Построение сценариев принятия решений на основе будстреп-моделирования

Лилия Ю.Уразаева
Санкт-Петербургский архитектурно-строительный университет
Санкт-Петербург, Россия
Delovoi2004@mail.com

Наталья Н. Дацун
Пермский государственный национальный исследовательский университет
Пермь, Россия
nndatsun@inbox.ru

***Аннотация*. Работа посвящена обоснованию возможности использования будстреп-методов при компьютерном моделировании сценариев развития динамических процессов в сложных системах. Обсуждается анализ границ применимости будстреп-анализа, предлагается процедура адаптивной генерации выборок с учетом развития динамических процессов, приведены примеры использования будстреп-моделирования для принятия решений.**

Ключевые слова: принятие решений; будстреп-моделирование; анализ рисков; построение сценариев развития процессов; сложные системы

I. Введение

Проектирование сложных систем, работающих в условиях неопределенности требует управления, основанного на многоступенчатых процедурах принятия решений.

Характерной особенностью таких систем управления является необходимость приспособления к изменяющимся условиям и возможность оценки и учета влияния неопределенностей.

Для выбора оптимального управления можно разработать сценариев принятия решений на основе будстреп-моделирования

В системном анализе принято выделять два вида адаптивного управления: прямое (непосредственное) и косвенное (опосредованное).

В случае прямого адаптивного управления внешняя информация непосредственно используются в процессе принятия решений для изменения траектории управления.

Опосредованные методы позволяют провести оценку диапазона изменения входных и выходных параметров с целью использования полученных результатов для настройки системы принятия решений при управлении сложной системой.

В настоящее время имеется большое число публикаций посвящено рассмотрению применению классических методов адаптивного управления в теоретических исследованиях и в практических приложениях.

Особое внимание привлекают задачи разработки процедур управления, в условиях наличия изменяющихся во времени неопределенностей, имеющих нелинейные эффекты воздействия на результат.

Ввиду неполноты информации при решении подобных задач авторами было предложено использование будстрtп-анализа для получения необходимых оценок в условиях малочисленных выборок при построении сценариев развития системы.

Для получения полной картины о возможностях применения будстреп-методов при принятии решений и оценке рисков был выполнен обзор научных работ, посвященных различным аспектам приложения методов размножения выборки и аппроксимации данных.

На основе рассмотренных публикаций можно отметить, что управление сложными системами с большим числом нелинейных взаимосвязей и неопределенностью может также проводиться с помощью оценки рисков состояния устойчивости системы при различных сценариях.

Таким образом, адаптивное управление находит применение в различных реальных приложениях. Проблема недостатка данных для оценки рисков принятия обоснованных решений решается методами будстреп-анализа и аппроксимации зависимостей.

II. Описание метода и анализ работ

A. Описание будстреп-метода

Бутстреп-метод был разработан в 1979 г. Б. Эфроном как продолжение метода складного ножа. Будстреп-метод используется в случае малочисленности выборочных совокупностей для получения оценок параметров генеральной совокупности.

В качестве искомых оценок параметров закона распределения случайной величины в генеральной совокупности могут выступать выборочная средняя, выборочный коэффициент корреляции и т.д.

Очевидно, выборочные оценки являются случайными величинами и могут меняться от выборки к выборке. В реальных задачах объем выборки может быть небольшим по объективным причинам, а именно, из-за сложности или трудоемкости измерений, дороговизны эксперимента и т.д.

Для случая малых выборок и был предложен метод будстреп-анализа, который состоит в размножении выборки(ресамплировании).

Принцип ресамплирования состоит в многократном случайном извлечении данных из выборки малого объема с целью получения выборки большого объема для проведения дальнейших расчетов.

Исходная выборка размножается за счет мнимых выборочных данных, полученных из исходной выборки малого объема.

Особенность генерации будстреп-выборки состоит в случайном выборе элементов выборки с возвращением. При генерации выборки без возвращения выбранного элемента элементов имеем метод складного ножа.

По каждой созданной выборки определяется оценка параметра генеральной совокупности, затем происходит усреднение полученных оценок параметров распределения, в итоге оценки усредняются.

Будстреп-метод можно рассматривать как реализацию метода Монте-Карло в компьютерном эксперименте.

В случае динамических процессов, возникающих в задачах адаптивного управления, может оказаться недостаточно данных для проведения прогнозирования, некоторые необходимые данные могут быть пропущены или могут среди них могут присутствовать аномальные наблюдения.

С помощью будстреп-метода можно построить эмпирический закон распределения, проверить гипотезу о виде распределения исследуемой случайной величины. Если известные законы распределения не подходят, то можно приближенно восстановить вид функции распределения.

