Racetrack Betting and Consensus of Subjective Probabilities

Lawrence D. Brown[1] and Yi Lin[2]

University of Pennsylvania and University of Wisconsin, Madison

Abstract

In this paper we consider the dynamic process of race track betting. We show that there is a close connection between the dynamic race track betting process and the pari-mutuel method for constructing consensus of subjective probabilities considered in Eisenberg and Gale. This enables us to show that there exists a unique equilibrium point for the betting process. We further show that the dynamic betting process converges to this equilibrium point almost surely. Therefore the sequential race track betting gives a natural approach to inducing the consensus probabilities in Eisenberg and Gale. These consensus probabilities are different from the average of the subjective probabilities which is used in the conventional way of combining individually held opinions into a collective group statement. We compare these probabilities and this leads to a potential explanation of the favorite-longshot bias consistently observed in the studies of race track betting.

Key words and phrases. Equilibrium probabilities, favorite-longshot bias, pari-mutuel betting.

1 Introduction

1.1 Some previous literature

Racetrack betting uses a pari-mutuel system. In the win bet market, the sums wagered on each horse are pooled, and the racetrack takes a percentage of the total pool. The remaining amount is distributed to those people who bet on the winning horse. Hence the payoff on each unit amount of wager depends on the proportion of total wager bet on each horse. Similar but more complicated systems are used for place betting, show betting, and some other possible types of betting. For simplicity, in this article we only consider win betting.

Numerous authors have analyzed horse racing data, and a negative relationship between rate of return and track odds has been noted consistently: favorites (horses with high proportion of total wager bet on them, or, in other words, low track odds) have positive expected

[1]Supported in part by NSF Grant DMS-9971751

[2]Supported in part by NSF Grant DMS-0134987

returns (after adjusting for track take), while longshots have negative expected returns. This is called the favorite-longshot bias. See Griffith (1949), McGlothlin (1956), Fabricand (1965), Ali (1977), Snyder (1978), Hausch, Ziemba, and Rubinstein (1981), Asch, Malkiel, and Quandt (1982), Henery (1985), Ziemba and Hausch (1984), and Brown, D'Amato, and Gertner (1994). Exceptions to this bias are Busche and Hall (1988) and Busche (1994) for Hong Kong and Japan.

Several explanations of this bias have been proposed in the literature. One obvious possibility is that the bias may reflect a taste for low probability, high payoff gambles — a form of risk love. Weitzman (1965) and Ali (1977) estimated the utility function of the representative bettor and showed it to be convex. See also Brown, D'Amato, and Gertner (1994). Another explanation is that the bias may represents a general systematic bias in probability perception, see Thaler and Ziemba (1988). Many of these analyses assume homogeneous beliefs on the part of the bettors, whereas in reality the perceptions of the bettors may differ. Ali (1977) considered races with two horses and assumed that the risk-neutral bettors hold heterogeneous beliefs about the winning probability of each horse. Ali supposed that each bettor's belief of the winning probability of the horses is a draw from a distribution that has the true underlying probability as its median value, and all bettors wager an identical amount. He showed that the favorite-longshot bias will be generated in this case. Blough (1994) extended Ali's analysis beyond two horses to an arbitrary number. He proposed a concept of consensus probability, which generalized the concept of median in a certain way, and showed that, if the consensus probability of the bettors matches the true probability, then with a restriction on the beliefs of bettors, the favorite-longshot bias results. Both of the above analyses proceed as if the bettors know (or can predict) the market established final odds.

1.2 The dynamic process

In this paper, we assume heterogeneous beliefs among the bettors, and consider a sequential dynamic model of the betting process. Consider a race with J horses. Assume the bettors are risk neutral. Suppose at the end of the betting process, the fraction of total wager placed on horse j is P_j, $j = 1, ..., J$. Denote $P = (P_1, P_2, ..., P_J)^T$. Let the racetrack take percentage be r. The net return on a dollar bet on horse j if that horse wins is $(1-r)/P_j - 1$. This net return is the payoff odds on horse j. Hence the proportion of total wager bet on each horse can be inferred from the published payoff odds on each horse. Of course the final proportion can be calculated only when the betting is over and the final odds are available. In the case when the bettors have homogeneous beliefs, the commonly perceived winning probabilities of the horses would be identical to P. This is why P is often called the subjective probability of the bettors. In this paper we use the name market probability (used in Blough (1994)) for P, since we consider heterogeneous beliefs. The favorite-longshot bias refers to the phenomenon that the market probability is lower than the objective probability for the favorites, higher than the objective probability for the longshots.

Now assume the bettors all wager an identical amount, say, one dollar, and the bettors come to the window in sequence. We assume that the bettor arriving at a given instant of time is randomly chosen from a population of bettors in a fashion made more precise below. Suppose right before the n-th bettor is to place his (or her) wager, the current proportions of money bet on each horse are represented by the vector $P^{(n-1)} = (P_1^{(n-1)}, ..., P_J^{(n-1)})^T$. (This is an idealization of the actual situation at the track in which the "current" payoff odds are announced, but with a small time lag.)

