MIS Quarterly

RESEARCH ARTICLE

MODELING FIXED ODDS BETTING FOR FUTURE EVENT PREDICTION

Weiyun Chen
Department of Educational Information Technology, Faculty of Education, East China Normal University, Shanghai, CHINA {weiyun.chen@qq.com}

Xin Li
Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong, CHINA {xin.li.phd@gmail.com}

Daniel Zeng
Department of Management Information Systems, University of Arizona, Tucson, AZ 85721 U.S.A., and State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, CHINA {zeng@email.arizona.edu}

Appendix

Theorem 1: If $\alpha_l(p) = \beta_l(1\text{-}p)$, then $f(x;p,\theta) = f(1\text{-}x;1\text{-}p,\theta)$.

Proof:

$$f(x;p,\theta) = \sum_{l=1}^{D} \frac{\lambda_l}{\sum_{j=1}^{D} \lambda_j} \cdot Beta\left(x;\alpha_l(p),\beta_l(p)\right)$$

$$= \sum_{l=1}^{D} \frac{\lambda_l}{\sum_{j=1}^{D} \lambda_j} \cdot Beta\left(1\text{-}x;\beta_l(p),\alpha_l(p)\right)$$

$$= \sum_{l=1}^{D} \frac{\lambda_l}{\sum_{j=1}^{D} \lambda_j} \cdot Beta\left(1\text{-}x;\alpha_l(1-p),\beta_l(1-p)\right) = f(1-x;1-p,\theta)$$

■

Theorem 2: $E[Beta(x;\ \alpha_l(p),\beta_l(p))]$ is strictly increasing with p.

Proof:

$$E[Beta(x;\ \alpha_l(p),\beta_l(p))] = \frac{\alpha_l(p)}{\alpha_l(p)+\beta_l(p)} = \frac{1}{1+\dfrac{\beta_l(p)}{\alpha_l(p)}}$$

Since p is within (0, 1), both $\alpha_l(p)$ and $\beta_l(p)$ are positive.

$$\frac{\partial \dfrac{\beta_l(p)}{\alpha_l(p)}}{\partial p} = \frac{\alpha_l(p)\dfrac{\partial \beta_l(p)}{\partial p} - \beta_l(p)\dfrac{\partial \alpha_l(p)}{\partial p}}{\alpha_l(p)^2} = \frac{\alpha_l(p)\left[-u_{l1} - \sum_{h=2}^{Z} u_{lh}(1-p_{\mathrm{A}})^{h-1}\right] - \beta_l(p)\left[u_{l1} + \sum_{h=2}^{Z} u_{lh} p_{\mathrm{A}}^{h-1}\right]}{\alpha_l(p)^2}$$

Clearly, $\dfrac{\partial \dfrac{\beta_l(p)}{\alpha_l(p)}}{\partial p} \leq 0$ and the equal sign holds only if all u_{lh}=0 when p is within range (0, 1).

If all u_{lh}=0, both $\alpha_l(p)$ and $\beta_l(p)$ equal 0, which is in conflict with our assumptions.

Thus $\dfrac{\beta_l(p)}{\alpha_l(p)}$ is strictly decreasing in p, and $\dfrac{\alpha_l(p)}{\alpha_l(p)+\beta_l(p)}$ is strictly increasing with p. ■

Theorem 3: If parameters $q, p, \rho, \gamma, \tau, o_A, o_B$ are all positive, there exists and only exists one belief value $c \in (0,1)$, called balance belief hereafter, satisfying $U(c, o_A) = U(1 - c, o_B)$.

Proof: When q, p, ρ, γ, τ are positive, both $w^+(\cdot)$ and $w^-(\cdot)$ are strictly increasing functions. Accordingly, the utility function $U(x, o)$ is a strictly increasing function and $U(1 - x, o)$ is a strictly decreasing function in belief x. Given $o_A > 0$ and $o_B > 0$, $[U(x, o_A) - U(1-x, o_B)]$ is strictly increasing. It is easy to verify that $U(x = 0, o_A) < 0 < U(x = 1, o_B)$ and $U(x = 1, o_A) > 0 > U(x = 0, o_B)$. Thus, $[U(x, o_A) - U(1-x, o_B)] < 0$ for x = 0 and $[U(x, o_A) - U(1 - x, o_B)] > 0$ for x = 1. As such, there must exist one and only one balance belief $x = c$, satisfying $U(c, o_A) = U(1 - c, o_B)$. ■

Theorem 4: For sufficiently large m_i, maximizing equation 7 reduces to solving $PA(p_i,\theta) = s_{iA}$.

