Price fluctuations: Lecture 6

Zempléni, András

Department of Probability Theory
Faculty of Sciences
ELTE

Bootstrap (Efron, 1979)

- Resampling method, for investigating the variances of our estimators, to check the fit of the models
- There are many different versions worked out since then, it is one of the most quickly developing area of the statistics
- Its advantage: it is flexible with respect of conditions on the distribution of the sample/statistics

Bootstrap method - introduction

- $\mathbf{X_i^*} = \{X_1^*, \ldots, X_m^*\}$ - sampling with replacement from the original sample
- P_* is its distribution
- in general $m = n$
- Difficulties in practice:
  1. $\underline{x} \Longrightarrow \hat{P}$ depends on the chosen model
  2. $\hat{P} \longrightarrow \underline{x}^*$ the many repetitions are computer-intensive

The i.i.d. bootstrap

- Let $X_1, X_2, \ldots$ be i.i.d. random variables with (unknown) distribution function, F
- $T_n = t_n(\mathcal{X}_n; F)$ the random variable (statistics) of interest, its distribution: G_n
- Our aim: to estimate the distribution of G_n
- Bootstrap method:
  - For given $\mathcal{X}$, we take an m-element sample with replacement $\mathcal{X}_m^* = \{X_1^*, \ldots, X_m^*\}$
  - The distribution of X_i^*: $F_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}$
  - $T_{m,n}^* = t_m(\mathcal{X}_m^*; F_n)$
  - Repetitions $\Rightarrow \hat{G}_{m,n}$

Fundamental theorem (Efron)

- In the case above, if $\sigma^2 = Var(X_i)$ is finite, and the statistics is the standardized sample mean

$$T_n = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}$$

then

$$\lim_{n\to\infty} \sup_x |P_*(T^*_{n,n} \leq x) - \Phi(x)| = 0$$

a.s.

- The proof is based on the Berry-Esséen theorem (it gives the speed of the convergence at the central limit theorem), the convergence in $\sup_x |\hat{G}_{n,n}(x) - G_n(x)|$ can be even quicker than that of the classical normal approximation.
- The approach can be generalized to many cases (smoothness is required)

A counter-example

- In some cases the estimator is not consistent (Singh, 1981):

Definition

$\{X_n\}n \geq 1$ is m-dependent for an $m \geq 0$, if $\{X_1, \ldots, X_k\}$ and $\{X_{k+m+1}, \ldots\}$ are independent for every $k \geq 0$.

- Notation: $\sigma_m^2 = Var(X_1) + 2\sum_{i=1}^{m-1} Cov(X_1, X_{1+i})$
- Let the statistics to be estimated: $T_n = \sqrt{n}(\overline{X}_n - \mu)$
- Its bootstrap counterpart: $T_{n,n}^* = \sqrt{n}(\overline{X}_n^* - \overline{X}_n)$

Theorem

Let $\{X_n\}n \geq 1$ be a stationary m-dependent sequence of r.v.s, $EX_1 = \mu$, $\sigma^2 = Var(X_1) \in (0, \infty)$, $\sum_{i=1}^{m-1} Cov(X_1, X_{1+i}) \neq 0$ and $\sigma_m^2 \neq 0$. Then

$$\lim_{n \to \infty} \sup_x |P_*(T_{n,n}^* \leq x) - P(T_n \leq x)| \neq 0$$

a.s.

Correction in case of confidence intervals

- A refinement is appropriate to the naive, empirical quantiles. The BCA (bias-corrected and accelerated)-method for determining the limits:

$$\hat{F}^{-1}\left\{\Phi\left(z_0 + \frac{z^\alpha + z_0}{1 - a(z^\alpha + z_0)}\right)\right\}$$

- where $\hat{F}^{-1}$ is the empirical distribution function of the bootstrap statistics
- z^α is the standard normal α-quantile
- z_0 is a term for correcting the bias
- a corrects the acceleration of the increase of the variance
- If $a = 0$ and $z_0 = 0$, the value is exactly $F^{-1}(\alpha)$, the empirical α-quantile.

