Dose-Response Signal Estimation and Optimization for Salesforce Management

Kush R. Varshney and Moninder Singh
Business Analytics and Mathematical Sciences
IBM Thomas J. Watson Research Center
1101 Kitchawan Road, Yorktown Heights, NY, USA

***Abstract*—Estimating generalizable relationships between actions and results from historical samples, especially when there is a level of noise or randomness in that signal, is an important problem to address before making decisions on actions to take. Many business analytics problems require the optimal assignment of limited resources to actions and activities to maximize some result or objective such as profit. We present a novel approach to solving this class of analytics problems by modeling the relationship between resource effort and expected return as a dose-response signal and formulating its causal estimation as one of kernel regression. The estimated expected value and variance of the result are then used to optimize resource allocation so as to maximize expected response while minimizing the risk around response subject to business constraints. We apply this approach to the task of optimally assigning salespeople to enterprise clients using real-world data, and show that profit can be substantially increased with fewer salespeople and reduced risk.**

I. Introduction

In running a business, there are many decisions to be made and actions to be taken with unknown and uncertain outcomes. Should the business outsource its information technology operations, and what effect will there be on key financial indicators such as earnings or expenses [1], [2]? Should the business offer salary increases to valued employees, and what effect will there be on the voluntary attrition rate [3]? Should pre-sales activities such as proposal and contract preparation be performed by salespeople or by a specialized group, and what effect will there be on the fraction of sales opportunities that are won [4]? What fraction of sales teams attending to clients should have technical job roles and what fraction non-technical job roles, and what effect will there be on revenue earned [5]? Business analytics is the science of using data, signals, and predictive modeling rather than gut instinct to drive such decisions.

The terminology of treatment doses and measured responses is often used in health and medicine, but applies equally well to any other application domain in which actions are taken and the effects of the action are to be inferred, including in business applications [5]. In these applications, understanding, and especially estimating quantitative causal relationships from observational studies is paramount [6], [7]. This inference involves taking historical samples of dose-response pairs and finding a mapping (if any) between the dose and response that holds in general.

There are two distinct types of treatments: discrete (including binary) and continuous [8], [9]. In the business analytics examples above, outsourcing and having pre-sales activities performed by a specialized group are both discrete treatments. The amount of salary increase and the fraction of the sales team having a technical job role are both continuous treatment doses. In this paper, we focus on continuous treatment doses and on the estimation of dose-response signals, i.e. estimating the response as a function of the continuous dose along with the uncertainty surrounding the function.

Furthermore, once a dose-response signal has been estimated with its inherent uncertainty, a portfolio optimization of doses may be performed to yield optimal responses subject to risk constraints and application-specific constraints [10]. Hedging bets and balancing risk and reward are common themes. Investors construct diversified stock portfolios to maximize expected return while minimizing the variance of the return [11]–[14]. Venture capitalists fund a collection of startups in the same way. Even sponsors sign multiple top athletes as brand ambassadors to hedge against the inevitable failure of some of them [15].

However, in typical portfolio settings, the action of investing or sponsoring is not a treatment dose that has a causal relationship with the response. Moreover, beyond a simple budget constraint, there typically are not additional application-specific constraints in investing. On the other hand, in health, medicine, and other dose-response settings, portfolio thinking is not typically employed because poor or even catastrophic outcomes for some patients balanced out by very good outcomes for other patients is not desirable. In this paper, we examine the class of problems in which actions have direct influence on results (with associated uncertainty) in a continuous dose-response signal relationship, and in which portfolio optimization of results is desired to maximize return subject to risk and under potentially complicated resource allocation constraints. This class of problems is under-studied in the literature because of the separate contexts and applications in which the two subproblems of estimation and optimization have arisen. We put forth this new joint machine learning and optimization paradigm motivated by business analytics problems and develop techniques towards its solution.

