Removing Bias for Fluctuating Asymmetry in Meristic Characters

John R. Young

Fluctuating asymmetry (FA) in meristic characters is a random difference in count of a character between the right and left sides of an organism. Although FA has been used as an index of developmental instability for more than 30 years, it has not been closely related to the intensity of factors commonly thought to induce instability. The poor relationship could be due to use of biased FA metrics based on differences in bilateral counts, bias introduced by counting errors, low sample sizes, or a truly low correspondence between FA and stressor intensity. I demonstrate why FA metrics based on count differences are inherently biased for meristic characters, develop an unbiased maximum likelihood metric for FA ($\hat{\sigma}_A$), and establish a general method to remove bias caused by counting errors. A likelihood ratio test based on the new metric is nearly as powerful as the F-test, and has a Type I statistical error closer to the nominal level. When applied to bilateral counts of sternopleural bristles in *Drosophila*, the new metric confirmed previous results based on count differences but removed the concern that the FA estimates were biased by the character mean. Apparent differences in pectoral fin ray FA between rainbow and cutthroat trout disappeared after correction for counting errors.

Key Words: Developmental instability; Maximum likelihood; Stress.

1. INTRODUCTION

Fluctuating asymmetry (FA) in meristic characters is a random difference in count of a character between sides of an organism: count on right side – count on left side ($r - l$), when that character would typically have the same count on both sides (Van Valen 1962). In order for the asymmetry to be classified as FA, $(r - l)$ must have a mean of 0 and be normally distributed (Palmer 1994). The most commonly used metric for FA is $\overline{|r - l|} = \sum |r_i - l_i|/n$, which has many desirable statistical properties when used for morphometric data (Palmer and Strobeck 1992).

Although FA has been used as an index of developmental instability for more than 30 years, it has not been a reliable indicator of the intensity of factors commonly thought to induce instability, particularly when used on meristic characters. Swain (1987), using simulated data and actual asymmetry data from the fish *Rivulus marmoratus*, demonstrated that

John R. Young is Senior Scientist, ASA Analysis and Communication, Inc., 921 Pike Street, P.O. Box 303, Lemont, PA 16851 (E-mail: *jyoung@asaac.com*).

$\overline{|r-l|}$ is not independent of the character mean when the amount of asymmetry is relatively low. Other FA metrics based on $(r-l)$ are similarly affected. Since 1987, this measurement problem has been elaborated by Palmer (1994), but despite theoretical advances in the study of FA, no solution to the measurement bias has been described (Palmer and Strobeck 2003).

An example of the uncertainty that can arise in FA studies as a result of the "Swain effect" was provided by Kristensen et al. (2003). They raised genetically identical groups of *Drosophila mercatorum* at temperatures ranging from 25°C to 40°C. Although FA of sternopleural bristles, as measured by $\overline{|r-l|}$ and $s^2_{(r-l)}$, increased at the higher test temperatures, none of the high-temperature treatments (35°C to 40°C) had significantly greater FA than the control sample raised at 25°C. However, two morphometric characters based on wing dimensions showed significantly higher FA at many of the high-temperature treatments. The potential bias in FA estimates for the meristic character clouds the interpretation of the difference in results for the meristic and morphometric characters.

Counting errors have also affected the results of meristic FA studies. Errors in character counts are extremely detrimental to FA studies because they inflate the amount of apparent asymmetry, leading to biased results (Palmer 1994). A few studies have reported undertaking replicate counts in order to assess the potential bias (e.g., Kristensen et al. 2003; Weiner and Rago 1987), but most have not reported replicating counts. Hubert and Alexander (1995) used three different observers to count meristic characters of hatchery-reared rainbow (*Oncorhynchus mykiss*) and cutthroat trout (*Oncorhynchus clarkii*). Observer 2 found unequal pectoral fin ray counts in 13% of 50 rainbow trout and 21% of 50 cutthroat trout. Observer 3 reported unequal counts in 14% and 24% of the same fish. Observer 2 would have estimated $\overline{|r-l|} = 0.28$ for rainbow trout and 0.88 for cutthroat, while the data from observer 3 produce values of 0.26 for rainbow trout and 0.55 for cutthroat. Hubert and Alexander (1995) correctly concluded that counting errors may limit the usefulness of FA. Although the bias of measurement error in morphometric characters can be removed through ANOVA (Palmer 1994), no general method of dealing with errors exists for meristic data.

Small sample sizes have also plagued FA studies. In many instances, investigators attempted to assess FA differences using sample sizes in the range of 50 to 100, or even fewer, from each population. Unless FA differences are relatively large, and counting error is low, power of statistical tests with sample sizes below 100 is likely to be very low.

In this manuscript I (1) develop a maximum likelihood metric that is nearly unbiased by the "Swain effect"; (2) provide, when replicate counts are available, a way to estimate and remove the bias associated with counting errors; (3) compare the power of likelihood ratio tests based on the metric with power of the F-test; and (4) demonstrate the metric on actual FA data.