The motivation and application of the BC-formula

- If applying a monotonic transformation $m(\vartheta)$ to our estimator, the result is normally distributed:

$$m(\hat{\vartheta}) \sim N\left(m(\vartheta) - z_0(1 + am(\vartheta)), 1 + am(\vartheta)\right).$$

- Then, because of the monotonicity $P(\hat{\vartheta} < \vartheta) = \Phi(z_0)$, z_0 can be easily estimated
- The estimation of a can be got from the skewness of the derivative of the loglikelihood function

Examples

- Confidence interval for the correlation:
  - the standard interval (it is based on the normality of the empirical correlation coefficient) is symmetric – not always realistic for small samples
  - Boostrap quantiles may be biased
  - The BCA method can give an asymmetric interval, its coverage probability is usually more exact
- Similar problems arise in extreme-value applications (VaR estimation)
- It is a question, whether the parametric or the nonparametric bootstrap is worth using (the parametric is based on an assumed model, like normality of the sample and it gives usually a more cautious - wider - interval)
- Bootstrap may fail in high quantile estimation (e.g. for the upper end-point estimation), because the limit is a random variable instead of ϑ

The m out of n bootstrap

- If the usual bootstrap does not work, quite often it helps if we take samples of size $m < n$
- In this case even sampling without replacement is possible, which may have better properties
- Bickel and Sakov (2008) give an algorithm for finding the optimal m - valid for the original sampling scheme (with replacement). The result is $m \sim n$, if a sample of n is good as well.

Example

- Let X_i be an i.i.d. sequence with expectation μ and standard deviation σ
- We test the hypothesis $\mu = 0$ using the statistics $\sqrt{n}\overline{X}_n$
- It is a good bootstrap algorithm to take samples from the "residuals" $X_i - \overline{X}_n$
- If we consider the bootstrap distribution of $\sqrt{n}\ \overline{X}_n^*$ then its quantiles are not consistent (as we shall see on the next slide)
- For a fixed m, when $n \to \infty$ the limit distribution of $\sqrt{m}\ \overline{X}_m^*$ depends on m (it is constant for all m just for the normal distribution)

Example/2

- $\sqrt{m}(\overline{X}^*_m - \overline{X}_n) \longrightarrow N(0, \sigma)$ if $n, m \to \infty$
- So $\sqrt{m}\,\overline{X}^*_m \sim N(\sqrt{m}\,\overline{X}_n, \sigma)$ if $m \to \infty$
- $\sqrt{m}\,\overline{X}_n = \sqrt{m/n}\sqrt{n}\,\overline{X}_n \longrightarrow N(0, \sqrt{\lambda}\sigma)$ where $\lambda = \lim m/n$
- We get the correct result in case of $m/n \to 0$ (otherwise there is additional randomness in the limit)

The choice of m

- The bootstrap distribution does not change much near the optimum
- If m is too large or too small, then the bootstrap distributions are different
- Thus the algorithm:
  1. Let $m_j = [q^j n]$ $(0 < q < 1)$
  2. Let us determine the distribution of $T^*_{m_j,n}$ for all m_j by simulation
  3. Choose the m for which $\hat{m} = \text{argmin}\rho(T^*_{m_j,n}, T^*_{m_{j+1},n})$ (where ρ is a metrics, consistent with the convergence in distribution - e.g. Kolmogorov-Smirnov metrics)

Applications for the dependent case

Circular block bootstrap (CBB)

1. $Y_t = X_{t \mod (n)}$ (we continue the time series from the beginning, n is the length of the series)
2. Let $i_1, i_2, \ldots i_m$ be a sample with replacement from the uniform distribution over $\{1, \ldots, n\}$
3. For a given block size b let us create $n' = mb$ ($n' \approx n$) pseudo-observations:
$$Y^*_{(k-1)b+j} = Y_{i_k+j-1}, \text{ where } j = 1, \ldots, b; \quad k = 1, \ldots, m$$
4. Computation of the statistics of interest from the pseudo-observations:
$$\overline{Y}^*_{n'} = (n')^{-1}(Y^*_1 + \ldots + Y^*_{n'})$$

Selecting the block-size (Politis & White)