In the remainder of the paper, we focus on the specific salesforce analytics application described as follows. Large enterprises must decide how to best deploy their salesforce amongst their client base [16]. This is one of the most important questions when running a large business because salespeople bring in the revenue that the business needs to operate and profit. Specifically, we would like to determine the amount of sales effort that should be allocated to each client to maximize total expected profit while minimizing the variance of the profit. The most interesting aspect of the problem is the inherent uncertainty and randomness in the relationship between the sales effort expended on a client and the corresponding profit earned, but also the possibility of there being a signal within that noise.

We look at this relationship as a dose-response signal [8], [9] and formulate its estimation as one of kernel regression. Both the expected value of the dose-response function and the variance are learned from historical data using the Nadaraya–Watson formulation [17]–[19], and used to optimize the sales effort allocated to each client so as to have large expected profit but also small variance. Further constraints are also incorporated in the optimization, such as requiring the total allocated effort to be less than the size of the salesforce. We discuss the difficulty of carrying out this optimization and describe how it can be approximated as a knapsack optimization problem and approximately solved.

Before setting out on this endeavor, it was not clear whether there even was a pattern to be recognized between (suitably normalized) sales effort and profit. As we show in empirical results on corporate data, a nice relationship does exist and salesforce deployment optimization based on it can significantly increase profits without high risk. The practical contribution of the research is in translating a mathematically ill-defined business question into a statistical signal processing formulation and putting all the components together into a unified predictive and prescriptive salesforce analytics solution.

Sales resource allocation is a problem that has been studied in the management literature for a long time, including considering risk, but the majority of that research has been focused on the optimization aspect and has not considered continuous dose-response functions [20]. The closest related work in this application domain is from the marketing literature and is quite simplistic, limited to linear regression for estimation and heuristics for optimization, does not consider risk, and does not provide optimal effort allocations at an individual client level [21]. Moreover, as discussed earlier, the combination of dose-response signal inference and portfolio-based risk-sensitive optimization is a novel combination that has not been considered in the information processing and inference literature before, neither in business applications nor other applications.

The remainder of the paper is organized as follows. In Section II, we detail the salesforce analytics problem setup, including notation and available data. Section III describes the kernel regression formulation. Section IV describes the formulation for portfolio optimization. Empirical results on two sets of real-world enterprise data showing the efficacy of the approach are presented in Section V. Section VI summarizes the contributions and provides directions for future research.

II. Problem Description

In this section, we describe the problem specifically in the context of allocating the effort of sales teams to clients for concreteness, while maintaining that the same general mathematical problem arises in several other business problems. Each sales resource expends a certain amount of effort selling to each client. This can be measured in full-time equivalents (FTEs) where 1 FTE equals 100% effort of 1 seller over a 1 year period. Thus, a seller dedicating 2 months solely to a single client expends $1/6$ FTEs on that client. If, however, the seller distributes her effort equally across two clients for 2 months, then the effort for each client is $1/12$ FTEs. Often, teams of salespeople cover clients; the total effort spent on a client is the sum of the effort spent by each salesperson.

Let the effort spent by the salesforce on client i be denoted $e_i \in [0, \infty)$, $i = 1, \dots, n$, where n is the number of clients. Also, let $p_i \in \mathbb{R}$ be the profit from client i, where profit is revenue from the client minus costs and expenses. We would like to find a dose-response relationship between sales effort and profit; however, we must consider one additional aspect to the problem first. Not all clients are the same; some are larger and some are smaller. Without even considering the details of the sales effort, it is known beforehand that clients will produce revenue in a certain range due to various reasons, such as client size, prior client-company relationship, and market growth. This prior knowledge of revenue size, known as aspirational revenue, denoted $a_i \in [0, \infty)$, is determined by combining knowledge from subject matter experts with estimation techniques [22]. The response that we are really interested in is the ratio of profit to aspirational revenue $y_i = p_i/a_i$, capturing the so-called lift, or performance above or below aspiration, provided by the salesforce deployment.

