APPLICATIONS OF CHANCE CONSTRAINED OPTIMIZATION IN O.M.

APPLICATIONS OF CHANCE CONSTRAINED OPTIMIZATION IN OPERATIONS MANAGEMENT

By MICHAEL R. METEL, B.E.Sc. M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (Computational Science and Engineering) 2016

McMaster University, Hamilton, Ontario

TITLE: Applications of Chance Constrained Optimization in Operations Management

AUTHOR: Michael R. Metel, B.E.Sc. (University of Western Ontario) M.Sc. (McMaster University)

SUPERVISORS: Professors Antoine Deza and Kai Huang

NUMBER OF PAGES: ix, 61

Abstract

In this thesis we explore three applications of chance constrained optimization in operations management. We first investigate the effect of consumer demand estimation error on new product production planning. An inventory model is proposed, whereby demand is influenced by price and advertising. The effect of parameter misspecification of the demand model is empirically examined in relation to profit and service level feasibility, and conservative approaches to estimating their effect on consumer demand is determined. We next consider optimization in Internet advertising by introducing a chance constrained model for the fulfillment of guaranteed display Internet advertising campaigns. Lower and upper bounds using Monte Carlo sampling and convex approximations are presented, as well as a branching heuristic for sample approximation lower bounds and an iterative algorithm for improved convex approximation upper bounds. The final application is in risk management for parimutuel horse racing wagering. We develop a methodology to limit potential losing streaks with high probability to the given time horizon of a gambler. A proof of concept was conducted using one season of historical race data, where losing streaks were effectively contained within different time periods for superfecta betting.

Acknowledgements

I would like to thank my supervisors, Antoine Deza and Kai Huang, for their guidance and support, members of my supervisory committee, Bartosz Protas and Sanzheng Qiao, as well as my colleagues in the Advanced Optimization Laboratory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Chance constrained optimization is a branch of mathematical programming which considers the uncertainty of input data. In order to accommodate this randomness, affected constraints must be satisfied with a given probability. A general chance constrained program can be formulated as

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & \mathbb{P}(F(x,\xi) \leq 0) \geq 1-\alpha \\ & x \in X \end{aligned} \tag{1.1}$$

where α is an error tolerance, X is a deterministic feasible region and ξ is a random vector. Chance constrained programs are in general difficult to solve due to their non-convexity and the numerical integration required to evaluate the chance constraint, see Pagnoncelli et al. [50]. Exceptions include joint chance constraints of the form $\mathbb{P}(Tx \geq \xi) \geq 1-\alpha$ where ξ follows a log-concave distribution, which are convex, and individual chance constraints with random data following a multivariate normal distribution of the form $\mathbb{P}(x^T\xi \leq 0) \geq 1-\alpha$, which can be formulated as second order conic constraints [53].

1.1 Approximation techniques

There are two general techniques for approximating (1.1): the sample approximation, using Monte Carlo sampling to generate an approximate mixed-integer program (MIP), and convex approximations, which rely on probability inequalities to generate convex regions which either restrict or relax the feasible region.

1.1.1 Sample approximations

(1.2) is a sample approximation of (1.1). N independent samples of ξ are generated, denoted as ξ_k for $k = 1, ..., N$. The first three constraints in (1.2) enforce that $F(x, \xi) \leq 0$ in at least $(1-\epsilon)N$ scenarios, which approximates the chance constraint $\mathbb{P}(F(x, \xi) \leq 0) \geq 1 - \epsilon$.

$$\begin{aligned} \min \quad & f(x) && && (1.2)\\ \text{s.t.} \quad & F(x,\xi_k)z_k \leq 0 && k = 1, ..., N \\ & \sum_{k=1}^{N} z_k \geq (1-\epsilon)N \\ & z \in \{0,1\}^N \\ & x \in X \end{aligned}$$

We can obtain lower and upper bounds with high probability by solving (1.2) with an appropriate choice for N and ϵ. Luedtke and Ahmed [39] gave results for achieving a lower bound for general chance constrained programs and upper bounds under certain conditions for the distribution of ξ and properties of $F(x, \xi)$. Calafiore and Campi [14] determined the number of samples required to calculate a feasible solution with high probability with $\epsilon = 0$, and an improved result allowing non-zero ϵ was given by Campi and Garatti [15] under the assumption that (1.2) with $\epsilon = 0$ is feasible for any finite number of samples N.

1.1.2 Convex approximations

Pintér [51] first proposed using classic probability inequalities (e.g. Chebyshev, Bernstein, Hoeffding) to conservatively approximate chance constraints. Nemirovski and Shapiro [48] further developed convex conservative approximations, resulting in a general class of approximations, and what is called the Bernstein approximation assuming the random variables are independent. Ahmed [2] developed a class of convex relaxations, resulting in lower bounds to (1.1) and the Bernstein relaxation, using a similar technique.

1.2 Solution techniques

Assuming $f(x)$ is a convex function and X is a convex region, convex approximations result in convex optimization programs which can be efficiently solved using interior point methods. Sample approximations result in MIPs which are challenging. Solution techniques assuming ξ has a finite distribution (which includes sample approximations of general chance constrained programs) have been developed, using valid inequalities and branch-and-cut algorithms. Luedtke et al. [40] developed strong valid inequalities assuming only a random right-hand side of the form $\mathbb{P}(Tx \geq \xi) \geq 1 - \alpha$, with a deterministic matrix T. Some of the results of [40] were further developed into a specialized branch-and-cut algorithm in [38] to solve chance-constrained programs for finite distributions with equal probabilities. Using a partial ordering of the set of scenarios, a precedence constrained knapsack polyhedron was defined by Ruszczyński in [55]. Valid inequalities were developed for general chance constrained programs with a finite distribution and used in a branch-and-cut algorithm.

1.3 Thesis overview

The purpose of this thesis is to make contributions in modeling with chance constraints and their solution techniques, with a focus on applications in operations management. In the next chapter, an inventory model is proposed where demand is influenced by price and advertising, and the effect of consumer demand estimation error on new product production planning is investigated. In Chapter 3, we introduce a chance constrained optimization model for the fulfillment of guaranteed display Internet advertising campaigns. We discuss and present theoretical and computational features of the model via Monte Carlo sampling and convex approximations. This work has been previously published in *Omega: The International Journal of Management Science*, see [21]. Chapter 4 considers risk management in parimutuel horse racing wagering. Through the use of a chance constraint, we incorporate the time horizon of the gambler, enforcing an upper bound on potential losing streaks with high probability.

Chapter 2

Imperfect demand estimation for new product production planning

2.1 Introduction

In this chapter we investigate the effect of model parameter estimation error to determine best practices when estimating the effect of price and advertising in relation to inventory management [70] for new products. We consider a single period inventory model with a minimum service level constraint [17], with the objective of maximizing profit under consumer demand uncertainty.

The Bass model [5] is a differential equation which is widely used to forecast new product adoption. Extensions have been made, such as the generalized Bass model [6], which incorporates both price and advertising. For this paper we use an approximation of the piecewise-diffusion model (PDM) of Nui [49], which extends the original Bass model by incorporating demand uncertainty, as well as price and advertising, resulting in a superior fit compared to previous models in empirical testing.

In the paper of Lim et al. [33], the misestimation of supply chain disruption probabilities was investigated. It was found that overestimating disruption probabilities reduces the expected cost when compared to underestimation. When faced with estimation uncertainty, this presents the managerial insight that having a bias towards overestimation prevents excessive costs. In the problem setting of this paper, the estimation uncertainty lies in the consumer demand model. It is unclear what the effect of the estimation error of the consumer demand's response to price and advertising is on profit and service level feasibility, and when faced with uncertainty, what the conservative approach to estimation would be. In an attempt to answer these questions, we conduct an empirical study, whereby the optimal solution is found for our inventory model under what is considered to be the true consumer demand dynamics, after which optimal solutions are found under biased responses to price and advertising to determine the effect of misestimation.

We present an overview of the PDM and briefly trace its roots in Section 2.2. In Section 2.3 the inventory optimization problem is presented as well as the formulation of its approximation. Included are details of the calibration of the PDM, the problem instances which we are interested in, as well as a justification of our approximation in terms of confidence intervals of the error. Details of the computational experiments are described in Section 2.4, with a commentary on the results. The conclusion is summarized in Section 2.5, with the results of the experiment graphically presented in the appendix in Section 2.6.

2.2 Piecewise-diffusion model (PDM)

The Bass model proposes that the number of adopters through time, $N(t)$, can be modeled by the differential equation $\frac{d}{dt}N(t) = (m - N(t))(p + \frac{q}{m}N(t))$, where m is the market size, p is the coefficient of innovation, which is the consumer's intrinsic desire to purchase the product, and q is the coefficient of imitation, which models the influence of existing adopters on the consumer, whose solution is $N(t, m, p, q) = m\frac{1-e^{-(p+q)t}}{1+(q/p)e^{-(p+q)t}}$.

The stochastic Bass model (SBM) assumes that consumer adoption follows a pure birth process, where $A_m(t)$ is the cumulative number of adopters by time t. The transition rate from adoption j to $j + 1$ is $\lambda_{mj} = (m - j)(\alpha + \frac{\beta}{m-1}j)$, where α and β, the intrinsic adoption rate and the induction rate, can be interpreted in the same manner as p and q in the Bass model. Let this be referred to as an SBM with specification $\{m, \alpha, \beta\}$. A central limit theorem is derived in [49], where it is proved that as $m \to \infty$, $\frac{A_m(t)-N(t,m,\alpha,\beta)}{\sqrt{\psi(t,m,\alpha,\beta)}}$ converges in distribution to a standard normal random variable, where $\psi(t, m, \alpha, \beta) = m\frac{(1+\beta/\alpha)e^{-2(\alpha+\beta)t}}{[1+(\beta/\alpha)e^{-(\alpha+\beta)t}]^4}\{e^{(\alpha+\beta)t}-1+2(\frac{\beta}{\alpha})(\alpha+\beta)t+(\frac{\beta}{\alpha})^2(1-e^{-(\alpha+\beta)t})\}$, so that for m sufficiently large, we can approximate $A_m(t)$ as normal with mean $N(t, m, \alpha, \beta)$ and variance $\psi(t, m, \alpha, \beta)$.

The piecewise stochastic Bass model assumes a sequence of time intervals where adoption levels are observed. Let a be the total number of adopters up to the present time. The model assumes that of the total available potential adopters $m - a$, only $(m - a)\pi$ are true prospects, where π is the participation fraction, and the remainder are dormant. We can simulate the demand up to time t under this formulation as an SBM with specification $\{(m - a)\pi, \hat{\alpha}, \hat{\beta}\}$, where $\hat{\alpha} = \alpha + \frac{\beta}{m-1}a$ and $\hat{\beta} = \frac{(m-a)\pi-1}{m-1}\beta$.

The PDM incorporates the central limit theorem result, as well as an additional variance component δ^2 to capture exogenous disturbance and model misspecification. The demand over time t is approximated as normal with mean $\mu = N(t, (m-a)\pi, \hat{\alpha}, \hat{\beta})$ and variance $\sigma^2 = \psi(t, (m-a)\pi, \hat{\alpha}, \hat{\beta}) + \delta^2 t$. Using the PDM, we are able to influence future demand by the choice of the product price p and advertising spending v. Their effect is modeled by setting $\pi = \pi_m\{1 - [(1 - \frac{\pi_p}{\pi_m})e^{-\gamma_p v}]^{(\frac{p}{p_{ref}})^{-\eta}}\}$ and replacing β by $\beta[1 + \gamma_b(v_0 + v)]$, where π_m is the maximum possible participation fraction, π_p is the value of π when $p = p_{ref}$, which is a calibration reference price, η controls price sensitivity, γ_p controls the impact of v, and γ_b scales the increase in influence of existing adopters from the aid of the advertising over the product's life.

The sales trajectory of room air conditioners from 1949-1961, Table 4.2 of the appendix, was used in the empirical study in [49], showcasing the superior fit of the PDM compared to the Bass and generalized Bass models, with a reduction in the sum of squared errors of 94.3% and 84.4% respectively. This dataset has been used extensively in the literature, including the papers describing these two past models. Its popularity stems from having a very generic pattern, with a sales trajectory following what is expected for new products. The PDM was fit to the historical data using maximum likelihood estimation, while the latter two were fit using nonlinear-least squares.[1] In this paper we utilize the actual history parameterization, fit to this dataset, which is in Table 2.2 of the appendix.

[1] For new products with no sales history, this process is not possible, which in part motivated this research.

