Methods for estimating peak physiological performance and correlating performance measures

Austen W. Head

Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94305-4065, USA

ahead@stanford.edu

Johanna S. Hardin

Department of Mathematics, Pomona College, 610 N. College Ave., Claremont, CA 91711, USA

jo.hardin@pomona.edu

Stephen C. Adolph

Department of Biology, Harvey Mudd College, 301 Platt Boulevard, Claremont, CA 91711, USA

adolph@hmc.edu

Estimates of animal performance often use the maximum of a small number of laboratory trials, a method which has several statistical disadvantages. Sample maxima always underestimate the true maximum performance, and the degree of the bias depends on sample size. Here, we suggest an alternative approach that involves estimating a specific performance quantile (e.g., the 0.90 quantile). We use the information on within-individual variation in performance to obtain a sampling distribution for the residual performance measures; we use this distribution to estimate a desired performance quantile for each individual. We illustrate our approach using simulations and with data on sprint speed in lizards. The quantile method has several advantages over the sample maximum: it reduces or eliminates bias, it uses all of the data from each individual, and its accuracy is independent of sample size. Additionally, we address the estimation of correlations between two different performance measures, such as sample maxima, quantiles, or means. In particular, because of sampling variability, we propose that the correlation of sample means does a better job estimating the correlation of population maxima than the estimator which is the correlation of sample maxima.

maximum, quantile, performance, correlation, repeated samples, mean, attenuation, repeatability

1. Introduction

Many studies of the functional ecology of animals involve estimating maximum performance in the laboratory. For example, studies of locomotion and feeding performance usually involve estimating maximum speed (e.g., Hertz et al. 1983) or maximum bite force (e.g., Herrel et al., 2005). Note that within-individual variability is expected and can be considered a special type of measurement error. By taking repeated measures, we hope to understand the true distribution of an individual's performance. Typically, maximum performance of each individual is estimated by using the largest value from a small number of trials by that individual. This procedure is statistically biased: on average, it underestimates the true maximum of the individual by an unknown amount. Moreover, the magnitude of the bias depends on sample size, since a larger sample size, on average, will yield a higher sample maximum (Gaines and Denny 1993). The problems involving the statistical estimation of maximum performance from a handful of trials have been described in several recent studies (Losos et al. 2002; Adolph and Pickering 2008), but statistical remedies are not yet available in the ecological literature. There is a well-developed statistical theory for estimating extreme values (Gumbel 1958), which has recently been applied to ecological problems (e.g., Gaines and Denny 1993; Denny and Gaines 2000; Katz et al. 2005). However, these methods require much larger sample sizes than are typically available in performance studies.

In this paper we propose an improved procedure for estimating performance from relatively small samples. Our method involves estimating defined performance quantiles (e.g., the 0.90 quantile) for each individual, rather than using the single largest value as an estimate of maximum performance. We first show that the conventional procedure for valuing performance (the sample maximum) already estimates a performance quantile rather than the true maximum of the individual, but that the particular quantile being estimated (a) is not identified and (b) is highly dependent on sample size. We then show how performance data can be used to explicitly estimate a specific performance quantile. We discuss the statistical advantages that the estimates of performance quantiles have over estimating maximum performance. Finally we recommend that the correlation of true maxima of individuals (e.g., correlation of maximum speed at two different temperatures) should be estimated by the correlation of the sample means (as opposed to the correlation of sample maxima).

As evidence that our methods are important for studies on physiological traits with very few repeated measures, we illustrate that interest in maximal performance based on repeated samples is apparent in the literature. For example, the maximum sprint speed of lizards usually taken from three to six runs per lizard is used in Bauwens et al. 1995, Bennett 1980, Bennett 1990, Cejudo and Márquez 2001, Hertz et al 1983, Holem et al. 2006, Losos 1990, Losos and Walton 1993, van Berkum 1986, van Berkum et al. 1989, and Van Damme et al. 2001. The maximum force with which lizards are able to cling to a wall taken from two to four samples is used in Irschick et al. 1996, Losos 1990, Losos et al. 1993, and Zani 2001. The distance that lizards are able to jump based on the maximum of two samples is used in Losos et al. 1993. Lizards' minimum prey catching and handling time of five samples is found in Van Damme et al. 1991. Similar studies with interest in maxima of repeated samples of individuals have been done with fish (Tudorache et al. 2007), and rodents (Djawdan and Garland 1988 and Dohm et al. 1996).