In the treatment dose, the sales effort, we also take aspirational revenue into account. Asking how many FTEs per dollar of aspirational revenue to allocate to a client is a more meaningful question than simply asking how many FTEs. This normalization by the size of the client account allows us to compare small and large clients on an even basis. In real corporate data, the ratio e_i/a_i spans several orders of magnitude and thus we take its logarithm for improved estimation, giving a dose variable $d_i = \log_{10}(e_i/a_i)$. With normalized effort as the dose, we can apply the estimated dose-response signal to make predictions on a new client with a given aspirational revenue. In Section III we discuss how to estimate the dose-response function $\hat{y}(d)$.

III. Estimation Formulation

Given historical samples of effort spent, profit earned, and revenue aspired, for different clients of the company, we can construct training samples $\{(d_1, y_1), \dots, (d_n, y_n)\}$. The estimator we use in estimating the dose-response function from samples is the Nadaraya–Watson estimator, a form of kernel regression. The functional form of the estimate is:

$$\hat{y}(d) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{d_i - d}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{d_i - d}{h}\right)}, \tag{1}$$

where $K(\cdot)$ is a kernel function, such as a Gaussian kernel or Epanechnikov kernel, and h is a bandwidth. The bandwidth can be set using standard methods from the literature including plug-in rules of thumb [19].

The inherent uncertainty in the dose-response relationship is captured by estimating the variance surrounding $\hat{y}(d)$ [19]:

$$\hat{\sigma}^2(d) = \frac{\sum_{i=1}^{n} y_i^2 K\left(\frac{d_i - d}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{d_i - d}{h}\right)} - \hat{y}(d)^2. \qquad (2)$$

Note that this variance is of the phenomenon itself, not of estimation error. Given $\hat{y}(d)$ and $\hat{\sigma}^2(d)$, optimization can be applied to a client set to determine the best doses.

IV. Optimization Formulation

The estimated dose-response function and its variance can now be applied to optimize doses for a new portfolio of m clients with aspirational revenues $\{a_1, \dots, a_m\}$. These clients are assumed to be similar to those from which the relationships were learned; one such scenario involves learning based on the previous year's efforts and profits, and optimizing effort allocations using the current year's aspirational revenues.

The optimization problem we pose is as follows.

$$\begin{aligned} \text{maximize} \quad & \sum_{j=1}^{m} a_j \hat{y}(d_j) - \lambda \sum_{j=1}^{m} a_j^2 \hat{\sigma}^2(d_j) \\ \text{subject to} \quad & \sum_{j=1}^{m} a_j 10^{d_j} \leq S, \end{aligned} \qquad (3)$$

where S is the total number of salespeople in the salesforce. The objective balances risk and reward using the parameter λ. The value of λ is related to an upper bound for the variance, which can be seen by viewing the objective as a Lagrangian form. The constraint ensures that total allocated effort is less than the salesforce size because it is not feasible to allocate more effort than what can be handled by the existing salespeople.

The kernel regression estimates $\hat{y}(d)$ and $\hat{\sigma}^2(d)$ have no convexity or concavity guarantees, making the m-dimensional constrained optimization problem (3) difficult to carry out in practice. To approximate (3), we can do one-dimensional unconstrained optimizations for each of the m clients, finding d_j^* to maximize $a_j \hat{y}(d_j) - \lambda_1 a_j^2 \hat{\sigma}^2(d_j)$. Then, if constraint (3) is not binding, i.e. the sum of the $a_j 10^{d_j^*}$ values is less than or equal to S, we have the solution to the original problem.