2.3 Optimization model

We consider an inventory model with zero lead time, variable ordering cost c, and salvage price $s < c$. At the beginning of the time period, we set the price p of our product, we determine the amount of advertising spending v, a product order o is placed and received, and then the consumer demand D is realized. We want to maximize profit subject to satisfying D with probability $1 - \theta$. Our sales over the period will be $\min\{o, D\}$, with our excess supply equal to $\max\{o - D, 0\}$. The optimization problem is as follows.

$$\begin{aligned} & \max\ \mathbb{E}(p \min\{o, D\} + s \max\{o - D, 0\} - co - v) \\ \text{s.t.}\ & \mathbb{P}(o - D \geq 0) \geq 1 - \theta \\ & p, v, o \geq 0 \end{aligned} \tag{2.1}$$

2.3.1 Program formulation

We use a sample average approximation (SAA) to approximate the objective function [9]. We approximate the expected sales $\mathbb{E}(\min\{o, D\})$ as $\{\frac{1}{N}\sum_{j=1}^{N} r_j : r_j \leq o, r_j \leq \mu + \sigma z_j\}$, where z_j is a standard normal sample, and the expected excess supply $\mathbb{E}(\max\{o - D, 0\})$ as $\{o - \frac{1}{N}\sum_{j=1}^{N} r_j\}$. The chance constraint can be written as $o \geq \mu + \sigma\Phi^{-1}(1 - \theta)$, where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution [12]. The objective contains bilinear terms, but for a fixed value of p, the objective becomes linear, so we only consider a finite number of values for p.

$$\max \frac{(p-s)}{N}\sum_{j=1}^{N} r_j + (s-c)o - v \tag{2.2}$$

$$\begin{aligned}
\text{s.t.}\quad & o \geq \mu + \sigma\Phi^{-1}(1-\theta) \\
& r_j \leq o \quad j = 1, ..., N \\
& r_j \leq \mu + \sigma z_j \quad j = 1, ..., N \\
& o \geq 0 \\
& p_{min} \leq p \leq p_{max} \\
& 0 \leq v \leq v_{max} \\
& p \in \mathbb{Z}.
\end{aligned}$$

Note that μ and σ are non-convex functions of p and v. We approximate μ and σ as piecewise linear functions using the logarithmic disaggregated convex combination (DLog) model of Vielma et al. [62]. A Delaunay triangulation is used to segment the price and advertising domain into a set of triangles $\mathcal{T}$. DLog requires $\lceil log_2|\mathcal{T}|\rceil$ binary variables, enforcing a convex combination of the vertices of a single triangle to represent the values of μ and σ.

2.3.2 Inventory scenarios and model calibration

We considered product costs of 60% and 80% of the historical 1949 price of a room air conditioner, \$410, $c_1 = \$246$ and $c_2 = \$328$, with $p_{min} = \$350$, $p_{max} = \$450$, $v_{max} =$ \$100MM, and $s = 0.1c$. Our test instances consist of values of $\theta = \{1, 0.5, 0.25, 0.05\}$.

Our piecewise linear approximation of μ and σ was constructed in the following manner. Beginning with the extreme points of our domain as vertices, we iteratively added vertices by taking a Delaunay triangulation of the current vertex set and finding the triangle with the centroid with the largest Euclidian norm of the percentage error between μ and σ, and their piecewise approximations, $\hat{\mu}$ and $\hat{\sigma}$. This error was then compared to the error of the midpoint of each edge of the triangle, with the point with the largest error added to the vertex set, with p rounded to the nearest integer. The approximation was limited to the use of 10 binary variables. Taking 20,000 samples to construct empirical distribution functions of the percentage error of $\hat{\mu}$ and $\hat{\sigma}$, confidence intervals were found using the Dvoretzky-Keifer-Wolfowitz inequality [46]. For an empirical distribution function with n samples, $F_n(x)$, and any $x \in \mathbb{R}$, $\mathbb{P}(F_n(x) - F(x) > \epsilon) \leq e^{-2n\epsilon^2}$ for every $\epsilon \geq \sqrt{\frac{1}{2n}\ln 2}$. This implies that $F(x) \geq (F_n(x) - \epsilon)(1 - e^{-2n\epsilon^2})$. Taking a value of $\epsilon = 0.014$, confidence intervals were found to be $\mathbb{P}(|\frac{\mu-\hat{\mu}}{\mu}| \leq 0.0024) \geq 0.95$ and $\mathbb{P}(|\frac{\sigma-\hat{\sigma}}{\sigma}| \leq 0.00050) \geq 0.95$.

The SAA sample size, $N = 15,000$, was chosen to ensure the sample problem optimal objective z^*_N is close to the true optimal objective value z^* with high probability. We consider the convergence of the most challenging problem, when $c = c_2$ and $\theta = 0.05$. We are interested in bounding the error due to sampling, so let $z(p, v, o) = (p - s)(\hat{\mu}F(o) - \hat{\sigma}^2 f(o)) + po(1 - F(o)) + soF(o) - co - v$, which is the objective value of (2.1) using $\hat{\mu}$ and $\hat{\sigma}$, where $F(x)$ and $f(x)$ are the cumulative and probability distribution functions of $D \sim N(\hat{\mu}, \hat{\sigma}^2)$. A confidence interval for the optimality gap was calculated based on the technique of Mak et al. [44]. $M = 20$ instances of (2.2) were solved with optimal values z^{*i}_N and objective values of $z(p^*_i, v^*_i, o^*_i)$. Let $\hat{\mu}^N_Z$ and $\hat{\sigma}^N_Z$, and $\hat{\mu}_Z$ and $\hat{\sigma}_Z$, equal the sample mean and standard deviations of z^{*i}_N and $z(p^*_i, v^*_i, o^*_i)$, respectively. When solving a single instance of (2.2) the $1-\alpha$ confidence interval of the optimality gap is estimated as $[\hat{\mu}_Z + t_{\frac{\alpha}{2},M-1}\hat{\sigma}_Z \leq z^* \leq \hat{\mu}^N_Z + t_{1-\frac{\alpha}{2},M-1}\hat{\sigma}^N_Z]$, where $t_{\frac{\alpha}{2},M-1}$ is the $\frac{\alpha}{2}$-critical value of the t-distribution with $M-1$ degrees of freedom. The percentage error optimality gap confidence interval with $\alpha = 0.05$ was found to be $\mathbb{P}(\frac{z_N^* - z(p,v,o)}{z(p,v,o)} \leq 0.0183) \geq 0.95$. Virtually all of the error came from the z_N^{*i}, as each problem instance found the same optimal p_i^*, v_i^* and o_i^* up to 15 significant digits.

2.4 Computational experiments

We are interested in the effect of parameter misspecification on profit and feasibility. In particular, if the effect is asymmetrical, this gives guidance when having to estimate consumer behaviour for new products with no prior history. We observe the effect of underestimating and overestimating the influence of price, η, and the influence of advertising, γ_b and γ_p, which we denote as the vector γ. Given what we consider a true parameter value x_0 from Table 2.2, we repeated the process described in the previous section, estimating μ and σ by a piecewise linear function and solving (2.2) for the optimal values p^*, v^*, and o^* for $x = \{-0.6x_0, -0.3x_0, x_0, 0.3x_0, 0.6x_0\}$, then the expected profit and the feasibility assuming x_0 was observed. All computing was conducted on a Windows 7 Home Premium 64-bit, Intel Core i5-2320 3GHz processor with 8 GB of RAM. The implementation was done in Matlab R2012a interfaced with Gurobi 6.0 using YALMIP [37] dated November 27, 2014. The results are shown graphically in Figures 1 to 4. Figure 1 displays the profit when optimizing over different values of η. Figure 2 displays the optimal order quantity o^* in relation to the minimum order quantity m required for feasibility, presented as a percentage difference, $\frac{o^*-m}{m} \times 100$. Figures 3 and 4 display the same for varying γ.

2.4.1 Results

When η is underestimated, the price is increased to take advantage of the subdued decrease in demand. As a result, too much is ordered, significantly decreasing profit. When η is overestimated, price is decreased slightly with an assumed exaggerated increase in demand, again resulting in an excessive order given the price. An interesting observation is regardless of over or underestimating η, the solution is feasible. So with the focus now only on profit, from Figure 1, we conclude that it is better to err on the side of overestimating the effect of price on consumer demand, which decreases profit at a lower rate than underestimating.

When γ is underestimated, the effect of advertising on demand is underestimated, resulting in an insufficient quantity of product ordered and an infeasible solution. The opposite effect occurs when γ is overestimated, resulting in an excessive, but feasible order. With no regard to service levels, we observe from Figure 3 that it is more profitable to underestimate rather than overestimate the value of γ, but given a service level constraint, it is best to overestimate to ensure feasibility.

2.5 Conclusion and future research

This chapter has examined the effect of over and underestimating the influence of price and advertising on consumer demand in the context of production planning. This is of particular interest for new products with no prior sales history to aid in decision making. From an empirical study, we have found that the error is asymmetrical, and were able to determine prudent manners to estimation. We see the potential for future work stemming from this research. We have focused on a single period model in order to capture the relationship between demand factors and profit and feasibility as clearly as possible, but the extension to a multi-stage inventory model with service level constraints would be interesting from a modeling and computational aspect. We examined the two factors of consumer demand which we felt are of most interest to business managers, but future research could examine the effect of misestimating other factors, such as the maximum participation fraction π_m, which is closely related to the estimation of the market size.

2.6 Appendix

Year	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961
Sales (M)	96	195	238	365	1045	1230	1270	1828	1586	1673	1660	1580	1500
Price ($)	410	370	365	388	335	341	320	293	310	279	269	275	259
Advertising ($MM)	0	0.615	1.198	3.196	5.34	14.372	9.391	13.61	16.785	9.238	5.863	3.923	1.493

Table 2.1: Room air conditioner data from 1949-1961

m (10^3)	a_0 (10^3)	π_p	α	β	δ	η	π_m	γ_p	γ_b
53,291	744	0.005191	0	19.14	39.52	6.218	0.04195	0.009746	0.3704

Table 2.2: Actual history parameterization

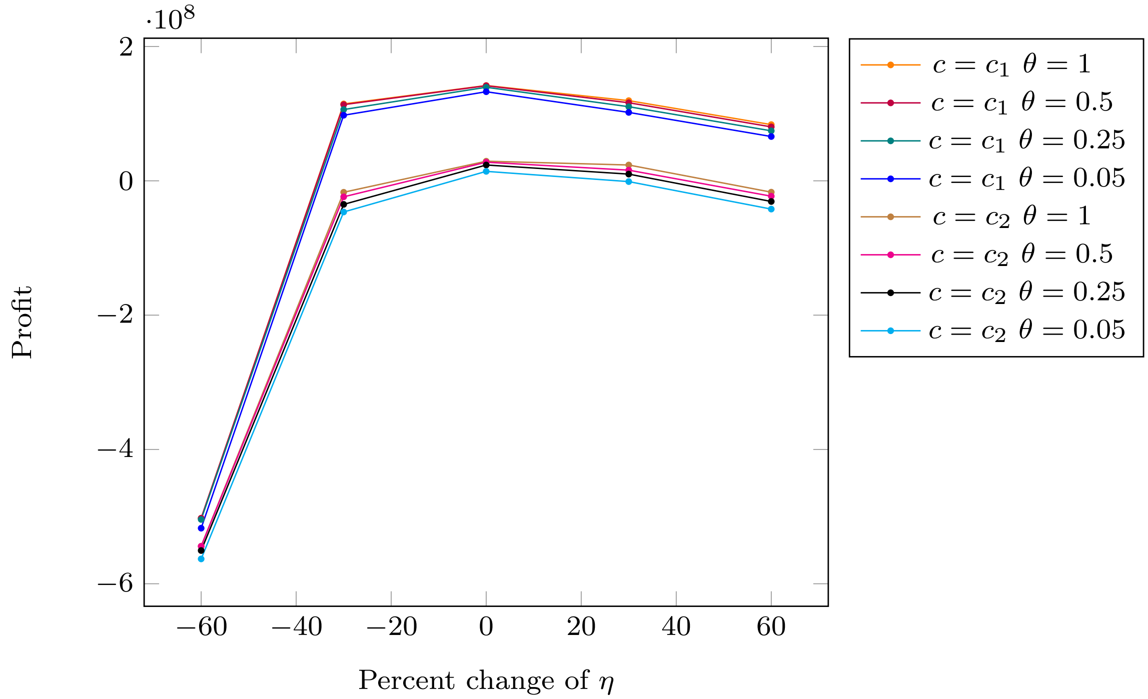

Figure 2.1: Expected profit when optimizing over different values of η.