Proof:

$Lc(p,\theta)$ in equation (7) is continuous and differentiable. Since $0 < p_i < 1$, the value of p_i that maximizes $Lc(p,\theta)$, if it exists, must satisfy the first-order condition $\dfrac{\partial Lc(p_i,\theta)}{\partial p_i} = 0$.

$$\begin{aligned}
&\frac{\partial Lc(p_i,\theta)}{\partial p_i} \\
&= m_i s_{iA} \frac{1}{PA(p_i,\theta)} \frac{\partial PA(p_i,\theta)}{\partial p_i} + m_i(1-s_{iA}) \frac{1}{1-PA(p_i,\theta)} \frac{-\partial PA(p_i,\theta)}{\partial p_i} + R_i \frac{1}{p_i} + (1-R_i)\frac{-1}{1-p_i} \\
&= m_i \frac{\partial PA(p_i,\theta)}{\partial p_i}\left(\frac{s_{iA}}{PA(p_i,\theta)} - \frac{(1-s_{iA})}{1-PA(p_i,\theta)}\right) + \left[\frac{R_i}{p_i} - \frac{(1-R_i)}{1-p_i}\right] \\
&= m_i \frac{\partial PA(p_i,\theta)}{\partial p_i}\left(\frac{s_{iA}(1-PA(p_i,\theta)) - (1-s_{iA})PA(p_i,\theta)}{PA(p_i,\theta)(1-PA(p_i,\theta))}\right) + \left[\frac{R_i(1-p_i)-(1-R_i)p_i}{p_i(1-p_i)}\right] \\
&= m_i \frac{\partial PA(p_i,\theta)}{\partial p_i}\left(\frac{s_{iA} - PA(p_i,\theta)}{PA(p_i,\theta)(1-PA(p_i,\theta))}\right) + \frac{R_i - p_i}{p_i(1-p_i)} = 0
\end{aligned}$$

Namely: $[PA(p_i,\theta) - s_{iA}]\dfrac{\partial PA(p_i,\theta)}{\partial p_i} = \dfrac{PA(p_i,\theta)(1-PA(p_i,\theta))(R_i - p_i)}{m_i p_i (1-p_i)}$

$$\lim_{m_i \to +\infty} [PA(p_i,\theta)\text{-}s_{iA}] \frac{\partial PA(p_i,\theta)}{\partial p_i} = \lim_{m_i \to +\infty} \left[\frac{PA(p_i,\theta)(1-PA(p_i,\theta))(R_i - p_i)}{m_i p_i (1-p_i)} \right] = 0$$

According to Theorem 4, $\frac{\partial PA(p_i,\theta)}{\partial p_i} > 0$, we obtain $\lim_{m_i \to +\infty} [PA(p_i,\theta)\text{-}s_{iA}] = 0$. ■

Theorem 5: $I_c(\alpha_l(p),\beta_l(p))$ is strictly decreasing in p, where $I_c(\alpha_l(p),\beta_l(p))$ is the regularized incomplete Beta function $\int_0^c Beta(x;\alpha_l(p),\beta_l(p))dx$.

Proof: Based on the chain rule of multivariable calculus, $\frac{\partial I_c(\alpha(p),\beta(p))}{\partial p} = \frac{\partial I_c(\alpha(p),\beta(p))}{\partial \alpha}\frac{\partial \alpha}{\partial p} + \frac{\partial I_c(\alpha(p),\beta(p))}{\partial \beta}\frac{\partial \beta}{\partial p}$.

$$\frac{\partial I_c(\alpha,\beta)}{\partial \alpha} = [\log(c) - \varphi(\alpha) + \varphi(\alpha+\beta)] I_c(\alpha,\beta) - \frac{\Gamma(\alpha)\Gamma(\alpha+\beta)}{\Gamma(\beta)} c^\alpha \sum_{k=0}^{\infty} \frac{(\alpha)_k(\alpha)_k(1-\beta)_k c^k}{k!\,\Gamma(k+1+\alpha)\Gamma(k+1+\beta)}$$

where $(\cdot)_k$ is the Pochhammer symbol specified as $(x)_0 = 1$; $(x)_n = x(x+1)(x+2)\ldots(x+n\text{-}1)$.

Since $\varphi(\alpha) = \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{k+\alpha-1}\right) - r$, $\varphi(\alpha+\beta) - \varphi(\alpha) = \sum_{k=1}^{\infty}\left(\frac{1}{k+\alpha+\beta-1} - \frac{1}{k+\alpha-1}\right) < 0$ when $\alpha > 0$ and $\beta > 0$.

Since $c<1$, $\log(c) < 0$ and $[\log(c) - \varphi(\alpha) + \varphi(\alpha+\beta)] < 0$.

Since $I_c(\alpha,\beta) > 0$ if $0 < c < 1$, we have $[\log(c) - \varphi(\alpha) + \varphi(\alpha+\beta)]I_c(\alpha,\beta) < 0$.

