Section 7: Nonparametric uncertainty estimation

Section 7. 21 Nov. 2001

Nonparametric methods of uncertainty estimination

δ-method, interpenetrating samples, jackknife, bootstrap, cross-validation

Techniques broadly applicable, complex statistics
Generally justified by asymptotics
There exist singular and inappropriate cases
NECESSARY

δ-method AKA method of linearization, propagation of error, Taylor series method

Gauss (1815)

Basically one approximates functions by Taylor expansions (usually linear) of basic random variables.

Rao, Section 6a.2

Sequence of k-dimensional statistics

$$\underset{\sim}{T}_n = (T_{1n}, \ldots, T_{kn}) \qquad n = 1, 2, \ldots$$

e.g. $(\bar{X}, \bar{Y})$

27 Nar 2001

Derived statistic

$$g(\underset{\sim}{T}_n) = g(T_{1n}, \ldots, T_{kn})$$

eg. $g(\bar{X}, \bar{Y}) = \dfrac{\bar{Y}}{\bar{X}}$

Suppose $\boxed{\sqrt{n}\,(\underset{\sim}{T}_n - \underset{\sim}{\theta}) \xrightarrow{d} N_k(0, \underset{\sim}{\Sigma})}$ $\xrightarrow{L}$

Write

$$g(\underset{\sim}{T}_n) = g(\underset{\sim}{\theta}) + \frac{\partial g(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} \cdot (\underset{\sim}{T}_n - \theta) + \ldots$$

The ~~entity~~ of principal interest is typically $g(\underset{\sim}{\theta})$

Theorem. If $g(\,)$ has a continuous first derivative

$$\boxed{\sqrt{n}\,\{g(\underset{\sim}{T}_n) - g(\theta)\} \xrightarrow{d} N\left(\underset{\sim}{0}, \frac{\partial g^T}{\partial \theta} \underset{\sim}{\Sigma} \frac{\partial g}{\partial \theta}\right)}$$

7 Nov. 2001

Result suggests estimating the covariance matrix of $g(\underset{\sim}{T}_n)$ by

$$\left(\left.\frac{\partial g}{\partial \theta}\right|_{\theta = T_n}\right)^{T} \hat{\underset{\sim}{\Sigma}} \left(\frac{\partial g}{\partial \theta}\right)$$

but...

Procedure "works" provided $\underset{\sim}{T}_n$ is a neighborhood of θ

Gives an approximating distribution.
Could use for confidence intervals.

7 Nov. 2001

Example where not too useful (Cramér's book)
{ but more later } Mathematica)

$$r = \frac{\sum'(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum'(x_i-\bar{x})^2\sum'(y_i-\bar{y})^2}}$$

$$= \frac{\bar{t}_4 - \bar{t}_1\bar{t}_2}{\sqrt{(\bar{t}_3-\bar{t}_1^2)(\bar{t}_4-\bar{t}_2^2)}} = g(\bar{t}_1, \bar{t}_2) \ldots$$

$t_1 = x/\sqrt{n}$ $\quad t_3 = x^2$ $\quad t_5 = y^2$

$t_2 = y/\sqrt{n}$ $\quad t_4 = xy$

Answer. An estimate of var r

$$\frac{r^2}{4n}\left[\frac{\hat\mu_{40}}{\hat\mu_{20}^2} + \frac{\hat\mu_{04}}{\hat\mu_{02}^2} + \frac{2\hat\mu_{22}}{\hat\mu_{20}\hat\mu_{02}} + \frac{4\hat\mu_{22}}{\hat\mu_{11}} - \frac{4\hat\mu_{31}}{\hat\mu_{11}\hat\mu_{20}} - \frac{4\hat\mu_{13}}{\hat\mu_{11}\hat\mu_{02}}\right]$$

But in normal case $(1-r^2)^2/n$

$$\hat\mu_{gh} = \frac{1}{n}\sum'(x_i-\bar{x})^g(y_i-\bar{y})^h$$

27 Nov 2001

Remarks

1. Is n large enough?
2. Is g sufficiently linear? (differentiable?)
3. Is the algebra correct?
4. Is it approximately normal? (CI's etc)
5. Note this is not giving expected values
6. The derivatives appearing may be approximated by finite differences

$$\frac{\partial g}{\partial \theta_j} = \frac{g(\theta_1, \ldots, \theta_{j-1}, \theta_j + \delta, \theta_{j+1}, \ldots, \theta_p) - g(\theta_1, \ldots, \theta_j - \delta, \ldots \theta_p)}{2\delta}$$

7. While often forgotten "learning" the bias of an estimate can be important.

Example where useful.