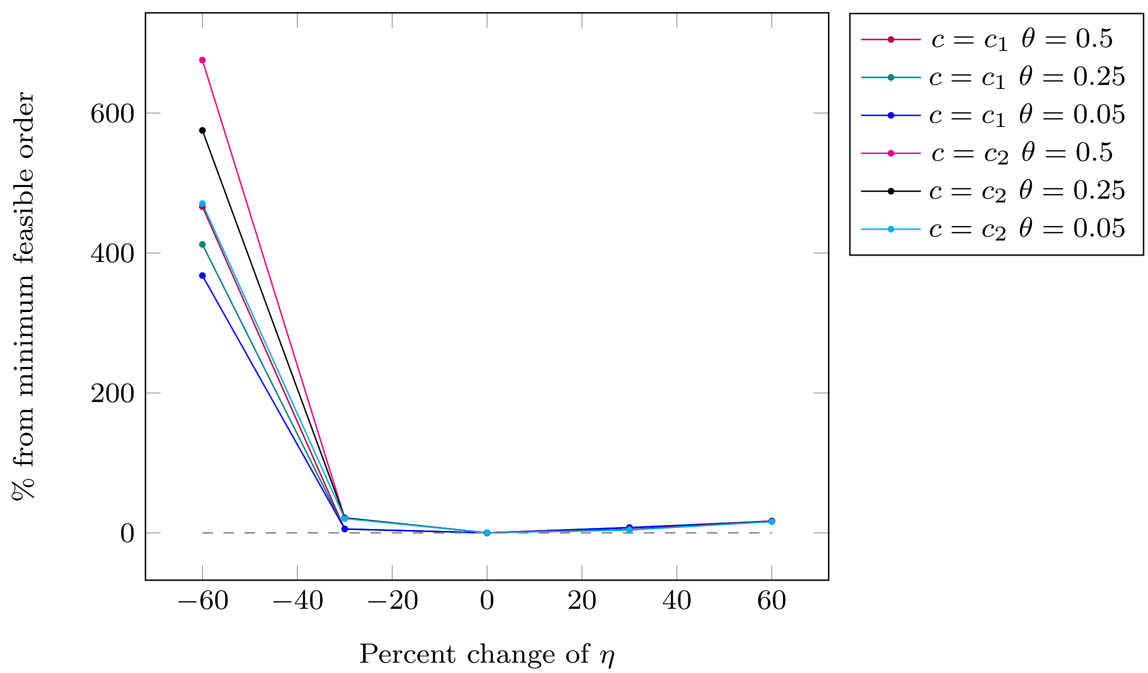

Figure 2.2: Percentage from minimum feasible order when optimizing over different values of η.

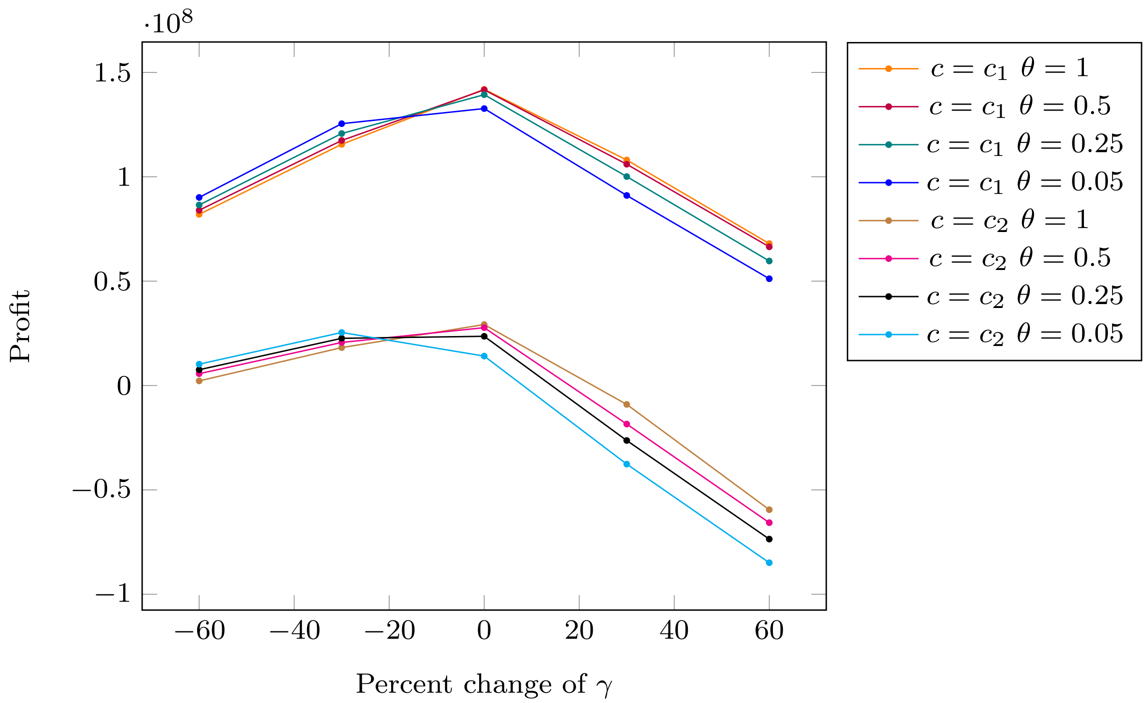

Figure 2.3: Expected profit when optimizing over different values of γ.

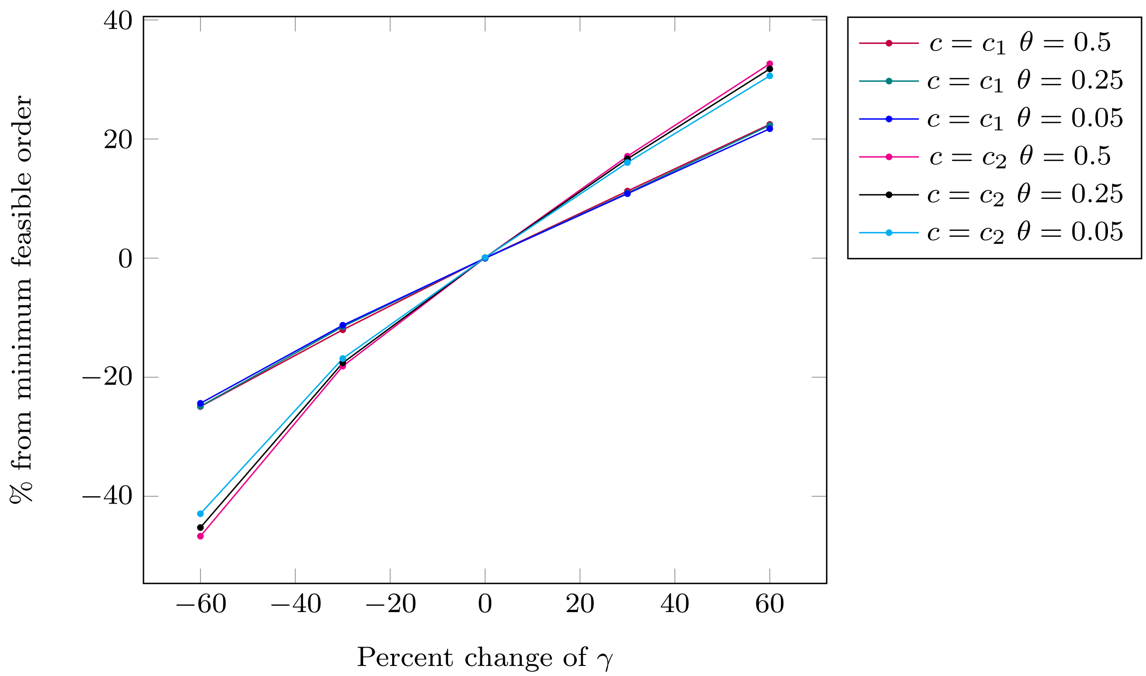

Figure 2.4: Percentage from minimum feasible order when optimizing over different values of γ.

Chapter 3

Chance constrained optimization for targeted Internet advertising

3.1 Introduction

Internet advertising has witnessed growth of 15% in 2012, reaching $36.6 billion in the United States [26]. This field is markedly different from traditional media used by advertisers such as radio, television and newspaper. Information such as a user's profile, data input and past Internet activity allow marketers to display their advertisements to targeted audiences, resulting in an efficient use of their advertising budget and an improved experience for users.

Our work is interested in the planning of guaranteed display Internet advertising by an ad network, which acts as an intermediary between website publishers and advertisers. Advertisers purchase an advertising campaign from the ad network consisting of a guaranteed campaign goal, which is the number of ads to be displayed, and a set of viewer types, which describes who to show the campaign's ads to. Guaranteed display advertising campaigns are typically for brand awareness where the industry practice is for ad networks to maximize representativeness, which is accomplished by displaying ads of each campaign as proportionally as possible to all targeted viewer types, see Yang et al. [69].

Quadratic optimization programs for this problem have recently been developed by Turner [61] and Yang et al. [69]. In particular, Turner showed that performance metrics are maximized using a proposed allocation methodology assuming the viewer supply follows a certain distribution. Our work addresses the uncertainty in viewer supply using a chance constrained framework. Bharadwaj et al. [8] presented an extension to [69] tangential to our research, using a two-stage stochastic program with recourse, with the second stage selling or purchasing ads on the spot market if the realized supply is greater or less than expected. An alternative objective to spread ads across campaigns is to maximize entropy, see Tomlin [60]. We pursue the quadratic objective function approach motivated in part by the availability of advanced and efficient solvers.

We introduce the model in Section 3.2 and formulate the joint chance constrained optimization program to solve the ad network's problem. Section 3.3 discusses how lower and upper bounds can be found through sample approximations. In Section 3.4, a convex approximation program is presented which can be used to find lower and upper bounds under different Internet viewer distribution assumptions. The results of a computational substantiation of the introduced bounds is discussed in Section 3.5. The conclusion and future research ideas can be found in Section 3.6 and a nomenclature in Section 3.7. The appendix in Section 3.8 contains the results of the computational substantiation.

3.2 Chance constrained optimization model

3.2.1 Definitions and notation

An online ad network is an aggregator of display ad slots, which it sells to advertisers in partnership with website publishers. For each guaranteed display advertising campaign, the ad network displays ads to a targeted set of viewers that fit certain criteria, such as by demographic or interest. Advertisers are able to choose their targeted set of viewers from the set V of viewer types, which partitions the publishers' viewers by a predefined set of attributes. Namely, the supply of viewers is modeled as a $|V|$-dimensional random variable with mean vector μ and covariance matrix Σ. Let $\mathcal{S}_v$ denote the supply of incoming ad slots across all websites in the ad network loaded by individuals of viewer type $v \in V$, with μ_v and σ_v being its mean and standard deviation respectively. Let K denote the set of advertising campaigns. For a campaign $k \in K$, the campaign goal g_k is the number of ads to be displayed to viewers, which we assume is given. For research concerning optimal campaign goal sizes, see [1]. The subset of viewer types $V_k \subseteq V$ are the viewer types targeted by advertiser k. The subset of campaigns $K_v \subseteq K$ are the campaigns which target viewer type v.

This problem can be viewed as a stochastic transportation problem with each viewer type as a source with random supply and each advertising campaign as a sink with known demand. Each time a user loads a website affiliated with the ad network, a decision must be made as to which advertisement to display. This paper focuses on the high level planning stage at the beginning of each optimization time period, determining what proportion of ads from each viewer type to allocate to each applicable campaign. The decision variables of the ad network are p_{vk}, the proportion of each viewer type v's supply allocated to each campaign $k \in K_v$. Another means of planning, especially when dealing with campaigns over short time periods, is by allocating ads to one minute time slots, whereby all visitors during each time period are shown the same ads, see [20].

3.2.2 Chance constrained optimization program

We introduce an optimization program to find the proportion allocations, p_{vk}, for all viewer types and targeting campaigns, with an explanation following.

$$\begin{aligned}
\min \sum_{k\in K} \frac{w_k}{|V_k|} \sum_{v\in V_k} (p_{vk} - q_k)^2 & & & \text{(CC)}\\
\text{s.t.} \sum_{k\in K_v} p_{vk} & \leq 1 & \forall v \in V & \\
\mathbb{P}\Big(\sum_{v\in V_k} \mathcal{S}_v p_{vk} \geq g_k,\ \forall k \in K\Big) & \geq 1-\alpha & & \\
q_k & = \frac{1}{|V_k|}\sum_{v\in V_k} p_{vk} & \forall k \in K & \\
p_{vk} & \geq 0 & \forall k \in K, v \in V_k &
\end{aligned}$$

The first constraint ensures that no more than 100% of a viewer type's supply is allocated. The second constraint models the idea of guaranteed campaign fulfillment, which is interpreted as fulfillment with high probability. In particular, the second constraint ensures that all campaigns are fulfilled with a probability of at least $1-\alpha$, where $\alpha < 0.5$ is the un-fulfillment tolerance. The fourth constraint ensures that proportions are non-negative.

Chance constrained programming has been used in many different fields such as finance [32] and water resource management [64]. We model campaign fulfillment using a chance constraint for two reasons. The first is that the success of an advertising campaign is unlikely to change dramatically if g_k ads or $(1-\epsilon)g_k$ ads are displayed for some small percentage ϵ, whereas strictly requiring the former may significantly limit the number of advertising campaigns the ad network can accept. With the parameter α, the ad network is able to balance advertiser satisfaction with the total number of advertising campaigns executed. The second, more fundamental reason is that robust solutions are unlikely to exist without making strong assumptions on the underlying distribution of Internet viewers. $\mathbb{P}(\cup_{k\in K}\{\sum_{v\in V_k}\mathcal{S}_v < g_k\}) = 0$ is a necessary condition for the existence of a robust solution. For distribution assumptions of viewer type supply where this condition does not hold, e.g., normal, Poisson, log-normal, there exists a minimal $\hat{\alpha} > 0$ such that $\alpha \geq \hat{\alpha}$ for (CC) to be a feasible program.