Since $\Gamma(x)>0$ if $x>0$, we have $\frac{\Gamma(\alpha)\Gamma(\alpha+\beta)}{\Gamma(\beta)} c^\alpha \sum_{k=0}^{\infty} \frac{(\alpha)_k(\alpha)_k(1-\beta)_k c^k}{k!\,\Gamma(k+1+\alpha)\Gamma(k+1+\beta)} > 0$

Thus, $\frac{\partial I_c(\alpha,\beta)}{\partial \alpha} < 0$ when $0 < c < 1$.

Similarly, we can prove $\frac{\partial I_c(\alpha,\beta)}{\partial \beta} > 0$ when $0 < c < 1$.

It is clear $\frac{\partial \alpha}{\partial p} > 0$ and $\frac{\partial \beta}{\partial p} < 0$ when $0 < p < 1$.

Thus, $\frac{\partial I_c(\alpha(p),\beta(p))}{\partial p} < 0$, i.e., $I_c(\alpha(p),\beta(p))$ is strictly decreasing in p. ■

Formula for AIC

The value of AIC criteria is computed as

$$AICc = 2t + 2t(t+1)/(N-t-1) - 2Lc(\theta)$$

where N denotes the number of data instances and t denotes the number of parameters, which is

$$t = \begin{cases} D(1+Z) & \text{D>1} \\ Z & \text{D=1} \end{cases}$$

according to equations (1) and (2).

Input: $\{o_{iA}, o_{iB}, R_i, s_{iA}, o_{iB},\}, i \in \{1, \ldots, H\}, k_{max}, \varepsilon$.
Output: estimated optimal parameter θ^*.
1. Compute balance belief $c_1, c_2, \ldots, c_H$ for each betting game according to (9), using a root finding algorithm.
2. Initialize a simplex SX which consists of $J + 1$ sets of parameters θ_j in the parameter space Θ, where J is the dimension of the parameter space.
3. While:
4. For each θ_j :
5. Solve p_{iA} from each betting according to (10) using a root finding algorithm.
6. Compute the $Lc(_j)$ according to (12).
7. If $\left(\max\left(Lc^*\left(\theta_j\right)\right) - \min\left(Lc^*\left(\theta_j\right)\right) \geq \varepsilon\right)$ or IterationCount < k_{max}:
8. Update vertices in SX using *Reflect*, *Expand*, *Outside contraction*, *Inside contraction*, or *Shrink* operations.
9. If operation <> "Shrink":
IterationCount = IterationCount +1
10. Else:
11. Return the parameter set corresponding to $\max\left(Lc^*\left(\theta_j\right)\right)$ as θ^*.

Figure A1. Maximum Likelihood Estimation Using a Nelder–Mead Method

Detailed Belief Distribution Estimation Procedure

Sina 2008 Olympic Games Dataset

For each setting of D and Z, varying from 1 to 3, respectively, we numerically obtained the optimal parameters $\theta^* \in \Theta$. Table A1 reports the log likelihood and *AIC* values for these models. Generally, the model's likelihood converges when D and Z are larger than 2. The model with $D = 2$ and $Z = 2$ is the model with the maximum likelihood. The estimated belief distribution function is given as:

$$f(x;p) = 0.74 * Beta(x; 3.76 * p + 0.1 * p^2, 3.76 * (1-p) + 0.1 * (1-p)^2)$$
$$+ 0.26 * Beta(x; 0.11 * p + 66.32 * p^2, 0.11 * (1-p) + 66.32 * (1-p)^2)$$

For the AIC criteria, we combined the three components with smallest *AICc* ($D = 1, Z = 1$; $D = 1, Z = 2$; and $D = 1, Z = 3$). The estimated belief distribution function is given as

$$f(x;p) = 0.67 * Beta(x; 5.98 * p, 5.98 * (1-p))$$
$$+ 0.24 * Beta(x; 5.69 * p + 0.16 * p^2, 5.69 * (1-p) + 0.16 * (1-p)^2)$$
$$+ 0.09 * Beta(x; 5.95 * p + 0.02 * p^3, 5.95 * (1-p) + 0.02 * (1-p)^3)$$

Table A1. Log Likelihood and AIC Values (Sina 2008 Olympic Games)

	Z = 1		*Z* = 2		*Z* = 3	
D	**Log likelihood**	**AIC**	**Log likelihood**	**AIC**	**Log likelihood**	**AIC**
1	-101.12	204.26	-101.11	206.29	-101.12	208.38
2	-101.03	210.30	-100.79	214.11	-101.05	219.01
3	-101.03	214.59	-100.80	220.74	-100.82	227.67