Ratio estimate, as in survey sampling

$$\hat{R} = \bar{y}/\bar{x} \qquad \text{(sampling without replacement)}$$

$$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R = (\bar{y} - R\bar{x})(\bar{X} + (\bar{x} - \bar{X}))^{-1}$$

Writing cap letters for population values

$$\hat{R} - R = \frac{\bar{y} - R\bar{x}}{\bar{X}}\left[1 - \frac{\bar{x} - \bar{X}}{\bar{X}} + \left(\frac{\bar{x} - \bar{X}}{\bar{X}}\right)^2 + \dots\right]$$

$$\simeq \frac{\bar{y} - R\bar{x}}{\bar{X}} \qquad \text{linear approx}$$

$$\simeq \frac{\bar{y} - R\bar{x}}{\bar{X}} - \frac{\bar{y} - R\bar{x}}{\bar{X}}\left(\frac{\bar{x} - \bar{X}}{\bar{X}}\right) \qquad \text{quadratic}$$

$$\operatorname{var} \hat{R} \simeq \operatorname{var}\left(\frac{\bar{y} - R\bar{x}}{\bar{X}}\right) = \frac{1-f}{n\bar{X}^2} \sum_{i=1}^{N} (y_i - Rx_i)^2/(N-1)$$

$$f = \frac{n}{N}$$

bias $E(\hat{R} - R) \simeq E\left(\frac{\bar{y} - R\bar{x}}{\bar{X}}\right) - E\,\frac{(\bar{y} - R\bar{x})(\bar{x} - \bar{X})}{\bar{X}^2}$

$$\simeq -\frac{1-f}{n\bar{X}^2}\left(\rho\, S_y S_x - R S_x^2\right)$$

14 April 04

Convergence in distribution (law)

Sequence of r.v.'s $\{X_n\}$

$$F_n(x) = \text{Prob}\{X_n \le x\}$$

$X_n \xrightarrow{d} X$ $\quad (X_n \xrightarrow{L} X)$ $\quad$ (weakly)

if $F_n(x) \to F(x)$ at all continuity points of F

equivalent to

$\int g \, dF_n \to \int g \, dF$ for all bounded continuous g

Doesn't always hold if g is unbounded.

e.g. $X_n = \mu + \sigma Z + \frac{1}{n} C$ $\quad$ C: Cauchy

$$X_n \xrightarrow{d} X \sim N(\mu, \sigma^2)$$

$$EX_n = \infty, \quad EX = \mu$$

$\text{var}\, h(X_n) \not\to \text{var}\, h(X)$ generally

14 April 04

Can use more terms for a (possibly) better approximation:

$$g(T_n) = g(\theta) + g'(\theta)(T_n - \theta) + \frac{1}{2} g''(\theta)(T_n - \theta)^2 + \dots$$

$\vec{\text{ave}}\, g(T_n)$?

$$E\, T_n = \theta + \frac{b_\theta}{n} + \dots \qquad \text{biased}$$

$$\text{var}\, T_n = \frac{\sigma^2}{n} + \dots$$

$$E(T_n - \theta)^2 = \left(\frac{b_\theta}{n}\right)^2 + \frac{\sigma^2}{n} + \dots$$

$$\vec{\text{ave}}\, g(T_n) = g(\theta) + g'(\theta)\frac{b_\theta}{n} + \frac{1}{2} g''(\theta)\frac{\sigma^2}{n} + \dots$$

14 April 04

Variance stabilizing transformation

Correlation coefficient

$$\sqrt{n}(r-\rho) \xrightarrow{d} N(0, (1-\rho^2)^2)$$

Look for $g(\cdot)$ such that variance of the large sample distribution is approximately constant