2. Quantiles and sample maxima

Quantiles are specific values that divide a statistical distribution into regions with specified area. Quantiles are most frequently reported as percentiles; for example, the 95th percentile (or 0.95 quantile) of a distribution of bite forces would be the smallest value that exceeds 95% of the values in the distribution. Measurements of maximum performance implicitly attempt to measure the 1.0 quantile of an individual's distribution of performances. In reality, using the maximum of a finite sample estimates a q quantile with $q < 1.0$; the specific quantile estimated depends on the sample size and the underlying distribution of performance values for an individual.

Several common quantile estimators demonstrate that sample maxima in fact do not estimate the maximum. For example, using the maximum value from a sample of size N, drawn from normally distributed data, yields (on average) an estimated quantile given approximately by

$$p(N) = \frac{N - 3/8}{N + 1/4} \tag{1}$$

(Hyndman and Fan 1996). Thus, the maximum of N = 5 measurements would estimate the 0.88 quantile of performance for that individual (on average), while the maximum of N = 10 data estimates the 0.94 quantile (typically, in work on physiological performance measures, N ranges from 3 to 6). With this framework, one could choose a per-individual sample size and obtain a corresponding quantile estimate of performance. However, such a method restricts the choice of quantiles to the discrete values available for a given value of N. As an alternative, we propose a method that involves pooling the data from all individuals in the sample, then applies the pooled data to estimate a specific performance quantile for each individual.

2.1 Using population data to obtain quantile estimates for individuals

We assume the population has constant within-individual variability, and therefore we propose the following quantile estimate: let

$$p_{q,i} = \bar{x}_i + s_w z_q \tag{2}$$

where $\bar{x}_i$ is the sample mean of lizard i, s_w is the pooled within-individual standard deviation (pooled across all individuals in the sample),

$$s_w = \sqrt{\frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i - 1)}} \tag{3}$$

and z_q is the z-score of the *q* quantile. For example, choosing $z_{0.9} = 1.28$ would yield a $p_{0.9,i}$ that estimates the 0.90 quantile performance value for individual i. We use the pooled standard deviation, as is typically done with ANOVA, as it is the best information of within individual variability given the small sample sizes. Our estimate is based on normality of error terms (as is our experience with real data, see Adolph and Pickering (2008)). However, if the residuals are not normally distributed, data can be transformed to better fit the quantile estimation method.

Quantile estimation has four major advantages over the sample maximum. First, quantile estimation allows for use of any quantile regardless of the number of samples. While the max of 5 samples is fixed at estimating the 0.88 quantile, using our quantile estimation allows for researchers to pick a standardized quantile (e.g., 0.90) to estimate. Second, by using quantile estimation, studies with different sample sizes can be meaningfully compared without any sample size adjustments. Since a typical practice in animal performance studies is to use the maximum of anywhere from three to six samples (see examples in the introduction), resulting estimates of the 1.0 quantile will actually range from the 0.81 to the 0.90 quantile of the distribution of performances, on average. Third, a quantile estimate is an approximately unbiased estimator of the true quantile of interest (under the assumption that the data are normally distributed); that is, in the long run, the average quantile estimate is equal to the true value of the quantile. Fourth, the quantile method allows one to estimate the same quantile for each individual before summarizing (e.g., averaging) individual trials. This is particularly important in cases where the number of trials differs across individuals (whose sample maxima would not be comparable).

2.2 Examples

Below we give two examples which demonstrate the performance of our estimator when applied to actual data (example 1) and simulated data (example 2). Note that these examples are not typical uses of the estimator (see recommended practices in the conclusion of this article), but rather they serve as an assessment of the usefulness of the method.

1. Sample maximum vs. estimator: empirical data

We illustrate our quantile method with laboratory data on sprint performance from two lizard species, *Sceloporus occidentalis* (Adolph and Pickering 2008) and *Sceloporus graciosus* (H. Groves and S. C. Adolph, unpublished data). Briefly, individual lizards were sprinted on a 2.5 m laboratory racetrack, 20 trials per individual (performed at two temperatures, 20°C and 35°C). We subtracted each individual's mean sprint speed from each trial, yielding a distribution of residual speeds for each individual; we then combined data across individuals (separately for the two species at each of the two different temperatures) and calculated the pooled within-individual standard deviation s_w for each species-temperature combination. Using the real data, we compared the average sample maxima with the theoretical values predicted by Eq. 2 given a particular sample size (Fig. 1). The sample data closely match the prediction of Eq. 2, indicating that the quantile estimator performs well for these empirical data.

