals

$$\hat{\phi}_1 = \sum_{i=1}^n e_i^2/(n-p) \quad \text{where} \quad e_i = (y_i - \hat{\mu}_i)/\hat{V}_i^{1/2}$$

$$\hat{\phi}_2 = \hat{\phi}_1/(1+\bar{s}) \quad \text{where} \quad s_i = (y_i - \hat{\mu}_i)/\hat{V}_i$$

- $\mathbb{E}(e_i) \approx 0 \qquad \mathbb{E}(e_i^2) \approx \frac{n-p}{n}\phi \qquad \mathbb{E}(e_i e_j) \approx 0$

- Set $u_i = \frac{n}{n-p} e_i^2$

  Estimate $\mathbb{E}(U_i) \approx \phi$ by $\hat{\phi}_1 = \frac{1}{n}\sum_{i=1}^n u_i$ (sample mean)

  Resampling the "observations" u_i

Single-Fit Bootstrapping

- Resampling the residuals

  $\hat{\phi}_1^* = \sum_{i=1}^n e_i^{*2}/(n-p)$

  $\hat{\phi}_2^* = \hat{\phi}_1^*/(1+\bar{s})$

  e_i^* randomly sampled from the e_i $(i = 1, \ldots, n)$

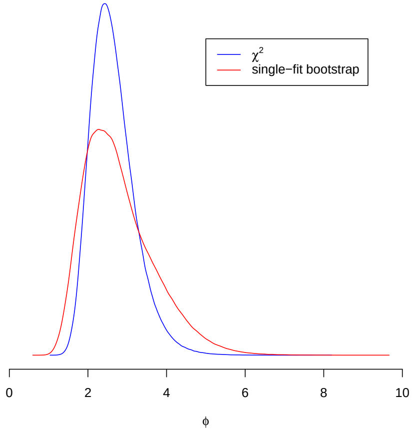
- No refitting of the model required

- What about the s_i?

  Use of $s_i^* = e_i^*/\hat{V}_i^{1/2}$ can lead to $\hat{\phi}_2^* < 0$

Single-Fit Bootstrapping

- Flamingo data (615 adults; 1982–1987; Camargue, France)
- Confidence distributions for ϕ based on $\hat{\phi}_2$



Single-Fit Bootstrapping

Simulations from mark-recapture model

- $\text{var}(Y_{ij}) = \phi\mu_{ij}$ (release-cohort i, capture-history j)
- 6-year study, 100 newly-ringed individuals released each year
- $\phi = 2.5$, $S = 0.95$, $P = 0.75$
- Results for 90% interval using $\hat{\phi}_2$

	Coverage	Median Width
χ^2	0.84	1.65
Single-fit bootstrap	0.87	1.83

Single-Fit Bootstrapping

Simulations from mark-recapture model

- $\text{var}(Y_{ij}) = \phi\mu_{ij}$ (release-cohort i, capture-history j)
- 6-year study, 10 newly-ringed individuals released each year
- $\phi = 2.5$, $S = 0.95$, $P = 0.75$
- Results for 90% interval using $\hat{\phi}_2$

	Coverage	Median Width
χ^2	0.50	1.25
Single-fit bootstrap	0.88	2.63

Single-Fit Bootstrapping

Simulations from negative binomial model

- $\text{var}(Y_i) = \phi\mu_i \quad \log(\mu_i) = \beta_0 + \beta_1 x_{1i}$
- $n = 30$, $\phi = 2 - 5$, $\beta_0 = -2$, $\beta_1 = 4$ $(0.1 < \mu_i < 8)$
- Coverage for 90% interval using $\hat{\phi}_2$

	ϕ			
	2	3	4	5
χ^2	0.79	0.70	0.61	0.61
Single-fit bootstrap	0.89	0.87	0.84	0.84

Single-Fit Bootstrapping

- Two settings where single-fit bootstrapping is useful
  1. Likelihood-based model with large n and $\theta = g(\beta)$ non-linear
  2. Estimation of a scale parameter (e.g. overdispersion in GLMs)
- θ involves location and scale parameters? (Fletcher et al 2005)
- Already used in practice? (e.g. Brooks and Morgan 2000)
- Other settings?

Single-Fit Bootstrapping

David Fletcher

University of Otago