2. METHODS

Swain (1987) used Falconer's (1981) conceptual developmental model for meristic characters. The character state (meristic count) is determined by the value of an underlying continuous distribution called the liability (L). There are thresholds along the liability that

determine the translation of the continuous liability into integer character states. Swain considered the realized liability to be the sum of the population mean liability (M), a fixed genotype effect (G), and the developmental error (E) : $L = M + G + E$, where E is distributed $N(0, \sigma^2)$. Although Swain suggested that developmental noise could be assessed more accurately by estimating σ than through the commonly employed $(r - l)$ metrics, thus far implementation of this approach to measuring FA has not been reported. The proposed model uses this general approach, but some of the nomenclature has been revised for clarity:

2.1 DEFINITIONS

μ = population mean liability for a bilateral character

GE_i = $N(0, \sigma_{GE}^2)$ genetic and/or environmental effect on the liability for organism i

L_i = organismic liability for the character for organism $i = \mu + GE_i$

A_{ij} = independent $N(0, \sigma_A^2)$ developmental disturbance on side j of organism i

L_{ij} = realized liability for the character on side j of organism $i = L_i + A_{ij}$

R_{ij} = random amount by which L_{ij} is rounded to an integer value

Y_{ij} = true character state on side j of organism $i = L_i + A_{ij} + R_{ij}$

E_{ij} = discrete counting error on side j of organism i

O_{ij} = observed count of character on side j of organism $i = Y_{ij} + E_{ij}$

Using this conceptual model, the problem with FA metrics based on $(r - l)$ for meristic characters is readily apparent. In the absence of counting error:

$$(r - l)_{\text{meristic}} = (L_i + A_{ir} + R_{ir}) - (L_i + A_{il} + R_{il}) = A_{ir} + R_{ir} - A_{il} - R_{il} \quad (2.1)$$

$$
\begin{aligned}
\text{var}(r - l)_{\text{meristic}} = \ & \text{var}(A_{ir}) + \text{var}(R_{ir}) + \text{var}(A_{il}) \\
& + \text{var}(R_{il}) + 2[\text{cov}(A_{ir}, R_{ir}) + \text{cov}(A_{il}, R_{il}) \\
& - \text{cov}(A_{ir}, A_{il}) - \text{cov}(R_{ir}, R_{il}) - \text{cov}(A_{ir}, R_{il}) - \text{cov}(A_{il}, R_{ir})].
\end{aligned}
$$
$$(2.2)$$

Under the assumptions inherent in the model $\text{var}(A_{ir}) = \text{var}(A_{il}) = \sigma_A^2, \text{cov}(A_{ir}, A_{il}) = 0$, $\text{var}(R_{ir}) = \text{var}(R_{il}) = \text{var}(R_{ij})$, $\text{cov}(A_{ir}, R_{ir}) = \text{cov}(A_{il}, R_{il})$, and $\text{cov}(A_{ir}, R_{il}) = \text{cov}(A_{il}, R_{ir})$, therefore:

$$\text{var}(r - l)_{\text{meristic}} = 2\left[\sigma_A^2 + \text{var}(R_{ij}) + 2\text{cov}(A_{ir}, R_{ir}) - 2\text{cov}(A_{ir}, R_{il}) - \text{cov}(R_{ir}, R_{il})\right]. \quad (2.3)$$

To the extent the variance and covariance terms involving the R_{ij} are nonzero, $\text{var}(r - l)$ will not provide an unbiased estimate of σ_A^2. Additionally, as these variances and covariances change with μ, σ_{GE}^2, or σ_A^2, the bias will also change. In contrast, for morphometric characters measured without error:

$$\text{var}(r - l)_{\text{morphometric}} = \text{var}(A_{ir}) + \text{var}(A_{il}) + 2\text{cov}(A_{ir}, A_{il}) = 2\sigma_A^2. \quad (2.4)$$

Therefore, $s_{(r-l)}$ for morphometric characters would be $\sigma_A\sqrt{2}$. Because the most common metric for FA ($|r - l|$) is an estimate of $s_{(r-l)}$ (Palmer 1994), all FA metrics will be expressed as standard deviations rather than variances to facilitate comparisons.

An unbiased estimator of meristic FA must be uncontaminated by the variances and covariances of R_{ij}. If the developmental disturbances on each side of the organism are independent from each other and from the genetic/environmental effect, then the probability of a true bilateral state (y_{ir}, y_{il}) is:

$$\Pr(y_{ir}, y_{il}) = \int_{L_{il}=y_{il}-0.5}^{y_{il}+0.5} \int_{L_{ir}=y_{ir}-0.5}^{y_{ir}+0.5} \int_{L_i=-\infty}^{\infty} f(L_{ir}, L_{il}|L_i) f(L_i) dL_i dL_{ir} dL_{il},$$
(2.5)

where $y_{ir} \geq 1$, $y_{il} \geq 1$;

$$f(L_{ir}, L_{il}|L_i) = \frac{1}{2\pi\sigma_A^2} \exp\left[\frac{(L_{ir}-L_i)^2 + (L_{il}-L_i)^2}{-2\sigma_A^2}\right]$$
(2.6)

$$f(L_i) = \frac{1}{\sqrt{2\pi\sigma_{GE}^2}} \exp\left[\frac{(L_i-\mu)^2}{-2\sigma_{GE}^2}\right].$$
(2.7)

Simply stated, the true number of meristic elements on the right and left side are the rounded values of independent $N(0, \sigma_A^2)$ random deviations from the organismic liability, where the organismic liability is itself a $N(\mu, \sigma_{GE}^2)$ random variable.