Let the subjective estimate by the n-th bettor of the objective winning probability be represented by the vector $Q^{(n)} = (Q_1^{(n)}, ..., Q_J^{(n)})^T$. We assume that $Q^{(n)}$ is a realization of a random vector $Q = (Q_1, ..., Q_J)^T$, with $Q_1 + ... + Q_J = 1$. Since we assume the bettors are risk neutral, the n-th bettor is going to choose the horse that maximize his (or her) subjective expected payoff based on the current proportion $P^{(n-1)}$. That is, he (or she) is going to maximize

$$\frac{Q_j^{(n)}(1-r)}{P_j^{(n-1)}}$$

over j. Here r is the racetrack take percentage. This is equivalent to minimizing $\frac{P_j^{(n-1)}}{Q_j^{(n)}}$ over j. After the n-th bettor places the wager, the proportions of wager placed on the horses are updated to $P^{(n)}$, and the $(n+1)$st bettor comes along.

We show that this dynamic betting process converges to a unique equilibrium set of odds. These limiting odds depend on the distribution of Q. They correspond in the usual way to an equilibrium probability vector, P^*.

For a given distribution of Q one may also compute the average vector having coordinates $P_i = E(Q_i)$, $i = 1, ..., J$. The values of P_i are often interpreted as the "true" odds as collectively determined by the betting population corresponding to the distribution of Q. We then study the relation between P and the equilibrium vector P^*. In several reasonable situations it turns out that the betting equilibrium P^* corresponds to underbetting of favorites and overbetting of longshots. This is qualitatively similar to the observed phenomenon, and may represent part of the explanation for this phenomenon.

1.3 Eisenberg and Gale equilibrium

Before we move on to study this dynamic betting process, we briefly review a pari-mutuel method of pooling subjective probabilities introduced by Eisenberg and Gale (1959), which we shall see is closely related to the race tracking betting process we study in this paper. Eisenberg and Gale (1959) considered a pari-mutuel model for pooling opinions among I individuals B_1, ..., B_I concerned with a race of J horses, with B_i having a budget b_i, $i = 1, ..., I$. It is convenient to assume $\sum_{i=1}^{I} b_i = 1$. Let (p_{ij}) be the $I \times J$ subjective probability matrix. That is, p_{ij} is the probability, in the opinion of B_i, that the j-th horse will win the race. We assume that each column of the matrix (p_{ij}) contains at least one positive entry. If this were not true then for some horse j, $p_{ij} = 0$ for all i, and none of the B_i's would bet on horse j. We could then eliminate horse j from consideration.

Assume each bettor bets in a way that maximizes his subjective expectation. Let β_{ij} be the amount that B_i bets on horse j. Then

$$\sum_{j=1}^{J} \beta_{ij} = b_i, \quad i = 1, ..., I; \tag{1}$$

and the total amount bet on horse j is

$$\pi_j = \sum_{i=1}^{I} \beta_{ij}. \tag{2}$$

Notice $\sum_{j=1}^{J} \pi_j = 1$. At an equilibrium state, each B_i is maximizing his expectation. Eisenberg and Gale (1959) wrote this condition as

$$\text{if } \ \mu_i = \max_s \frac{p_{is}}{\pi_s} \ \text{ and } \ \beta_{ij} > 0, \ \text{ then } \ \frac{p_{ij}}{\pi_j} = \mu_i, \tag{3}$$

which states that B_i bets only on those horses for which his expectation is a maximum.

Eisenberg and Gale (1959) referred to nonnegative numbers π_j and β_{ij} that satisfy the conditions (1), (2), and (3) as equilibrium probabilities and bets. They proved that equilibrium probabilities exist and are unique. Norvig (1967) presented a deterministic algorithm for reaching this equilibrium as the limit of iterated bids in the case of a finite population of bettors.

Now let us come back to the dynamic race track betting process described earlier. First consider the situation where the random subjective probability vector $Q = (Q_1, ..., Q_J)^T$ is supported on a finite set of points. Assume this finite set consists of I support points $p_i = (p_{i1}, ..., p_{iJ})$, $i = 1, ..., I$, with probability b_i, $i = 1, ..., I$, respectively. We define an equilibrium state of the betting process as one in which (1), (2), and (3) are satisfied for some nonnegative numbers β_{ij}'s and π_i's. By the results in Eisenberg and Gale (1959), such an equilibrium state exists, and the π_i's at the equilibrium are unique. Notice, however, here b_i's are the probability masses of the support points, not the budget of a bettor. Roughly speaking, β_{ij} is the proportion of people that has subjective probabilities p_i and has bet on horse j, and π_j is proportional to the total money betted on horse j. The meaning of (1), (2), and (3) are then clear: at the equilibrium state, every bettor bets only on a horse for which his expectation is a maximum.

