Chapter 1 : Bootstrapping (statistics) - Wikipedia

*Bootstrap Methods And Their Application (Cambridge Series in Statistical and Probabilistic Mathematics) [A. C. Davison] on www.nxgvision.com *FREE* shipping on qualifying offers. This book gives a broad and up-to-date coverage of bootstrap methods, with numerous applied examples.*

Bootstrap statistics Bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample, most often with the purpose of deriving robust estimates of standard errors and confidence intervals of a population parameter like a mean , median , proportion , odds ratio , correlation coefficient or regression coefficient. It may also be used for constructing hypothesis tests. It is often used as a robust alternative to inference based on parametric assumptions when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors. The bootstrap allows to replace the samples with low weights by copies of the samples with high weights. Jackknife resampling Jackknifing, which is similar to bootstrapping, is used in statistical inference to estimate the bias and standard error variance of a statistic, when a random sample of observations is used to calculate it. Historically this method preceded the invention of the bootstrap with Quenouille inventing this method in and Tukey extending it in Quenouille invented this method with the intention of reducing the bias of the sample estimate. The basic idea behind the jackknife variance estimator lies in systematically recomputing the statistic estimate, leaving out one or more observations at a time from the sample set. From this new set of replicates of the statistic, an estimate for the bias and an estimate for the variance of the statistic can be calculated. Instead of using the jackknife to estimate the variance, it may instead be applied to the log of the variance. This transformation may result in better estimates particularly when the distribution of the variance itself may be non normal. For many statistical parameters the jackknife estimate of variance tends asymptotically to the true value almost surely. In technical terms one says that the jackknife estimate is consistent. The jackknife is consistent for the sample means , sample variances , central and non-central t-statistics with possibly non-normal populations , sample coefficient of variation , maximum likelihood estimators , least squares estimators, correlation coefficients and regression coefficients. It is not consistent for the sample median. In the case of a unimodal variate the ratio of the jackknife variance to the sample variance tends to be distributed as one half the square of a chi square distribution with two degrees of freedom. The jackknife, like the original bootstrap, is dependent on the independence of the data. Extensions of the jackknife to allow for dependence in the data have been proposed. Another extension is the delete-a-group method used in association with Poisson sampling. Comparison of bootstrap and jackknife[edit] Both methods, the bootstrap and the jackknife, estimate the variability of a statistic from the variability of that statistic between subsamples, rather than from parametric assumptions. For the more general jackknife, the delete-m observations jackknife, the bootstrap can be seen as a random approximation of it. Both yield similar numerical results, which is why each can be seen as approximation to the other. Although there are huge theoretical differences in their mathematical insights, the main practical difference for statistics users is that the bootstrap gives different results when repeated on the same data, whereas the jackknife gives exactly the same result each time. Because of this, the jackknife is popular when the estimates need to be verified several times before publishing e. On the other hand, when this verification feature is not crucial and it is of interest not to have a number but just an idea of its distribution, the bootstrap is preferred e. Whether to use the bootstrap or the jackknife may depend more on operational aspects than on statistical concerns of a survey. The jackknife, originally used for bias reduction, is more of a specialized method and only estimates the variance of the point estimator. This can be enough for basic statistical inference e. The bootstrap, on the other hand, first estimates the whole distribution of the point estimator and then computes the variance from that. While powerful and easy, this can become highly