$\mathcal{F}_{-\infty}^{0} = \{X_n : n \leq 0\}$, $\mathcal{F}_{k}^{\infty} = \{X_n : n \geq k\}$

Definition

$\{X_t$: $t \in \mathbb{Z}\}$ is strongly mixing, if $\alpha_X(k) \longrightarrow 0$ $(k \to \infty)$, where
$\alpha_X(k) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{k}^{\infty}\}$

Theorem

Let us suppose that $E|X_t|^{6+\delta} < \infty$, $\sum_{k=1}^{\infty} k^2 (\alpha_X(k))^{\frac{\delta}{6+\delta}} < \infty$

for a suitable δ>0.
Let $b = o(N^{1/2})$, so if $N \to \infty$ then $b \to \infty$.
This implies $MSE(\sigma^2_{b,\overline{X}}) = \frac{G^2}{b^2} + D\frac{b}{n} + o(b^{-2}) + o(\frac{b}{n})$

where $D = \frac{4}{3}g^2(0)$ *and* $G = \sum_{k=-\infty}^{\infty} |k|R(k)$

$g(\cdot)$ is the spectral density function
$R(\cdot)$ is the autocovariance function

Selecting the block-size/2

- The optimal block-size: $b_{opt} = [(\frac{2G^2}{D})n^{1/3}]$
- Question: how to estimate G and D
- $\hat{D} = \frac{4}{3}\hat{g}^2(0)$
- $\hat{G} = \sum_{k=-M}^{M} \lambda(\frac{k}{M})|k|\hat{R}(k)$, where

$$\hat{R}(k) = N^{-1} \sum_{k=1}^{N-|k|} (X_i - \overline{X}_N)(X_{i+|k|} - \overline{X}_N)$$

$$\lambda(t) = \begin{cases} 1 & \text{if } |t| \in [0, 1/2] \\ 2(1-|t|) & \text{if } |t| \in [1/2, 1] \\ 0 & \text{otherwise} \end{cases}$$

$M = 2\hat{m}$, where $\hat{m}$ is the index, from where the correlogram "essentially" equals 0

Parametric bootstrap

- Till now, we have not used any models
- If we have a reliable model, it is worth using also in the bootstrap
- In the simplest case the samples are simulated from the fitted parametric model and the statistics is calculated for these
- When the sample size is small, it is often better than the nonparametric
- It is often used e.g. for linear models, when the residuals are simulated and these are added to the fitted values
- Selection may be based on the aim of the investigation
  - Model selection: nonparametric bootstrap
  - Model reliability: parametric bootstrap

A simple example for parametric bootstrap

- Question: can the shape parameter of the fitted gamma distribution be equal to 1?
- Bootstrap samples are taken from the exponential distribution (this is the $\Gamma(1, \lambda)$ distribution).
- Statistics: the ML estimator of the shape parameter for these samples
- Bootstrap *p*-value: the proportion of cases with estimators that were further away from 1 than the estimator for the observed case

AR-sieve bootstrap

- Condition: the process is stationary and estimable with an AR(p) model: $X_t - \mu_X = \sum_{j=1}^{p} \phi_j (X_{t-j} - \mu_X) + \varepsilon_t, \quad t \in \mathbb{Z}$, where $\mu_X = EX_t$, $(\varepsilon_t)_{t\in\mathbb{Z}}$ i.i.d., E(ε_t)=0 and ε_t is independent from $\{X_s; s < t\}$
- Estimation of parameters and errors:
  - $\hat{p}$=? $\longrightarrow$ AIC
  - $\hat{\mu}_X = n^{-1} \sum_{t=1}^{n} X_t$
  - $\hat{\phi}_1, \ldots, \hat{\phi}_{\hat{p}}$=? $\longrightarrow$ Yule-Walker method
  - $R_t = X_t - \sum_{j=1}^{\hat{p}} \hat{\phi}_j X_{t-j}$, where $t = \hat{p} + 1, \ldots, n$ from this we have $\hat{\varepsilon}_t = R_t - \overline{R}_t$, where $t = \hat{p} + 1, \ldots, n$
- The steps of constructing the bootstrap sample:
  - ε_t^*: a random element from the set $\{ \hat{\varepsilon}_{\hat{p}+1}, \ldots, \hat{\varepsilon}_n \}$
  - Let $(X^*_{-u}, \ldots, X^*_{-u+\hat{p}-1}) = (\hat{\mu}_X, \ldots, \hat{\mu}_X)$ (initial values, u is large)
  - $X_t^* = \hat{\mu}_X + \sum_{j=1}^{p} \hat{\phi}_j (X^*_{t-j} - \hat{\mu}_X) + \varepsilon_t^* \quad t \in \mathbb{Z}$
- The bootstrap sample: $\{ X_1^*, \ldots, X_n^* \}$