При моделировании поведения динамических процессов используют аппроксимации, построенные с помощью рядов Грама-Шарлье и другие методы [1–30].

При использовании рядов Грама-Шарлье коэффициенты ряда рассчитываются по статистическим моментам изучаемого параметра, в результате приближенного характера вычислений могут наблюдаться определенные искажения, противоречащие смыслу и свойствам восстанавливаемого закона распределения.

Таким образом, бутстреп позволяет получить скорректированные на смещение оценки статистических моментов случайной величины, а аппроксимирующие кривые Грама-Шарлье, – восстановить неизвестную функцию распределения случайной величины, используя полученные при помощи бутстрепа-метода значения.

Совместное применение статистических методов позволяет построить неизвестную функцию плотности распределения и оценить имеющие место риски.

B. Постановка задачии модификация метода

Пусть требуется оценить сценарии прогноза некоторого экономического показателя по имеющимся данным за предыдущие моменты времени. Будем считать, что показатель имеет теоретическое предельное максимальное значение и теоретическое предельное минимальное значение. В разные моменты времени показатель принимает значения между максимальным и минимальным значениями. Пусть известны отдельные значения показателя в отдельные моменты времени на временном отрезке [t_1,t_n], заданы таблично в табл. 1.

ТАБЛИЦА I ИСХОДНЫЕ ДАННЫЕ

t_i	y_i
t_1	y_1
t_2	y_2
….	…
t_n	y_n

Требуется с помощью будстреп-моделирования построить сценарии развития некоторого показателя с учетом ограниченности значений показателя, и текущего поведения показателя.

Будем рассматривать наряду с рядом исходных значений показателя, ряд полученный вычитанием заданных значений показателя от максимального значения, заданы таблично в табл. 2.

Если рассматривать значения исходного показателя как прибыль, то полученный временной ряд можно рассматривать как упущенную прибыль в контексте нашего примера о прибыли.

ТАБЛИЦА II ВТОРОЙ РЯД

t_i	$w_i=y_{max}-y_i$
t_1	w_1
t_2	w_2
….	…
t_n	w_n

Прогноз будем строить на основе ресамплирования с учетом коэффициента доверия к прошлым данным, отношения размаха колебаний между текущими данными к предельному размаху колебаний.

Ресамплирование будем проводить для первого исходного временного ряда и второго временного ряда (полученного вычитанием заданных значений из предельных) ряда, затем с помощью моделируемого в ресамплинге коэффициента доверия (значение от 0 до 1) будем использовать данные исходного ряда, а данные полученного второго ряда с коэффициентом, равным единице минус коэффициент доверия.

Отношение текущего размаха колебаний к предельному размаху колебаний будет использовано для построения возможных сценариев развития процесса.

Для прогноза можно на основе многократного ресамплирования построить следующие сценарии:

1) сценарий, определяемый как реализация потенциала, то есть максимального размаха случайной величины с учетом случайности, то есть умноженного на коэффициент доверия потенциала, множитель, представляющий собой случайное число от 0 до 1;

2) сценарий, определяемый как реализация случайной величины, равной фактическому размаху с учетом случайности, умноженного на коэффициент доверия к фактическому размаху, случайный множитель, изменяющийся от 0 до 1, с равномерным распределением;

3) сценарий, определяемый как сумма результатов сценариев 1 и 2, со случайными множителями, имеющими равномерное распределение.

На основе исходных полученных сценариев можно осуществить прогноз развития исследуемого показателя.

III. Результаты расчетов и их анализ

На основе предложенного алгоритма проведено численное моделирование с помощью бесплатного средства эконометрического моделирования Гретл.

Для численного эксперимента можно также составить свою программу для ремсамплинга, или использовать другие математические пакеты или даже табличных процессор. Возможности Gretl достаточно широки, программный продукт постоянно обновляется. Для проведения расчетов можно использовать также бесплатный пакет статистических расчетов и графики R, математический пакет Octave www.octave.org: Octave - это бесплатный аналог известного пакета Matlab. Удобен для обработки результатов и проведения исследований Julia(julialang.org).

Для расчетов в пакете Gretl были сформированы новые переменные, представляющие линейные комбинации фактических значений наблюдений и отклонений фактических наблюдений от возможного максимального значения со случайными коэффициентами, изменяющимися от 0 до 1, имеющими равномерное распределение.