computer intensive. However, the bootstrap variance estimator is not as good as the jackknife or the balanced repeated replication BRR variance estimator in terms of the empirical results. Furthermore, the bootstrap variance estimator usually requires more computations than the jackknife or the BRR. Thus, the bootstrap is mainly recommended for distribution estimation. It should only be used with smooth, differentiable statistics e. This could become a practical disadvantage. This disadvantage is usually the argument favoring bootstrapping over jackknifing. More general jackknifes than the delete-1, such as the delete-m jackknife or the delete-all-but-2 Hodgesâ€”Lehmann estimator , overcome this problem for the medians and quantiles by relaxing the smoothness requirements for consistent variance estimation. Usually the jackknife is easier to apply to complex sampling schemes than the bootstrap. Complex sampling schemes may involve stratification, multiple stages clustering , varying sampling weights non-response adjustments, calibration, post-stratification and under unequal-probability sampling designs. Theoretical aspects of both the bootstrap and the jackknife can be found in Shao and Tu , [7] whereas a basic introduction is accounted in Wolter Subsampling statistics Subsampling is an alternative method for approximating the sampling distribution of an estimator. The two key differences to the bootstrap are: The advantage of subsampling is that it is valid under much weaker conditions compared to the bootstrap. In particular, a set of sufficient conditions is that the rate of convergence of the estimator is known and that the limiting distribution is continuous; in addition, the resample or subsample size must tend to infinity together with the sample size but at a smaller rate, so that their ratio converges to zero. While subsampling was originally proposed for the case of independent and identically distributed iid data only, the methodology has been extended to cover time series data as well; in this case, one resamples blocks of subsequent data rather than individual data points. There are many cases of applied interest where subsampling leads to valid inference whereas bootstrapping does not; for example, such cases include examples where the rate of convergence of the estimator is not the square root of the sample size or when the limiting distribution is non-normal. Cross-validation statistics Cross-validation is a statistical method for validating a predictive model. Subsets of the data are held out for use as validating sets; a model is fit to the remaining data a training set and used to predict for the validation set. Averaging the quality of the predictions across the validation sets yields an overall measure of prediction accuracy. Cross-validation is employed repeatedly in building decision trees. One form of cross-validation leaves out a single observation at a time; this is similar to the jackknife. Another, K-fold cross-validation, splits the data into K subsets; each is held out in turn as the validation set. For comparison, in regression analysis methods such as linear regression , each y value draws the regression line toward itself, making the prediction of that value appear more accurate than it really is. Cross-validation applied to linear regression predicts the y value for each observation without using that observation. This is often used for deciding how many predictor variables to use in regression. Without cross-validation, adding predictors always reduces the residual sum of squares or possibly leaves it unchanged. In contrast, the cross-validated mean-square error will tend to decrease if valuable predictors are added, but increase if worthless predictors are added. Exact test A permutation test also called a randomization test, re-randomization test, or an exact test is a type of statistical significance test in which the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points. In other words, the method by which treatments are allocated to subjects in an experimental design is mirrored in the analysis of that design. If the labels are exchangeable under the null hypothesis, then the resulting tests yield exact significance levels; see also exchangeability. Confidence intervals can then be derived from the tests. The theory has evolved from the works of Ronald Fisher and E. Pitman in the s. To illustrate the basic idea of a permutation test, suppose we have two groups A.