The objective of the ad network is to maximize representativeness by allocating each campaign k's ads across all $v \in V_k$ proportionally to the supply, which is achieved by having $p_{vk} = p_{v'k}$ for all $v, v' \in V_k$. Objectives of the following general form have been proposed for guaranteed advertising campaigns, see [61, 69],

$$\min \sum_{k\in K}\sum_{v\in V_k} w_{vk}(p_{vk} - \frac{g_k}{\bar{\mu}_k})^2$$

where the w_{vk}'s are weights, $\bar{\mu}_k = \sum_{v\in V_k}\mu_v$ is the total expected supply from the viewer types targeted by campaign k, and $\frac{g_k}{\bar{\mu}_k}$ is the target proportion. The objective maximizes weighted representativeness of campaigns, assuming the ad network is constrained to fulfill campaigns in expectation. Given the chance constraint, an ideal feasible allocation is unknown a priori. We propose to minimize the variance of each campaign's allocation proportions. The objective is then

$$\sum_{k\in K}\frac{w_k}{|V_k|}\sum_{v\in V_k}(p_{vk}-q_k)^2$$

where q_k is the mean of the proportions allocated to campaign k from viewer types in V_k, enforced in the third constraint, and the weights w_k represent the campaign's priority to the ad network. For example, assume campaign k targets 5 viewer types, and a feasible solution to (CC) includes the vector of proportions allocated to campaign k, $p_k = [0.2, 0.3, 0.1, 0.4, 0]$. Since $q_k = \frac{0.2+0.3+0.1+0.4+0}{5} = 0.2$, the variance of p_k is then $\frac{1}{5}\sum_{v=1}^{5}(p_{vk}-0.2)^2 = 0.02$. The objective attempts to set $p_k = [\phi, \phi, \phi, \phi, \phi]$ for some unknown ϕ, which would achieve perfect representativeness for campaign k, with a variance of 0.

In the following section, we present results enabling the construction of sample approximations which, when solved, achieve lower and upper bounds with high probability for (CC).

3.3 Sample approximations

3.3.1 SA program

The following program, (SA), is a finite approximation to (CC). For $i = 1, ..., N$, the $\mathcal{S}_v^i$'s are independently sampled supply scenarios. The binary variable $x_i = 1$ enforces the fulfillment of all campaign goals in scenario i. The second and third constraints require that all campaign goals are satisfied in at least $\lceil(1-\xi)N\rceil$ scenarios, which approximates the joint chance constraint $\mathbb{P}(\sum_{v\in V_k}\mathcal{S}_v p_{vk} \geq g_k,\ \forall k \in K) \geq 1-\xi$.

$$\begin{aligned}
\min \sum_{k\in K} \frac{w_k}{|V_k|} \sum_{v\in V_k} (p_{vk}-q_k)^2 & & & \qquad\qquad \text{(SA)}\\
\text{s.t.} \sum_{k\in K_v} p_{vk} & \leq 1 & & \forall v \in V\\
\sum_{v\in V_k} \mathcal{S}_v^i p_{vk} & \geq x_i g_k & & \forall k \in K, i=1,...,N\\
\sum_{i=1}^{N} x_i & \geq \lceil (1-\xi)N \rceil & & \\
q_k & = \frac{1}{|V_k|}\sum_{v\in V_k} p_{vk} & & \forall k \in K\\
p_{vk} & \geq 0 & & \forall k\in K, v \in V_k\\
x_i & \in \{0,1\} & & \forall i = 1,...,N
\end{aligned}$$

We can obtain lower and upper bounds with high probability by solving (SA) with an appropriate choice for N and ξ.

3.3.2 SA lower bound

Assume (CC) is a feasible program with optimal objective value $z^{(CC)*}$ and optimal solution $p_{vk}^{(CC)*}$. Property 1 determines the probability of $p_{vk}^{(CC)*}$ being feasible in (SA), implying the optimal objective value of (SA), $z^{(SA)*} \leq z^{(CC)*}$. When (CC) is not feasible, $z^{(SA)*} \leq z^{(CC)*}$, using the convention $z^{(CC)*} = \infty$.

Property 1 (Luedtke and Ahmed [39])**.** $\mathbb{P}(z^{(SA)*} \leq z^{(CC)*}) \geq \sum_{i=0}^{\lfloor \xi N \rfloor} \binom{N}{i}\alpha^i(1-\alpha)^{N-i}$.

3.3.3 SA upper bound

Property 2 requires that the objective is convex, the deterministic feasible region is convex and closed, and that the chance constraint mapping is closed and convex. Let (RSA) be the robust version of (SA) with $\xi = 0$. This implies all $x_i = 1$ converting (SA) into a convex quadratic program. Property 2 gives the probability that $z^{(\mathrm{RSA})*} \geq z^{(CC)*}$.

Property 2 (Calafiore and Campi [14])**.**
$\mathbb{P}(z^{(RSA)*} \geq z^{(CC)*}) \geq 1 - \binom{N}{|V_K|}(1-\alpha)^{N-|V_K|}$, where $|V_K| = \sum_{k=1}^{|K|} |V_k|$ is the number of decision variables.

3.3.4 A branching scheme for the branch-and-bound algorithm

In this subsection we discuss an aspect of the algorithm used to find sample approximation lower bounds, which enabled us to solve larger scale problems. Assuming we are in the midst of solving (SA), we must solve the following program, (SA_m), at node

m of the Branch-and-Bound algorithm.

$$
\begin{aligned}
\min \sum_{k \in K} \frac{w_k}{|V_k|} \sum_{v \in V_k} (p_{vk} - q_k)^2 & & & \qquad (\text{SA}_m) \\
\text{s.t. } \sum_{k \in K_v} p_{vk} & \leq 1 & & \forall v \in V \\
\sum_{v \in V_k} \mathcal{S}_v^i p_{vk} & \geq x_i g_k & & \forall k \in K, i = 1, ..., N \\
\sum_{i=1}^{N} x_i & = \lceil (1-\xi) N \rceil & & \\
x^T \text{diag}(X_m^1) & = \mathbf{1}^T \text{diag}(X_m^1) & & \\
x^T \text{diag}(X_m^0) & = \mathbf{0} & & \\
q_k & = \frac{1}{|V_k|} \sum_{v \in V_k} p_{vk} & & \forall k \in K \\
p_{vk} & \geq 0 & & \forall k \in K, v \in V_k \\
x_i & \in [0, 1] & & \forall i = 1, ..., N
\end{aligned}
$$

where X_m^1 and X_m^0 are binary vectors of length N which indicate the x_i set to one and zero at node m of the branching tree. We use an equality in the constraint $\sum_{i=1}^N x_i = \lceil (1-\xi)N \rceil$, as for any integral optimal solution with excess x_i^*'s equal to 1 not enforced by X_m^1 can be set to 0 with no effect to the optimal solution or objective value.

After solving (SA_m), assume that $p^{(\text{SA}_m)*}$ is not feasible in (SA) and the optimal objective value, $z^{(\text{SA}_m)*}$ is less than the current upper bound. Thus, we want to branch on one of the x_i for $i \in \{l : X_m^1(l) = 0, X_m^0(l) = 0\}$. We use the following heuristic which finds the scenario j with the constraint which is the farthest from being satisfied on a percentage basis: $j = \underset{i:X_m^1(i)=0, X_m^0(i)=0}{\text{argmin}} \min_{k=1,\ldots,|K|} \frac{\sum_{v\in V_k} \mathcal{S}_v^i p_{vk}^{(\text{SA}_m)*}}{g_k}$. For the path with $x_j = 1$, the branching tree can be effectively pruned as enforcing scenario j will likely enforce other scenarios, and the path with $x_j = 0$ will lead to near optimal solutions as x_j is a promising candidate for one of the $\lfloor \xi N \rfloor$ scenarios to discard.

3.4 Convex approximations

In this section we present convex constraints which can replace the joint chance constraint in (CC) to achieve bounding convex programs. Let $\mathcal{S}_k$ be the vector of the viewer types' supply which campaign k targets, with μ_k being the $|V_k|$-dimensional mean vector and Σ_k being the $|V_k| \times |V_k|$ covariance matrix. In addition, let p_k be the $|V_k|$-dimensional vector of proportions allocated to campaign k from viewer types in V_k.

3.4.1 Distribution-free bounds

This subsection assumes that we only have estimates for the first two moments with no knowledge of the underlying distribution. We present lower and upper bounds based on classic probability inequalities.

Property 3 (Distribution-free Lower Bound). *Any feasible solution of (CC) satisfies the constraints $p_k^T \mu_k \geq (1-\alpha) g_k$ for $k \in K$.*

Proof. Assume there exists a $k' \in K$ with $p_{k'}^T \mu_{k'} < (1-\alpha) g_{k'}$, then $\mathbb{P}(p_k^T \mathcal{S}_k \geq g_k \ \forall k \in K) \leq \mathbb{P}(p_{k'}^T \mathcal{S}_{k'} \geq g_{k'}) \leq \frac{p_{k'}^T \mu_{k'}}{g_{k'}} < 1-\alpha$, where the second inequality follows from Markov's inequality. □

Property 4 (Distribution-free Upper Bound)**.** *The constraints*

$$g_k - p_k^T \mu_k + \sqrt{\frac{1-\alpha_k}{\alpha_k}} \sqrt{p_k^T \Sigma_k p_k} \leq 0 \qquad \forall k \in K$$

where $\sum_{k \in K} \alpha_k = \alpha$, $\alpha_k > 0$, *form a conservative approximation of* $\mathbb{P}(p_k^T \mathcal{S}_k \geq g_k \ \forall k \in K) \geq 1 - \alpha$.

Proof. Following the reasoning in [48, Sec. 2], assume $\mathbb{P}(p_k^T \mathcal{S}_k < g_k) \leq \alpha_k \ \forall k \in K$, then $\mathbb{P}(\cup_{k \in K} \{p_k^T \mathcal{S}_k < g_k\}) \leq \sum_{k \in K} \mathbb{P}(p_k^T \mathcal{S}_k < g_k) \leq \alpha$, implying $\mathbb{P}(p_k^T \mathcal{S}_k \geq g_k \ \forall k \in K) \geq 1 - \alpha$. To show $\mathbb{P}(p_k^T \mathcal{S}_k < g_k) \leq \alpha_k$, we use the one-sided Chebyshev inequality, $\mathbb{P}(Y \leq \mathbb{E}(Y) - b) \leq \frac{\text{Var}(Y)}{\text{Var}(Y) + b^2}$ for a random variable Y and constant $b > 0$,

$$\begin{aligned}
\mathbb{P}(p_k^T \mathcal{S}_k < g_k) &\leq \mathbb{P}(p_k^T \mathcal{S}_k \leq g_k) \\
&\leq \frac{p_k^T \Sigma_k p_k}{p_k^T \Sigma_k p_k + (p_k^T \mu_k - g_k)^2} \\
&\leq \frac{p_k^T \Sigma_k p_k}{p_k^T \Sigma_k p_k + \frac{1-\alpha_k}{\alpha_k} p_k^T \Sigma_k p_k} \\
&= \alpha_k.
\end{aligned}$$

□

3.4.2 Bounds assuming a normal distribution

The normal distribution has been proposed in the literature for modeling viewer type supply, see [8]. This subsection presents convex approximations under the assumption that $\mathcal{S}_k$ follows a multivariate normal distribution, so that $p_k^T \mathcal{S}_k \sim N(p_k^T \mu_k, p_k^T \Sigma_k p_k)$. Let F_k denote the cumulative distribution function of $p_k^T \mathcal{S}_k$.

3.4.2.1 Normal lower bound

Requiring each campaign's probability of fulfillment to be at least $1-\alpha$ is necessary for feasibility in (CC), resulting in a convex relaxation. The chance constraint for each campaign is equivalent to a second-order cone constraint [12],

$$
\begin{aligned}
\mathbb{P}(p_k^T \mathcal{S}_k \geq g_k) = & 1 - F_k(g_k) \geq 1-\alpha \\
F_k(g_k) & \leq \alpha \\
g_k & \leq F_k^{-1}(\alpha) \\
g_k & \leq p_k^T \mu_k + n_\alpha \sqrt{p_k^T \Sigma_k p_k},
\end{aligned}
$$

where n_α is the α percentile of a standard normal random variable.

3.4.2.2 Normal upper bound

As in Property 4, an upper bound can be found by requiring $\mathbb{P}(p_k^T \mathcal{S}_k \leq g_k) \leq \alpha_k$ for all $k \in K$. Assuming the viewer supply follows a normal distribution, we can then use the constraints found in the previous subsection with α replaced by α_k.