If the constraint is binding, then we may express the problem as a knapsack optimization problem in which we decide not to sell to some of the clients [23]. This zero-one knapsack optimization problem is:

$$\begin{aligned} \text{maximize} \quad & \sum_{j=1}^{m} x_j a_j \hat{y}(d_j^*) - \lambda \sum_{j=1}^{m} x_j a_j^2 \hat{\sigma}^2(d_j^*) \\ \text{subject to} \quad & \sum_{j=1}^{m} x_j a_j 10^{d_j} \leq S, \quad x_j \in \{0, 1\}, \end{aligned} \qquad (4)$$

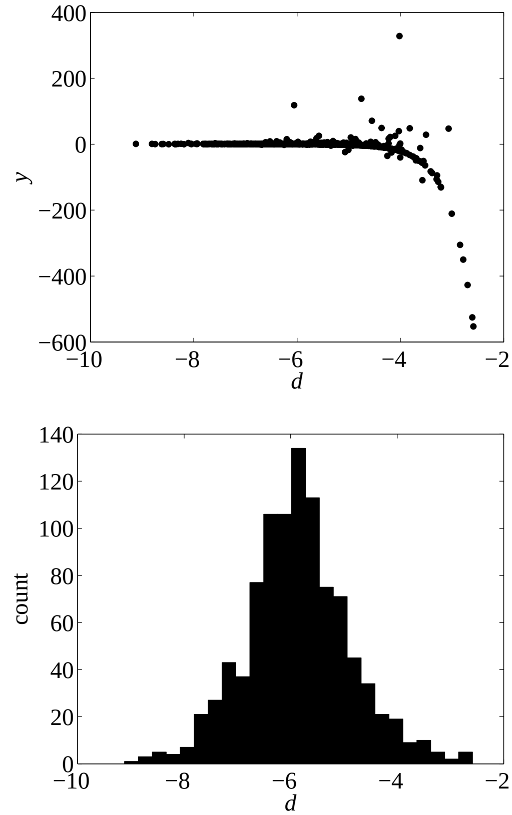

Fig. 1. Historical data for business unit A.

where the x_j are the decision variables by which we decide not to sell to clients. There are several ways to approximately solve problem (4) efficiently [23]. Thus overall, we have proposed a new formulation to find the optimal doses that balance expected return and the variance of the return for a new portfolio based on historical dose-response samples. We apply this formulation to real-world data from the salesforce management problem in the next section.

V. Empirical Results on Real-World Data

In this section, we present empirical results for subsets of clients from two different business units of a large multinational technology company using revenue and sales effort over a one year period.

In historical data, business unit A had $n = 980$ clients, had total sales effort $S = 364.69$ FTEs, and achieved $\$4.33 \times 10^8$ in total profit from all clients. The (d_i, y_i) samples are shown in Fig. 1 along with a histogram of the doses. Using (1) and (2) on the data shown in Fig. 1, we obtain $\hat{y}(d)$ and $\hat{\sigma}^2(d)$. The estimate of the dose-response relationship is plotted in Fig. 2 surrounded by the estimated inherent uncertainty. The estimated dose-response curve shows a nice, clear pattern. The first interesting thing to note is that the company lost money, i.e. $\hat{y} < 0$, when lots of sales effort per dollar of aspirational revenue ($d \gtrsim -5$) was put in. This makes business sense because overstaffing implies high expenses,

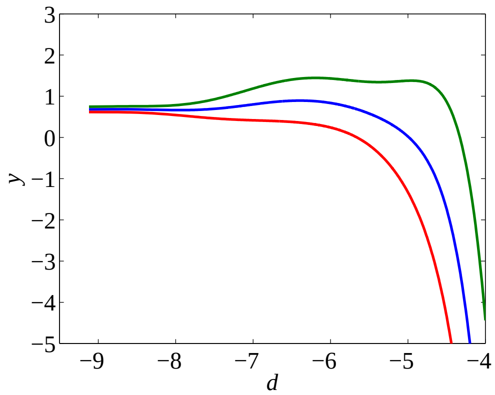

Fig. 2. Dose-response estimates for business unit A ($\hat{y}(d) \pm \hat{\sigma}(d)/10$).

resulting in an overall loss. The second interesting thing to note is that there is an intermediate response-maximizing dose: $d = -6.4$ or $10^{-6.4}$ FTEs per dollar of aspirational revenue. The third interesting thing to note in the learned relationships is that variance tends to increase with increasing d; specifically, $d = -6.4$ has much higher variance than some of the smaller values of d.