DOWNLOAD PDF BOOTSTRAP METHODS AND THEIR APPLICATIONS

Chapter 2 : [PDF/ePub Download] bootstrap methods and their application eBook

*Figure: Summary plots for nonparametric bootstrap simulations. Top: normal probability plots of t * and z * = (t * âˆ' t)/v * 1/2. Line on the top left has intercept t and slope v 1/2, line.*

Show Context Citation Context Second, for each prospective memory subdomain, prospective memory cue type, and study setting, robust statistical techniquesâ€" counts and sign testsâ€"were used to determine if the specific prospective Organizational Structure as a Determinant of Performance: Evidence from Mutual Funds by Felipe A. Csaszar - Strategic Management Journal " This article develops and tests a model of how organizational structure influences organizational performance. Organizational structure, conceptualized as the decision-making structure among a group of individuals, is shown to affect the number of initiatives pursued by organizations and the omissio Organizational structure, conceptualized as the decision-making structure among a group of individuals, is shown to affect the number of initiatives pursued by organizations and the omission and commission errors Type I and II errors, respectively made by organizations. The empirical setting is more than , stock-picking decisions made by mutual funds. Mutual funds offer an ideal and rare setting to test the theory, since there are detailed records on the projects they face, the decisions they make, and the outcomes of these decisions. The findings suggest that organizational structure has relevant and predictable effects on a wide range of organizations. In particular, the article shows empirically that increasing the consensus threshold required by a committee in charge of selecting projects leads to more omission errors, fewer commission errors, and fewer approved projects. Applications include designing organizations that achieve a given mix of exploration and exploitation, as well as predicting the consequences of centralization and decentralization. This work constitutes the first large-sample empirical test of the model by Sah and Stiglitz With the increased occurrence of outbreaks of H5N1 worldwide there is concern that the virus could enter commercial poultry farms with severe economic consequences. We analyse data from four recent outbreaks of highly pathogenic avian influenza HPAI in c We analyse data from four recent outbreaks of highly pathogenic avian influenza HPAI in commercial poultry to estimate the farm-to-farm reproductive number for HPAI. The reproductive number is a key measure of the transmissibility of HPAI at the farm level because it can be used to evaluate the effectiveness of the control measures. In these outbreaks the mean farm-to-farm reproductive number prior to controls ranged from 1. Enhanced bio-security, movement restrictions and prompt isolation of the infected farms in all four outbreaks substantially reduced the reproductive number, but it remained close to the threshold value 1 necessary to ensure the disease will be eradicated. Our results show that depending on the particular situation in which an outbreak of avian influenza occurs, current controls might not be enough to eradicate the disease, and therefore a close monitoring of the outbreak is required. The method we used for estimating the reproductive number is straightforward to implement and can be used in real-time. It therefore can be a useful tool to inform policy decisions. To obtain proper parametric bootstrap intervals would involve generating infection times and trees according to the underlying epidemic model, which we do not wish to specify completely. Stochastic simulation of the impact of antiretroviral therapy by Ronald H. Gray, Xianbin Li, Maria J. Gangeb, David Serwaddac, Nelson K. A stochastic simulation model estimated HIV incidence, probabilities of transmission per coital act and the reproductive number R0 with A Model inputs included Rakai data on HIV transmission probabilities per coital act by HIV viral load, age and gender, and sexual behaviors. Component projection models estimated the numbers of HIV-infected persons over 20 years. The model incidence [1. The variance of the parameters was uniformly low Appendix III. The total number of seroconversions, total person years PY at risk among HIV-negative persons and total number of sexual acts were Structural equation models SEM are very popular in many disciplines. The par-tial least squares PLS approach to SEM offers an alternative to covariance based SEM, which is especially suited for situations when data is not normally distributed. PLS path modelling is referred to as softâ€"modelingâ€"technique with minimum demands regarding measurement scales, sample sizes and residual distributions. Different setups for the estimation of factor scores can be used. Furthermore it contains modular methods for computation of bootstrap confidence intervals, model parameters and several quality indices. Various plot functions help to evaluate the model. The well known mobile phone dataset from marketing research is used to demonstrate the features of the package.

Chapter 3 : Resampling (statistics) - Wikipedia

Bootstrap Methods and their Application (Cambridge Series in Statistical and Probabilistic Mathematics) - Kindle edition by A. C. Davison, D. V. Hinkley. Download it once and read it on your Kindle device, PC, phones or tablets.