2. Sample maximum vs. estimator: simulated data

To evaluate the quantile estimator we simulated experimental data by drawing samples from a normal distribution and comparing the sample maximum to the quantile estimate. These simulations showed that the quantile estimator (especially when using a pooled estimate of the variance) is more precise and better defined than the sample maximum (Fig. 2).

3. Correlations

A common question of interest is to determine whether two different performance measures are correlated (for example, maximum speed at two different temperatures). Assuming that the within-individual variances of performance are constant (as in, for example, Adolph and Pickering 2008) then the correlation of two performance measures based on any measure of location will be identical. That is, the correlation of any two location parameters (including the minimum, maximum, mean, or any quantile) is equal to the correlation of any other two location parameters. Because means are accurate, precise, and efficient, the correlation of means provides a very desirable estimation of the correlation of any location parameters. In short: the correlation of any two parameter measurements A_X and B_Y (that are both measures of location in a distribution) should be estimated with the same equation as for the correlation of means. Note that the invariance is only in the case of correlation, and obviously it is not relevant in the case of simply estimating the extrema (as seen above).

As Spearman (1904) first noted, the magnitude of correlation coefficients are underestimated in cases where there is measurement error (e.g., within-individual variance can be considered a type of measurement error). This bias of sample correlations towards zero is called attenuation. Correlation is mathematically defined as

$$cor(A,B) = \frac{cov(A,B)}{\sqrt{var(A)var(B)}}, \tag{4}$$

and using this definition we can write any correlation in terms of another as

$$cor(A,B) = cor(a,b)\frac{cov(A,B)}{cov(a,b)}\frac{\sqrt{var(a)var(b)}}{\sqrt{var(A)var(B)}}. \tag{5}$$

Specifically, we can estimate the correlation of true measures A and B for each individual with the correlation of sample measures a and b for each individual. If we are interested in the correlation of the true means of individuals, we can approximate it with the sample means. Using Eq. 5 and making basic assumptions of independence of errors, it is relatively straightforward to derive a correction for the attenuation in the correlation of sample means:

$$\hat{cor}(\mu_{X_i},\mu_{Y_i}) = \hat{cor}(\bar{x}_i,\bar{y}_i)\sqrt{\left(\frac{s_{Xb}^2+\frac{s_{Xw}^2}{N_X}}{s_{Xb}^2}\right)\left(\frac{s_{Yb}^2+\frac{s_{Yw}^2}{N_Y}}{s_{Yb}^2}\right)} = \hat{cor}(\bar{x}_i,\bar{y}_i)\sqrt{\left(1+\frac{1-rep_X}{N_X rep_X}\right)\left(1+\frac{1-rep_Y}{N_Y rep_Y}\right)} \tag{6}$$

In which $\bar{x}_i$ is the sample mean of the ith individual for performance trait X, s^2_{Xb} is the pooled between-individual sample variance, s^2_{Xw} is the pooled sample variance within individuals, N_X is the sample size, and rep_X is the sample repeatability (intraclass correlation). We put hats over the correlations to denote that the equation is an estimate using the sample variances instead of true variances. Adolph and Hardin (2007) give equivalent expressions to those in Eq. 6 and remark how correlations of sample maxima are also attenuated (although they did not propose a correction). Since there is a correction factor for means (above) we recommend that regardless of what the true location measures of interest (e.g., mean, min, max, quantile) is, the correlation of sample means with the correction factor above should be used to estimate the correlation of the true parameters. Adolph and Hardin (2007) discuss the allocation of sample size between and within individuals based on repeatability for the correlation of sample means.

3.1 Examples with Simulated Data

To demonstrate that the correction factor applied to means outperforms correlations on any raw estimates of location, we simulated data from both symmetric (normal) as well as right skewed (chi squared) distributions to model both within and between variability.