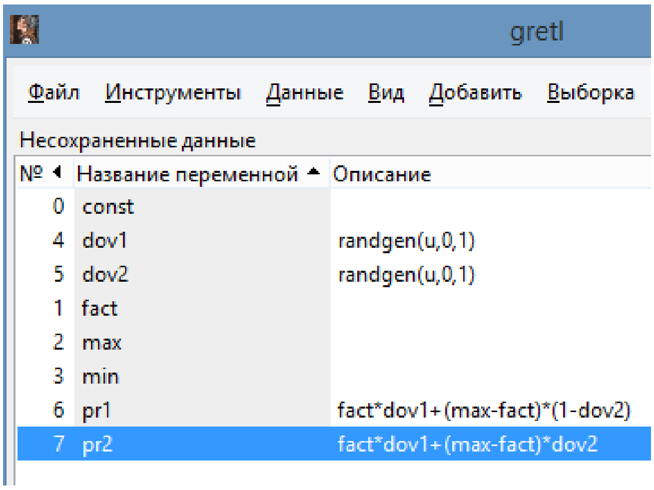

Рис. 1. Окно со списком переменных для ресамплирования

При большем объеме результатов ресамплирования можно получить результаты различных сценариев развития процесса. Для выбора конкретных сценариев для проведения анализа и прогнозирования надо исходить их анализа конкретной предметной области т исследуемого процесса.

Результаты численного эксперимента показывают, что данная методика дает определенный сдвиг прогнозных данных по отношению к исходным данным.

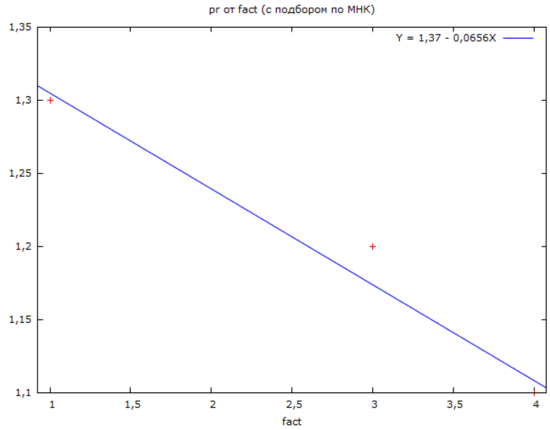

Рис. 2. Взаимосвязь между фактическими и прогнозными переменными после ресамлирования k=50

В случае реальных данных этот сдвиг возможен, так как прогнозные значения не обязательно должны совпадать с фактическими значениями.

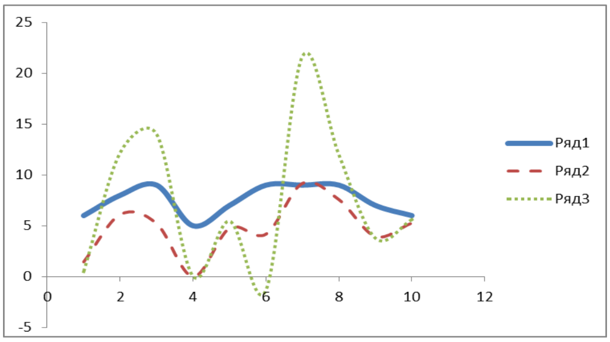

Рис. 3. График сценариев(сплошная линия факт), использовались различные комбинации фактических значений со случайными коэффициентами доверия(экспорт в Excel)

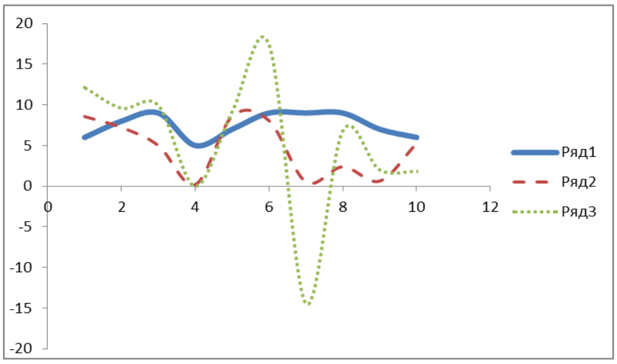

Рис. 4. График сценариев(сплошная линия факт), при других значения случайных коэффициентов доверия(экспорт в Excel)

IV. Выводы

Ресамплирование можно применять для прогнозирования, но при этом необходимо учитывать конкретные особенности исследуемых временных рядов.

Полученные при ресамплировании сценарии можно использовать как альтернативные пути развития исследуемого процесса.

Список литературы

Athreya, K.B. (1986). Bootstrap of the mean in the infinite variance case. Ann. Stat. 14, 724-731.

Azzalini, A. and Hall, P. (2000). Reducing variability using bootstrap methods with quantitative constraints. Biometrika, 87, 895-906.

Babu, G.J. (1984). Bootstrapping statistics with linear combination of Chi-square as weak limit. Sankhya A. 46, 85-93.

Babu, G.J. and Singh, K. (1983). Inference on means using the bootstrap. Ann. Stat. 11, 999-1003.