Since counts of meristic elements can differ from the true character state due to counting errors, the probability of an observed bilateral state (o_{ir}, o_{il}) is:

$$\Pr(o_{ir}, o_{il}) = \sum_{e_{il}=-\infty}^{o_{il}} \sum_{e_{ir}=-\infty}^{o_{ir}} \Pr(E_{ir} = e_{ir})$$
$$\times \Pr(E_{il} = e_{il}) \Pr(Y_{ir} = o_{ir} - e_{ir}, Y_{il} = o_{il} - e_{il}).$$
(2.8)

If the distributional form of E is known, then maximum likelihood methods can be used to derive estimates for the parameters. However, for many meristic characters, particularly those with large values, it may be difficult to determine the true values of y_{ij} and e_{ij}, but differences between replicated counts can be used to determine the distribution and parameter values for E_{ij}:

Let D_{ij} = difference between two counts of a character on side j of organism i

$$= (Y_{ij} + E_{ij1}) - (Y_{ij} + E_{ij2}) = E_{ij1} - E_{ij2}.$$

The distribution of a difference between two identically distributed independent random variables can be determined by the cumulative distribution function method (Mood et al. 1974):

$$\Pr(D_{ij} = d) = \sum_{x=-\infty}^{\infty} \Pr(E_{ij1} = x) \Pr(E_{ij2} = x - d).$$
(2.9)

If a set of replicated counts are available, the actual distribution of the count differences can be compared to candidate discrete distributions. One possible distribution for count errors

would be that the absolute values of the errors would have a geometric distribution:

$$
\Pr\left(E_{ij}=x\right) \begin{aligned} &= q && \text{for} \quad x=0 \\ &= \frac{q(1-q)^{|x|}}{2} && \text{for} \quad x=-\infty,\dots,-3,-2,-1,1,2,3,\dots\infty, \end{aligned} \tag{2.10}
$$

where $0 \le q \le 1$. A similar adjustment to the Poisson would be a second possibility:

$$
\Pr(E_{ij}=x) \begin{aligned} &= \exp(-\lambda) && \text{for} \quad x=0 \\ &= \frac{\exp(-\lambda)\lambda^{|x|}}{2|x|!} && \text{for} \quad x=-\infty,-3,-2,-1,1,2,3,\dots,\infty \end{aligned} \tag{2.11}
$$

where $\lambda \ge 0$

These two candidates were chosen without any theoretical justification other than they are discrete and symmetric around 0, thus possibly appropriate for situations where counting errors impose no directional bias.

Maximum likelihood estimates of the counting error parameter are found by determining the value of q or λ that minimizes

$$
\begin{aligned}
&LL_1(q \text{ or } \lambda | d) \\
&= -\sum_{ij}\sum_{k=1}^{K_i-1}\log\left(\sum_{x=-\infty}^{\infty}\Pr\left(E_{ij1}=x|q \text{ or } \lambda\right)\Pr\left(E_{ij2}=x-d_{ijk}|q \text{ or } \lambda\right)\right),
\end{aligned} \tag{2.12}
$$

where $d_{ijk}=o_{ijk+1}-o_{ijk}$, which produces $2(K_i-1)$ values of d_{ijk} for an organism with K_i replicated counts.

Once the count error distribution and its parameter are determined, the maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}_{\mathrm{GE}}$, and $\hat{\sigma}_A$ are found by minimizing the weighted log-likelihood:

$$
\begin{aligned}
&LL_2\left(\mu,\sigma_{\mathrm{GE}},\sigma_A|o_r,o_l,\hat{q} \text{ or } \hat{\lambda}\right) \\
&= -\sum_{i=1}^{N}\sum_{k=1}^{K_i}\frac{1}{K_i}\log\left[\sum_{e_{il}=-\infty}^{o_{il}}\sum_{e_{ir}=-\infty}^{o_{ir}}\Pr\left(E_{irk}=e_{irk}|\hat{q} \text{ or } \hat{\lambda}\right)\right. \\
&\qquad\qquad \times \Pr\left(E_{ilk}=e_{ilk}|\hat{q} \text{ or } \hat{\lambda}\right) \\
&\qquad\qquad \times \Pr\left(Y_{ir}=o_{irk}-e_{irk},Y_{il}=o_{ilk}-e_{ilk}|\mu,\sigma_{\mathrm{GE}},\sigma_A\right)].
\end{aligned} \tag{2.13}
$$

The $1/K_i$ weighting term ensures that each of the N organisms in the sample is weighted equally when estimating μ, σ_{GE}, and σ_A, regardless of how many replicate counts are made. In situations where no replicate counts are made, the usual assumption that $\Pr(E_{ij}=0)=1$ is required, and the parameters are estimated using only LL_2.

Minimization of the likelihood functions was accomplished with the NLIN procedure in SAS, using the Gauss method. Starting values for the σ parameters were derived from a one-way ANOVA with the right and left counts treated as replicate observations: starting value for $\sigma_{\mathrm{GE}} = \sqrt{SS_i/df_i}$ (SS_i = sum of squares between individuals); starting value for $\sigma_A = \sqrt{SS_{\mathrm{error}}/df_{\mathrm{error}}}$. The sample mean of the observed count was used as the starting value for μ, and a constant (0.5) was used as the starting value for the error parameter. Instead of evaluating counting errors over the interval $(-\infty, \infty)$, they were evaluated over a lesser range $(-5, 5)$ to reduce computation time. All parameters were restricted to be ≥ 0.

