Integration of Statistical Techniques in the Design Curriculum

Farinaz Koushanfar
Electrical and Computer Engineering Dept.
Rice University
farinaz@rice.edu

Miodrag Potkonjak
Computer Science Dept.
University of California, Los Angeles
miodrag@cs.ucla.edu

Abstract

Traditionally, design and analysis techniques have been using abstractions and models that are well captured using probabilistic models and are treated using standard algorithmic techniques. However, technology progress and application diversification, created a need for considering the impact of non-gaussian and non-uniform uncertainty distributions that can only be addressed using non-parametric statistical techniques and robust algorithms. This trend provided an impetus for creation of a new graduate level course that aims to teach students sound statistical techniques, to enable them to obtain hands-on experience using statistical software, and to develop and apply the most effective ways to address design and manufacturing uncertainties. We summarize the syllabus of the course that integrates statistical techniques in the design curriculum, identifying the major obstacles and suggesting future improvements.

1   Rationale

Modeling of complex large-scale deep submicron designs, multimedia applications, wireless low power distributed systems with lossy links and many other microelectronics devices and systems require accurate statistical models. Although, the current norm is to employ simple probabilistic models that facilitate fast and sometimes even closed-form analysis, it is clear that they are adequate tools for the emerging complex systems. The large-scale integrated systems are subject to complex uncertainties that require application of non-parametric statistical techniques.

Parametric statistical techniques are the ones that assume the form of the functional relation between the modeled parameters and the observed data is known a priori. The objective is to find the parameters of the functional dependency that minimize a certain error norm. For example, a parametric model may assume that $y = ax + b$, where y is the modeled variable, x is observed variable and the goal is to find parameters a and b. On the other hand nonparametric statistical models (we use the terms nonparametric and data-driven interchangeably) are statistical models that are built upon actual data while they pose either very mild or no assumptions on the relationship among the parameters of the functional relations. For example, it may be assumed that y is a monotonic function of x.

The goal of the course is to prepare students to use parametric models, to nurture the importance of valid model assumptions, to use the power of non-parametric models, to select the best predictors, to build a system of interacting models, and to apply model evaluations and validations.

An alternative approach for learning the parametric and nonparametric modeling techniques would be to direct students to the courses in the statistics departments. We practiced this approach for a number of years, with a very limited success. Many students had troubles following the material due to the lack of background and knowledge. Others complained that they do not see a connection between their research and what is presented in the classes and that the courses emphasize mathematics at the expense of ignoring the algorithmic and practical skills.

2   Examples

Here, we describe three emerging applications of statistical modeling in modern designs:

1. Process variations. Industrial miniaturization of CMOS devices has been driven by the quest for increasing computational speed and device density, while lowering cost-per-function, as predicted by Moore's law [3, 4]. CMOS feature sizes have decreased to the point that the device parameters are highly variable and complex. Recent International Technology Roadmap For Semiconductors (ITRS) has identified Silicon complexity and its associated non-recurring engineering (NRE) costs as major challenges to continuation of the semiconductor roadmap [2]. For example, the magnitude of variations in the gate length is predicted to increase from 35% in a 130nm technology to 60% in a 70nm technology [5]. Traditional device models are inefficient. There

is a need for more complex statistical device models in circuit simulating tools such as SPICE [7].

2. Run time. The paradigm shift in device models renders the traditional CAD tools that are based on deterministic models insufficient. These tools typically use corner-case models that assume worst-case values for process parameters. CMOS variations result in high variability in the delay and power consumption of the VLSI circuits. There is a need for statistical CAD-tools that model the performance (e.g., timing and delay) as a statistical distribution rather than a deterministic quantity.

3. Wire length prediction. One of our covered topics is *a priori prediction of wire lengths* that we address by novel nonparametric statistical methods. We first collect a large amount of data from a benchmark set using different placement parameters within three different layout tools from Cadence and UCLA. Next, the students are introduced to a system of rigorous statistical modeling and validation that establishes quantitative bounds on the accuracy of a priori wire length predictions. For each design, we produced a large number of competitive implementations by varying the placement parameters. The data is used for development of non-parametric data-driven statistical models for individual nets, total wire lengths and the distribution of wire lengths in terms of their quantitative predictability. The models can guide the designers to select the most suitable parameters for a given design and given tools.

3   Pedagogical challenges

Our main dilemma was finding a suitable course organization. One option is to first present the statistical methods and then their engineering applications. The other is to discuss design and engineering problems and demonstrate how different statistical techniques can help. After some experimentation, we settled on a middle road solution where in one third of a semester (quarter) we present statistical techniques while keep illustrating their application of engineering and design problems, and consequently go through a number of selected design engineering problems and show how the illustrated statistical techniques can ensure much better solutions. We prepared data sets from several engineering problems (e.g., layout placement, gate-delay variability, scheduling, power grid) and gave homework sets to students that included using of the R statistical package [6]. Our goal was to acquire the students with hands-on experience with the introduced tools and techniques.

4   Syllabus

In our latest offering, the content and time allocation was that approximately 40% of time was spent on non-parametric statistics; 10% on each of statistical timing analysis, manufacturing variability, lossy wireless links, and variable load on WWW servers and general purpose computing; and 20% on new and emerging topics that require statistical treatment including nanoelectronics, nanophotonics, biotechnology, and MEMS/NEMS systems. In three different offerings so far at UCLA, Rice Univ., and UIUC[1] we have covered in technical details various subsets of the following general key topics [1]: (i) parametric and nonparametric regression; (ii) hypothesis testing, model selection (reduction) and bias-variance trade-offs; (iii) linear and nonlinear classification including principal component analysis (PCA) along with model assessment; (iv) support vector machines; (iv) kernel methods; (v) independent component analysis (ICA); (vi) model inference including maximum likelihood (ML), Bayesian inference, and expectation maximization (EM); (vii) additive models, generalized additive models; (viii) Classification and regression trees (CART); and (ix) additive trees and Boosting.

5   Conclusion

We introduced a new graduate-level course on application of statistical techniques for design curriculum. The emphasis was on practical and accurate non-parametric statistical techniques and hand-on skill development using the R statistical package on actual data from several design problems. The course has received a warm acceptance from graduate students at three different universities.

References

[1] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.

[2] International Technology Roadmap for Semiconductors (ITRS). http://www.itrs.net/Links/2005ITRS/Home2005.htm.

[3] G. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):114–117, 1965.

[4] S. Roy and A. Asenov. Where do the dopants go? *Science*, 309(5733):388–390, 2005.

[5] A. Srivastava, D. Sylvester, and D. Blaauw. *Statistical Analysis and Optimization for VLSI: Timing and Power*. Series on ICs and Systems. Springer, 2005.

[6] The R Project for Statistical Computing. http://www.r-project.org/.

[7] The Spice Page at UC Berkeley. http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE.

---

[1]Prof. Koushanfar was a visiting scholar at UIUC, 2005-06.

COMPUTER SOCIETY