Beran, R. (1984). Prepivoting to reduce level errors of confidence sets. Biometrika. 74, 151-173. Beran, R. (1990) Refining bootstrap simultaneous confidence sets. Jour. Amer. Stat. Assoc. 85, 417-428.

Bickel, P.J. (2003). Unorthodox bootstraps (invited papers). J. of Korean Stat. Soc. 32, 213-224. Bickel, P.J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. Ann. Stat. 9, 1196- 1217.

Bickel, P.J. and Freedman, D (1984). Asymptotic normality and the bootstrap in stratified sampling. Ann. Stat. 12, 470-482.

Boos, D.D. and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances. Technometics. 31, 69-82.

Boos, D.D. and Munahan, J.F. (1986). Bootstrap methods using prior information. Biometrika. 73, 77-83.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. Ann. Stat. 16, 1709-1722. Breiman, L. (1996). Bagging predictors. Machine Learning, 26, 123-140.

Buhlmann, P. (1994). Bootstrap empirical process for stationary sequences. Ann. Stat. 22, 995- 1012.

Buhlmann, P. (2002). Sieve bootstrap with variable length – Markov chains for stationary categorical Time series (with discussions) Jour. Amer. Stat. Assoc. 97, 443-455.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in Cox model. Jour. Amer. Stat. Assoc. 89, 1290-1302.

Collins, M.A., Millard-Stafford, M.L., Sparling, P.B., Snow, T.K., Rosskopf, L.B., Webb, S.A., Omer, J. (1999). Evaluating BOD POD(R) for assessing body fat in collegiate football players. Medicine and Science in Sports and Exercise. 31,1350-56.

Davison, A.C. and Hinkley, D. V. (1988). Saddle point approximations in resampling method. Biometrika. 75, 417-431. DiCiccio, T.J. and Romano, J.P. (1988). A review of bootstrap confidence intervals (with discussions). J. R. Stat. Soc. B. 50, 538-554.

Eaton, M.L. and Tyler, D.E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. Ann. Stat. 19, 260-271.

Efron, B. (1979). Bootstrap methods: Another look at jackknife. Ann. Stat. 7, 1-26. Efron, B. (1987). Better bootstrap confidence intervals (with discussions). Jour. Amer. Stat. Assoc. 82, 171-200.

Efron, B. (1992). Jackknife-after-bootstrap standard errors and influences functions (with discussions). J.R. Stat. Soc. B. 54, 83-127.

Efron, B. (1994). Missing data, imputation and the bootstrap (with discussions). Jour. Amer. Stat. Assoc. 89, 463-479.

Efron, B. and Tibshirani, R.J. (1993). An introduction to the bootstrap, Chapman and Hall New York.

Freedman, D.A. (1981) Bootstrapping Regression models. Ann. Stat. 9, 1281-1228.

Hall, P. (1989). On efficient bootstrap simulation. Boimetrika. 76, 613-617. Hall, P. (1992). Bootstrap confidence intervals in nonparametric regression. Ann. Stat. 20, 695- 711.

Hinkley, D.V. (1988). Bootstrap methods (with discussions). J. Roy. Stat. Soc. B, 50, 321-337.

Kunch, H.R. (1989). The jackknife and bootstrap for general stationary observations. Ann. Stat. 17, 1217-1241.

Lahiri, S.N. (1993). Bootstrapping the studentized sample mean of Lattice variables. J. Mult. Analy. 45, 247-256.

Lahiri, S.N. (1993). On the moving block bootstrap under long range dependence. Stat. Prob. Letters. 18, 405-413.

Liu, R.Y. and Singh, K. (1992). Efficiency and Robustness in re sampling. Ann. Stat. 20, 370- 384. Liu, R.Y. and Singh, K. (1992). Moving block jackknife and bootstrap capture weak dependence. EXPLORING THE LIMITS OF BOOTSTRAP, R. Lepage and L. Billard edited. Wiley, N.Y.

Lunneborg, EE. (2000). Data analysis by resampling: concepts and apllications. Duxbury press. Mamman, e. (1992). When does bootstrap work. Asymptotoc results and simulations. Springer Verlag, N.Y. Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap. Jour. Amer. Stat. Assoc. 89, 1303-1313.

Rubin, D.B. (1981). The Bayesian bootstrap. Ann. Stat. 9, 130-134. Shao, J. and Tu, D. (1995). THE JACKKNIFE AND BOOTSTRAP, Springer, Verlag, N.Y. Singh, K. (1981). On Asymptotic accuracy of Efron’s bootstrap. Ann. Stat. 9, 1187-1195.

Singh, K (1998). Breakdown theory for bootstrap quantiles. Ann. Stat. 26, 1719-1732. Singh, K. and Xie M. (2003). Bootlier-plot-Bootstrap based outlier detection plot. Sankhya, 65, 532-559.