In addition, a likelihood ratio test (Mood et al. 1974) was constructed to compare the values of $\hat{\sigma}_A$ for two or more populations. For two populations, the sum of the log-likelihoods over both populations when μ_1, μ_2, σ_{GE1}, σ_{GE2}, (q_1 or λ_1), (q_2 or λ_2), and σ_{A1} and σ_{A2} are estimated is compared to the sum of the log-likelihoods when only a single σ_A parameter is estimated. This test methodology can be easily expanded for multiple populations (Hilborn and Mangel 1997), adding 1 degree of freedom for each additional population.

The unbiasedness of the FA metric ($\hat{\sigma}_A$) over a range of combinations of μ, σ_{GE}, and σ_A, was evaluated by estimating the parameters for a "sample" of 10,000 bilateral counts whose frequencies exactly fit the conceptual model. Using this "sample," if the metric and maximum likelihood algorithm are unbiased, the resulting value of $\hat{\sigma}$ will match σ_A.

Performance of $\hat{\sigma}_A$, in terms of unbiasedness and linearity with true σ_A, was compared to estimators developed from two other commonly used FA metrics: $\overline{|r - l|}$, and the sample standard deviation of $(r–l)$, $s_{(r-l)} = \sqrt{\sum(r - l)^2/(n - 1)}$ (Palmer and Strobeck 1992). As shown in Equation (2.4), $s_{(r-l)} = \sigma_A\sqrt{2}$, so $s'_{(r-l)} = s_{(r-l)}/\sqrt{2}$ would be an unbiased estimator for σ_A for a morphometric character. In the absence of directional asymmetry or antisymmetry, $\overline{|r - l|}$ is also an unbiased estimate of $s_{(r-l)}$ (Palmer 1994), thus $\overline{|r - l|}' = \overline{|r - l|}/\sqrt{2}$ should also estimate σ_A.

To assess the effect of counting errors, and the ability of the estimation method to eliminate bias in $\hat{\sigma}_A$, samples of size 200 were simulated for two populations: ($\mu_1 = \mu_2 = 20.25$, $\sigma_{\text{GE1}} = \sigma_{\text{GE2}} = 0.5$, $\sigma_{A1} = 0.5$ and $\sigma_{A2} = 0.25$). A Poisson count error distribution with λ ranging from 0 to 0.2 was assumed, and either 0% or 25% replicate measurements. Parameter estimates were made for each simulated population and differences in σ_A were tested with the likelihood ratio test at Type I error $= 0.05$. The simulations were repeated 250 times to estimate the mean $\hat{\sigma}_A$ and power of the test.

Power and Type I error rates of the likelihood ratio test, without counting errors, were examined by generating 500 sets of pairs of random samples of size 50 or 100 from two populations when $\mu_1 = \mu_2 = 20.25$, and $\sigma_{\text{GE1}} = \sigma_{\text{GE2}} = 0.5$ and $\sigma_{A1} = 0.5$ and $0.1 \leq \sigma_{A2} \leq 0.9$, and determining what fraction of the tests were statistically significant (nominal Type I error rate $= 0.05$). Results for the likelihood ratio test were compared to those of the F-test, calculated using $s^2_{(r-l)}$ (Palmer 1994).

These methods for estimating the FA parameters were applied to the sternopleural bristle count data of Kristensen et al. (2003), without error correction, and to the pectoral fin ray counts of Hubert and Alexander (1995), with error correction.

3. RESULTS

For the distribution generated from Equation (2.5) with $\mu = 20.0$, $\sigma_{\text{GE}} = 0.25$, only $\hat{\sigma}_A$ gave unbiased estimates σ_A over the range $0.025 \leq \sigma_A \leq 5$ (Figure 1). $\overline{|r - l|}'$ under-estimated σ_A over the entire range, and $s'_{(r-l)}$ overestimated it when $\sigma_A < 1$, but converged to σ_A at higher values ($\sigma_A > 1$).