Weighted (wild) bootstrap

- Here we do not take a sample, but weigh the original sample (practically the likelihood function is weighted)
- Formally: $Z_i^{(k)}$ are the weights, $E(Z_i^{(k)}) = 0$ and $Var(Z_i^{(k)}) = 1$ where $i = 1, \ldots, n$, $k = 1, \ldots, N$ (N is the number of boostrap repetitions)
- In the classical case $\underline{Z}$ has multinomial distribution
- The first application was for the regression: $\hat{y}_i^* = \hat{y}_i + Z_i \varepsilon_i$
- It is worth to use in the heteroscedastic cases
- Further applications: goodness-of-fit for copulas (we come back to this approach later)

- Nonparametric bootstrap methods often underestimate the uncertainty
- Parametric bootstrap is used the most
- Here a more cautious approach is also shown: the median of the profile likelihood intervals for the bootstrap samples

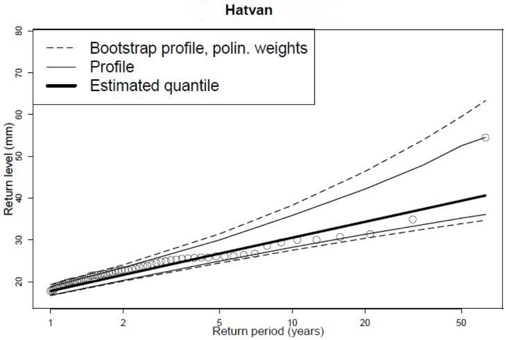

Figure: Different confidence intervals for the return level

The method of Hall and Weissman

- $m << n$ is needed in extreme-value models and at the same time we have to simplify the problem to the estimation of not so extreme quantiles
- Fine tuning is possible by parameters (s, t)
- The aim: $D_1(t, n, x) := E\left\{(F_{\hat{\theta}(t)}(x) - F(x))^2\right\} \to \min_t$
- If the $1 - p$-quantile is to be estimated, then it can be rewritten: $D_2(t, n, x) := D_1(t, n, \overline{F}^{-1}(p)) = E\left\{(\overline{F}_{\hat{\theta}(t)}(\overline{F}^{-1}(p)) - p)^2\right\} \to \min_t$
- The bootstrap estimators $\hat{D}_1(t, m, y) = E'\left\{\left(F_{\hat{\theta}^*(t)}(y) - \hat{F}(y)\right)^2\right\}$ and $\hat{D}_2(t, m, q) = E'\left\{\left(F_{\hat{\theta}^*(t)}\left(\hat{\overline{F}}^{-1}(q)\right) - q\right)^2\right\}$.
- One has to pay attention that the ratio $\log(x)/\log(n)$ should asymptotically not change when we use the transform from (n, x) to the pair (m, y).

Hivatkozások

- Efron, B. and Tibshirani, R.J.: An Introduction to the Bootstrap (1993)
- Kysely, J. : A cautionary note... (2008)
- Bickel, P.J., Götze, F. and van Zwet, W.R.: Resampling fewer than n observations (1997)
- Lahiri, S.N.: Resampling methods for dependent data (Springer, 2003)
- Bickel, P.J. and Sakov, A.: On the Choice of *m* in the *m* Out of *n* Bootstrap and its Application to Confidence Bounds for Extrema (2008)
- Politis, D. N. and White, H.: Automatic Block-Length Selection for the Dependent Bootstrap (2004)