History[edit] The bootstrap was published by Bradley Efron in "Bootstrap methods: As the population is unknown, the true error in a sample statistic against its population value is unknown. As an example, assume we are interested in the average or mean height of people worldwide. We cannot measure all the people in the global population, so instead we sample only a tiny part of it, and measure that. Assume the sample is of size N; that is, we measure the heights of N individuals. From that single sample, only one estimate of the mean can be obtained. In order to reason about the population, we need some sense of the variability of the mean that we have computed. The bootstrap sample is taken from the original by using sampling with replacement e. This process is repeated a large number of times typically 1, or 10, times , and for each of these bootstrap samples we compute its mean each of these are called bootstrap estimates. We now can create a histogram of bootstrap means. This histogram provides an estimate of the shape of the distribution of the sample mean from which we can answer questions about how much the mean varies across samples. The method here, described for the mean, can be applied to almost any other statistic or estimator. Discussion[edit] This section includes a list of references , related reading or external links , but its sources remain unclear because it lacks inline citations. Please help to improve this section by introducing more precise citations. June Advantages[edit] A great advantage of bootstrap is its simplicity. It is a straightforward way to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution, such as percentile points, proportions, odds ratio, and correlation coefficients. Bootstrap is also an appropriate way to control and check the stability of the results. Although for most problems it is impossible to know the true confidence interval, bootstrap is asymptotically more accurate than the standard intervals obtained using sample variance and assumptions of normality. The apparent simplicity may conceal the fact that important assumptions are being made when undertaking the bootstrap analysis e. Recommendations[edit] The number of bootstrap samples recommended in literature has increased as available computing power has increased. If the results may have substantial real-world consequences, then one should use as many samples as is reasonable, given available computing power and time. Increasing the number of samples cannot increase the amount of information in the original data; it can only reduce the effects of random sampling errors which can arise from a bootstrap procedure itself. Moreover, there is evidence that numbers of samples greater than lead to negligible improvements in the estimation of standard errors. Since the bootstrapping procedure is distribution-independent it provides an indirect method to assess the properties of the distribution underlying the sample and the parameters of interest that are derived from this distribution. When the sample size is insufficient for straightforward statistical inference. If the underlying distribution is well-known, bootstrapping provides a way to account for the distortions caused by the specific sample that may not be fully representative of the population. When power calculations have to be performed, and a small pilot sample is available. Most power and sample size calculations are heavily dependent on the standard deviation of the statistic of interest. If the estimate used is incorrect, the required sample size will also be wrong. One method to get an impression of the variation of the statistic is to use a small pilot sample and perform bootstrapping on it to get impression of the variance. However, Athreya has shown [20] that if one performs a naive bootstrap on the sample mean when the underlying population lacks a finite variance for example, a power law distribution , then the bootstrap distribution will not converge to the same limit as the sample mean. As a result, confidence intervals on the basis of a Monte Carlo simulation of the bootstrap could be misleading. Athreya states that "Unless one is reasonably sure that the underlying distribution is not heavy tailed , one should hesitate to use the naive bootstrap". Types of bootstrap scheme[edit] This section includes a list of references , related reading or external links , but its sources remain unclear because it lacks inline citations. June Learn how and when to remove this template message In univariate problems, it is usually acceptable to resample the individual observations with replacement "case resampling" below unlike subsampling , in which resampling is without replacement and is valid under much weaker conditions compared to the bootstrap. In small samples, a parametric bootstrap approach might be preferred. For other problems, a smooth bootstrap will likely be preferred. For regression problems, various other alternatives are available. Bootstrap comes in handy when there is no analytical form or normal theory to help estimate the distribution of the statistics of interest, since bootstrap methods can apply to most random quantities, e. There are at least two ways of performing case resampling. The Monte Carlo algorithm for case resampling is quite simple. First, we resample the data with replacement, and the size of the resample must be equal to the size of the original data set. Then the statistic of interest is computed from the resample from the first step. We repeat this routine many times to get a more precise estimate of the Bootstrap distribution of the statistic. This can be computationally expensive as there are a total of.

DOWNLOAD PDF BOOTSTRAP METHODS AND THEIR APPLICATIONS

Chapter 4 : Bootstrap Methods and Their Application - A. C. Davison, D. V. Hinkley - Google Books

This book gives a broad and up-to-date coverage of bootstrap methods, with numerous applied examples, developed in a coherent way with the necessary theoretical basis.