The $\hat{\sigma}_A$ metric proved to be much less sensitive to the value of μ than are the asymmetry statistics based on $(r - l)$ (Figure 2). When μ varied from 20.0 to 20.5, with $\sigma_{\text{GE}} = 0.25$ and true $\sigma_A = 0.1$, σ_A was essentially constant even as μ approached 20.5. Slight deviations
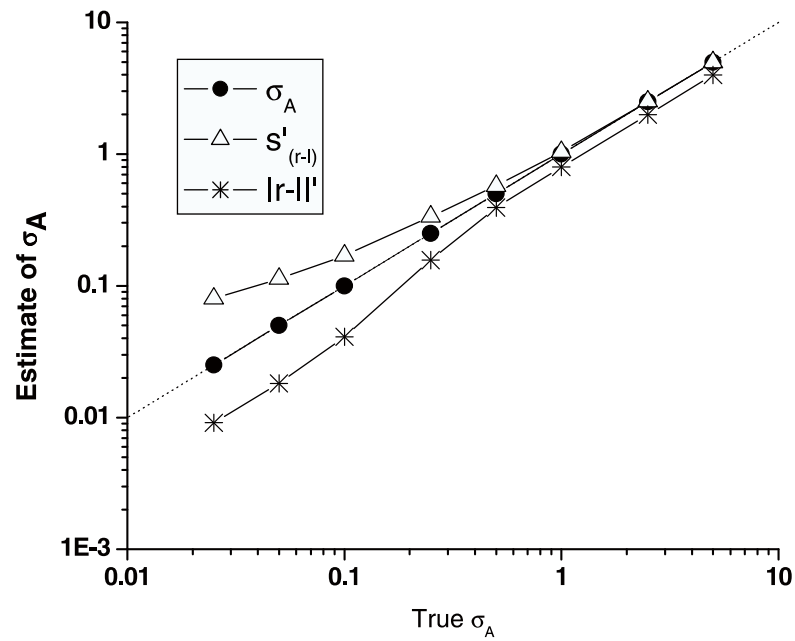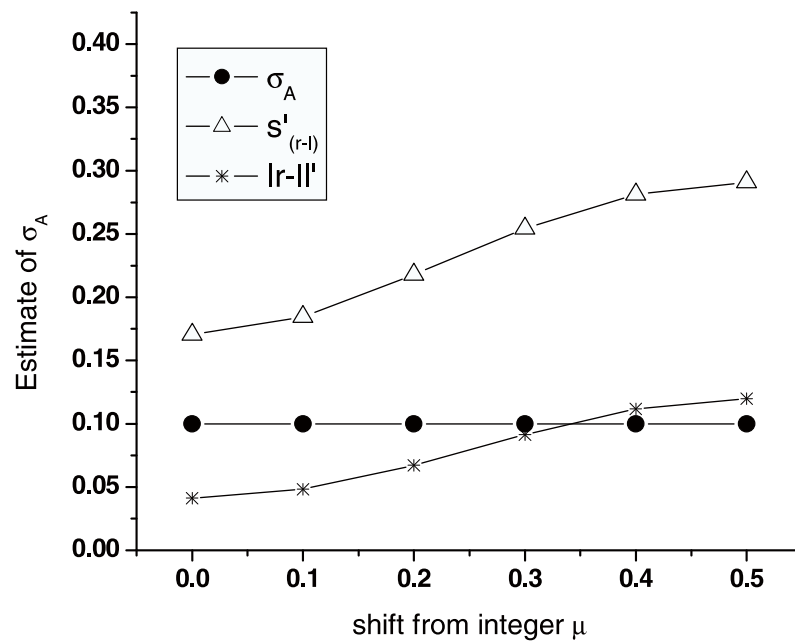
Figure 1.   Asymmetry index values as a function of true σ_A. Values estimated from 10,000 pairs of right and left counts distributed exactly in accordance with $\mu = 20.0$, $\sigma_{GE} = 0.25$, and the true σ_A. (Diacritical marks on $\hat{\sigma}_A$ and $\overline{|r - l|}'$ are omitted in the legend.)

Figure 2.   Asymmetry index values as a function of shift from integer value of μ. Values estimated from 10,000 pairs of right and left counts distributed exactly in accordance with $20.0 \leq \mu \leq 20.5$, $\sigma_{GE} = 0.25$, and $\sigma_A = 0.1$. (Diacritical marks on $\hat{\sigma}_A$ and $\overline{|r - l|}'$ are omitted in the legend.)
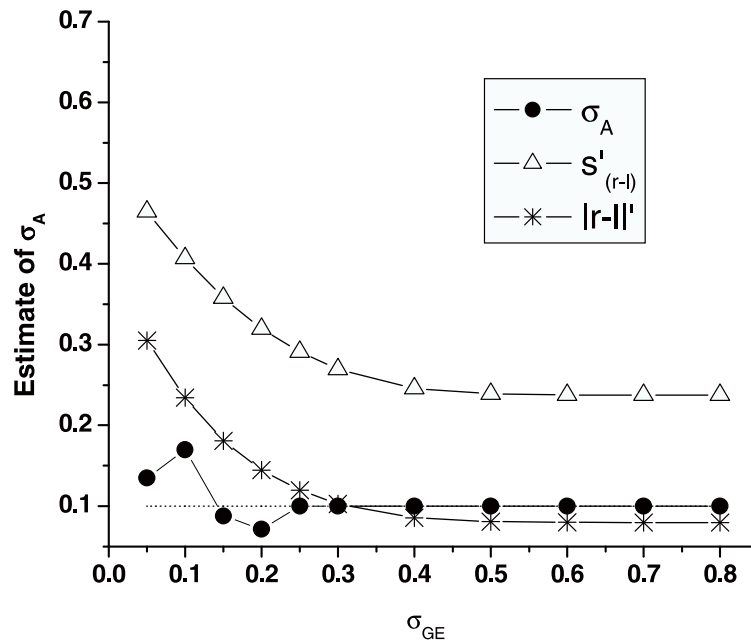
Figure 3.   Asymmetry index values as a function of σ_{GE}. Values estimated from 10,000 pairs of right and left counts distributed exactly in accordance with $\mu = 20.5$, $\sigma_A = 0.10$, and σ_{GE}. (Diacritical marks on $\hat{\sigma}_A$ and $\overline{|r - l|}'$ are omitted in the legend.)

from the true σ_A could be due as much to the numerical algorithm used to estimate $\hat{\sigma}_A$ as to an inherent bias in the estimator. In contrast, $s'_{(r-l)}$ and $\overline{|r - l|}'$ both increased as μ departed further from an integer value. $\overline{|r - l|}'$ was negatively biased when the shift from an integer value of μ was ≤ 0.33 and positively biased at higher values. $s'_{(r-l)}$ always overestimated σ_A, and the degree of bias increased as μ departed further from an integer value.