In Figure 3 we see that the effect of attenuation is greater for small per-individual sample sizes, and the correction coefficient nicely removes the bias from attenuation (especially for larger values of N where s^2 is a better estimate of σ^2). If individual performance values are symmetrically (normally) distributed, then correlations involving mean performance show less attenuation than correlations involving sample minima or maxima. Indeed, if the measure of interest is in the shorter tail of a skewed distribution (e.g., the maximum performance in a left skewed distribution for trained athletes), the correlation of the extreme value will outperform the correlation of the uncorrected mean (we see this by looking at the plot for the minimum of replicates in a right skewed distribution). However, because we cannot correct for attenuation of correlations involving minima or maximum, we continue to recommend using the correlation of the mean values as the optimal estimator. Overall, the corrected mean correlation outperforms uncorrected correlations for any sample size or distribution. (See Adolph and Hardin (2007) and Adolph and Pickering (2008) for complete examples on real data.)

4. Conclusions

For physiological performance measures whose variability is reasonably described by a normal distribution, we recommend the following:

- Quantiles should be used to estimate peak performance instead of sample extrema (min or max).

- Corrected correlations of sample means should be used to estimate correlation of any two location values (mean, min, max, quantile).

With larger sample sizes (N>10), the problems with estimating extrema (as well as the problems with correlation attenuation) become less serious. However, we believe our methods are still valuable for larger sample sizes. Though attenuation decreases, adjusting the correlations in large sample sizes will not hurt the estimates; additionally, as seen in Eq. 1, measuring quantiles instead of maxima will lead to estimates of the same quantities even for different sample sizes.

In our work we have addressed the issue of using extrema as measures of individual performance in studies of ecomorphology and functional ecology. Because of their bias, sample minima and maxima do not accurately measure true minima or maxima of performance traits of individual performance. We recommend, instead, using sample quantiles to estimate true quantiles of individual performance. Not only are the sample quantiles unbiased, but they are also more precisely defined and easier (than extrema) to use for cross-study comparisons. When estimating correlations involving measures of location (e.g., min, max, quantile, mean), we recommend using the correlation of means corrected for attenuation because of the invariance property of correlation and the efficiency property of sample means.

AWH was funded through grants from the Howard Hughes Medical Institute and Pomona College Summer Undergraduate Research Program. The data used in Figure 1 was collected by Trevor Pickering (*Sceloporus occidentalis* data) and Holden Groves (*Sceloporus graciosus* data). Computing resources were available through an NIH AREA grant (#1 R15 AG021907-01A1).

Adolph, S. C., and J. Hardin. 2007. Estimating phenotypic correlations: correcting for bias due to intraindividual variability. Functional Ecology, 21:178–184.

Adolph, S. C., and T. Pickering. 2008. Estimating maximum performance: Effects of intraindividual variation. The Journal of Experimental Biology, 211:1336–1343.

Bauwens, Jr., D., T. Garland, A. M. Castilla, and R Van Dame. 1995. Evolution of sprint speed in lacertid lizards: Morphological, physiological, and behavioral covariation. Evolution, 49:848–863.

Bennett, A.F. 1980. Thermal dependence of lizard behaviour. Animal Behaviour, 28:752–762.

Bennett, A.F.1990. Thermal dependence of locomotor capacity. American Journal of Physiology, 259:R253–R258.

Cejudo, D. and R. Márquez. 2001. Sprint performance in the lizards *Gallotia simonyi* and *Gallotia stehlini* (*Lacertidae*): Implications for species management. Herpetologica, 57:87–98.

Denny, M., and S. Gaines. 2000. Chance in biology: using probability to explore nature. Princeton University Press, Princeton, New Jersey, USA.

Djawdan, M. and T. Garland, Jr. 1988. Maximal running speeds of bipedal and quadrupedal rodents. Journal of Mammalogy, 69:765–772.

Dohm, M.R., J.P. Hayes, and T. Garland, Jr. 1996. Quantitative genetics of sprint running speed and swimming endurance in laboratory house mice (*Mus domesticus*). Evolution, 50:1688–1701.

Gaines, S. D., and M. W. Denny. 1993. The largest, smallest, highest, lowest, longest, and shortest: extremes in ecology. Ecology 74:1677–1692.

Gumbel, E. J. 1958. Statistics of extremes. New York, Columbia University Press, 1958.

Herrel, A., J. Podos, S. K. Huber and A. P. Hendry. 2005. Bite performance and morphology in a population of Darwin's finches: implications for the evolution of beak shape. Functional Ecology 19: 43-48.