3.4.3 Summary

The above convex approximations can be obtained by solving the following convex approximation program (CA) with the proper choice of parameters u_k and h_k as summarized in Table 3.1.

$$
\begin{aligned}
\min \sum_{k\in K} \frac{w_k}{|V_k|} \sum_{v\in V_k} (p_{vk} - q_k)^2 & & & \qquad \text{(CA)} \\
\text{s.t.} \sum_{k\in K_v} p_{vk} & \leq 1 & & \forall v \in V \\
p_k^T \mu_k - u_k \sqrt{p_k^T \Sigma_k p_k} & \geq h_k & & \forall k \in K \\
q_k & = \frac{1}{|V_k|} \sum_{v\in V_k} p_{vk} & & \forall k \in K \\
p_{vk} & \geq 0 & & \forall k \in K, v \in V_k
\end{aligned}
$$

Bound	$\mathbf{u_k}$	$\mathbf{h_k}$
Distribution-free lower bound	0	$(1-\alpha)g_k$
Distribution-free upper bound	$\sqrt{\frac{1-\alpha_k}{\alpha_k}}$	g_k
Normal lower bound	$-n_\alpha$	g_k
Normal upper bound	$-n_{\alpha_k}$	g_k

Table 3.1: Parameter values for (CA)

(CA) was solved using a primal-dual interior point algorithm. The algorithm used generalized logarithm barriers to solve for points on the central path, see [12, Ch. 11.6-11.8]. To form the modified KKT conditions, the Jordan algebra for second-order cones is used to express the complementary slackness conditions of the second-order cone constraints, see [3]. The system of equations to solve for the Newton steps was simplified so that only a system involving the step of p, Δp, was required to be solved, with closed form expressions for the remaining dual variable steps in terms of their current value, p, and Δp. In order to maintain stability, a universal step size was found such that all variables remained feasible. The central path parameter t is updated to equal a multiple of the reciprocal of the maximum error of the modified

KKT conditions involving t. The algorithm quits when the maximum error of the modified KKT conditions is less than or equal to a small ϵ times the current objective function value. In order to find an initial feasible solution, we begin with a function which spreads proportions relative to the campaign's goal size to expected targeted supply, while satisfying the first and fourth constraints of (CA). The Big M method is then used to find an initial solution feasible in the second set of constraints.

3.4.4 Setting the α_k's

We now present an iterative method to calculate upper bounds. When finding a distribution-free upper bound, (CA) is first solved with the α_k's set equal to $\frac{\alpha}{|K|}$, as proposed in [48, Sec. 2]. Letting p_k^* be the optimal solution, with optimal objective value $z^{(CA)*}$, the approximating constraint of Property 4 can be rearranged as $\alpha_k \geq \frac{p_k^{*T}\Sigma_k p_k^*}{p_k^{*T}\Sigma_k p_k^* + (g_k - p_k^{*T}\mu_k)^2} = \hat{\alpha}_k$. For any k for which this constraint is not tight, we can set $\alpha_k = \hat{\alpha}_k$. As these tighter constraints are valid for (CA), resolving the optimization problem with the tighter constraints, (TCA), will result in an objective value $z^{(TCA)*} = z^{(CA)*}$. Assuming there was at least one constraint in (CA) which had slack, $\sum_k \alpha_k < \alpha$ in (TCA). The total slack $s = \alpha - \sum_k \alpha_k$ can be added evenly to all α_k's of the originally tight constraints in (CA). Solving this relaxation of (TCA), (RTCA), will result in an objective value $z^{(RTCA)*} \leq z^{(TCA)*}$. This process of redistributing slack among the α_k's is iterated until the improvement in the objective value becomes sufficiently small. The same process is used for the normal upper bound, where for all approximating constraints with slack, α_k is updated to equal $F_k(g_k)$. The algorithm for solving the distribution-free upper bound is presented below, with the necessary changes to solve for the normal upper bound in the comments.

Algorithm 1 Calculating the distribution-free upper bound

$\alpha_k = \frac{\alpha}{|K|}\ \forall k \in K$
$u_k = \sqrt{\frac{1-\alpha_k}{\alpha_k}}\ \forall k \in K$ {$u_k = -n_{\alpha_k}$ for the normal upper bound.}
$[z^*, p^*] = CA(u, g)$
$z^*_{old} = \infty$
$I = 0^{|K|}$ {Indicator vector with k^{th} entry set to 1 when slack found in constraint associated with campaign k.}
while $z^*_{old} - z^* > 0$ **do**
$s = 0$ {Stores total slack across all constraints.}
for k=1:$|K|$ **do**
if $\alpha_k > \hat{\alpha}_k$ **then**
$s = s + \alpha_k - \hat{\alpha}_k$ {$\hat{\alpha}_k = F_k(g_k)$ for the normal upper bound.}
$\alpha_k = \hat{\alpha}_k$
$I_k = 1$
end if
end for
if $s > 0$ & $\sum_{j=1}^{|K|} I_j < |K|$ **then**
for k=1:$|K|$ **do**
if $I_k = 0$ **then**
$\alpha_k = \alpha_k + \frac{s}{|K| - \sum_{j=1}^{|K|} I_j}$
end if
end for
end if
$z^*_{old} = z^*$
$u_k = \sqrt{\frac{1-\alpha_k}{\alpha_k}}\ \forall k \in K$ {$u_k = -n_{\alpha_k}$ for the normal upper bound.}
$[z^*, p^*] = CA(u, g)$
end while

3.5 Computational substantiation

In this section we compare the solutions of the sample and convex approximations. All testing was conducted on a Windows 7 Home Premium 64-bit, Intel Core i5-2320 3GHz processor with 8 GB of RAM. All coding was done in Matlab R2012a interfaced with CPLEX 12.4 using YALMIP [37] dated 13-Feb-2013. Ten random test problems were generated. For each test problem, the number of campaigns and viewer types were chosen randomly between $5, ..., 10$ and $10, ..., 20$. Campaign targeting was achieved by generating a $|K| \times |V|$ matrix of Bernoulli random variables with $p = 0.5$, with cell $(i, j) = 1$ indicating that campaign i targets viewer type j. If there was a campaign or viewer type which was not assigned at least one viewer type or campaign, then a random cell in the appropriate row or column was set to 1. A random vector of viewer type means were generated, with each mean following a uniform distribution between [1000,10000]. Given the mean, μ_v, σ_v^2 was randomly generated uniformly within $[0.25, 0.5] \times \mu_v$. A random correlation matrix was generated using the random Gram matrix approach [25]. For each campaign, $g_k = U_{[0.5,0.75]} \sum_{v \in V_k} \frac{\mu_v}{|K_v|}$, where $U_{[0.5,0.75]}$ is uniform between $[0.5, 0.75]$. For all campaigns $w_k = 1$.

The sample approximation parameters for each test problem were chosen so that the optimal solution is between the bounds with a probability of at least 99%. For the first five problems $\alpha = 0.1$, with the lower bound parameters chosen as $N = 108$ and $\epsilon = 1.76\alpha$, and for the remaining five problems $\alpha = 0.05$, with the lower bound parameters chosen as $N = 102$ and $\epsilon = 2.16\alpha$. For both cases, $\sum_{i=0}^{\lfloor \xi N \rfloor} \binom{N}{i} \alpha^i (1-\alpha)^{N-i} = 0.995$. The value of N for the upper bound is problem instance specific, and was set as the minimum N such that $1 - \binom{N}{|V_K|}(1-\alpha)^{N-|V_K|} \geq 0.995$. The average value of N was 1487 for $\alpha = 0.1$, and 3371 for $\alpha = 0.5$, and the average value of $1 - \binom{N}{|V_K|}(1-\alpha)^{N-|V_K|}$

was 0.9952 over all 10 problem instances. We tested all bounds sampling the viewer supply from a normal distribution.

Let the probability of fulfillment (PF) equal $\mathbb{P}(\sum_{v\in V_k}\mathcal{S}_v p_{vk} \geq g_k, \forall k \in K)$. This probability is estimated for all solutions by generating 100,000 supply scenarios. Indicator variables, $\mathbb{1}_{\{\sum_{v\in V_k}\mathcal{S}_v^i p_{vk}\geq g_k,\, \forall k\in K\}}$, for each scenario i were generated and treated as a Bernoulli sample. The 99% one-sided confidence interval of the probability of fulfillment, $\hat{PF}$, was then estimated, $\mathbb{P}(PF \geq \hat{PF}) = 0.99$.

Results for each test problem are displayed in the appendix. Objective values were multiplied by 1000 for readability. For the sample approximation bounds, the lower bound objective, the lower bound solution's $\hat{PF}$, the lower bound computation time using the branching heuristic of Subsection 3.3.4, the lower bound computation time solving directly with CPLEX, the upper bound objective, the upper bound solution's $\hat{PF}$, and the upper bound computation time are presented from left to right in Table 3.2 . When computing the lower bound directly with CPLEX, a time limit of $20 * T_H$ was set, after which CPLEX would quit, where T_H is the computation time using the branching heuristic.

For the convex approximation bounds, the lower bound objective, the lower bound solution's $\hat{PF}$, and the lower bound computation time comprise columns 2-4 of Tables 4.3 and 4.5. With $\alpha_k = \frac{\alpha}{|K|}$, the upper bound objective, the upper bound solution's $\hat{PF}$, and the computation time follow in columns 5-7. Using the algorithm of Subsection 3.4.4, the upper bound objective, the upper bound solution's $\hat{PF}$, and the computation time are displayed in columns 8-10.

The average optimality gap for the sample approximation bounds was 43%, with an average computation time of 320 seconds using the heuristic. The heuristic required on average 282 seconds to solve for the lower bound, with CPLEX requiring over an order of magnitude more time on average.

The average optimality gap and computation time for the distribution-free bounds was 385% and 0.21 seconds using the algorithm. The large optimality gap for the convex approximations is to be expected as we are finding bounds taking into account the most extreme possible distributions. The average improvement of the distribution-free upper bound using the algorithm was 16%. We see there is a trade-off between time and solution quality when deciding between sample and convex approximations.

With specific knowledge of the distribution, much tighter bounds can be found. The average optimality gap and computation time for the normal bounds was 11% and 0.15 seconds using the algorithm. Examining the normal upper bounds, we can see these solutions are close to optimality, with $\hat{PF}$ in excess of $1-\alpha$ on average by only 2.4%. The average improvement of the normal upper bound using the algorithm was 4%.

The sample approximation and distribution-free bounds are very conservative for our problem data, with $\hat{PF} > 0.99$ and $\hat{PF} = 0.99995$ for all instances, respectively. In practice, high quality solutions can be attained using the sample approximation technique. Beginning with the theoretical upper bound value of N, (SA) can be iteratively solved, decreasing the value of N until a near optimal solution is found with $\hat{PF}$ sufficiently close to $1-\alpha$.

3.6 Conclusion and future research

This paper presented a chance constrained optimization model for guaranteed displayed Internet advertising campaigns. A sample approximation program with a branching heuristic was developed, as well as convex approximations under Normal and distribution-free viewer supply assumptions, with an iterative method for improving feasible solutions. Log-normal and Poisson distributions have also been proposed to model viewer supply, see [8, 22]. Convex approximations under these assumptions is an area of potential future research.

3.7 Nomenclature

α Campaign un-fulfillment tolerance.

μ Mean vector of viewer type supply.

μ_k Mean vector of the viewer types' supply which campaign k targets.

μ_v Mean supply from viewer type v.

$\mathcal{S}_k$ Vector of the viewer types' supply which campaign k targets.

$\mathcal{S}_v$ Supply of viewer type v.

Σ Covariance matrix of viewer type supply.

Σ_k Covariance matrix of the viewer types' supply which campaign k targets.

σ_v Standard deviation of viewer type v's supply.

ξ Campaign un-fulfillment tolerance for (SA).

g_k Campaign goal of campaign k.

K Set of advertising campaigns.

K_v Set of campaigns which target viewer type v.

N Number of viewer type supply scenarios for (SA).

p_k Vector of proportions allocated to campaign k from viewer types in V_k.

p_{vk} Proportion of viewer type v's supply allocated to campaign $k \in K_v$.

V Set of viewer types.

V_k Set of viewer types targeted by campaign k.

w_k Campaign k's priority weighting.