1.4 Dynamic equilibrium (continuous case)

We then turn to the continuous case in which the random subjective probability vector $Q = (Q_1, ..., Q_J)^T$ has a continuous distribution with density function f supported on

$$\Theta = \{q = (q_1, ..., q_J)^T : | \sum_{j=1}^{J} q_j = 1; q_j \geq 0, j = 1, 2, ..., J\}.$$

(It is in this sense that the n-th bettor is randomly chosen from an infinite population.) The concept of equilibrium probabilities can be extended naturally to this situation. We define an equilibrium state of the betting process as one in which there is a joint density $g(j, q)$ (with respect to the product measure of the counting measure and the Lebesgue measure on Θ), and equilibrium probabilities $\pi = (\pi_1, ..., \pi_J)$ such that

$$\sum_{j=1}^{J} g(j, q) = f(q), \quad \text{for all } \; q \in \Theta, \tag{4}$$

$$\int_{\Theta} g(j, q) dq = \pi_j, \quad \text{for all } \; j, \tag{5}$$

$$\text{if } \; \mu_q = \max_s \frac{q_s}{\pi_s} \; \text{ and } \; g(j, q) > 0, \; \text{ then } \; \frac{q_j}{\pi_j} = \mu_q, \tag{6}$$

These are clearly the continuous counterparts of (1), (2), and (3).

Proposition 1.1 *If equilibrium probabilities satisfying (4), (5), and (6) exist, they are unique.*

The proof of Proposition 1.1 is parallel to the proof in Eisenberg and Gale (1959) of the uniqueness of the equilibrium probabilities satisfying (1), (2), and (3), with β_{ij} replaced by $g(j, q)$, μ_i replaced by μ_q, p_{ij} replaced by q_j, b_i replaced by $f(q)$. Therefore the proof is skipped here.

It is possible to prove the existence of the equilibrium probabilities in the continuous case with a proof that is similar to the one in Eisenberg and Gale (1959) for discrete case. However, it turns out in the continuous case it is much easier to proceed with a fixed point argument, as will be seen in Section 2, where we show that there exists a unique set of equilibrium probabilities and the betting process converges to the equilibrium probabilities almost surely. In Section 3 we compare the equilibrium probabilities with the average probabilities. This leads to an alternative explanation of the favorite-longshot bias in racetrack betting. Conclusions are given in Section 4.

2 Convergence of the betting process

In this section we concentrate on the continuous case. Define $\phi : \Theta^0 \longrightarrow \Theta^0$ by $\phi(q) = (\phi_1(q), ..., \phi_J(q))$, where $q = (q_1, ..., q_J)^T \in \Theta^0$, Θ^0 is the interior of Θ, and

$$\phi_j(q) = Pr\{q_j/Q_j = \min_{1\leq k\leq J} q_k/Q_k\} = \int_{\{q_j/t_j<q_k/t_k,\forall k\neq j\}} f(t_1, ..., t_J)dt_1...dt_J.$$

Then ϕ is a continuous mapping. If the current proportions of the wager bet on the horses are represented by q, then $\phi_j(q)$ is the probability of the next bettor betting on horse j.

Proposition 2.1 *There exists a fixed point of ϕ in Θ^0.*

Proof: Define

$$\eta = \min_{1\leq j\leq J} Pr\{Q_j = \max_{1\leq k\leq J} Q_k\}$$

Then $0 < \eta \leq 1/J$.

Let $\underline{\Theta} = \{q = (q_1, ..., q_J)^T : | \sum_{j=1}^J q_j = 1; q_j \geq \eta, j = 1, 2, ..., J\}$.

Define $\phi^* : \quad \Theta \longrightarrow \Theta$ as

$$\phi^*(q) = \phi(q) \qquad \text{if} \quad q \in \underline{\Theta}$$

$$\phi^*(q) = \phi(\underline{q}) \qquad \text{if} \quad q \in \Theta \setminus \underline{\Theta}$$

where $\underline{q}$ is the point where $\partial\underline{\Theta}$ and the line connecting q and $(\frac{1}{J}, ..., \frac{1}{J})^T$ intersects.

Then ϕ^* is continuous from Θ to itself, and therefore by Brouwer's fixed point theorem, (see, for example, Theorem 4.2.5 of Istratescu (1981)) there exists a fixed point $q^* \in \Theta$ of the mapping ϕ^*., i.e. $\phi^*(q^*) = q^*$ and $q^* \in \Theta$.

Now we show that q^* has to be in $\underline{\Theta}$. For any $q \in \Theta\setminus\underline{\Theta}$, we have $\underline{q} \in \partial\underline{\Theta}$ by the definition of $\underline{q}$. Hence $\underline{q}_{j_0} = \eta$ for some j_0. Then $q_{j_0} < \eta$. Also, since $\underline{q} \in \underline{\Theta}$, for any $k \neq j_0$, we have $\underline{q}_k \geq \eta = \underline{q}_{j_0}$. Now we have

$$\begin{aligned}
\phi^*_{j_0}(q) &= \phi_{j_0}(\underline{q}) \\
&= Pr\{\underline{q}_{j_0}/Q_{j_0} = \min_{1\leq k\leq J} \underline{q}_k/Q_k\} \\
&= Pr\{\cap_{k\neq j_0}\{\underline{q}_{j_0}/Q_{j_0} < \underline{q}_k/Q_k\}\} \\
&\geq Pr\{\cap_{k\neq j_0}\{Q_{j_0} > Q_k\}\} \\
&= Pr\{Q_{j_0} = \max_{1\leq k\leq J} Q_k\} \\
&\geq \eta \\
&> q_{j_0}
\end{aligned}$$

This means $\phi^*(q) \neq q$, and so any point in $\Theta \setminus \underline{\Theta}$ can not be a fixed point of ϕ^*, therefore $q^* \in \underline{\Theta}$. So we have

$$\phi(q^*) = \phi^*(q^*) = q^*$$

That is, q^* is a fixed point of ϕ. Q.E.D.