Fig. 3 shows the results of the optimization. At $\lambda = 0$, all clients are assigned at the peak value of $10^{-6.4}$ FTEs per dollar of aspirational revenue. This dose solution results in an expected total profit of $\$4.51 \times 10^8$ using 203.46 FTEs. The S constraint is not binding in this case. The total profit variance for this solution is 3.44×10^{17}. In the $\lambda = 0$ case, the company is able to achieve $\$1.84 \times 10^7$ more in profit using 161.23 fewer FTEs of effort in expectation by optimizing the deployment of the sales resources. As more importance is given to controlling risk with increasing λ, the total variance drops quite a bit. For example, at $\lambda = 5 \times 10^{-9}$, the total variance is only 6.94×10^{15} while still having an expected total profit of $\$3.65 \times 10^8$. The optimal dose distribution (Fig. 3) has mass mostly at the peak of $\hat{y}(d)$, but with some smaller doses as well.

Business unit B had $n = 453$ clients, sales effort $S = 982.11$, and total profit $\$4.48 \times 10^8$. Its dose-response data samples and dose histogram are shown in Fig. 4. The kernel regression estimate for this business (Fig. 5) exhibits the same general shape as the one for business unit A. This time the maximum of $\hat{y}(d)$ is at -6.1. However, if the corresponding number of FTEs is allocated to all clients, then the total effort across all clients will be 4782.08 FTEs, which violates the S constraint. Since the constraint is binding for this business, we must apply the knapsack optimization (for small values of λ). Applying Dantzig's greedy heuristic to solve the knapsack component, we obtain the solution paths seen in Fig. 6. The historical profit can be quadrupled through an optimal allocation. Looking at the optimal allocation for $\lambda = 5 \times 10^{-9}$, the total expected profit is $\$1.61 \times 10^9$ with a variance of 3.47×10^{17} and requires 702.79 fewer FTEs to achieve. Both examples, but especially the one for business unit B, show how significant an improvement can be made in a company's profit