Hertz, P.E., R. B. Huey, and E. Nevo. 1983. Homage to Santa Anita: Thermal sensitivity of sprint speed in agamid lizards. Evolution, 37:1075–1084.

Holem, R.R., W.A. Hopkins, and L.G. Talent. 2006. Effect of acute exposure to malathion and lead on sprint performance of the western fence lizard (*Sceloporus occidentalis*). Archives of Environmental Contamination and Toxicology, 51:111–116.

Hyndman, R. J., and Y. Fan. 1996. Sample quantiles in statistical packages. The American Statistician, 50:361–365.

Irschick, D.J., C.C. Austin, K. Petren, R.N. Fisher, J.B. Losos, and O. Ellers. 1996. A comparative analysis of clinging ability among pad-bearing lizards. Biological Journal of the Linnean Society, 59:21–35.

Katz, R.W., G.S. Brush, and M.B. Parlange. 2005. Statistics of extremes: modeling ecological disturbances. Ecology 86: 1124–1134.

Losos, J.B. 1990. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: An evolutionary analysis. Ecological Monographs, 60:369–388.

Losos, J. B., D. A. Creer, and J. A. Schulte II. 2002. Cautionary comments on the measurement of maximum locomotor capabilities. Journal of Zoology, London 258: 57-61.

Losos, J.B., B. M. Walton, and A.F Bennett. 1993. Trade-offs between sprinting and clinging ability in Kenyan chameleons. Functional Ecology, 7:281–286.

Spearman, C. 1904. The proof and measurement of association between two things. The American Journal of Psychology, 15(1):72–101.

Tudorache, C., R. Blust, and G. De Boeck. 2007. Swimming capacity and energetics of migrating and non-migrating morphs of three-spined stickleback *Gasterosteus aculeatus L.* and their ecological implications. Journal of Fish Biology, 71:1448–1456.

van Berkum, F. H. 1986. Evolutionary patterns of the thermal sensitivity of sprint speed in *Anolis* lizards. Evolution, 40:594–604.

van Berkum, F.H., R. B. Huey, J. S. Tsuji, and T. Garland. 1989. Repeatability of individual differences in locomotor performance and body size during early ontogeny of the lizard *Sceloporus occidentalis*. Functional Ecology, 3:97–105.

Van Damme, R., D. Bauwens, and R. F. Verheyen. 1991. The thermal dependence of feeding behaviour, food consumption and gut-passage time in the lizard *Lacerta vivipara Jacquin*. Functional Ecology, 5:507–517.

Zani, P.A. 2001. Clinging performance of the western fence lizard, *Sceloporus occidentalis*. Herpetologica, 57:423–432.

Figure 1. Comparison of sample maximum and theoretical quantile estimates. Using the mean centered residual values calculated from the real data, we randomly sampled N residual speeds from the distribution and recorded the maximum value of each of these N trials. Note, the data were collected from two species of lizard, *Sceloporus occidentalis* (*S.o.*) and *Sceloporus graciosus*

(*S.g.*), at two different temperatures, 20°C and 35°C. We repeated the resampling procedure 10,000 times for each value of N from 2 to 10 and plotted the average maximum values for each N (solid dots). Additionally, our plot shows the quantile we would expect to estimate through the process of taking the maximum value from a sample of size N (solid line). For example, the maximum of a sample size of N=2 should estimate the 0.72 quantile (Eq. 1). Using the quantile q=0.72 within Eq. 2 gives the value for the solid line at N=2. The empirical data show that on average, the sample maximum increases with sample size, and we know that the sample maximum always underestimates the true maximum. The quantile estimation function, on the other hand, gives an excellent approximation of the sample maximum for a given sample size.

Figure 2. Distribution of sample maxima and estimated quantiles from simulated performance data drawn from a normal distribution. One sample consists of 1 individual with N (3, 5, or 10) trials; 50,000 samples were taken. For each sample, we calculated (1) the individual's maximum, (2) the 0.95 quantile for each individual, using Eq. 2, and (3) the 0.95 quantile for each individual using a pooled estimate of the variance within Eq. 2. To find the pooled estimate of the variance, we randomly sampled 19 additional variances (as if we had a sample of 20 individuals) and averaged the 20 variances. Even for a reasonably large number of trials (N=10), the maximum observation is typically much lower than even the true 0.95 quantile. Additionally, the pooled estimate of variability increases the precision of the estimate. Even without pooling the variability, the un-pooled estimate of the 0.95 quantile gives a reasonable estimate of the true quantile.