The maximum severity of the bias was examined by setting μ at 20.5 and varying σ_{GE} with $\sigma_A = 0.1$ (Figure 3). $\hat{\sigma}_A$ was able to estimate σ_A accurately when σ_{GE} was greater than 0.25, but below 0.25 fluctuated above and below 0.1, suggesting problems with the NLIN algorithm rather than inherent bias. It may be possible to adjust the algorithm to obtain accurate σ_A estimates in this region. $s'_{(r-l)}$ and $\overline{|r - l|}'$ both were substantially positively biased at low values of σ_{GE} but trended toward the true σ_A as σ_{GE} increased. $\overline{|r - l|}'$ became negatively biased as σ_{GE} exceeded 0.3.

Poisson distributed counting errors adversely affected $\hat{\sigma}_A$, even for λ as low as 0.05. When no replicate counts were available, the bias to $\hat{\sigma}_A$ increased linearly with increasing λ (Figure 4), but was slightly more severe for the population with true $\sigma_A = 0.25$ so that the estimated FA for the two populations appeared to converge somewhat as λ increased. However, when replicate counts were available for 25% of the individuals, the estimation process (Equations (2.12) and (2.13)) was able to remove the bias introduced by the counting errors so that $\hat{\sigma}_A$ remained at the true parameter value.

Counting error also affected the power of the likelihood ratio test to detect differences
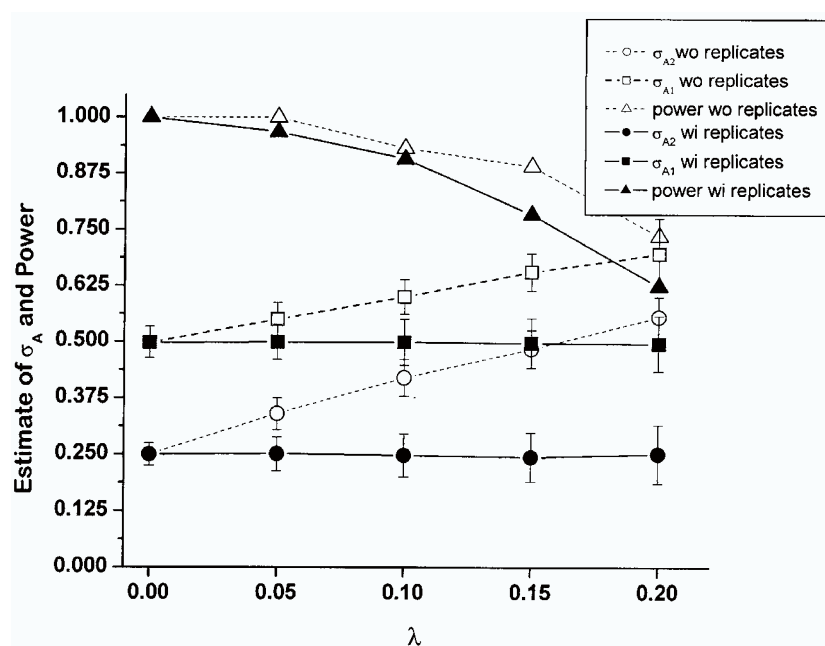
Figure 4. Mean estimated $\hat{\sigma}_A$ and power of likelihood ratio test for two populations with $\mu = 20.25$, $\sigma_{GE} = 0.5$, $\sigma_{A1} = 0.5$, and $\sigma_{A2} = 0.25$ based on 250 simulations. Absolute counting error is Poisson distributed with $0 \leq \lambda \leq 0.2$. Sample size for each population is 200, with either 0 or 25% replicate counts.

in the two populations (Figure 4). With no counting errors, the test detected the difference between two populations with $\sigma_{A1} = 0.5$ and $\sigma_{A2} = 0.25$ in 100% of the trials. As counting error increased, the fraction of significant differences declined, both with and without error correction. The decline in power was slightly more severe when the replicate counts were available to remove bias.

Power curves based on 500 simulated two-sample tests (without counting errors) at each value of σ_{A2} indicated the F-test was slightly more powerful than the likelihood ratio test (Figure 5), but the likelihood ratio test was the closest to the nominal Type I error rate (0.05) when the null hypothesis ($\sigma_{A1} = \sigma_{A2}$) was true. Type I error for the F-test was approximately twice the nominal level. Neither test was very powerful for a sample size of 50. Even with sample size of 100, 80% power was only achieved when the difference in σ_A was approximately 0.2 or greater.