3.8 Appendix

Sample Approximation Bounds							
#	**LB**				**UB**		
	z	$\hat{PF}$	T_H (s)	T_C (s)	z	$\hat{PF}$	T (s)
1	0.08656	0.67382	394.12189	7885.28954	0.29815	0.99883	36.11891
2	38.47802	0.83258	271.55118	5432.76618	43.57778	0.99931	12.98950
3	1.39652	0.72996	118.50722	2372.77228	1.98864	0.99448	20.11220
4	147.26848	0.78323	642.50342	5631.70639	159.05476	0.99384	12.74514
5	1.55731	0.76499	228.41868	4570.70080	2.91692	0.99728	20.72935
6	100.38367	0.76192	139.77481	1454.69702	117.07098	0.99687	43.10787
7	209.21887	0.78952	375.03623	3746.96343	219.04089	0.99844	65.37173
8	29.24341	0.83597	69.03840	1128.23085	32.06909	0.99838	73.24367
9	0.00000	0.99995	0.45513	1.50496	0.00000	0.99995	36.52844
10	313.32316	0.80525	578.03043	9679.22504	338.78909	0.99755	57.58939

Table 3.2: Results for Sample Approximation Bounds

	Distribution-free Bounds								
#	LB			$UB_{\frac{\alpha}{\lvert\mathbf{K}\rvert}}$			$UB_{\mathbf{ALG}}$		
	z	$\hat{PF}$	T (s)	z	$\hat{PF}$	T (s)	z	$\hat{PF}$	T (s)
1	0.02433	0.10208	0.04318	1.56683	0.99995	0.13366	0.70294	0.99995	0.33721
2	35.11906	0.16493	0.00708	51.48740	0.99995	0.01638	49.67440	0.99995	0.19283
3	1.07339	0.11786	0.00652	4.90718	0.99995	0.02245	3.27246	0.99995	0.27213
4	134.91058	0.05268	0.00575	199.83478	0.99995	0.01840	194.25912	0.99995	0.17045
5	0.91310	0.17005	0.00610	9.19147	0.99995	0.01769	4.96429	0.99995	0.24479
6	88.71204	0.09186	0.00600	219.27627	0.99995	0.01562	198.94794	0.99995	0.24656
7	200.41163	0.01049	0.00738	276.35676	0.99995	0.02342	266.59733	0.99995	0.20504
8	26.67377	0.10453	0.00721	54.57600	0.99995	0.02209	52.12026	0.99995	0.24255
9	0.00000	0.99995	0.01565	0.00000	0.99995	0.01508	0.00000	0.99995	0.02982
10	290.49479	0.05637	0.00673	503.95350	0.99995	0.01890	503.95350	0.99995	0.03657

Table 3.3: Results for Distribution-free Bounds

	Normal Bounds								
#	LB			$UB_{\frac{\alpha}{\lvert\mathbf{K}\rvert}}$			$UB_{\mathbf{ALG}}$		
	z	$\hat{PF}$	T (s)	z	$\hat{PF}$	T (s)	z	$\hat{PF}$	T (s)
1	0.08488	0.68915	0.13207	0.16233	0.96111	0.02831	0.13225	0.90751	0.11698
2	37.80323	0.72256	0.04230	39.48617	0.93148	0.01569	39.27912	0.91504	0.04832
3	1.41510	0.70816	0.02301	1.71973	0.96048	0.01950	1.61711	0.91542	0.21837
4	144.25999	0.61781	0.01768	150.74349	0.92082	0.01747	150.16448	0.90605	0.05213
5	1.53287	0.74523	0.01855	2.11183	0.96274	0.01825	1.85374	0.90628	0.05421
6	100.78383	0.80557	0.01657	107.56587	0.96945	0.02045	106.72161	0.96050	0.06145
7	208.52184	0.73829	0.02060	213.06374	0.96324	0.01995	212.60627	0.95321	0.21879
8	29.15067	0.83168	0.02237	30.56063	0.97911	0.02198	30.08620	0.95146	0.07682
9	0.00000	0.99995	0.01484	0.00000	0.99995	0.01449	0.00000	0.99995	0.02862
10	311.30318	0.73201	0.02364	323.19042	0.95818	0.02235	323.18883	0.95787	0.24979

Table 3.4: Results for Normal Bounds

Chapter 4

Managing losses in exotic horse race wagering

Since the mid 1980's, horse racing has witnessed the rise of betting syndicates akin to hedge funds profiting from statistical techniques similar to high frequency traders in stock exchanges [30]. This is possible as parimutuel wagering is employed at racetracks, where money is pooled for each bet type, the racetrack takes a percentage, and the remainder is disbursed to the winners in proportion to the amount wagered.

Research on horse racing stems in large part due to the fact that it can be viewed as a simplified financial market. Research on important economic concepts such as utility theory [65], the efficient market hypothesis [4], and rational choice theory [54] can be conducted in a straight forward manner, given horse racing's discrete nature, fixed short term contract lengths and attainable sets of historical data for empirical study.

Optimization in the horse racing literature can be traced back to Isaacs [27] deriving a closed form solution for the optimal win bets when maximizing expected profit. Hausch et al. [24] utilized an optimization framework to show inefficiencies in the place and show betting pools using win bet odds to estimate race outcomes. In particular, they used the Kelly criterion [31], maximizing the expected log utility of wealth, and found profitability when limiting betting to when the expected return was greater than a fixed percentage. More recently, Smoczynski and Tomkins [56] derived a simple procedure for optimal win bets under the Kelly criterion through analysis of the Karush-Kuhn-Tucker optimality conditions.

Having found a favourable opportunity in a gambling setting, such as betting on the outcome of flipping a biased coin, the Kelly criterion answers the question of how much to wager. For example, if the probability of heads is $\mathbb{P}(H) = 0.6$ and we have an initial wealth w, we can determine how much to wager, x, by solving $\max 0.6 \log(w + x) + 0.4 \log(w - x)$, which tells us to bet $x^* = 0.2w$. Kelly style betting is widely recognized both in academia [42] and in practise, being used professionally in blackjack [16], general sports betting [67], and in particular horse race betting [66]. Positive aspects of the Kelly criterion are that it asymptotically maximizes the rate of return of one's wealth, and assuming one can wager any fraction of money, it never risks ruin. The volatility of wealth through time is too large for most though, as $\mathbb{P}(w_t \leq \frac{w_0}{n} | t > 0) \approx \frac{1}{n}$ [58], e.g. there's approximately a 10% chance your wealth in the future will be 10% of what it currently is using the Kelly criterion. As a result, many professional investors choose to employ a fractional Kelly criterion [59], which has been shown to possess favourable risk-return properties by MacLean et al. [43], with betting half the Kelly amount being popular among gamblers [52].

There are several different types of wagers one can place on horses, including what are known as exotic wagers, which include the exactor, triactor and superfecta, which require the bettor to pick the first two, three and four finishers in order, respectively. The exotic wagers are popular among professional gamblers, as superior knowledge of the outcome of a race is better rewarded, and the more exotic the bet, the higher the advantage one can attain [7]. For this reason we focus on the superfecta bet, the most exotic wager placed on a single race.

4.1 Time horizon

In recognition of the similarities between parimutuel horse race betting and financial markets, we see superfecta betting being most similar to the purchase of deep out of the money options, with the general trend of a successful strategy being small steady losses through time with infrequent large gains. Speaking of his experience as a key member of a Hong Kong horse racing gambling syndicate, C. X. Wong [66] states that investing in horse racing is more stressful than in the stock market, and that for professional groups wagering in exotic pools it is normal not to have a winning wager once in three months. Once the losing streak terminates a large profit is achieved, but in the interim, there will be various sources of pressure. Doubt in the system may set in leading to the potential for irrational decisions to be made, based not on statistical findings but emotion.

It would be ideal to have a mechanism to control losing streaks, not only to avoid failure but to determine if a losing streak is in range with the current strategy or if an investigation into the system is warranted. As this is a form of risk management, we consider such methods from stock portfolio management. The most famous framework is mean-variance portfolio optimization based on the work of Markowitz [45], where one maximizes the expected return subject to a constraint which limits the variance in portfolio returns. One of the criticisms of this model is that the use of variance as a measure of risk penalizes both positive and negative deviations in the same manner. Given the expected positive skewness of superfecta returns this would be particularly problematic for our application.

A popular risk measure proposed to replace variance is the value at risk (VaR) [13], which estimates the maximum amount a portfolio could lose over a given time period at a given confidence level $1-\alpha$. Maximizing the Kelly criterion subject to a VaR constraint has been considered previously by MacLean et al. [41] in the context of allocating investment capital to stocks, bonds and cash over time. Let S represent the set of top four horse finishers with each $s \in S$ corresponding to a sequence of 4 horses, with $x = \{x_s\}$ being our decision variables dictating how much to wager on each outcome s, and $P(x)$ being the random payout given our decision vector x. Let the outcome probability of s be denoted as π_s, with $\pi_x = \sum_{s\in S} \pi_s \mathbb{1}_{\{x_s>0\}}$ being the probability of having a winning bet. We can now limit our betting strategy's VaR to be no greater than v by enforcing the chance constraint $\mathbb{P}(P(x) - \sum_{s\in S} x_s \geq -v) \geq 1-\alpha$.

VaR calculations typically use a small α, being concerned with large potential losses near the tail of the distribution. Tail risk is not a concern in our setting as the most that could possibly be lost is the amount we wager, which we expect to occur most of the time, in fact, a VaR constraint with $v > 0$ in our setting corresponds to a betting limit for $\alpha < 1 - \pi_x$.

Though risk measures concerning tail losses seem unapplicable, a VaR constraint with $v = 0$ enables the control of losing streaks. Let τ be the gambler's time horizon, for which we desire to set as the limit for potential losing streaks with high probability. For a betting decision x, let $\tilde{\pi}_x = \mathbb{P}(P(x) - \sum_{s\in S} x_s \geq 0)$ and $B_x \sim \text{binomial}(\tau, \tilde{\pi}_x)$ be the random number of times money is not lost repeating the race τ times with the same wager x. In order to enforce the gambler's time horizon, we require that $\mathbb{P}(B_x \geq 1) \geq 1 - \alpha$, which implies $\tilde{\pi}_x \geq 1 - \alpha^{\frac{1}{\tau}}$. Assuming independence between races, limiting betting decisions to those which have a VaR of 0 with confidence of at least $1-\alpha^{\frac{1}{\tau}}$ ensures that a non-negative return on a race will occur with a probability of at least $1 - \alpha$ over the next τ races.

4.2 Optimization model

A conceptual optimization model is displayed below. Using the Kelly criterion, the objective is to maximize the expected log of wealth, where w is the current wealth of the gambler.

$$
\begin{aligned}
&\max \mathbb{E}\log(P(x) + w - \sum_{s\in S} x_s) \\
\text{s.t.}\quad & \sum_{s\in S} x_s \leq w \\
& \mathbb{P}(P(x) - \sum_{s\in S} x_s \geq 0) \geq 1 - \alpha^{\frac{1}{\tau}} \\
& x_s \geq 0 \;\; s \in S
\end{aligned}
$$

4.3 Case study

The optimization model was tested using historical race data from the 2013-2014 season at Flamboro Downs, Hamilton, Ontario, Canada. This amounted to a total of 1,168 races. Race results, including the payouts, pool sizes, and final win bet odds were collected from TrackIT [57]. Handicapping data, generated by CompuBet [18], was collected from HorsePlayer Interactive [68]. The first 70% of the race dataset was used to calibrate the race outcome probabilities and payout model, with the remaining 30% of races used for out of sample testing.

4.3.1 Estimating outcome probabilities and payouts

The multinomial logistic model, first proposed by Bolton and Chapman [10], is the most widely used method of estimating the probability of each horse winning a race. Given a vector of handicapping data on each horse h, v_h, the horses are given a value $V_h = \beta^T v_h$, and assigned winning probabilities $\pi_h = \frac{e^{V_h}}{\sum_{i=1}^n e^{V_i}}$. A three factor model was used, including the log of the public's implied win probabilities from the win bet odds, $\log \pi_h^p$, and the log of two CompuBet factors, which were all found to be statistically significant at the $\alpha = 0.05$ level. The analysis was performed using the *mlogit* package [19] in *R*. Details of the handicapping data and the statistical estimation can be found in the subsection *Estimating win probabilities* in the appendix.

The Harville model [23] assumes the probability that a horse finishes m^{th} equals the probability that it wins against the horses that didn't finish $1^{st}, .., m-1^{th}$. The conditional probabilities are $\pi_{ij|i} = \frac{\pi_j}{1-\pi_i}$, $\pi_{ijk|ij} = \frac{\pi_k}{1-\pi_i-\pi_j}$, and $\pi_{ijkl|ijk} = \frac{\pi_l}{1-\pi_i-\pi_j-\pi_k}$, where for example, $\pi_{ijk|ij}$ is the probability estimate of horses i, j, and k finishing first, second, third, given horses i and j finished first and second. Multiplying together with π_i, $\pi_{ijkl} = \frac{\pi_i\pi_j\pi_k\pi_l}{(1-\pi_i)(1-\pi_i-\pi_j)(1-\pi_i-\pi_j-\pi_k)}$. This model was found to be biased towards favourite horses by Lo [34] and Lo and Bacon-Shone [35]. We use the improved approximation derived by Lo and Bacon-Shone [36], $\pi_{ijkl} = \pi_i \frac{\pi_j^{\lambda_1}}{\sum_{s\neq i}\pi_s^{\lambda_1}} \frac{\pi_k^{\lambda_2}}{\sum_{s\neq i,j}\pi_s^{\lambda_2}} \frac{\pi_l^{\lambda_3}}{\sum_{s\neq i,j,k}\pi_s^{\lambda_3}}$, where λ_1,

λ_2 and λ_3 are calibrated to the historical race data. As the log-likelihood is separable, optimal λ_i's were determined individually using multinomial logistic regression. The results of the statistical estimation can be found in the subsection *Estimating superfecta probabilities* in the appendix.