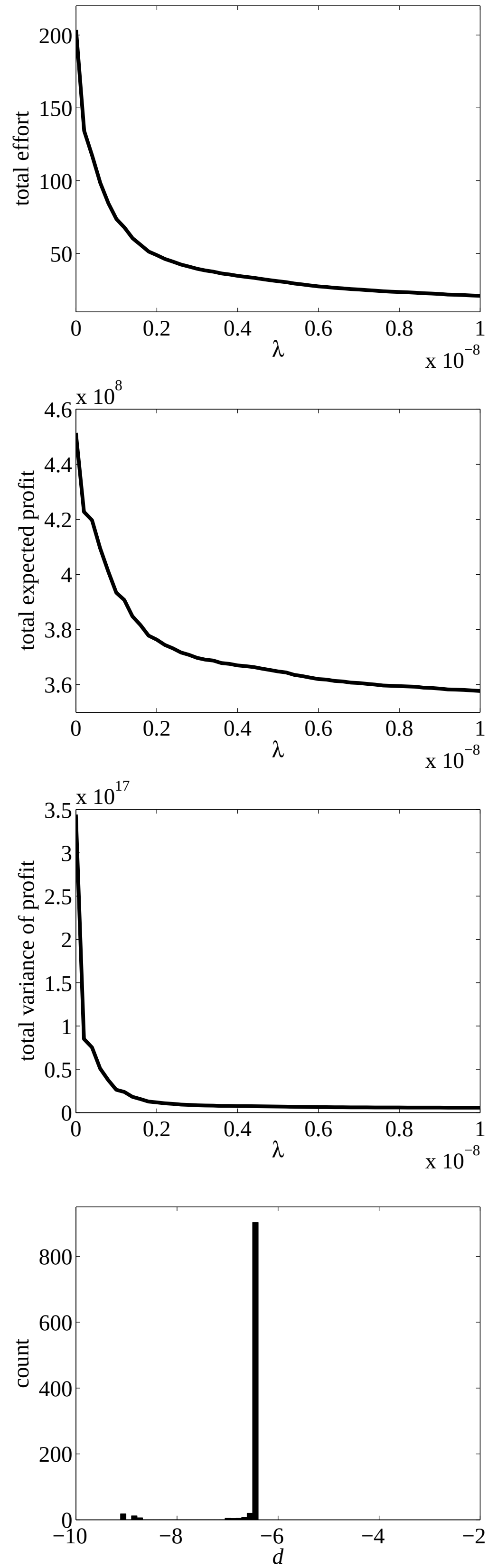

Fig. 3. Total effort, expected profit and variance of profit from optimal doses as a function of λ, and the optimal doses for $\lambda = 5 \times 10^{-9}$ for business unit A.

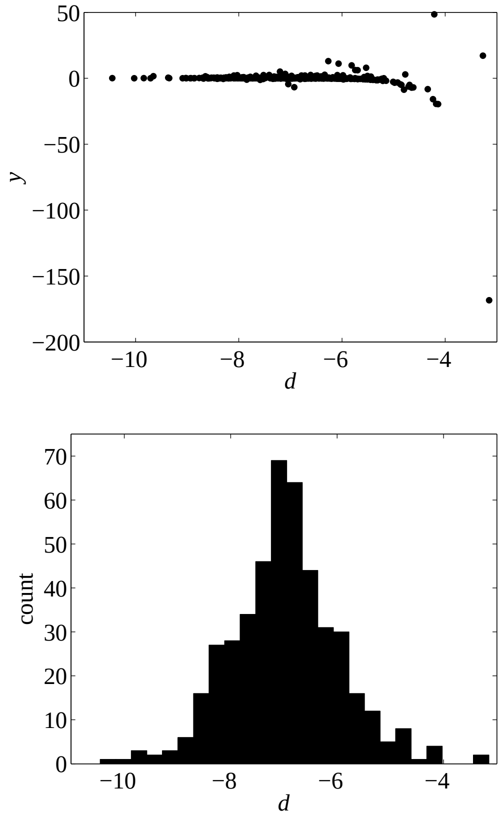

Fig. 4. Historical data for business unit B.

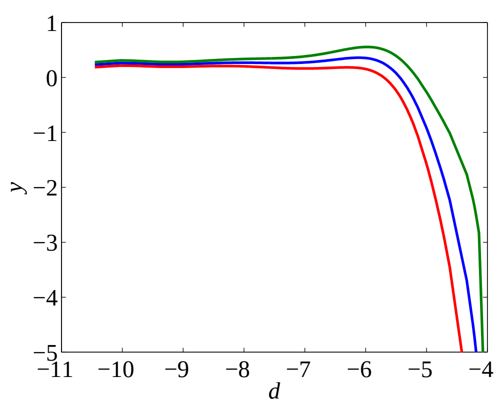

Fig. 5. Dose-response estimates for business unit B ($\hat{y}(d) \pm \hat{\sigma}(d)/10$).

by using the proposed approach considering dose-response and portfolio-based optimal salesforce allocations.

VI. Conclusion

We have presented a novel approach to the fairly general problem of optimally assigning resources to business activities having uncertain but partly predictable outcomes with the aim of maximizing returns while minimizing the associated risk. Motivated by the application of salesforce management, we formulate this problem as one of estimating a dose-