Wish $[g'(\rho)]^2 (1-\rho^2)^2 = c$

Take $g'(\rho) = \frac{c}{1-\rho^2}$

$$\begin{aligned} g(\rho) &= \int \frac{c}{1-\rho^2} d\rho \\ &= \frac{c}{2} \int \left(\frac{1}{1+\rho} + \frac{1}{1-\rho} \right) d\rho \\ &= \frac{c}{2} \left(\log(1+\rho) - \log(1-\rho) \right) \\ &= c \tanh^{-1} \rho \end{aligned}$$

Claims: makes variable more gaussian
relationships more additive

14 April 04

An oddity r^2 when $\rho = 0$

$$\operatorname{var} r^2 \sim [2\rho]^2 \frac{1}{n}(1-\rho^2)^2$$

$$= 0 \quad \text{when } \rho = 0$$

Need more terms in expansion

$$\operatorname{var} r^2 \sim \frac{2}{n^2} \quad \text{when } \rho = 0$$

Singular $\left.\frac{\partial g}{\partial \theta}\right|_{\theta_0}$ does occur in practical situations

e.g. estimating MI

Distributions become χ^2, not normal

There are <u>functional forms of these results</u>, e.g. using Frechet or Grateaux derivatives

Consider $\bar{y}/\bar{x}$

Suppose c.d.f. of (x,y) is $F(x,y)$ and empirical c.d.f. is

$$F_n(x,y) = \frac{1}{n}\#\{i \mid x_i \le x, y_i \le y\}$$

Then

$$\theta = \frac{\mu_y}{\mu_x} = \frac{\iint y \, dF(x,y)}{\iint x \, dF(x,y)}$$

and

$$\hat{\theta} = \frac{\bar{y}}{\bar{x}} = \frac{\iint y \, dF_n(x,y)}{\iint x \, dF_n(x,y)}$$

One is considering

$$\theta = t(F) \text{ and } \hat{\theta} = t(F_n) \qquad (*)$$

Might use a <u>density estimate</u> f_n instead of F_n

Using $(*)$ defines $\hat{\theta}$, consistently, for all n

3.2.2. *Example 2: likelihood ratio statistic*

The likelihood ratio statistic is asymptotically χ^2. Large sample approximations for this statistic can be improved through a Bartlett correction. The resultant statistic has the same moments as a χ^2-variable up to order $n^{-3/2}$ (Lawley, 1956). In this example the asymptotic expansion of the likelihood ratio statistic and its expected value are derived by using the tools developed in this paper.

In[18]: = Like = 2(Sum[l[i, th], {i, n}] − Sum[l[i, t], {i, n}])

In[19]: = AsyExpLike = AsymptoticExpansion[Like, 2]

$$\text{Out[19]} = \frac{z_1^2}{I_2} + \frac{z_1^3 I_{001}}{3I_2^3\sqrt{n}} + \frac{z_1^2 z_{01}}{I_2^2\sqrt{n}} + \frac{z_1^3 z_{001}}{3I_2^3 n} + \frac{z_1^2 z_{01}^2}{I_2^3 n} + \frac{z_1^4 I_{0001}}{12 I_2^4 n} + \frac{z_1^3 z_{01} I_{001}}{I_2^4 n} + \frac{z_1^4 I_{001}^2}{4 I_2^5 n}$$

The expected value of the likelihood ratio statistic is then

In[20]: = Expand[Expect[AsyExpLike, 2]/.
Union[BartRepl[2, I_2],BartRepl[3, I_3],BartRepl[4, I_4]]]

$$\text{Out[20]} = 1 + \frac{I_{21}}{I_2^2 n} + \frac{I_{02}}{I_2^2 n} + \frac{I_{101}}{I_2^2 n} + \frac{I_{0001}}{4I_2^2 n} + \frac{2I_{11}^2}{I_2^3 n} + \frac{2I_{11} I_{001}}{I_2^3 n} + \frac{5I_{001}^2}{12 I_2^3 n}$$

```
Haldane[ExpandSum[3]]
```