The superfecta payout function for sequence s is approximately $P_s(x) = x_s \frac{(Q+\sum_{u\in S} x_u)(1-t)}{Q_s+x_s}$, where Q is the superfecta pool size, Q_s is the amount wagered on sequence s by other gamblers, and $t = 24.7\%$ is the track take at Flamboro Downs. The payout per dollar wagered is rounded down to the nearest nickel, termed breakage, but this is unlikely to be significant and is omitted from the formula. The only information available to bettors is the value of Q. The approach taken to estimate Q_s is motivated by the work of Kanto and Renqvist [29] who fit the win probabilities of the Harville model to the money wagered on quinella bets using multinomial maximum likelihood estimation. The minimum superfecta bet allowed in practice is \$0.2 with \$0.2 increments. The amount wagered on sequence s is $Q_s = \frac{Q(1-t)}{P_s}$, where P_s is the amount paid on a \$1 wager. Let $n = 5Q_s$ be the number of bets placed on s out of $N = 5Q$, which we assume follows a binomial distribution. We model the public's estimate of superfecta outcome probabilities using a discount model with the public's implied win probabilities, so for $s = \{i, j, k, l\}$, $\pi^p_s = \frac{(\pi^p_i)^{\theta_1}}{\sum(\pi^p_h)^{\theta_1}} \frac{(\pi^p_j)^{\theta_2}}{\sum_{h\neq i}(\pi^p_h)^{\theta_2}} \frac{(\pi^p_k)^{\theta_3}}{\sum_{h\neq i,j}(\pi^p_h)^{\theta_3}} \frac{(\pi^p_l)^{\theta_4}}{\sum_{h\neq i,j,k}(\pi^p_h)^{\theta_4}}$. Let $\pi^p_{s,u}$ and $\pi^p_{s,l}$ represent the numerator and denominator of π^p_s. The likelihood function, using data from R historical races assumed to be independent, with w_r being the winning sequence in race r, is $\mathcal{L}(\theta) \propto \Pi^R_{r=1}(\pi^p_{w_r})^{n_r}(1-\pi^p_{w_r})^{N_r-n_r}$. The log-likelihood is a difference of concave functions, $\log \mathcal{L}(\theta) \propto \sum^R_{r=1} n_r \log(\pi^p_{w_r,u}) + (N_r - n_r)\log(\pi^p_{w_r,l} - \pi^p_{w_r,u}) - N_r \log(\pi^p_{w_r,l})$. This function was minimized twice using *fminunc* in Matlab, the first with an initial guess that the public uses the Harville model, $\theta_i = 1$, the second assuming that the public believes superfecta outcomes are purely random, $\theta_i = 0$, with both resulting in the same optimal solution. Statistical estimation results can be found in the subsection *Estimating public's superfecta probabilities* in the appendix. Given our estimate of the public's estimate of the probability of outcome s, π_s^p, we take $Q_s = \pi_s^p Q$, the expected amount wagered on s.

4.3.2 Optimization formulation

We now formulate the optimization program as it will be solved. Given its success in practice, we wager (approximately) half the Kelly amount. This is accomplished by first dividing the optimal solution in half, then rounding each bet to the closest multiple of \$0.2 to generate a valid wager. The chance constraint is implemented using binary variables, z_s, which indicate that a bet will be placed on outcome s by equalling 1 when $x_s \geq 0.4$, which can be implemented by the constraint $0.4z_s \leq x_s$. We then enforce the chance constraint by $\sum_{s\in S} \pi_s z_s \geq 1 - \alpha^{\frac{1}{\tau}}$. Note that given the expected large superfecta payouts, $\tilde{\pi}_x \approx \pi_x$, and so we have made the simplification of replacing $\tilde{\pi}_x$, requiring simply that $\pi_x \geq 1 - \alpha^{\frac{1}{\tau}}$.

$$
\begin{aligned}
\max \; & \sum_{s\in S} \pi_s \log\Big(x_s \frac{(Q + \sum_{u\in S} x_u)(1-t)}{Q_s + x_s} + w - \sum_{u\in S} x_u\Big) && (4.1)\\
\text{s.t.} \; & \sum_{s\in S} x_s \leq w \\
& \sum_{s\in S} \pi_s z_s \geq 1 - \alpha^{\frac{1}{\tau}} \\
& \left(\frac{Q_s + 0.4}{Q_s}\right)^{z_s} \leq \frac{Q_s + x_s}{Q_s} \quad s \in S \\
& z_s \in \{0,1\} \quad s \in S \\
& x_s \geq 0 \quad s \in S
\end{aligned}
$$

The objective function of (4.1) is not concave. We use the 1 to 1 mapping proposed by Kallberg and Ziemba [28], $y_s = \log(x_s + Q_s)$, resulting in the following program which is convex after relaxing the binary constraints on z_s. We have written the constraints $0.4z_s \leq x_s$ equivalently above as $\left(\frac{Q_s+0.4}{Q_s}\right)^{z_s} \leq \frac{Q_s+x_s}{Q_s}$ in order to achieve convex constraints after the change of variable.

$$
\begin{aligned}
\max \; & \sum_{s\in S} \pi_s \log(Q + w - (t + (1-t)Q_s e^{-y_s}) \sum_{u} e^{y_u}) && (4.2)\\
\text{s.t.} \; & \sum_{s\in S} e^{y_s} \leq w + Q \\
& \sum_{s\in S} \pi_s z_s \geq (1 - \alpha^{\frac{1}{\tau}}) \\
& z_s \ln\left(\frac{Q_s + 0.4}{Q_s}\right) \leq y_s - \log Q_s \quad \forall s \in S \\
& z_s \in \{0, 1\} \quad \forall s \in S \\
& y_s \geq \log(Q_s) \quad \forall s \in S
\end{aligned}
$$

4.3.3 Implementation

All computation was conducted on a Windows 7 Home Premium 64-bit, Intel Core i5-2320 3GHz processor with 8 GB of RAM, in Matlab R2016a using OPTI toolbox v2.16. For each race, IPOPT [63] was first used to solve (4.2) without the time horizon constraint. If $\sum_{s\in S} x_s = 0$, we do not bet on the current race and if $\sum_{s\in S} \pi_s z_s \geq 1-\alpha^{\frac{1}{\tau}}$ we take this as the solution. If $\sum_{s\in S} x_s > 0$ but $\sum_{s\in S} \pi_s z_s < 1 - \alpha^{\frac{1}{\tau}}$, we proceed to solve the full problem using Bonmin's [11] *B-Hyb* algorithm. None of the default stopping criteria was altered in OPTI's optimization settings, so the maximum execution time was limited to 1000 seconds, the maximum number of iterations to 1500 and the maximum function evaluations to 10000. With these settings it was not always guaranteed that the optimal solution was found. In order to improve convergence and solution quality, cuts were added as described in the following paragraph.

If $\pi_j > \pi_i$ and $Q_j \leq Q_i$ or $\pi_j \geq \pi_i$ and $Q_j < Q_i$ we say that s_j dominates s_i and we will have $z_j \geq z_i$ in the optimal solution. Likewise, if $\pi_j < \pi_i$ and $Q_j \geq Q_i$ or $\pi_j \leq \pi_i$ and $Q_j > Q_i$ we say that s_j is dominated by s_i and we will have $z_j \leq z_i$. Let D_j be the set of outcomes which s_j dominates and let B_j be the set of outcomes which dominate s_j. In order to maintain a manageable set of constraints, the constraints $z_j \geq z_i \; i \in D_j$ are added together to form the single constraint, $z_j \geq \frac{1}{|D_j|}\sum_{i\in D_j} z_i$, as are the constraints $z_j \leq z_i \; i \in B_j$, to form the single constraint $z_j \leq \frac{1}{|B_j|}\sum_{i\in B_j} z_i$. These cuts were added to the above formulation for all outcomes.

4.3.4 Results

Testing was done on a total of 350 races. Given our optimal betting solution, the realized payout was calculated by adjusting the published payout to account for our wagers and breakage. Four simulations were done with the gambler's initial wealth set to \$5,000. The wealth through time for all are plotted in Figure 4.1, with statistics displayed in Table 4.1. A preliminary simulation was done with $\tau = \infty$. The longest losing streak was found to be 55 races. Given this number, simulations were done with $\tau = 40$, 30 and 20, with $\alpha = 0.05$.

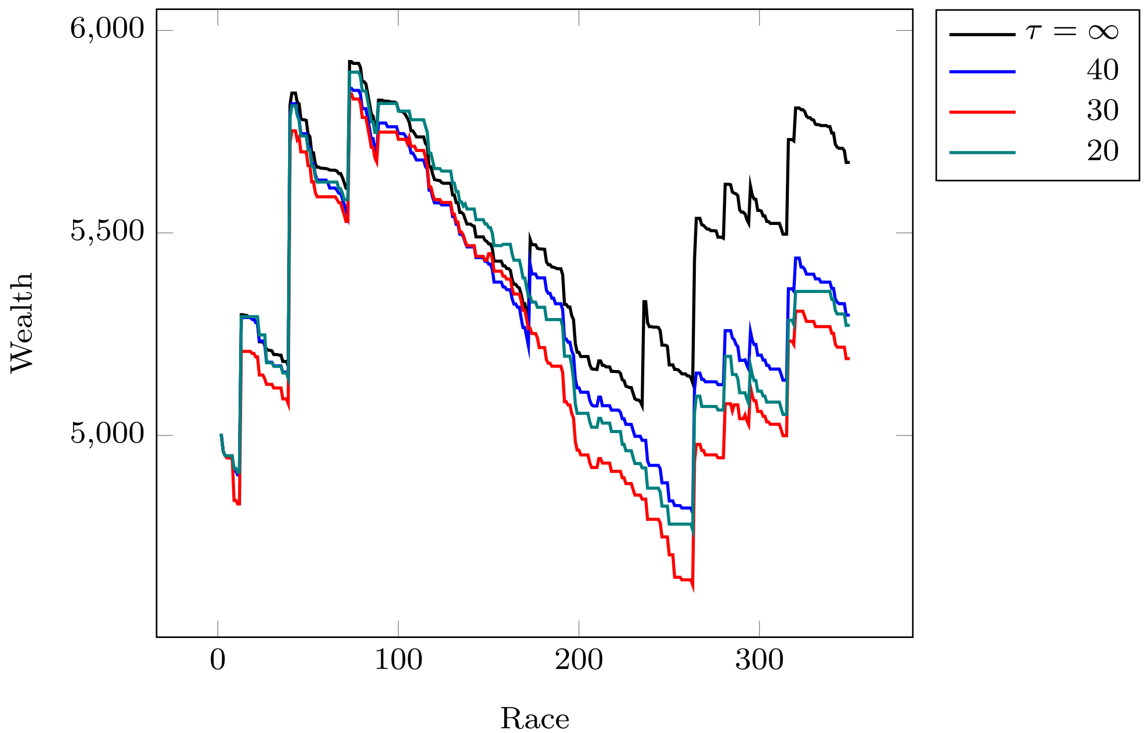

Figure 4.1: Wealth over the course of 350 races at Flamboro Downs.

τ	Tot ret (%)	σ ($\times 10^{-3}$)	Races bet	BPR	Loss streak	Num wins
∞	12.65	10.0	232	11.7974	55	14
40	5.77	9.8	148	19.0189	34	15
30	3.72	9.8	130	21.5646	26	17
20	5.29	9.5	105	25.3048	17	14

Table 4.1: Optimization results

Examining Table 4.1, Tot ret is the total return over the 350 races, σ is the standard deviation in race returns, Races bet is the total number of races actually bet on, BPR is the average bet per race, Loss streak is the maximum losing streak over races bet on and Num wins is the total number of races for which a profit was made. Through the use of a chance constraint, the length of losing streaks were successfully limited to the chosen time horizon, but we can see there is a trade off between risk and return, resulting in a reduction in profit using the chance constrained model. We also note that σ decreases with τ. An increase in σ using the chance constrained model would make it clearly undesirable given its inferior return, so we take this as positive evidence towards the soundness of this methodology.

4.4 Conclusion and future research

We have developed a methodology for limiting losing streaks given a gambler's time horizon through the use of chance constrained optimization, exemplified in exotic horse race wagering. Initial results using one season of historical racing data have been presented which show the viability of the method by effectively limiting losing streaks for different chosen time horizons. Certain approximations were used which could be addressed in future research. Point estimates of outcome probabilities, π_s, as well as the amount wagered on each outcome by the public, Q_s, were utilized. Taking into account the uncertainty of these estimates could improve results. Though the focus of this work has been on horse racing, we feel this general methodology is applicable for any gambling or investing setting which have low probability outcomes with high payouts, such as for investing in deep out of the money options.

4.5 Appendix

4.5.1 Estimating win probabilities

A number of factors and their logarithms were considered, displayed in Table 4.2 below. The domain of each factor is listed in brackets, but all were normalized to be between 0 and 1 for statistical use. The first six factors are from Compubet, with the other two from the race program and result.

Factor description	
Post	Starting position of the horse (1-9).
Pre	The quality of the data available for each horse (30-100).
Form	The overall success of this horse in recent starts (10-130).
Class	The horse's performance relative to the class of its competition in recent races (52.8-95).
Speed	An adjusted speed rating using complex statistical analysis, the daily track variant, track condition, and the track-to-track speed variant (113.3-128.1 seconds).
Driver Points	The driver's rating (4-39).
π_h^{ML}	The winning probability implied by the morning line odds.
π_h^m	The winning probability implied by the final winning bet odds.

