McMaster University

Advanced Optimization Laboratory

Title:

Chance constrained optimization for parimutuel horse race betting

Authors:

Antoine Deza, Kai Huang, and Michael R. Metel

AdvOL-Report No. 2015/2

March 2015, Hamilton, Ontario, Canada

Chance constrained optimization for parimutuel horse race betting

Antoine Deza[1], Kai Huang[2], and Michael R. Metel[3]

[1] *Advanced Optimization Laboratory, Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada,* deza@mcmaster.ca
[2] *DeGroote School of Business, McMaster University, Hamilton, Ontario, Canada,* khuang@mcmaster.ca
[3] *School of Computational Science and Engineering, McMaster University, Hamilton, Ontario, Canada,* metelm@mcmaster.ca

March 23, 2015

Abstract

We consider the time horizon of a gambler in the optimization of horse race betting through the use of chance constrained programming. The optimization problem is formulated as a mixed integer nonlinear program for which global optimal solutions are found using optimization tools. A novel approach to estimating superfecta payouts is presented using maximum likelihood estimation. A computational substantiation with historical race data found an increase in return of over 10% using the chance constrained model.

1 Introduction

Beginning in 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 on stock exchanges [12]. 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.

Optimization in the horse racing literature can be traced back to Isaacs deriving a closed form solution for the optimal win bets when maximizing expected profit in 1953 [9]. Hausch et al. [8] 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 [13], maximizing the expected log utility of wealth and found profitability when limiting the betting to when the expected return was greater than a fixed percentage. More recently, Smoczynski and Tomkins derived a simple procedure for the optimal win bets under the Kelly criterion using the KKT conditions [16]. Although the Kelly criterion maximizes the asymptotic rate of asset growth, the volatility of wealth through time is too large for most, resulting in many professional investors employing a fractional Kelly criterion [17], which has been shown to possess favourable risk-return properties by MacLean et al. [15]. We investigate a further manner of risk management in the form of a chance constraint, taking into account the time horizon of the bettor, which can be employed in conjunction with the Kelly criterion.

There are several different types of wagers one can place on horses, but in order to best display the effect of the chance constraint, we concentrate on the riskiest of bets on a single race, the superfecta, which requires the bettor to pick, in order, the first 4 finishers.

2 Optimization Model

2.1 Time Horizon

To motivate the discussion, we examine the 4 horse outcome probabilities of race 5 on March 20, 2014 at Flamboro Downs, Hamilton, Ontario, Canada. Information about the race dataset and how these probabilities are estimated can be found in Section 3. Let S represent the set of top 4 horse finishes with each $s \in S$ corresponding to a sequence of 4 horses. If we bet on this race an infinite number of times, then the average number of races before a superfecta bet on outcome s pays off would be $\frac{1}{\pi_s}$, where π_s is the outcome's probability. Summary statistics for the average wait time is in the following table.

Statistic	**Races**
min	141
max	566, 225
median	13, 600
mean	38, 192

Table 1: Average wait time statistics

The median wait time for a superfecta bet to payoff is then over 11 seasons with roughly 1,200 races per season. Assuming the horseplayer requires some form of regular income or desires to at least turn a profit every season, consideration of the likelihood of receiving a payoff is warranted. In particular, we can limit betting strategies to those which pay out with high probability over a number of races equal to the desired time horizon, τ. Let $x = \{x_s\}$ be our decision variables dictating how much to wager on each outcome s. For a betting decision $\hat{x}$, let $B_{\hat{x}} \sim \text{binomial}(\tau, \pi_{\hat{x}})$, where $\pi_{\hat{x}}$ is the probability of a payout. In order to enforce the gambler's time horizon, we require that $\mathbb{P}(B_{\hat{x}} \geq 1) \geq 1 - \alpha$, where α is our error tolerance, which is chosen arbitrarily small. Rearranging, we require $\pi_{\hat{x}} \geq 1 - \alpha^{\frac{1}{\tau}}$. Assuming independence between races, limiting betting decisions to having a payout probability of at least $1-\alpha^{\frac{1}{\tau}}$ ensures that a payout will occur with probability at least $1-\alpha$ over τ races.

2.2 Optimization Program

The objective is to maximize the exponential rate of return. Let $P_\rho(x)$ be the random payout given our decision vector x. The payout uncertainty stems from the result of the race, ρ, with S as its sample space. Let w be the current wealth of the gambler. Incorporating the gambler's time horizon through the use of a chance constraint, the optimization problem is below.

$$
\begin{aligned}
\max\ & \mathbb{E}\log(P_\rho(x) + w - \sum_{s\in S} x_s) \\
\text{s.t.}\ & \sum_{s\in S} x_s \le w \\
& \mathbb{P}(P_\rho(x) > 0 | \sum_{s\in S} x_s > 0) \ge 1-\alpha^{\frac{1}{\tau}} \\
& x_s \ge 0\ \forall s \in S
\end{aligned}
$$

The chance constraint is conditional on there being favourable bets to be placed, as we do not want to decrease our expected utility below $\log(w)$ to satisfy it. We assume that the frequency with which we are forced to abstain from gambling is sufficiently small so as not to significantly alter our effective time horizon.