There present variety of motives behind it due to which the readers stop reading the eBooks at their first most effort to make use of them. However, there exist some techniques that could help the readers to really have a nice and effective reading experience. Someone ought to adjust the suitable brightness of display before reading the eBook. As a result of this they have problems with eye sores and headaches. The very best option to overcome this severe difficulty would be to decrease the brightness of the screens of eBook by making specific changes in the settings. You may also adjust the brightness of screen depending on the type of system you are utilizing as there exists lot of the ways to correct the brightness. It is suggested to keep the brightness to possible minimum level as this will help you to increase the time which you can spend in reading and provide you great comfort onto your eyes while reading. An excellent eBook reader should be set up. It will be useful to have a good eBook reader to be able to truly have a good reading experience and high quality eBook display. You may also make use of free software that could offer the readers that have many functions to the reader than only a simple platform to read the wanted eBooks. You can even save all your eBooks in the library that is also supplied to the user by the software program and have a superb display of all your eBooks as well as access them by identifying them from their specific cover. Aside from offering a place to save all your precious eBooks, the eBook reader software even give you a high number of features to be able to boost your eBook reading experience in relation to the conventional paper books. You can also improve your eBook reading experience with help of options provided by the software program such as the font size, full display mode, the certain number of pages that need to be displayed at once and also alter the colour of the background. You must not make use of the eBook consistently for several hours without breaks. You must take appropriate breaks after specific intervals while reading. However, this will not mean that you need to step away from the computer screen every now and then. Continuous reading your eBook on the computer screen for a long time without taking any rest can cause you headache, cause your neck pain and suffer with eye sores and also cause night blindness. So, it is important to give your eyes rest for a while by taking breaks after particular time intervals. This can help you to prevent the problems that otherwise you may face while reading an eBook constantly. While reading the eBooks, you should prefer to read big text. So, raise the size of the text of the eBook while reading it on the monitor. Although this will definitely mean you will have less text on each page and greater number of page turning, you will be able to read your desired eBook with great convenience and have a great reading experience with better eBook display. It is suggested that never use eBook reader in full screen mode. It is suggested not to go for reading the eBook in full-screen mode. Even though it might look easy to read with full screen without turning the page of the eBook quite often, it put lot of stress in your eyes while reading in this mode. Constantly favor to read the eBook in the exact same span that will be similar to the printed book. This really is so, because your eyes are used to the span of the printed book and it would be comfy for you to read in exactly the same manner. Test out different shapes or sizes until you find one with which you will be comfortable to read eBook. By using different techniques of page turn you can also enhance your eBook encounter. Check out whether you can turn the page with some arrow keys or click a certain part of the screen, apart from using the mouse to manage everything. Try using the mouse if you are comfy sitting back. Lesser the movement you need to make while reading the eBook better is going to be your reading experience. This will definitely definitely help to make reading easier. By using all these powerful techniques, you can surely improve your eBook reading experience to a terrific extent. These tips will help you not only to prevent particular hazards which you may face while reading eBook consistently but also ease you to relish the reading experience with great relaxation.

Chapter 5 : The Statistical Bootstrap and Other Resampling Methods - Burns Statistics

Bootstrap methods are computer-intensive methods of statistical analysis, which use simulation to calculate standard errors, confidence intervals, and significance tests. The methods apply for any level of modelling, and so can be used for fully parametric, semiparametric, and completely nonparametric analysis.