Application of the maximum likelihood estimation method to the *Drosophila* data resulted in relatively small changes in the estimates of FA. $s'_{(r-l)}$, calculated by adjusting the values from Table 1 of Kristensen et al. (2003), was just slightly larger than $\hat{\sigma}_A$ except for a larger difference for the 37.5°C treatment. Similar to the original result, none of the temperature treatments had significantly different FA than the control group based on the likelihood ratio test. The Kristensen et al. (2003) estimates of σ_E^2 were nearly identical to $(\hat{\sigma}_{GE})^2$ for most treatments, except when σ_E^2 was negative $(\hat{\sigma}_{GE})^2$ approached zero and
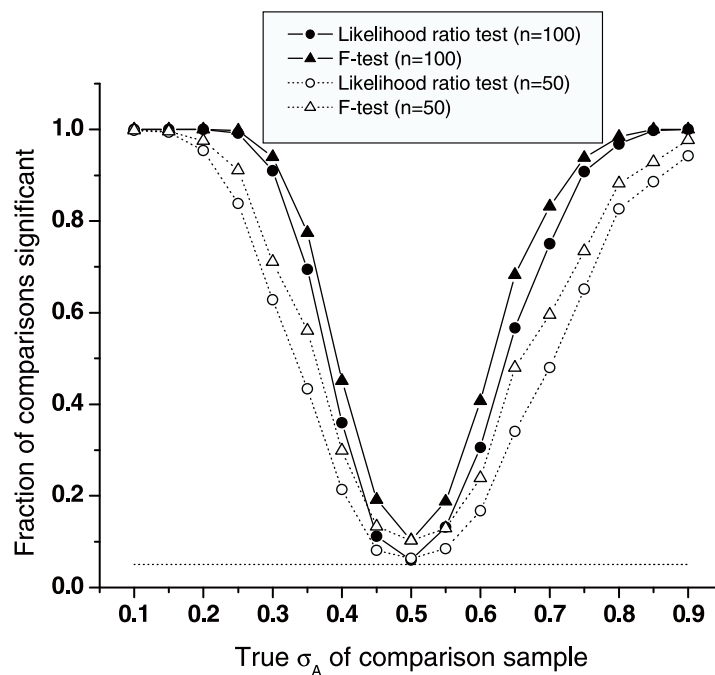
Figure 5. Estimated power of two-sample tests for differences in fluctuating asymmetry. Paired samples of size 50 or 100 were drawn randomly from model distributions with $\mu = 20.5$, $\sigma_{GE} = 0.5$, $\sigma_{A1} = 0.5$, $0.1 \leq \sigma_{A2} \leq 0.9$.

the NLIN algorithm failed to converge. Overall, there is little difference from the original results, but uncertainty due to possible bias by μ has been removed.

As an illustration of the adjustment for counting errors, the unadjusted data from observers 2 and 3 of Hubert and Alexander (1995) suggest that pectoral fins may be more asymmetrical in cutthroat trout than in rainbow trout. $\overline{|r - l|}'$ for the combined data was 0.19 for rainbow trout, and 0.50 for cutthroat (Table 1). The Poisson and geometric error models fit the distributions of count differences very well (Figure 6), with the Poisson model slightly better for rainbow trout and the geometric model slightly better for cutthroat trout (Table 1). The error-corrected estimates of $\hat{\sigma}_A$ were 0.18 for rainbow trout and 0.20 for cutthroat trout, with standard errors larger than the parameter estimates (Table 1). The adjusted estimates of $\hat{\sigma}_A$ thus refute the apparent difference in FA of the two species that was suggested by the usual FA metrics applied to the raw counts.

4. DISCUSSION

Since Swain (1987) originally described the bias of $(r - l)$-based metrics for meristic data, there have been advances in the theory and statistical analysis of fluctuating asymmetry (e.g., Björkland and Merilä 1997; Van Dongen 1999, 2000; Van Dongen et al. 1999; Houle 2000; Polack and Starmer 2001), but mostly applicable to morphometric data. A recent summary of the statistical analysis of FA data (Palmer and Strobeck 2003) provided no

Table 1.   Analysis of FA of pectoral fin asymmetry in rainbow and cutthroat trout using combined data for counters 2 and 3. Data from Hubert and Alexander (1995).

		Rainbow trout	Cutthroat trout
	$\overline{\lvert r-l \rvert}'$	0.19	0.50
	$s'_{(r-l)}$	0.37	1.00
LL_1	(geometric)	24.38	61.24
	$\hat{q}$	0.94	0.77
LL_1	(Poisson)	24.16	62.31
	$\hat{\lambda}$	0.07	0.30
LL_2		68.82	123.10
	$\hat{\mu}$	14.09	15.29
	$\hat{\sigma}_{GE}$	0.35	0.48
	$\hat{\sigma}_A$	0.18	0.20
	$se(\hat{\sigma}_A)$	0.20	0.33

means to remove the potential bias, nor any way to adjust for counting errors.

The proposed estimator $\hat{\sigma}_A$ has two advantages over indices based on $(r-l)$ for meristic characters. The primary advantage of $\hat{\sigma}_A$ is that it is unbiased by shifts in the mean liability away from integer values, except when the total amount of character variation is extremely low. Even then, it is much less biased than the traditional $(r-l)$ metrics. Also, to the extent that the prevailing normal model for the distribution of the liability is correct, $\hat{\sigma}_A$ changes linearly with the true parameter, whereas metrics derived from $(r-l)$ do not.