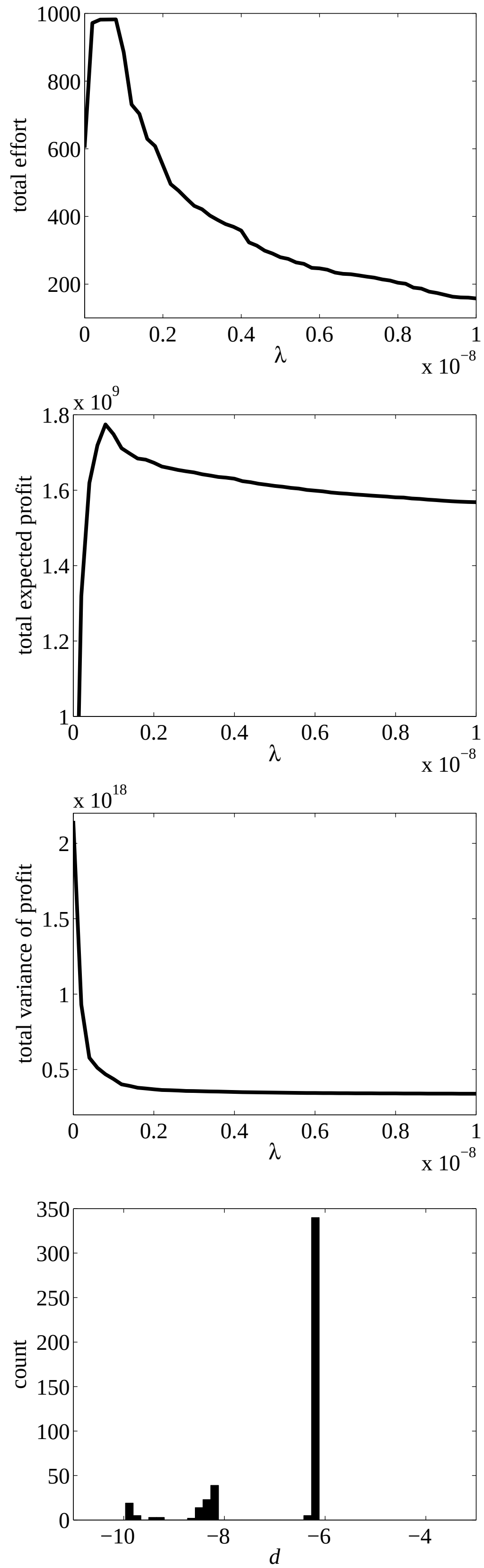

Fig. 6. Total effort, expected profit and variance of profit from optimal doses as a function of λ, and the optimal doses for $\lambda = 5 \times 10^{-9}$ for business unit B.

response function followed by resource allocation optimization using the estimated value and variance functions. Empirical results on two real-world datasets from the salesforce analytics domain show that substantial increases in profits from a set of clients can be achieved by using less sales effort and with lower risk by allocating sellers via this approach.

Going forward, we are exploring several extensions. First, we are investigating applying the analytics methodology proposed in this paper to other business problems besides sales effort assignment to clients for profitability, including a few mentioned in the first paragraph of Section I. Second, it is often infeasible to optimize over the entire client/seller space as it may entail expensive and logistically difficult wholesale movement of sellers. As such, we are incorporating more constraints in the optimization, such as disallowing a seller from covering clients in multiple geographic regions. Such an extension still fits within the multidimensional knapsack formulation. Other such application-specific constraints arise in other business problems as well.

Third, different types of sellers are skilled in different tasks and, thus, the composition of the sales team assigned to a client plays an integral part in determining the profit attained. The effect of sales team composition on the profit can be similarly modeled as a multi-dimensional dose-response function estimation problem. Multi-dimensional continuous treatment doses have not been studied in the literature. Another avenue of further research is to examine Gaussian process regression for the signal inference rather than kernel regression, which may be more amenable to multi-dimensional signal estimation and also allow the injection of prior knowledge [24]. Longer term planning horizons may also be considered in the optimization.

References

[1] D. A. Katz-Rogozhnikov, K. R. Varshney, A. Mojsilović, and M. Singh, "MCMC inference of the shape and variability of time-response signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Prague, Czech Republic, May 2011, pp. 3680–3683.

[2] K. R. Varshney and A. Mojsilović, "Business analytics based on financial time series," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 83–93, Sep. 2011.

[3] M. Singh, K. R. Varshney, J. Wang, A. Mojsilović, A. R. Gill, P. I. Faur, and R. Ezry, "An analytics approach for proactively combating voluntary attrition of employees," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Brussels, Belgium, Dec. 2012, pp. 317–323.