Proposition 2.2 *Any fixed point of ϕ in Θ^0 is an equilibrium probability vector.*

Proof: Let q^* be a fixed point of ϕ. Define

$$g(j, q) = 1(\frac{q_j^*}{q_j} = \min_s \frac{q_s^*}{q_s}) \cdot f(q),$$

and $\pi = q^*$ in the equations (4), (5), and (6). It is straightforward to check that (4) and (6) are satisfied. (5) is satisfied since q^* is the fixed point of ϕ. Q.E.D.

The above two propositions establish the existence of the equilibrium probability vector. Therefore the equilibrium probability vector exists and is unique by Proposition 1.1. This in turn implies that the fixed point of ϕ is unique by Proposition 2.2. We summarize these results in the following lemma.

Lemma 2.1 *ϕ has a unique fixed point in Θ^0, and this is the unique equilibrium probability vector satisfying (4), (5), and (6).*

Denote the unique fixed point of ϕ by P^*. The following theorem shows that the proportion of wager placed on the horses will converge to P^* almost surely.

Theorem 2.1 *As $n \to \infty$, $P^{(n)} \to P^*$ almost surely.*

To facilitate the proof of this theorem, we introduce the following lemma:

Lemma 2.2 *For any $q = (q_1, q_2, ..., q_J) \in \Theta$, we have*

$$\sum_{j=1}^{J} (\frac{q_j}{P_j^*} - 1)(\phi_j(q) - P_j^*) \leq 0$$

Proof: We will make use of the following standard result, whose proof we will omit:

Proposition 2.3 *If $a_1 \geq a_2 \geq ... \geq a_k \geq 0$ and $\sum_{i=1}^{l} b_i \leq 0$ for all $l \leq k$, then $\sum_{i=1}^{k} a_i b_i \leq 0$.*

Now without loss of generality, assume

$$\frac{q_1}{P_1^*} \geq \frac{q_2}{P_2^*} \geq ... \geq \frac{q_k}{P_k^*} \geq 1 \geq \frac{q_{k+1}}{P_{k+1}^*} \geq ... \geq \frac{q_J}{P_J^*}.$$

Now we show that $\sum_{j=1}^{k}(\frac{q_j}{P_j^*}-1)(\phi_j(q)-P_j^*)\leq 0$.

For any $l\leq k$, we have

$$\begin{aligned}
\sum_{j=1}^{l}\phi_j(q) &= Pr\{\min(q_1/Q_1,q_2/Q_2,...,q_l/Q_l)\leq \min(q_{l+1}/Q_{l+1},...,q_J/Q_J)\} \\
&= Pr\{\min(\frac{q_1/P_1^*}{Q_1/P_1^*},...,\frac{q_l/P_l^*}{Q_l/P_l^*})\leq \min(\frac{q_{l+1}/P_{l+1}^*}{Q_{l+1}/P_{l+1}^*},...,\frac{q_J/P_J^*}{Q_J/P_J^*})\} \\
&\leq Pr\{\min(\frac{q_l/P_l^*}{Q_1/P_1^*},...,\frac{q_l/P_l^*}{Q_l/P_l^*})\leq \min(\frac{q_l/P_l^*}{Q_{l+1}/P_{l+1}^*},...,\frac{q_l/P_l^*}{Q_J/P_J^*})\} \\
&= Pr\{\min(P_1^*/Q_1,...,P_l^*/Q_l)\leq \min(P_{l+1}^*/Q_{l+1},...,P_J^*/Q_J)\} \\
&= \sum_{j=1}^{l}\phi_j(P^*)=\sum_{j=1}^{l}P_j^*.
\end{aligned}$$

So we have, for any $l\leq k$,

$$\sum_{j=1}^{l}(\phi_j(q)-P_j^*)\leq 0.$$

By Proposition 2.3, we have

$$\sum_{j=1}^{k}(\frac{q_j}{P_j^*}-1)(\phi_j(q)-P_j^*)\leq 0$$

That $\sum_{j=k+1}^{J}(\frac{q_j}{P_j^*}-1)(\phi_j(q)-P_j^*)\leq 0$ can be shown in a similar fashion. So we have $\sum_{j=1}^{J}(\frac{q_j}{P_j^*}-1)(\phi_j(q)-P_j^*)\leq 0$. Q.E.D.