yields

$$\frac{a_{i_1 j_1}^2 a_{i_2 j_1}^2 a_{i_3 j_1}^2}{s_{i_1} s_{i_2} s_{i_3} t_{j_1}^3} + \frac{3 a_{i_1 j_1}^2 a_{i_2 j_1}^4}{s_{i_1} s_{i_2} t_{j_1}^2} + \frac{a_{i_1 j_1}^6}{s_{i_1} t_{j_1}} + \frac{a_{i_1 j_1}^2 a_{i_2 j_2}^2 a_{i_3 j_2}^2}{s_{i_1} s_{i_2} s_{i_3} t_{j_1} t_{j_2}^2} +$$
$$\frac{6 a_{i_1 j_1}^2 a_{i_2 j_1}^2 a_{i_2 j_2}^2}{s_{i_1} s_{i_2}^2 t_{j_1}^2 t_{j_2}} + \frac{a_{i_1 j_1}^2 a_{i_3 j_1}^2 a_{i_2 j_2}^2}{s_{i_1} s_{i_2} s_{i_3} t_{j_1}^2 t_{j_2}} + \frac{a_{i_1 j_1}^2 a_{i_2 j_1}^2 a_{i_3 j_2}^2}{s_{i_1} s_{i_2} s_{i_3} t_{j_1}^2 t_{j_2}} + \frac{3 a_{i_1 j_1}^2 a_{i_1 j_2}^4}{s_{i_1}^2 t_{j_1} t_{j_2}} +$$
$$\frac{3 a_{i_1 j_1}^2 a_{i_2 j_2}^4}{s_{i_1} s_{i_2} t_{j_1} t_{j_2}} + \frac{a_{i_1 j_1}^2 a_{i_1 j_2}^2 a_{i_1 j_3}^2}{s_{i_1}^3 t_{j_1} t_{j_2} t_{j_3}} + \frac{3 a_{i_1 j_1}^2 a_{i_2 j_2}^2 a_{i_2 j_3}^2}{s_{i_1} s_{i_2}^2 t_{j_1} t_{j_2} t_{j_3}} + \frac{a_{i_1 j_1}^2 a_{i_2 j_2}^2 a_{i_3 j_3}^2}{s_{i_1} s_{i_2} s_{i_3} t_{j_1} t_{j_2} t_{j_3}},$$

$$\frac{s_{i_1}(-2+s_{i_2})(-1+s_{i_2}) s_{i_2}(-2+t_{j_1})(-1+t_{j_1}) t_{j_1} t_{j_2}}{(-3+n)(-2+n)(-1+n)n} +$$
$$\frac{(-1+s_{i_1}) s_{i_1}(-2+s_{i_2})(-1+s_{i_2}) s_{i_2}(-3+t_{j_1})(-2+t_{j_1})(-1+t_{j_1}) t_{j_1} t_{j_2}}{(-4+n)(-3+n)(-2+n)(-1+n)n} +$$
$$\frac{s_{i_1}(-2+s_{i_2})(-1+s_{i_2}) s_{i_2}(-1+t_{j_1}) t_{j_1}(-1+t_{j_2}) t_{j_2}}{(-3+n)(-2+n)(-1+n)n} +$$
$$\frac{(-1+s_{i_1}) s_{i_1}(-2+s_{i_2})(-1+s_{i_2}) s_{i_2}(-2+t_{j_1})(-1+t_{j_1}) t_{j_1}(-1+t_{j_2}) t_{j_2}}{(-4+n)(-3+n)(-2+n)(-1+n)n} +$$
$$\frac{s_{i_1}(-3+s_{i_2})(-2+s_{i_2})(-1+s_{i_2}) s_{i_2}(-2+t_{j_1})(-1+t_{j_1}) t_{j_1}(-1+t_{j_2}) t_{j_2}}{(-4+n)(-3+n)(-2+n)(-1+n)n} +$$
$$\frac{(-1+s_{i_1}) s_{i_1}(-3+s_{i_2})(-2+s_{i_2})(-1+s_{i_2}) s_{i_2}(-3+t_{j_1})(-2+t_{j_1})(-1+t_{j_1}) t_{j_1}(-1+t_{j_2}) t_{j_2}}{(-5+n)(-4+n)(-3+n)(-2+n)(-1+n)n}.$$

 14 April 04

Replicated subsamples
Interpenetrating subsamples (Mahalanobis)

One computes $\hat{\theta}_i = t(\hat{F}_{im})$ for the i-th subsample.

$$S = \bigcup_i S_i \ , \quad S_i \cap S_{i'} = \emptyset \quad i \neq i'$$

Purposes
1. To estimate sampling variances when sample design is complicated and exact estimators are unavailable or cumbersome

2. To control field work

3. To measure components of nonsampling variances (e.g. enumerators)
Particularly useful for the study of correlated errors.