Table 4.2: Win probability considered factors.

Systematically removing the least significant factor with a significance $\alpha > 0.05$ resulted in the parameter estimation in Table 4.3.

π_h **Coefficients**		
Coefficient	**Estimate**	**P-Value**
$\log(\pi_h^m)$	1.08318	$< 2.2e-16$
$\log(Pre)$	0.42104	0.02577
$\log(Class)$	0.72842	0.01093

Table 4.3: Win Probability Coefficients

The McFadden [47] R^2 goodness of fit measure was used to compare the public's implied winning probabilities to the model's, where $R^2 = 1$ implies perfect predictive ability and $R^2 = 0$ means predictability is no better than random guessing. Using the last 30% of the racing data, $R^2_{\pi_h} = 0.218077$ and $R^2_{\pi_h^m} = 0.214455$. We see the model has a small positive "edge" of $\Delta R^2 = R^2_{\pi_h} - R^2_{\pi_h^m} = 0.0036$ over the general public.

4.5.2 Estimating superfecta probabilities

Below are the results of estimating the λ^i parameters.

Superfecta probability parameters		
Coefficient	**Estimate**	**P-Value**
λ^1	0.600548	$< 2.2e-16$
λ^2	0.384509	$< 2.2e-16$
λ^3	0.26239	$7.767e-13$

Table 4.4: Superfecta probability parameters

4.5.3 Estimating public's superfecta probabilities

Below are the results of estimating the θ^i parameters.

Superfecta probability parameters		
Coefficient	**Estimate**	**P-Value**
θ^1	1.2058	$< 2.2e-16$
θ^2	0.8215	$< 2.2e-16$
θ^3	0.5312	$< 2.2e-16$
θ^4	0.4146	$< 2.2e-16$

Table 4.5: Superfecta probability parameters

The p-values were found using the estimated Hessian at the optimal solution supplied by *fminunc* to calculate the observed Fisher information.

Chapter 5

Conclusion

There are many possible applications for chance constraints, with varying roles within optimization problems. In this thesis we have investigated three applications of chance constrained optimization in operations management. In Chapter 2, chance constraints served as a form of quality control, ensuring a certain level of consumer demand is satisfied with high probability. We examined the effect of over and underestimating the influence of price and advertising on demand for new product production planning, and from an empirical study, we were able to determine prudent methods of estimation. Chance constraints can be of necessity to capture realities of the problem. In Chapter 3, we examined chance constrained optimization for guaranteed display Internet advertising campaigns, where the randomness of Internet viewer supply hinders our ability to satisfy campaigns with certainty. Sample approximations were developed, with a branching heuristic accelerating lower bound computation time by over an order of magnitude, as well as convex approximations, with an iterative algorithm which effectively tightened upper bounds. Chapter 4 focused on risk management in exotic horse race wagering, with chance constraints enabling the consideration of the time horizon of a gambler. A proof of concept showing its viability was conducted using one season of historical race data, where the length of losing streaks were successfully limited for different time periods. It is quite rare when faced with managerial decisions to have all required information with certainty. This thesis has shown that with chance constraints, we are able to incorporate the reality of uncertainty into decision making.

Bibliography

[1] M. T. Ahmed and C. Kwon. Optimal contract-sizing in online display advertising for publishers with regret considerations. *Omega: The International Journal of Management Science*, 42(1):201–212, 2014.

[2] S. Ahmed. Convex relaxations of chance constrained optimization problems. *Optimization Letters*, 8(1):1–12, 2014.

[3] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[4] P. Asch, B. G. Malkiel, and R. E. Quandt. Market Efficiency in Racetrack Betting. *The Journal of Business*, 57(2):165–75, 1984.

[5] F. M. Bass. A New Product Growth for Model Consumer Durables. *Management Science*, 15(5):215–227, 1969.

[6] F. M. Bass, T. V. Krishnan, and D. C. Jain. Why the Bass Model Fits Without Decision Variables. *Marketing science*, 13(3):203–223, 1994.

[7] W. Benter. Computer Based Horse Race Handicapping and Wagering Systems: A Report. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba, editors, *Efficiency of Racetrack Betting Markets*, pages 183–198. World Scientific, 2008.

[8] V. Bharadwaj, M. A. Saunders, and J. A. Tomlin. Sensitivity of an Inventory Allocation Model for Online Advertising with Stochastic Supply. *Presentation at the 24th European Conference on Operational Research*, 2010.

[9] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 2011.

[10] R. N. Bolton and R. G. Chapman. Searching for Positive Returns at the Track: A Multinomial Logic Model for Handicapping Horse Races. *Management Science*, 32(8):1040–1060, 1986.

[11] P. Bonami, L. T. Biegler, A. R. Conn, G. Cornuéjols, I. E. Grossmann, C. D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, and A. Wächter. An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization*, 5(2):186 – 204, 2008.

[12] S. Boyd and L. Vanderberghe. *Convex Optimization*. Cambridge University Press, Cambridge, U.K., 2004.

[13] Paolo Brandimarte. *Numerical methods in finance and economics: a MATLAB-based introduction*. John Wiley & Sons, 2006.

[14] G. C. Calafiore and M. C. Campi. The Scenario Approach to Robust Control Design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[15] M. C. Campi and S. Garatti. A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011.

[16] B. Carlson. *Blackjack for Blood*. Pi Yee Press, 2001.

[17] F. Y. Chen and D. Krass. Inventory models with minimal service level constraints. *European Journal of Operational Research*, 134(1):120 – 140, 2001.

[18] CompuBet. CompuBet. `https://compubet.com`. Accessed: 2014-06-02.

[19] Y. Croissant. Estimation of multinomial logit models in R: The mlogit Packages. *R package version 0.2.4*, 2012.

[20] J. Deane and A. Agarwal. Scheduling online advertisements to maximize revenue under variable display frequency. *Omega: The International Journal of Management Science*, 40(5):562–570, 2012.

[21] A. Deza, K. Huang, and M. R. Metel. Chance constrained optimization for targeted internet advertising. *Omega: The International Journal of Management Science*, 53:90–96, 2015.

[22] S. Girgin, J. Mary, P. Preux, and O. Nicol. Managing advertising campaigns - an approximate planning approach. *Frontiers of Computer Science*, 6(2):209–229, 2012.

[23] D. A. Harville. Assigning Probabilities to the Outcomes of Multi-Entry Competitions. *Journal of the American Statistical Association*, 68(342):312–316, 1973.

[24] D. B. Hausch, W. T. Ziemba, and M. Rubinstein. Efficiency of the Market for Racetrack Betting. *Management science*, 27(12):1435–1452, 1981.

[25] R. B. Holmes. On Random Correlation Matrices. *SIAM Journal on Matrix Analysis and Applications*, 12(2):239–272, 1991.

[26] Interactive Advertising Bureau. IAB internet advertising revenue report: 2012 full year results. 2013.

[27] R. Isaacs. Optimal Horse Race Bets. *American Mathematical Monthly*, 60(5):310–315, 1953.

[28] J. G. Kallberg and W. T. Ziemba. Concavity Properties of Racetrack Betting Models. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba, editors, *Efficiency of Racetrack Betting Markets*, pages 99–107. World Scientific, 2008.

[29] A. Kanto and G. Rosenqvist. On the Efficiency of the Market for Double (Quinella) Bets at a Finnish Racetrack. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba, editors, *Efficiency of Racetrack Betting Markets*, pages 485–498. World Scientific, 2008.

[30] M. Kaplan. The High Tech Trifecta. *Wired Magazine*, 10:10–13, 2002.

[31] J. L. Kelly. A New Interpretation of Information Rate. *Information Theory, IRE Transactions on*, 2(3):185–189, 1956.

[32] S. X. Li. An insurance and investment portfolio model using chance constrained programming. *Omega: The International Journal of Management Science*, 23(5):577–585, 1995.

[33] M. K. Lim, A. Bassamboo, S. Chopra, and M. S. Daskin. Facility Location Decisions with Random Disruptions and Imperfect Estimation. *Manufacturing & Service Operations Management*, 15(2):239–249, 2013.

[34] V. S. Y. Lo. Application of Running Time Distribution Models in Japan. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba, editors, *Efficiency of Racetrack Betting Markets*, pages 237–247. World Scientific, 2008.

[35] V. S. Y. Lo and J. Bacon-Shone. A Comparison Between Two Models for Predicting Ordering Probabilities in Multiple-Entry Competitions. *The Statistician*, pages 317–327, 1994.

[36] V. S. Y. Lo and J. Bacon-Shone. Approximating the Ordering Probabilities of Multi-Entry Competitions by a Simple Method. *Handbook of Sports and Lottery Markets*, 2008.

[37] J. Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *IEEE International Symposium on Computer Aided Control Systems Design 2004*, pages 284–289, 2004.

[38] J. Luedtke. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming*, 146(1-2):219–244, 2014.

[39] J. Luedtke and S. Ahmed. A Sample Approximation Approach for Optimization with Probabilistic Constraints. *SIAM Journal on Optimization*, 19(2):647–699, 2008.

[40] J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2):247–272, 2010.

[41] L. C. MacLean, R. Sanegre, Y. Zhao, and W. T. Ziemba. Capital growth with security. *Journal of Economic Dynamics and Control*, 28(5):937–954, 2004.

[42] L. C. MacLean, E. O. Thorp, and W. T. Ziemba. *The Kelly Capital Growth Investment Criterion: Theory and Practice*, volume 3. World Scientific, 2011.

[43] L. C. MacLean, W. T. Ziemba, and G. Blazenko. Growth versus Security in Dynamic Investment Analysis. *Management Science*, 38(11):1562–1585, 1992.

[44] W.-K. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(12):47 – 56, 1999.

[45] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[46] P. Massart. The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *The Annals of Probability*, pages 1269–1283, 1990.

[47] D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

[48] A. Nemirovski and A. Shapiro. Convex Approximations of Chance Constrained Programs. *SIAM Journal of Optimization*, 17:969–996, 2006.

[49] S.-C. Niu. A Piecewise-Diffusion Model of New-Product Demands. *Operations Research*, 54(4):678–695, 2006.

[50] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample Average Approximation Method for Chance Constrained Programming: Theory and Applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009.

[51] J. Pintér. Deterministic Approximations of Probability Inequalities. *Zeitschrift für Operations Research*, 33(4):219–239, 1989.

[52] W. Poundstone. *Fortune's Formula*. Hill and Wang, New York, 2005.

[53] A. Prékopa. *Stochastic Programming*. Springer, 1995.

[54] R. N. Rosett. Gambling and rationality. *The Journal of Political Economy*, 73(6):595, 1965.

[55] A. Ruszczyński. Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming*, 93(2):195–215, 2002.

[56] P. Smoczynski and D. Tomkins. An explicit solution to the problem of optimizing the allocations of a bettor's wealth when wagering on horse races. *Mathematical Scientist*, 35(1):10–17, 2010.

[57] Standardbred Canada. TrackIT. `https://trackit.standardbredcanada.ca`. Accessed: 2014-06-02.

[58] E. O. Thorp. The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market. In S. A. Zenios and W. T. Ziemba, editors, *Handbook of Asset and Liability Management, Volume I*, pages 385–428. Elsevier, 2006.

[59] E. O. Thorp. Understanding the Kelly criterion. *Wilmott Magazine*, May 2008.

[60] J. A. Tomlin. An Entropy Approach to Unintrusive Targeted Advertising on the Web. In *Proceedings of the 9th International World Wide Web Conference*, pages 767–774, 2000.

[61] J. Turner. The Planning of Guaranteed Targeted Display Advertising. *Operations Research*, 60(1):18–33, 2012.

[62] J. P. Vielma, S. Ahmed, and G. L. Nemhauser. Mixed-Integer Models for Nonseparable Piecewise-Linear Optimization: Unifying Framework and Extensions. *Operations Research*, 58(2):303–315, 2010.

[63] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[64] S. Wang and G. H. Huang. An Integrated Approach for Water Resources Decision Making Under Interactive and Compound Uncertainties. *Omega: The International Journal of Management Science*, 44:32–40, 2014.

[65] M. Weitzman. Utility Analysis and Group Behavior: An Empirical Study. *Journal of Political Economy*, February, 1965.

[66] C. X. Wong. *Precision: Statistical and Mathematical Methods in Horse Racing*. Outskirts Press, Inc., Denver, Colorado, 2011.

[67] Stanford Wong. *Sharp Sports Betting*. Pi Yee Press, 2009.

[68] Woodbine Entertainment Group. Horseplayer Interactive. `http://www.horseplayerinteractive.com`. Accessed: 2014-06-02.

[69] J. Yang, E. Vee, S. Vassilvitskii, J. A. Tomlin, J. Shanmugasundaram, and T. Anastasakos. Inventory Allocation for Online Graphical Display Advertising. *Proceedings of the 1st International Conference on Operations Research and Enterprise Systems*, 2012.

[70] Paul Herbert Zipkin. *Foundations of Inventory Management*. McGraw-Hill New York, 2000.