3 Computational Substantiation

The optimization model was tested using historical race data from the 2013-2014 season at Flamboro Downs. 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 [3]. Handicapping data, generated by CompuBet [4], was collected from HorsePlayer Interactive [6]. The first 70% of the race dataset was used to calibrate the race outcome probabilities and payout models, with the remaining 30% of races used for out of sample testing.

3.1 Estimating Outcome Probabilities and Payouts

The multinomial logistic model, first proposed by Bolton and Chapman [1], was used to estimate win probabilities. 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. The analysis was performed using the *mlogit* package [5] in *R*. The discount model, derived by Lo and Bacon-Shone [14], was used to estimate the order probabilities, $\pi_{ijkl} = \pi_i \frac{\pi_j^{\lambda_1}}{\sum_{s\neq i}\pi_s^{\lambda_1}} \frac{\pi_k^{\lambda_2}}{\sum_{s\neq i,j}\pi_s^{\lambda_2}} \frac{\pi_l^{\lambda_3}}{\sum_{s\neq i,j,k}\pi_s^{\lambda_3}}$, where optimal λ_i's were determined using multinomial logistic regression.

Let Q and Q_s be the superfecta pool size, and the total amount wagered on sequence s. 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 [11] who fit the win probabilities of the Harville model [7] to the money wagered on Quinella bets using multinomial maximum likelihood estimation. The amount wagered on sequence s is $Q_s = \frac{Q(1-t)}{P_s}$, where $t = 24.7\%$ is the track take at Flamboro Downs and P_s is the \$1 payout. The minimum superfecta bet allowed in practice is \$0.2 with \$0.2 increments, so 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 outcome probabilities using the discount model with their implied win probabilities, so for $s = \{i,j,k,l\}$, $\pi_s^p = \frac{(\pi_i^p)^{\theta_1}}{\sum(\pi_h^p)^{\theta_1}} \frac{(\pi_j^p)^{\theta_2}}{\sum_{h\neq i}(\pi_h^p)^{\theta_2}} \frac{(pi_k^p)^{\theta_3}}{\sum_{h\neq i,j}(\pi_h^p)^{\theta_3}} \frac{(\pi_l^p)^{\theta_4}}{\sum_{h\neq i,j,k}(\pi_h^p)^{\theta_4}} = \frac{(\pi_s^p)^u}{(\pi_s^p)^l}$. The likelihood function, using data from R historical races assumed to be independent, with w_r being the winning sequence in race r, is $\mathcal{L}(\theta) \propto \Pi_{r=1}^R (\pi_{w_r}^p)^{n_r}(1-\pi_{w_r}^p)^{N_r - n_r}$. The negative log-likelihood is a difference of convex functions, $-\log\mathcal{L}(\theta) \propto \sum_{r=1}^R N_r \log((\pi_{w_r}^p)^l) - (n_r \log((\pi_{w_r}^p)^u) + (N_r - n_r)\log((\pi_{w_r}^p)^l - (\pi_{w_r}^p)^u))$. 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 solution. The payout function is $P_s(x) = x_s \frac{(Q+\sum_{u\in S} x_u)(1-t)}{Q_s + x_s}$, where we take $Q_s = \pi_s^p Q$, the expected amount wagered on s.

3.2 Optimization Program Formulation

Our optimization program now has the following form. When testing the model we round down the optimal solution to the nearest 0.2 to avoid overbetting. The z_s variables are used to indicate when $x_s \geq 0.2$, implying $P_{s,\xi_s}(x) > 0$ and $\bar{z}$ nullifies the chance constraint when $\sum_{s\in S} x_s = 0$.

$$
\begin{aligned}
\max \quad & \sum_{s\in S} \pi_s \log(x_s \frac{(Q + \sum_{u\in S} x_u)(1-t)}{Q_s + x_s} + w - \sum_{u\in S} x_u) \\
\text{s.t.} \quad & \sum_{s\in S} x_s \leq w\bar{z} \\
& \sum_{s\in S} \pi_s z_s \geq (1 - \alpha^{\frac{1}{\tau}})\bar{z} \\
& \left(\frac{Q_s + 0.2}{Q_s}\right)^{z_s} \leq \frac{Q_s + x_s}{Q_s} \quad && \forall s \in S \\
& \bar{z}, z_s \in \{0,1\} && \forall s \in S \\
& x_s \geq 0 && \forall s \in S
\end{aligned}
$$