[4] J. Wang, M. Singh, and K. R. Varshney, "Does selection bias blind performance diagnostics of business decision models? A case study in salesforce optimization," in *Proc. IEEE Int. Conf. Serv. Oper. Logist. Inform.*, Suzhou, China, Jul. 2012, pp. 416–421.

[5] K. R. Varshney, M. Singh, and J. Wang, "Decision trees for heterogeneous dose-response signal analysis," in *Proc. IEEE Statist. Signal Process. Workshop*, Ann Arbor, MI, Aug. 2012, pp. 904–907.

[6] P. W. Holland, "Statistics and causal inference," *J. Am. Stat. Assoc.*, vol. 81, no. 396, pp. 945–960, Dec. 1986.

[7] X. Su, J. Kang, J. Fan, R. A. Levine, and X. Yan, "Facilitating score and causal inference trees for large observational studies," *J. Mach. Learn. Res.*, vol. 13, pp. 2955–2994, Oct. 2012.

[8] K. Hirano and G. W. Imbens, "The propensity score with continuous treatments," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*. Chichester, UK: Wiley, 2005, pp. 73–84.

[9] C. A. Flores, "Estimation of dose-response functions and optimal doses with a continuous treatment," Dept. Econ., Univ. Miami, Coral Gables, FL, 2007.

[10] A. Salo, J. Keisler, and A. Morton, "Improving resource allocation with portfolio decision analysis," *INFORMS Analytics Mag.*, vol. 4, no. 6, pp. 23–26, Nov.–Dec. 2011.

[11] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, Mar. 1952.

[12] E. Jay, P. Duvaut, S. Darolles, and A. Chrétien, "Multifactor models," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 37–48, Sep. 2011.

[13] G. Ganesan, "A subspace approach to portfolio analysis," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 49–60, Sep. 2011.

[14] M. U. Torun, A. N. Akansu, and M. Avellaneda, "Portfolio risk in multiple frequencies," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 61–71, Sep. 2011.

[15] M. McCue, "Taking the plunge," *ESPN Mag.*, pp. 83–85, Mar. 19 2012.

[16] M. Baier, J. E. Carballo, A. J. Chang, Y. Lu, A. Mojsilović, M. J. Richard, M. Singh, M. S. Squillante, and K. R. Varshney, "Sales-force performance analytics and optimization," *IBM J. Res. Dev.*, vol. 56, no. 6, p. 8, Nov./Dec. 2012.

[17] E. A. Nadaraya, "On estimating regression," *Theory Probab. Appl.*, vol. 9, no. 1, pp. 141–142, 1964.

[18] G. S. Watson, "Smooth regression analysis," *Sankhyā Ser. A*, vol. 26, no. 4, pp. 359–372, Dec. 1964.

[19] W. Härdle, *Applied Nonparametric Regression*. Cambridge, UK: Cambridge University Press, 1994.

[20] A. A. Zoltners and P. Sinha, "Integer programming models for sales resource allocation," *Manage. Sci.*, vol. 26, no. 3, pp. 242–260, Mar. 1980.

[21] B. Skiera and S. Albers, "Prioritizing sales force decision areas for productivity improvements using a core sales response function," *J. Personal Selling & Sales Management*, vol. 28, no. 2, pp. 145–154, Spring 2008.

[22] R. Lawrence, C. Perlich, S. Rosset, J. Arroyo, M. Callahan, J. M. Collins, A. Ershov, S. Feinzig, I. Khabibrakhmanov, S. Mahatma, M. Niemaszyk, and S. M. Weiss, "Analytics-driven solutions for customer targeting and sales-force allocation," *IBM Syst. J.*, vol. 46, no. 4, pp. 797–816, Oct. 2007.

[23] G. B. Dantzig, "Discrete-variable extremum problems," *Oper. Res.*, vol. 5, no. 2, pp. 266–288, Apr. 1957.

[24] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.