Figure 3. Effect of within-individual sample size on the correlation of different measures of location of two performance traits. Bivariate data was simulated with a true correlation of 0.7, a between-individual variance of 0.5, and a within-individual variance of 0.1 (i.e., a repeatability of 0.83). The entire simulation was repeated 4 times: with normal within and between errors; with normal within and chi-square (skew right) between errors; with chi-square within and normal between errors; and with chi-square within and between errors. The different lines represent the average correlations of different measures of location (sample minima, sample maxima, sample means, and sample means corrected for attenuation as in Eq. 6 where k is the correction coefficient). Sample correlations are normalized to the true correlation (i.e., each sample correlation was divided by 0.7). These graphs demonstrate how the corrected correlation of means is much less biased than the raw correlation of means or the correlation of extrema (maxima or minima).

Figure 1:

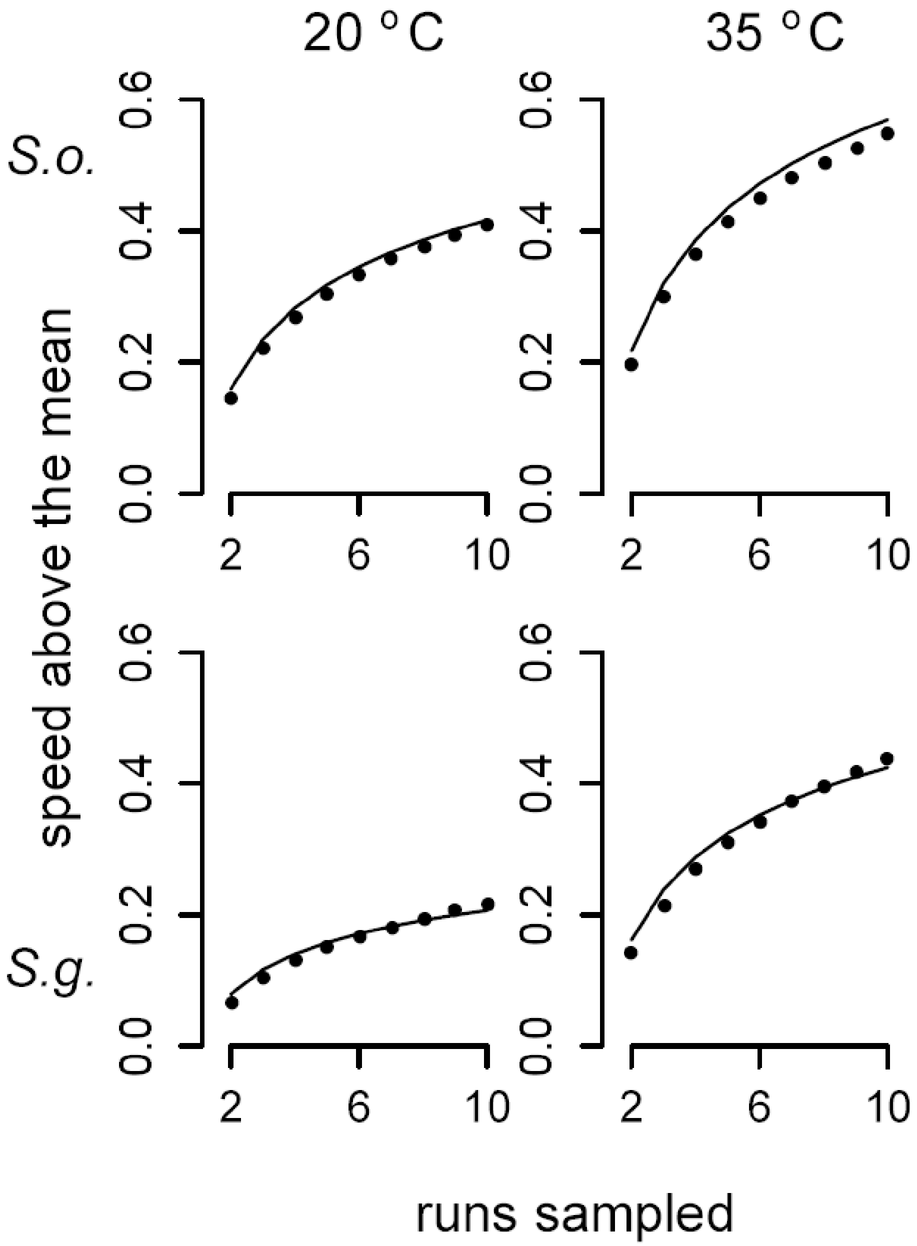

20 °C
35 °C
S.o.
S.g.
speed above the mean
0.0
0.2
0.4
0.6
2
6
10
runs sampled

Figure 2:

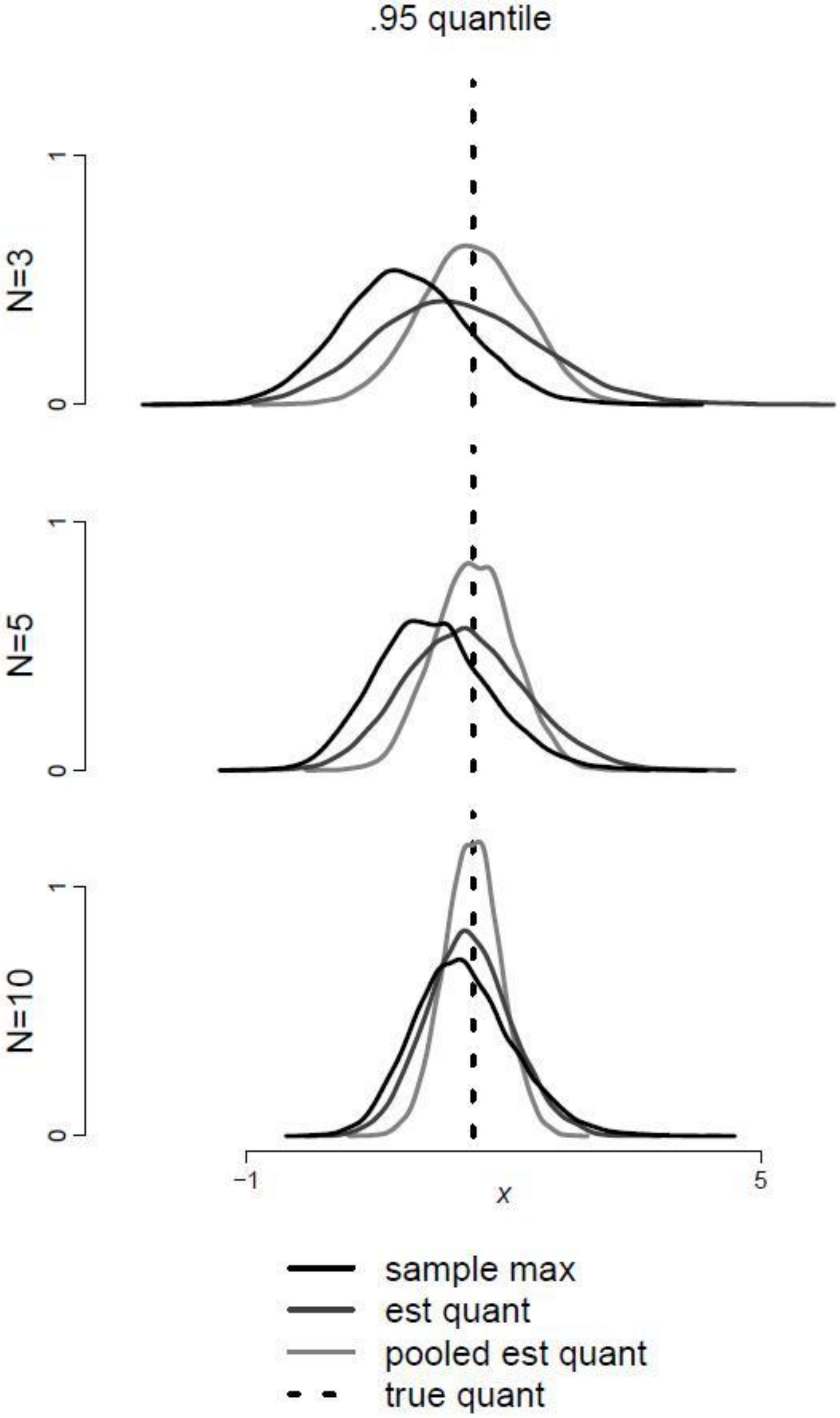
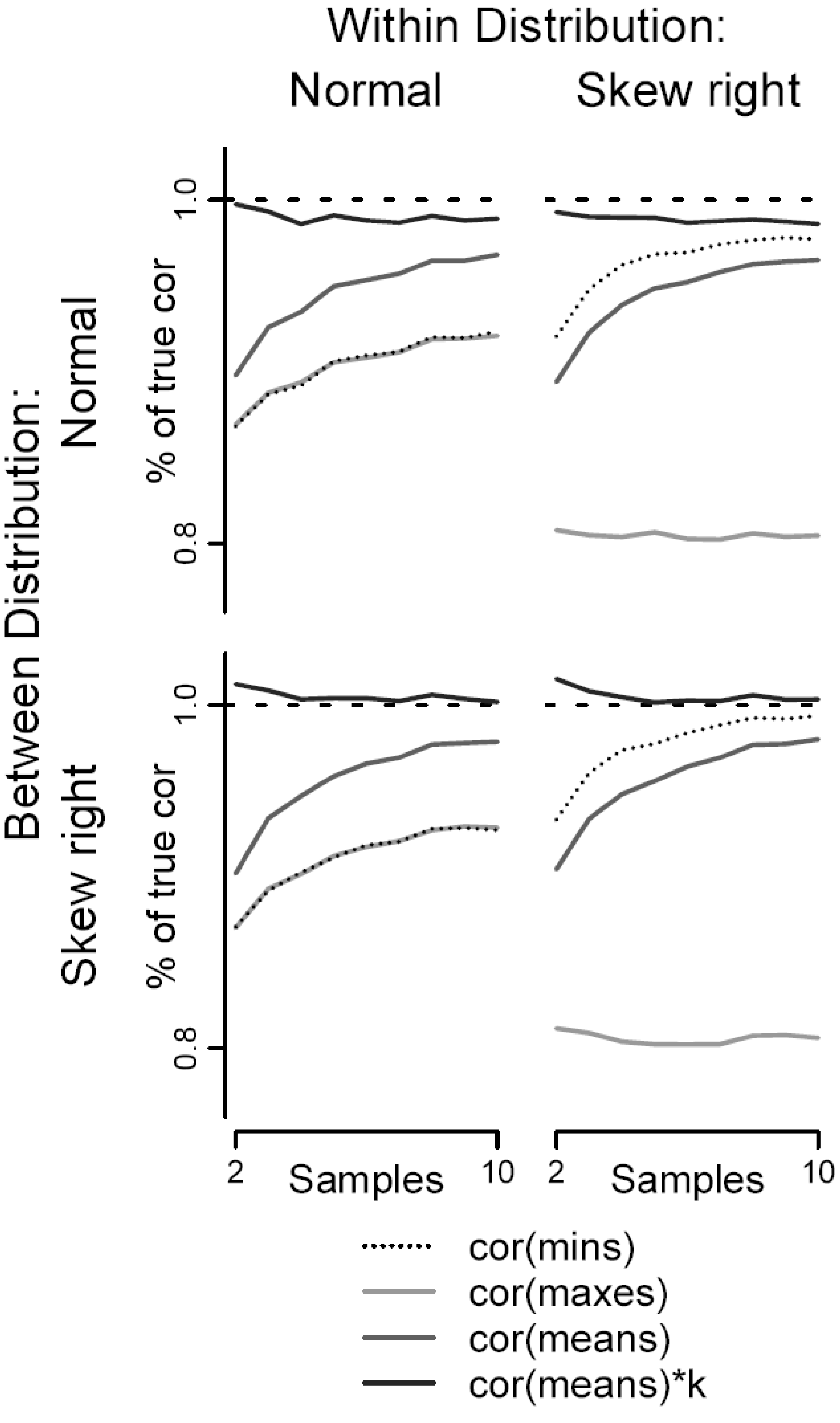

.95 quantile
N=3
1
0
N=5
1
0
N=10
1
0
−1
5
x
sample max
est quant
pooled est quant
true quant

Figure 3:


Within Distribution:
Normal
Skew right
Between Distribution:
Normal
Skew right
% of true cor
1.0
0.8
% of true cor
1.0
0.8
2
Samples
10
2
Samples
10
cor(mins)
cor(maxes)
cor(means)
cor(means)*k