Proof of Theorem 2.1: Define

$$X^{(n)}=\frac{C}{n}+\sum_{j=1}^{J}\frac{(P_j^{(n)}-P_j^*)^2}{P_j^*}$$

Where $C=\sum_{j=1}^{J}\frac{1}{P_j^*}$ is a nonrandom number. We intend to show that $X^{(n)}$ is a supermartingale with respect to the sequence of σ-fields $F^{(n)}$ generated by $\{P^{(i)}:i=1,...,n\}$.

Let $e^{(n)}$ be the J dimensional random column vector indicating which horse the n-th bet is on. That is, if the n-th bet is on horse i, then $e^{(n)}$ is the i-th unit vector. Then $E(e^{(n+1)}|F^{(n)})=\phi(P^{(n)})$, and

$$P^{(n+1)}=(nP^{(n)}+e^{(n+1)})/(n+1)=P^{(n)}+(e^{(n+1)}-P^{(n)})/(n+1)$$

and it is easy to see that $E(e^{(n+1)}|F^{(n)})=\phi(P^{(n)})$.

Hence we have

$$X^{(n+1)}=\frac{C}{n+1}+\sum_{j=1}^{J}\frac{(P_j^{(n+1)}-P_j^*)^2}{P_j^*}$$

$$
\begin{aligned}
&= \quad \frac{C}{n+1} + \sum_{j=1}^{J} \frac{(P_j^{(n)} - P_j^* + (e_j^{(n+1)} - P_j^{(n)})/(n+1))^2}{P_j^*} \\
&= \quad X^{(n)} - \frac{C}{n(n+1)} + \sum_{j=1}^{J} \frac{1}{P_j^*} \left[\frac{2}{n+1}(P_j^{(n)} - P_j^*)(e_j^{(n+1)} - P_j^{(n)}) + \frac{1}{(n+1)^2}(e_j^{(n+1)} - P_j^{(n)})^2 \right] \\
&< \quad X^{(n)} - \frac{C}{n(n+1)} + \sum_{j=1}^{J} \frac{1}{P_j^*} \left[\frac{2}{n+1}(P_j^{(n)} - P_j^*)(e_j^{(n+1)} - P_j^{(n)}) \right] + \frac{1}{(n+1)^2} \sum_{j=1}^{J} \frac{1}{P_j^*} \\
&= \quad X^{(n)} - \frac{C}{n(n+1)} + \sum_{j=1}^{J} \frac{1}{P_j^*} \left[\frac{2}{n+1}(P_j^{(n)} - P_j^*)(e_j^{(n+1)} - P_j^{(n)}) \right] + \frac{C}{(n+1)^2} \\
&< \quad X^{(n)} + \sum_{j=1}^{J} \frac{1}{P_j^*} \left[\frac{2}{n+1}(P_j^{(n)} - P_j^*)(e_j^{(n+1)} - P_j^{(n)}) \right]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\left[X^{(n+1)} - X^{(n)} | F^{(n)}\right] &< \sum_{j=1}^{J} \frac{1}{P_j^*} \left[\frac{2}{n+1}(P_j^{(n)} - P_j^*)(\phi_j(P^{(n)}) - P_j^{(n)}) \right] \\
&= \sum_{j=1}^{J} \frac{2}{(n+1)P_j^*} \left[(P_j^{(n)} - P_j^*)(\phi_j(P^{(n)}) - P_j^*) - (P_j^{(n)} - P_j^*)^2 \right]
\end{aligned}
$$

By Lemma 2.2, we have

$$
E\left[X^{(n+1)} - X^{(n)} | F^{(n)}\right] < -\frac{2}{n+1} \sum_{j=1}^{J} \frac{(P_j^{(n)} - P_j^*)^2}{P_j^*} \leq 0 \tag{7}
$$

Hence $X^{(n)}$ is a non-negative super-martingale, so it has to converge to some non-negative random variable almost surely. Denote the limit random variable by X. Also, since $X^{(n)}$ is a non-negative super-martingale, $E(X^{(n)})$ has a non-negative limit.

By (7) we have

$$
E\left[X^{(n+1)} - X^{(n)} | F^{(n)}\right] < -\frac{2}{n+1} \left[X^{(n)} - \frac{C}{n} \right]
$$

Therefore

$$
E\left[X^{(n+1)} - X^{(n)}\right] < -\frac{2}{n+1} \left[E(X^{(n)}) - \frac{C}{n} \right]
$$

This shows that the limit of $E(X^{(n)})$ has to be 0. (Suppose $E(X^{(n)}) \to \epsilon > 0$. Then

$$
E(X^{(n+k)}) - E(X^{(n)}) < \sum_{j=1}^{k} \left[\frac{2}{n+j} E(X^{(n+j-1)}) + \frac{2C}{(n+j)(n+j-1)} \right] \to -\infty
$$

as $k \to \infty$, a contradiction.)