Links Preliminaries The purpose of this document is to introduce the statistical bootstrap and related techniques in order to encourage their use in practice. The examples work in R â€" see Impatient R for an introduction to using R. However, you need not be a user to follow the discussion. On the other hand, R is arguably the best environment in which to perform these techniques. A tab separated file of the data is available at: The first few rows look like: If you want to be able to reproduce them exactly, there are two choices: You can save the. You can use the set. The Bootstrap The idea: We have just one dataset. The bootstrap creates a large number of datasets that we might have seen and computes the statistic on each of these datasets. Thus we get a distribution of the statistic. Our example data are log returns also known as continuously compounded returns. The log return for the year is the sum of the daily log returns. We can use the bootstrap to get an idea of the variability of that figure. There are daily returns in the year. One bootstrap sample is randomly sampled daily returns. The sampling is with replacement, so some of the days will be in the bootstrap sample multiple times and other days will not appear at all. Once we have a bootstrap sample, we perform the calculation of interest on it â€" in this case the sum of the values. Below is some simple code to perform this bootstrap with bootstrap samples. The command says to sample from the integers from 1 to , make the sample size and sample with replacement. The effect is that this. In the subsequent line we collect the annual return from each of the hypothetical years. We can then plot the distribution of the bootstrapped annual returns. It could have easily been anything from 0 to The actual annual return is squarely in the middle of the distribution. Bootstrapping smooths More than numbers can be bootstrapped. In the following example we bootstrap a smooth function over time. There are periods of low volatility, and periods of high volatility. Since the bootstrapping does not preserve time ordering, the bootstrap samples will not have volatility clustering. We will return to volatility clustering later. In finance this number is called the beta of the stock. This is often thought of as a fixed number for each stock. In fact betas change continuously, but we will ignore that complication here. The result of the call to lm is an object representing the linear regression. We then extract the coefficients from that object. To get just the beta, we can subscript for the second element of the coefficients: Basically, a number of hypothetical years are created. Bootstrap regression coefficients version 2 There is another approach to bootstrapping the regression coefficient. The response in a regression is identically equal to the fit of the regression plus the residuals the regression line plus distance to the line. We can take the viewpoint that only the residuals are random. Sample from the residuals of the regression on the original data, and then create synthetic response data by adding the bootstrapped residuals to the fitted value. The explanatory variables are still the original data. For each day we will create a new IBM return by adding the fit of the regression for that day to the residual from some day. This second method is performed below. An experiment with 50, bootstrap samples showed the two bootstrap densities to be almost identical. There are times, though, when the residual method is forced upon us. If we are modeling volatility clustering, then sampling the observations will not work â€" that destroys the phenomenon we are trying to study. We need to fit our model of volatility clustering and then sample from the residuals of that model. R Software For data exploration the techniques that have just been presented are likely to be sufficient. If you are using R, S-PLUS or a few other languages, then there is no need for any specialized software â€" you can just write a simple loop. Often it is a good exercise to decide how to bootstrap your data. You are not likely to understand your data unless you know how to mimic its variability with a bootstrap. If formal inference is sought, then there are some technical tricks, such as bias corrected confidence intervals, that are often desirable. Specialized software does generally make sense in this case. There are a number of R packages that are either confined to or touch upon bootstrapping or its relatives. This package incorporates quite a wide variety of bootstrapping tricks. A package of relatively simple functions for bootstrapping and related techniques. A package for permutation tests which are discussed below. This package is for Monte Carlo hypothesis tests, that is, tests using some form of resampling. This includes code for sampling rules where the number of samples taken depend on how certain the result is. Provides a method of bootstrapping a time series. A package containing a function for permutation tests of microarray data. This package produces approximately unbiased hypothesis tests via bootstrapping. A package of a few functions that perform or present bootstraps in simple situations, such as one and two samples, and linear regression. There are a large number of R packages that include bootstrapping. Examples include multtest that has the boot. The BurStMisc package has some simple functions for permutation tests. The Bootstrap More Formally Bootstrapping is an alternative to the traditional statistical technique of assuming a particular probability distribution. For example, it would be reasonably common practice to assume that our return data are normally distributed. This is clearly not the case. However, there is decidedly no consensus on what distribution would be believable. Bootstrapping outflanks this discussion by letting the data speak for itself. As long as there are more than a few observations the data will reveal their distribution to a reasonable extent. One way of describing bootstrapping is that it is sampling from the empirical distribution of the data. This produces observations that are close to a specific data observation rather than exactly equal to the data observation. If zero were given for sd, then it would be exactly the standard bootstrap sample. Some statistics are quite sensitive to tied values which are inherent in bootstrap samples. Smoothed bootstrapping can be an improvement over the standard bootstrap for such statistics. The usual assumption to make about data that are being bootstrapped is that the observations are independent and identically distributed. If this is not the case, then the bootstrap can be misleading. If you consider just location, returns are close to independent. However, independence is definitely shattered by volatility clustering. It is probably easiest to think in terms of predictability. The predictability of the returns is close to but not exactly zero. There is quite a lot of predictability to the squared returns though. The amount that the bootstrap is distorted by predictability of the returns is infinitesimal. Distortion due to volatility clustering could be appreciable, though unlikely to be overwhelming. There are a number of books that discuss bootstrapping. Here are a few: An Introduction to the Bootstrap by Efron and Tibshirani. Efron is the inventor of the bootstrap. Bootstrap Methods and their Application by Davison and Hinkley. Permutation Tests The idea: Permutation tests are restricted to the case where the null hypothesis really is null â€" that is, that there is no effect. If changing the order of the data destroys the effect whatever it is , then a random permutation test can be done. The test checks if the statistic with the actual data is unusual relative to the distribution of the statistic for permuted data. The test does a regression with the squared returns as the response and some number of lags most recent previous data of the squared returns as explanatory variables.