Sohu Entertainment Dataset

Table A2 shows the results on the Sohu entertainment event dataset, varying *D* and *Z* from 1 to 3. The model ($D = 3$ and $Z = 3$) is the model with the maximum likelihood. The estimated belief distribution function is as follows:

$$f(x;p) = 0.84 * Beta(x; 0.1*p + 0.1*p^2 + 1000*p^3, 0.1*(1-p) + 0.1*(1-p)^2 + 1000*(1-p)^3) \\ + 0.16*Beta(x; 7.54*p + 0.1*p^2 + 0.1*p^3, 7.54*(1-p) + 0.1*(1-p)^2 + 0.1*(1-p)^3)$$

For the AIC criteria, we combined the three components with smallest *AIC* ($D = 1, Z = 1$ and $D = 1, Z = 2$ and $D = 1, Z = 3$). The estimated belief distribution function is given as

$$f(x;p) = 0.62*Beta(x; 21.28*p, 21.28*(1-p)) \\ + 0.29*Beta(x; 1.29*p + 13.3*p^2, 1.29*(1-p) + 13.3*(1-p)^2) \\ + 0.09*Beta(x; p + 12.6*p^2 + p^3, (1-p) + 12.6*(1-p)^2 + (1-p)^3)$$

Table A2. Likelihood and AIC Values (Sohu Entertainment)

D	*Z* = 1		*Z* = 2		*Z* = 3	
	Log likelihood	AIC	Log likelihood	AIC	Log likelihood	AIC
1	-22.56	47.25	-22.20	48.79	-22.20	51.20
2	-22.26	53.90	-21.32	57.75	-21.32	64.40
3	-22.26	59.63	-21.32	68.14	-21.20	81.26

Sohu 2014 FIFA Dataset

Table A3 shows the results on the Sohu 2014 FIFA dataset, varying *D* and *Z* from 1 to 3. The model ($D = 1$ and $Z = 1$) is the model with the maximum likelihood. The estimated belief distribution is given as

$$f(x;p) = Beta(x; 42.18*p, 42.18*(1-p))$$

For the AIC criteria, we combine the three components with smallest *AIC* ($D = 1, Z = 1$; $D = 1, Z = 2$; and $D = 1, Z = 3$). The estimated belief distribution function is given as

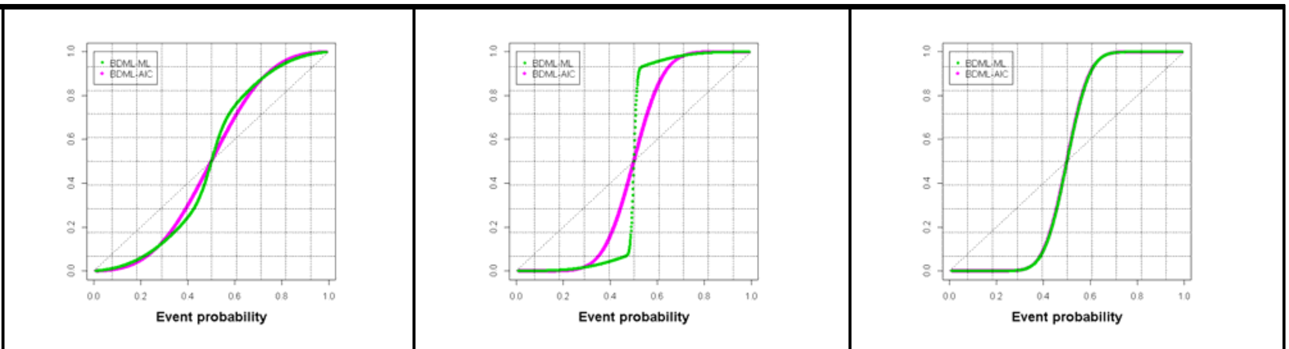
$$f(x;p) = 0.67*Beta(x; 42.18*p, 42.18*(1-p)) \\ + 0.24*Beta(x; 41.88*p, 41.88*(1-p)) \\ + 0.09*Beta(x; 41.98*p, 41.98*(1-p))$$

Table A-3: Likelihood and AIC Values (Sohu 2014 FIFA)

D	*Z* = 1		*Z* = 2		*Z* = 3	
	Log likelihood	AIC	Log likelihood	AIC	Log likelihood	AIC
1	-119.36	240.74	-119.36	242.79	-119.36	244.86
2	-119.36	246.95	-119.36	251.21	-119.36	255.56
3	- 119.36	251.21	-119.36	257.78	-119.36	264.59

With the estimated belief distribution function, we can calculate the percentage of people who consider the event may happen if there are no odds (or equal odds on both sides) in the prediction market, as illustrated in Figure A2. In all three datasets, the bettors' beliefs are more extreme than the actual event probability. If the event probability is less than 0.5, the final bet ratio will be lower than the event probability. If the event probability is higher than 0.5, the final bet ratio is higher than event probability.

People with a Belief >.5



Sina 2008 Olympic Games	Sohu Entertainment	Sohu 2014 FIFA

Figure A2. Bettor Belief Without Odds' Influence