By Fatou's lemma, $E(X) \leq \lim_{n\to\infty} E(X^{(n)}) = 0$. Since X is non-negative, we have $X = 0$ almost surely. That is, $X^{(n)}$ goes to 0 almost surely. Hence,

$$\sum_{j=1}^{J} \frac{(P_j^{(n)} - P_j^*)^2}{P_j^*} \longrightarrow 0, a.s.$$

This implies

$$\sum_{j=1}^{J} (P_j^{(n)} - P_j^*)^2 \longrightarrow 0, a.s.$$

That is, $P^{(n)} \longrightarrow P^*, a.s.$ Q.E.D.

Theorem 2.1 shows that the race track betting process induces the equilibrium probabilities (of an infinite population).

3 The Equilibrium Probabilities vs. the Average Probabilities

Theorem 2.1 establishes that in the betting process the market probability will converge to the equilibrium probability P^*, and this equilibrium probability is the unique fixed point of the mapping ϕ. Now let us investigate the properties of this equilibrium probability. To do this, we need some further assumption on the random vector Q. Recall that Q is a random vector supported on the set

$$\Theta = \{q = (q_1, ..., q_J)^T : |\sum_{j=1}^{J} q_j = 1; q_j \geq 0, j = 1, 2, ..., J\}.$$

We shall model the distribution of Q by a Dirichlet distribution. This is the most natural distribution on a set like Θ. In the present situation, the use of Dirichlet distribution can also be motivated by making the following two assumptions for the formation of the subjective estimate of the winning probabilities:

1. Each individual bettor assigns an "ability score" vector $(A_1, A_2, ..., A_J)$ for the horses, and calculates his (or her) subjective estimate of the winning probabilities as

   $$(\frac{A_1}{S_J}, ..., \frac{A_J}{S_J})$$

   where $S_J = \sum_{j=1}^{J} A_j$. We assume A_j, $j = 1, 2, ..., J$, are realizations of independent random variables.

2. The total score S_J is independent of the relative score vector $(\frac{A_1}{S_J}, ..., \frac{A_J}{S_J})$.

From these two assumptions, it follows that there is a constant c such that $cA_1, ..., cA_J$ are Gamma random variables. (See Johnson, Kotz, and Balakrishnan (1994), page 350.) Let the means of these Gamma random variables be denoted by $a_1, ..., a_J$, then we have that $(\frac{A_1}{S_J}, ..., \frac{A_J}{S_J})$ follows a Dirichlet distribution with parameters $(a_1, ..., a_J)$. Notice that by definition, $(a_1, ..., a_J)$ is proportional to the mean assigned "ability score" $(E(A_1), E(A_2), ..., E(A_J))$.

Hence, under the two assumptions made, the random vector Q follows a Dirichlet distribution with parameters $(a_1, ..., a_J)$. We will not use these two assumptions in later exposition, and will work directly with the Dirichlet distribution.

In a race of two horses, we have the following result:

Theorem 3.1 *Let $J = 2$, and suppose Q has a Dirichlet distribution with parameters a_1 and a_2. If $a_1 > a_2$, then $P_1^* < E(Q_1)$, and $P_2^* > E(Q_2)$.*

Remark 3.1 *Since in this case $E(Q_1) = \frac{a_1}{a_1+a_2} > \frac{a_2}{a_1+a_2} = E(Q_2)$, if we assume that $E(Q_1)$ and $E(Q_2)$ are identical to the objective winning probabilities, that is, the bettors get the objective winning probabilities right on average, the theorem then says the favorite is underbet and the longshot is overbet. Ali (1977) studied the two horse race with a rational expectation model instead of our dynamic wagering model. The results in Ali (1977) imply that $P_1^* <$ median(Q_1) under general conditions. Notice that since Q_1 follows a Beta distribution with parameters $a_1 > a_2$, the distribution of Q_1 is skewed to the left, and the mean of Q_1 is smaller than the median of Q_1. Hence the conclusion of Theorem 3.1 is stronger than that in the result of Ali (1977).*

Proof of Theorem 3.1: Note we just have to prove $P_1^* < \frac{a_1}{a_1+a_2}$.

Since Q follows Dirichlet distribution with parameters (a_1, a_2), we have Q_1 follows a Beta distribution with parameters (a_1, a_2), and Q_2 follows a Beta distribution with parameters (a_2, a_1). Now since $P_1^* = \phi_1(P^*) = Pr\{P_1^*/Q_1 < P_2^*/Q_2\} = Pr\{Q_1 > P_1^*\}$, we have (here $\Longleftrightarrow$ means "is equivalent to")

$$\begin{aligned}
P_1^* < \frac{a_1}{a_1+a_2} \quad &\Longleftrightarrow \quad Pr\{Q_1 > \frac{a_1}{a_1+a_2}\} < \frac{a_1}{a_1+a_2} \\
&\Longleftrightarrow \quad \frac{Pr\{Q_1 > \frac{a_1}{a_1+a_2}\}}{Pr\{Q_1 < \frac{a_1}{a_1+a_2}\}} < \frac{a_1}{a_2} \\
&\Longleftrightarrow \quad \frac{Pr\{Q_2 < \frac{a_2}{a_1+a_2}\}}{Pr\{Q_1 < \frac{a_1}{a_1+a_2}\}} < \frac{a_1}{a_2} \\
&\Longleftrightarrow \quad \int_0^{\frac{a_2}{a_1+a_2}} x^{a_2-1}(1-x)^{a_1-1}dx < \frac{a_1}{a_2}\int_0^{\frac{a_1}{a_1+a_2}} x^{a_1-1}(1-x)^{a_2-1}dx \\
&\Longleftrightarrow \quad \int_0^{\frac{a_2}{a_1+a_2}} x^{a_2-1}(1-x)^{a_1-1}dx < \int_0^{\frac{a_2}{a_1+a_2}} c^{a_1}y^{a_2-1}(1-cy^{\frac{a_2}{a_1}})^{a_2-1}dy
\end{aligned}$$