Chapter 6 : DiCiccio , Efron : Bootstrap confidence intervals

Bootstrap Methods and their Application / Edition 1 This book gives a broad and up-to-date coverage of bootstrap methods, with numerous applied examples, developed in a coherent way with the necessary theoretical basis.

But why not replace them altogether with more informative tools? The bootstrap affords a unique opportunity for obtaining a large amount of information very simply. The process of setting confidence intervals merely picks two points off a bootstrap histogram, ignoring much relevant information about shape and other important features. Graphics such as these provide a simple but powerful way to convey information lost in numerical summaries. The opportunities offered by dy namic graphics are also attractive, particularly when confidence information needs to be passed to a lay audience. Bootstrap methods and new graphical way s of presenting information offer, together, exciting prospects for conveying information about uncertainty. The horizontal dotted lines are quantiles of t t for all bootstrap replicates, while the solid lines join the corresponding quantiles for the subsets of bootstrap replicates in which each of the 20 observations did not appear. The x-axis shows empirical influence values lj which measure the effect on t of putting more mass on each of the observations separately. If F represents the empirical distribution function of the data, which puts mass n-1 on each of the observations y1 yn and t F is the corresponding statistic, we can write of Davison and Hinkley We have developed a library of bootstrap functions in S-PLUS which facilitates this ty pe of analysis. The library may be obtained by anony mous ftp to markov. ABC, use bootstrap calibration directly on the crude percentile-based procedures these methods refine, and which seem currently favored in published applications of the bootstrap, as any literature search confirms. In doing so, we retain the desirable properties of these basic procedures stability of length and endpoints, invariance under parametrization etc. The price is one of great computational expense, although, as is demonstrated by Lee and Young , there are approximations which can bring such bootstrap iteration within the reach of even a modest computational budget. An advantage of this solution lies in its simplicity: Which solution is best? To answer this requires a careful analysis of what we believe the bootstrap methodology to be. Our view is that willingness to use extensive computation to extract information from a data sample, by simulation or resampling, is quite fundamental. In other words, in comparing different methods, computational expense should not be a factor. All things being equal, we naturally look for computational efficiency, but things are hardly ever equal. How do the two solutions, that provided by DiCiccio and Efron and that involving the iterated percentile bootstrap, compare? There are two concerns here, theoretical performance and empirical performance, and the two might conflict. We demonstrate by considering the simple problem of constructing a two-sided nonparametric bootstrap confidence interval for a scalar population mean. The calibration method of Loh corresponds to the method of Beran when applied to a bootstrap confidence interval. For the confidence interval problem the method of Hall amounts to making an additive adjustment, estimated by the bootstrap, to the endpoints of the confidence interval, while the method of Beran amounts to making an additive adjustment, again estimated by bootstrapping, to the nominal coverage level of the bootstrap interval. The method of calibration described by DiCiccio and Efron in Section 7 of their paper is a subtle variation on the latter procedure, and one which should be used with care. DiCiccio and Efron use a method in which the bootstrap is used to calibrate separately the nominal levels of the lower and upper limits of the interval, rather than the overall nominal level. Theoretical and empirical evidence which we shall present elsewhere leads to the conclusion that, all things being taken into consideration, preference should be shown to methods which adjust nominal coverage, rather than the interval endpoints. We shall therefore focus on the question of how to calibrate the nominal coverage of a bootstrap confidence interval. With the percentile interval, the coverage error, of order n-1, of the coverage-corrected one-sided interval ty pically involves an odd poly nomial, and terms of that order will not cancel when determining the coverage error of the two-sided interval, which remains of order n On the face of it, therefore, we should be wary of the calibration method described by DiCiccio and Efron, although the problems with it do not arise with the ABC interval. Application of these methods to the intervals under consideration here allows closer examination of coverage error. What is immediately obvious from the table is that the order of coverage error only tells part of the story. Compare the coefficients of n-1 for the interval IPITb with the coefficients of n-2 for the other iterated intervals. However, if we focus on those intervals that ensure a coverage error of order n-2, it appears that the two ty pes of iterated ABC interval are not significantly different, but that the iterated percentile interval has a leading error term consistently and significantly smaller than that of the ABC method. This same conclusion is true for any nominal coverage in the range 0. Saddlepoint approximation for the Studentized mean, with an application to the bootstrap.