We use the 1 to 1 mapping proposed by Kallberg and Ziemba [10], $y_s = \log(x_s + Q_s)$, which results in the following program whose linear relaxation is convex.

$$\begin{aligned}
\max \quad & \sum_{s \in S} \pi_s \log(Q + w - (t + (1-t)Q_s e^{-y_s}) \sum_u e^{y_u}) \\
\text{s.t.} \quad & \sum_{s \in S} e^{y_s} \leq w\bar{z} + Q \\
& \sum_{s \in S} \pi_s z_s \geq (1 - \alpha^{\frac{1}{\tau}})\bar{z} \\
& z_s \ln\left(\frac{Q_s + 0.2}{Q_s}\right) \leq y_s - \log Q_s \quad \forall s \in S \\
& \bar{z}, z_s \in \{0, 1\} \quad \forall s \in S \\
& y_s \geq \log(Q_s) \quad \forall s \in S
\end{aligned}$$

3.3 Results

The model was tested 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. The gambler's wealth over the course of the races was calculated using the optimization program with and without the chance constraint, Opt^+ and Opt respectively. Initial wealth was set to \$1000, with the time horizon set to $\tau = 350$ and $\alpha = 0.01$. All testing 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 with the OPTI toolbox, using the IPOPT[18] and Bonmin[2] solvers.

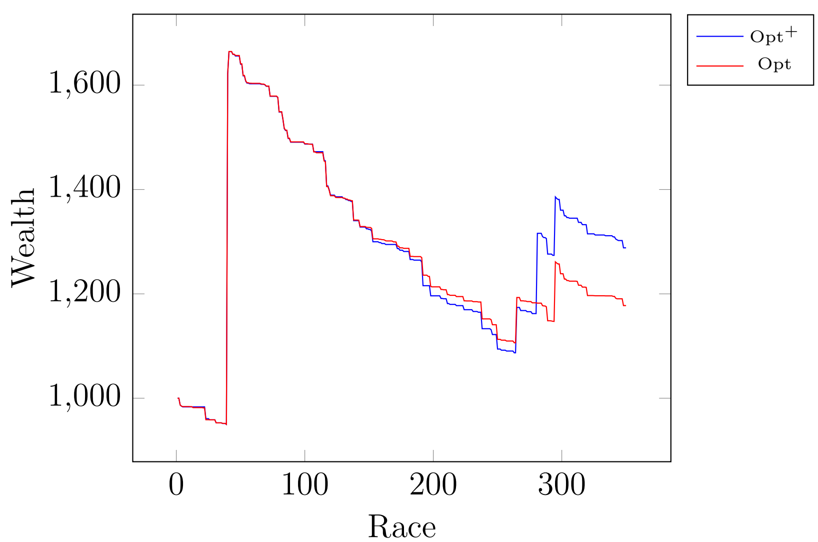

Figure 1: Wealth over the course of 350 races at Flamboro Downs.

The result in Figure 1 is intuitive, as Opt^+ attempts to mimic Opt, while generally having to take on extra bets to satisfy the chance constraint. This extra cost results in a lower wealth until one of these extra wagers does in fact payout, which occured at approximately race 280, resulting in a superior return of 28.8% compared to 17.8% for Opt.

4 Conclusion

We presented a chance constrained optimization model for parimutuel horse race betting, as well as a method for estimating superfecta bet payouts. Profitability was achieved when employing the Kelly criterion, with a superior return when taking into consideration the gambler's time horizon.

References

[1] 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.

[2] 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.

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

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

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

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

[7] D. A. Harville. Assigning probabilities to the outcomes of multi-entry competitions. *Journal of the American Statistical Association*, 68(342):312–316, 1973.

[8] D. B. Hausch, W. T. Ziemba, and M. Rubinstein. Efficiency of the market for racetrack betting. *Management Science*, 27(12):1435–1452, 1981.

[9] R. Isaacs. Optimal horse race bets. *American Mathematical Monthly*, 60(5):310–315, 1953.

[10] 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.

[11] 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.

[12] M. Kaplan. The high tech trifecta. *Wired Magazine*, 10:10–13, 2002.

[13] J. L. Kelly. A new interpretation of information rate. *Information Theory, IRE Transactions on*, 2(3):185–189, 1956.

[14] 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*, pages 51–65, 2008.

[15] L. C. MacLean, W. T. Ziemba, and G. Blazenko. Growth versus security in dynamic investment analysis. *Management Science*, 38(11):1562–1585, 1992.

[16] 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.

[17] E. Thorp. Understanding the Kelly criterion. *Wilmott Magazine (May 2008).*

[18] 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.