where the last step follows from a change of variable on the right hand side: $x = cy^{\frac{a_2}{a_1}}$ with $c = \frac{\frac{a_1}{a_1+a_2}}{(\frac{a_2}{a_1+a_2})^{\frac{a_2}{a_1}}}$.

Therefore, to prove the theorem, we only need to prove

$$(1-x)^{a_1-1} < c^{a_1}(1-cx^{\frac{a_2}{a_1}})^{a_2-1}, \qquad \forall 0 < x < \frac{a_2}{a_1+a_2}. \tag{8}$$

It is easy to check that $1 > cx^{\frac{a_2}{a_1}} > x$, $\forall 0 < x < \frac{a_2}{a_1+a_2}$. Therefore, to prove (8), it is sufficient to prove

$$(1-x)^{a_1} < c^{a_1}(1-cx^{\frac{a_2}{a_1}})^{a_2}, \qquad \forall 0 < x < \frac{a_2}{a_1+a_2}$$

or, equivalently,

$$a_1 \log(1-x) \leq a_1 \log c + a_2 \log(1-cx^{\frac{a_2}{a_1}}), \qquad \forall 0 < x < \frac{a_2}{a_1+a_2} \tag{9}$$

Let $s = \frac{a_2}{a_1}$, then $s < 1$. Let $t = \frac{x}{\frac{a_2}{a_1+a_2}} = x(1+s^{-1})$. After some simplification, (9) becomes

$$B(t) \leq 0, \qquad \forall 0 < t < 1 \tag{10}$$

where

$$B(t) = \log(1+s-st) - s\log(1+s^{-1}-s^{-1}t^s)$$

Since $B(1) = 0$, to prove (10), it is sufficient to show that $B'(t) \geq 0$, $\forall 0 < t < 1$. After some simplification, we can see this is equivalent to

$$C(t) \geq 0, \qquad \forall 0 < t < 1 \tag{11}$$

where

$$C(t) = (s+1-st)t^{s-1} - (1+s^{-1}-s^{-1}t^s)$$

Now it is easy to check that $C'(t) \leq 0$, $\forall 0 < t < 1$, and that $C(1) = 0$. Therefore (11) holds, and the theorem is proved. Q.E.D.

Corollary 3.1 *In a horse race with J horses, suppose Q follows a Dirichlet distribution with parameters $a_1, a_2, ..., a_J$. If $E(Q_1) > 1/2$, then $P_1^* < E(Q_1)$.*

Proof: Consider a horse race with two horses. Suppose the distribution of the subjective estimate of the winning probabilities is the same as that of random vector $(Q_1, Q_2+...+Q_J)^T$. Denote the limit vector in this race by $P^{**} = (P_1^{**}, P_2^{**})^T$. Since $Q = (Q_1, Q_2, ..., Q_J)^T$ follows Dirichlet distribution with parameters $a_1, a_2, ..., a_J$, we know $(Q_1, Q_2 + ... + Q_J)^T$ follows Dirichlet distribution with parameters $a_1, a_2 + ... + a_J$. Since $E(Q_1) > 1/2$, by Theorem 3.1, $P_1^{**} < E(Q_1)$.

Now $P_1^{**} = Pr\{P_1^{**}/Q_1 < P_2^{**}/(Q_2 + ... + Q_J)\} = Pr\{Q_1 > P_1^{**}\}$, and

$$\begin{aligned} P_1^* &= Pr\{P_1^*/Q_1 = \min_{1 \le k \le J} P_k^*/Q_k\} \\ &\le Pr\{P_1^*/Q_1 < (P_2^* + ... + P_J^*)/(Q_2^* + ... + Q_J^*)\} \\ &= Pr\{P_1^*/Q_1 < 1\} \\ &= Pr\{Q_1 > P_1^*\} \end{aligned}$$

So $P_1^* \le P_1^{**}$, and we have $P_1^* < E(Q_1)$. Q.E.D.

Corollary 3.1 says that if the bettors get the objective probability on average, and there is a heavy favorite in a J horse race, then the heavy favorite will be underbet. Blough (1994) studied the J horse race with a rational expectation model instead of our dynamic wagering model. Under some strong assumptions on the logs of the probability ratios, Blough compared the equilibrium probabilities with a certain generalization of the median probability. Our assumptions are more natural than the assumption employed in Blough (1994) and we compare the equilibrium probabilities with the mean probabilities.