DOWNLOAD PDF BOOTSTRAP METHODS AND THEIR APPLICATIONS

Chapter 7 : CiteSeerX â€" Citation Query Hinkley DV: Bootstrap methods and their applications

Bootstrap Methods and Their Application (Cambridge Series in Statistical and Probabilistic Mathematics, No 1) by A. C. Davison, D. V. Hinkley and a great selection of similar Used, New and Collectible Books available now at www.nxgvision.com

Transformation of the dependent cost variable is often used to solve the problems of heteroscedasticity and skewness in linear ordinary least square regression of health service cost data. However, transformation may cause difficulties in the interpretation of regression coefficients and However, transformation may cause difficulties in the interpretation of regression coefficients and the retransformation of predicted values. Aims of the Study: The study compares the advantages and disadvantages of different methods to estimate regression based cost functions using data on the annual costs of schizophrenia treatment. Annual costs of psychiatric service use and clinical and socio-demographic characteristics of the patients were assessed for a sample of patients with a diagnosis of schizophrenia ICD F The clinical characteristics of the participants were assessed by means of the BPRS 4. Diabetes Care by Richard J. We investigated whether the risk of MI or stroke being fatal in type 2 diabetes can be estimate We investigated whether the risk of MI or stroke being fatal in type 2 diabetes can be estimated using information available around the time diabetes is diagnosed. Multivariate logistic regres-sion was used to examine differences in risk factors, measured within 2 years of diagnosis of diabetes, between fatal and nonfatal MI. Similar analyses were performed for strokes 48 fatal that occurred in patients. Patients with fatal stroke had higher HbA1c than those with nonfatal stroke odds ratio 1. Other risk factors for MI case fatality included increased age, blood pressure, and urine albumin level. Details are available from the corresponding author on request. A simple method for finding molecular signatures from gene expression data. Spanish National Cancer Center. Signatures are a key feature in cancer research because they can provide insight into biological Signatures are a key feature in cancer research because they can provide insight into biological mechanisms and have potential diagnostic use. However, available methods to search for signatures fail to address key requirements of signatures and signature components, especially the discovery of tightly coexpressed sets of genes. Results We suggest a method with good predictive performance that follows from the biologically relevant features of signatures. After identifying a seed gene with good predictive abilities, we search for a group of genes that is highly correlated with the seed gene, shows tight coexpression, and has good predictive abilities; this set of genes is reduced to a signature component using using Principal Components Analysis. The process is repeated until no further component is found. We show that the suggested method can recover signatures present in the data, and has predictive performance comparable to state-of-the-art methods. Conclusions Our method is unique because it returns signature components that fulfill what are understood as biologically relevant features of signatures. Moreover, it can help identify cases where the data are inconsistent with the assumptions underlying the existence of a few, easily interpretable, signature components of coexpressed genes. However, this only assesses the effect of different random 26partitions in steps 1 and 5b. Further work in this area is in progress. If the model built from signatures is to be used for prediction, this raises the issue of w

Chapter 8 : Bootstrap Methods and Their Application by A.C. Davison

Preface The publication in of Bradley Efron s rst article on b o otstrap metho ds w as a ma jor ev en t in Statistics at once syn thesizing some of the earlier resam.

Chapter 9 : CiteSeerX â€" Citation Query Bootstrap Methods and Their Applications

Find helpful customer reviews and review ratings for Bootstrap Methods and their Application (Cambridge Series in Statistical and Probabilistic Mathematics) at www.nxgvision.com Read honest and unbiased product reviews from our users.