$\bar{Y}$ overall sample mean
$\bar{Y}_i$ mean of i-th subsample

Estimate var $\bar{Y}$ by

$$\frac{1}{I} \sum_i (\bar{Y}_i - \bar{Y})^2 / (I-1)$$

Faster

Anova

	SS	DF	MS
Between subsamples	$\sum_j \sum_i (\bar{Y}_i - \bar{Y})^2$	$I-1$	s_b^2
Within subsamples	$\sum_j \sum_i (Y_{ij} - \bar{Y}_i)^2$	$I(J-1)$	s_w^2
Total	$\sum_j \sum_i (Y_{ij} - \bar{Y})^2$	$IJ-1$	

Perhaps $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$ $\qquad \tau_i$ random

τ_i make values in i-th subsample correlated,

$$\text{corr}\{Y_{ij}, Y_{ij'}\} = \frac{\sigma_\tau^2}{\sigma_\epsilon^2 + \sigma_\tau^2}$$

$$E\, s_w^2 = \sigma_\epsilon^2$$

$$E\, s_b^2 = \sigma_\epsilon^2 + J\sigma_\tau^2$$

15 April 04

The jackknife. $n = IJ$, I groups of J

$\hat{\theta}$ based on all the data

$\hat{\theta}_{-i}$ " " ", but the i-th group

$$\hat{\theta}_{pi} = I\hat{\theta} - (I-1)\hat{\theta}_{-i}$$

$$\text{ave}\,(I\hat{\theta} - (I-1)\hat{\theta}_{-i}) = \theta + \frac{c}{(IJ)^2} + \cdots$$

$\bar{\theta} = \text{ave}\, \hat{\theta}_{pi}$ has reduced bias, in an asymptotic sense

$$\hat{\theta} - \bar{\theta} = -(I-1)\left[\hat{\theta} - \sum_i \hat{\theta}_{-i}/n\right]$$

estimates the bias

$$s^2 = \sum_i (\hat{\theta}_{pi} - \bar{\theta})^2/(I-1)$$

Estimate of var $\hat{\theta}$, var $\bar{\theta}$ provided by

$$\frac{s^2}{I}$$

J. W. Tukey, M. Quenouille

15 April 04

E.g. $\hat{\theta} = \dfrac{X_1 + \ldots + X_n}{n} = \overline{X}$

$$= \frac{J\overline{X}_1 + \ldots + J\overline{X}_I}{IJ}$$

$$= \frac{\overline{X}_1 + \ldots + \overline{X}_I}{I}$$

$$\hat{\theta}_{-i} = \frac{J\overline{X}_1 + \ldots + J\overline{X}_I - J\overline{X}_i}{(I-1)J}$$

$$= \frac{\overline{X}_1 + \ldots + \overline{X}_I - \overline{X}_i}{(I-1)J}$$

$$\hat{\theta}_{pi} = \overline{X}_i$$

$\overline{\theta} = \hat{\theta} = \overline{X}$ here

$\hat{\theta} - \overline{\theta} = 0$

$s^2 = \sum_i (\overline{X}_{i} - \overline{X})^2 / (I-1)$

var $\overline{X}$ estimated by s^2/I

Other justifications are asymptotic, e.g.
for $\hat{\theta} = g(\bar{A}, \bar{B}, \dots)$
(function of means)

Which asymptotics?
I fixed, $J \to \infty$ easy
$I \to \infty$, J fixed harder
e.g. $J = 1$

The estimate is inconsistent for the sample median when $J = 1$. (Not a regular enough functional)

Might compute for: histogram, qq plot, ...

There are also weighted jackknives, e.g. for regression

Tukey suggested $I = 10$

20 April 04

Advantages of the jackknife.

"Like the Boy Scout's knife, it can be used to do many jobs ..."

Just need a program to evaluate the estimate of interest, $\hat{\theta}$

8 Nov 01

Variants of the jackknife

1. drop 1, .., s

2. regression

3. vector-case

4. what really needed
$\widehat{\text{var}\,\theta} \sim \frac{\hat{J}^2}{n}$

5. which asymptotics?

6. bias correction via variance
(history)

7. missing values