RESEARCH REPORT

Serial No. 13

June 1992

THE SELECTION OF OLYMPIC ATHLETES

by

John Bacon-Shone

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS

Pokfulam Road, Hong Kong. Tel: (852) 859 2466 Fax: (852) 858 9041
E-Mail: STATIST@HKUCC.HKU.HK or STATIST@HKUCC.BITNET

The selection of Olympic athletes

JOHN BACON-SHONE

Social Sciences Research Centre and Department of Statistics, The University of Hong Kong, Pofulam Road, Hong Kong.

Abstract. The selection of athletes for the Olympics requires the ranking of athletes across events. We propose a simple methodology for doing so and give a theoretical justification under the assumption of Weibull running-times for athletes.

1 Introduction

The Olympic qualifying rules allow each country to send at least one athlete per event, regardless of qualifying standards. Given that most countries have limited funds, Olympic committees have the difficult task of selecting the best athletes across a range of events. The central olympic committee sets the qualifying standards for each event on the basis that only about 40 people in the world should be capable of achieving this standard. The standards and world records are shown for a number of athletic events in Table 1, separately for men and women. Previous statistical studies of athletic records have concentrated on prediction (e.g. Smith (1988)) and variation across countries and events (e.g. Dawkins (1989)).

2 Proposed Methodology

A simplistic approach would be to select athletes by ranking their performances scaled by the world record. Unfortunately, this would mean that some athletes who achieved the Olympic qualifying standard would be ranked lower than some who failed to meet the standard. Clearly this could be achieved by scaling relative to the qualifying standard, but the wide variations in the ratio of the qualifying standard to the world record still suggests that this would be unfair. It is clear that the ratio of qualifying standard to the world record measures the competitiveness of the sport and that this should be used to scale in some way the performances relative to the world record. Our proposal is to use the following formula:

$$Score = 100 \frac{\log(\frac{t_Q}{t_R})}{\log(\frac{t_B}{t_R})}$$

where t_Q = Olympic qualifying time
t_R = world record time
t_B = athlete's best time

In this scoring function a score of 100 or greater means that the athlete has achieved the qualifying time, and the higher the score, the better. As those who meet the standard would normally be selected anyway, a rational selection criteria is to select those with the highest scores.

3 Theoretical Justification

Clearly we require a failure (or success) time distribution that has two parameters in order to take account of both the overall event difficulty and the range of abilities taking part. The three obvious possibilities are the log-normal, gamma and Weibull distributions, but of these three, only the Weibull provides simple answers. If we assume that winning times follow a Weibull distribution and that P_R is the probability that time t_R is the winning time and P_Q is the probability that time t_Q is the winning time, then we have that

$$\exp[\,-\lambda(t_R)^{\alpha}] = P_R$$

and

$$\exp[\,-\lambda(t_Q)^{\alpha}] = P_Q$$

Thus

$$\log(\,-\log P_R) = \alpha \log\,(\frac{t_R}{\lambda})$$

and

$$\log(\,-\log P_Q) = \alpha \log\,(\frac{t_Q}{\lambda})$$

We can solve these equations to give

$$\alpha = \frac{\log\,(\frac{-\log P_R}{-\log P_Q})}{\log(\frac{t_R}{t_Q})}$$

and

$$\lambda = \frac{(-\log P_R)}{t_R^{\alpha}} \quad .$$

If t_B is an individual's best time, then P_B, the probability that this time will be sufficient to win,

$$= \exp [-\lambda(t_B)^{\alpha}]$$
$$= \exp [\log P_R(\frac{t_B}{t_R})^{\alpha}]$$

so that

$$\log (-\log P_B) = \log (-\log P_R) + \alpha \log (\frac{t_B}{t_R})$$
$$= \log (-\log P_R) + \frac{\log (\frac{-\log P_R}{-\log P_Q})}{\log (\frac{t_R}{t_Q})} \log(\frac{t_B}{t_R})$$

If we assume that P_R and P_Q are constant across events, then we have that

$$\log (-\log P_B) = a + b \frac{\log (\frac{t_B}{t_R})}{\log (\frac{t_Q}{t_R})}$$

which gives the result that our scoring function will rank individuals according to their chance of winning in the Olympics.

4 Example

If athlete A (male) has run the 100m in 10.5 seconds and athlete B (female) has run the 1500m in 4 minutes and 20 seconds, their scores will be:

$$athlete\ A = 100 \frac{\log (\frac{10.3}{9.92})}{\log (\frac{10.5}{9.92})} = 100\ x\ \frac{0.01633}{0.02468} = 66.2$$

while

$$athlete\ B = 100 \frac{\log (\frac{250.2}{232.47})}{\log (\frac{260}{232.47})} = 100\ x\ \frac{0.03192}{0.04861} = 65.7$$

This suggests that athlete A should be ranked marginally higher. Note that the log-ratios needed for scaling are shown in Table 1.

5 Discussion

The theory suggests that our score function is appropriate for ranking individual performances across sports as it ranks on the basis of the probability of winning the event. The major weaknesses in this approach are that, firstly, it treats the world record as an known constant rather than as an extreme random variable and secondly, that it assumes that P_R and P_Q are constant across events, which may not be too accurate given that some events are much more popular than others. However, at least this proposal is an attempt at rationality. Further refinement will require analysis of past Olympic events in order to assess how well the system works in predicting the chance of different individuals winning the event given their previous best performance.

Acknowledgements

The author is grateful to Bill Purves for suggesting the problem and providing the data.

References

Dawkins, B. (1989) Multivariate Analysis of National Track Records, The American Statistician, 43(2), pp.110-115.

Smith, R.L. (1988) Forecasting Records by Maximum Likelihood, Journal of the American Statistical Association, 83, pp.331-338.

Table 1. Olympic records and qualifying times for the 1994 Olympics

	Man			Woman		
	Record	Standard	log ratio	Record	Standard	log ratio
100m	9.92	10.30	.0163	10.49	11.40	.0361
200	19.72	20.80	.0232	21.34	23.00	.0325
400	43.29	45.90	.0254	47.60	52.00	.0384
800	1:41.73	1:46.20	.0187	1:53.28	2:01.30	.0297
1500	3:29.46	3:37.00	.0154	3:52.47	4:10.20	.0319
3000	-	-		8:22.62	8:56.00	.0279
5000	12:58.39	13:27.30	.0158	-	-	
10,000	27:08.23	28:07.00	.0154	30:13.74	32:50.00	.0359
100m hurdles	-	-		12.21	13.30	.0371
110m hurdles	12.92	13.80	.0286	-	-	
400m hurdles	47.02	50.00	.0267	52.94	56.50	.0283
Steeplechase	8:05.35	8:29.00	.0207	-	-	
10km walk	-	-		41:30	48:00	.0653
20km walk	1:18:13	1:24.00	.0315	-	-	
50km walk	3:37:41	4:05.00	.0519	-	-	

Recorded in hours:minutes:seconds.