We have also used the computer to calculate the equilibrium probability in $J > 2$ horse races without a heavy favorite. In the case of Dirichlet distributions with various parameters, the equilibrium probabilities all show the usual "underbet favorite, overbet longshot" bias. We list a few examples in the following table. (All numbers are rounded to the second digit after the decimal point.)

Equilibrium Probabilities: Dirichlet Distribution Case

Parameters	(3,4,5,6)	(3,3,10,10)	(3,3,3,4)
Average Prob.	(.17, .22, .28, .33)	(.12, .12, .38, .38)	(.23, .23, .23, .31)
Equilibrium Prob.	(.19, .23, .27, .31)	(.15, .15, .35, .35)	(.24, .24, .24, .29)

When we investigate the non-Dirichlet distribution case, the tendency shows in most cases, but is reversed in some of the cases. For example, in the $J = 2$ case, let $Q = (Q_1, Q_2)$ have the distribution of $(\frac{10G_1}{10G_1+G_2}, \frac{G_2}{10G_1+G_2})$, where $G_1 \sim Gamma(1)$, $G_2 \sim Gamma(b)$. We have the following table:

Equilibrium Probabilities: Non-Dirichlet Distribution Case

b	3	4	5	6	7	8
Aver. Prob.	(.66, .34)	(.60, 0.40)	(.55, .45)	**(.52, .48)**	(.49, .51)	(.46, .54)
Equi. Prob.	(.63, .37)	(.59, .41)	(.55, .45)	**(.53, .47)**	(.50, .50)	(.49, .51)

Note: the reverse happens when $b = 6$.

4 Conclusion

There is a close connection between the race track betting process and the pari-mutuel pooling of subjective probabilities. We showed that the race track betting process provides a natural tool to induce the unique set of equilibrium probabilities in the pari-mutuel method. The set of equilibrium probabilities is not identical to the average of the subjective estimate by the bettors. This means, even if the bettors get the true probability right on average, and even if the bettors are risk neutral, the proportion of money bet on the horses would not match the objective winning probabilities. This gives a possible explanation for the favorite-longshot bias, in addition to the risk-seeking bettor argument and various behavioral explanations. However, these explanations do not contradict each other, and we believe they each play a part in generating the favorite-longshot bias.

References

[1] Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. *Journal of political economics* **84**, 803-815.

[2] Asch, P., Malkiel, B., and Quandt R. (1982). Racetrack betting and informed behavior. *Journal of financial economics* **10**, 187-194.

[3] Blough, S. (1994). Differences of opinions at the racetrack. In D. Hausch, V. Lo and W.Ziemba (eds.) *Efficiency of racetrack betting markets*, 323-341.

[4] Brown, L. D., D'Amato, R., and Gertner, R. (1994). Racetrack betting: do bettors understand the odds? *Chance*, **7**, 17-23.

[5] Busche, K. (1994). Efficient market results in an Asian setting. In D. Hausch, V. Lo and W. Ziemba (eds.) *Efficiency of racetrack betting markets*, 615-616.

[6] Busche, K., and Hall, C. (1988). An exception to the risk preference anomaly. *Journal of business*, **61**, 337-346.

[7] Eisenberg, E. and Gale, D. (1959). Consensus of Subjective Probabilities: The Pari-mutuel Method. *Annals of Mathematical Statistics* **30**, 165-168.

[8] Fabricand, B. P. (1965). *Horse Sense*. David McKay Company Inc., New York.

[9] Griffith, R. M. (1949). Odds adjustments by American horse race bettors. *American journal of psychology* **62**, 290-294.

[10] Hausch, D., Ziemba, W., and Rubinstein, M. (1981). Efficiency of the market for racetrack betting. *Management science* **27**, 1435-1452.

[11] Henery, R. J. (1985). On the average probability of losing bets on horses with given starting price odds. *Journal of royal statistical society - A.* **148**, 342-349.

[12] Istratescu, V. I. (1981). *Fixed Point Theory: An Introduction.* Kluwer Boston Inc.

[13] Johnson, N. L., Kotz, S., and Balakrichnan, N. (1994). *Continuous Univariate Distributions, Volume 1.* John Wiley & Sons, Inc., New York.

[14] McGlothlin, W. H. (1956). Stability of choices among uncertain alternatives. *American journal of psychology*, **69**, 604-615.

[15] Norvig, T. (1967). Consensus of Subjective Probabilities: A Convergence Theorem. *Annals of Mathematical Statistics* **38**, 221-225.

[16] Snyder, W. W. (1978). Horse racing: testing the efficient markets model. *Journal of finance* **33**, 1109-1118.

[17] Thaler, R., and Ziemba, W. (1988). Parimutuel betting markets: Racetrack and lotteries. *Journal of economical perspective* **2**, 161-174.

[18] Weitzman, M. (1965). Utility analysis and group behavior: An empirical study. *Journal of political economics* **73**, 18-26.

[19] Ziemba, W. T. and Hausch, D. B. (1984). *Beat the Racetrack.* New York: Harcourt, Brace, Jovanovich.