The consistent overestimate of σ_A by $s'_{(r-l)}$ and variable bias of $\overline{\lvert r-l \rvert}'$ result from the variance and covariances of the rounding variable R_{ij}, which change with the magnitude of μ, σ_{GE}, and σ_A. These variances are inherent to estimators based on $(r-l)$, but are not a component of $\hat{\sigma}_A$. This additional variation in $(r-l)$ is perhaps also the reason for inflated Type I error of the F-test in the simulations.

Another important advantage of $\hat{\sigma}_A$ is that a method to correct for the bias introduced by counting errors is available if replicate counts are made. A small amount of counting error may be enough to seriously bias any FA metric that cannot be corrected for counting errors. The error-correction method has been demonstrated with two possible statistical models for counting errors, but additional models could also be used if the geometric or Poisson models prove unsuitable when tested against a broader selection of data. Finally, use of these methods would provide more accurate statistical tests for differences in asymmetry among samples. The likelihood ratio test produces nearly the nominal Type I error rate when two samples are drawn from the same population, and also has relatively high power to detect samples that actually have different asymmetry levels. The F-test based on the variance of $(r-l)$ has slightly higher power, but also an inflated Type I error rate.

Another advantage of $\hat{\sigma}_A$ is that it provides the opportunity to make FA more directly comparable across meristic and morphometric characters, especially if the $(r-l)$ metrics

used for morphometric characters are adjusted by $1/\sqrt{2}$ so that they also would estimate σ_A.

Use of $\hat{\sigma}_A$ as an index of fluctuating asymmetry, with adjustment for counting error, should improve our ability to estimate the asymmetry of real-world populations from meristic characters. The partitioning of the character variation into components due to inter-individual genetic and/or environmental variation (σ_{GE}), counting errors, and that due to fluctuating asymmetry (σ_A) should allow more accurate estimates of fluctuating asymmetry, more powerful tests for differences, and ultimately a better understanding of fluctuating asymmetry and its causes.
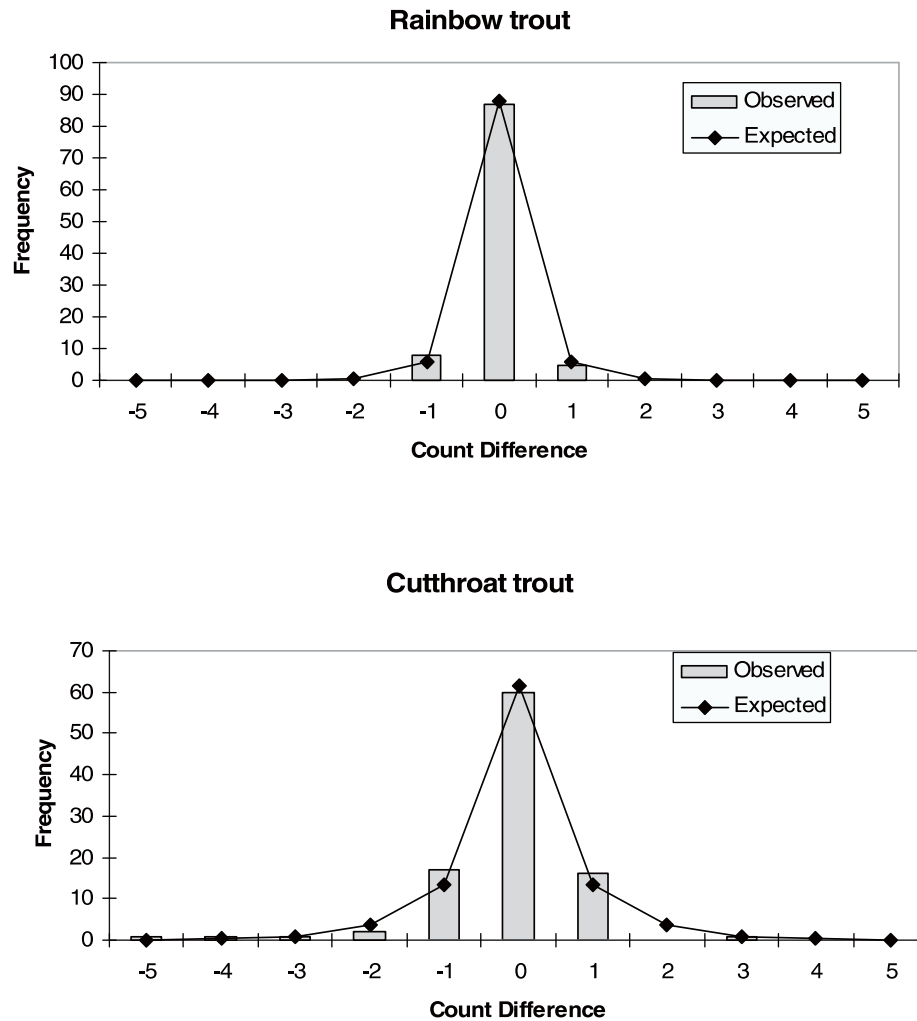

Figure 6.   Observed and expected distribution of count differences for rainbow and cutthroat trout pectoral fin rays. Expected distribution for rainbow trout is Poisson, $\hat{\lambda} = 0.07$. Expected distribution for cutthroat trout is geometric, $\hat{q} = 0.77$. Data from Hubert and Alexander (1995).

ACKNOWLEDGMENTS

REFERENCES

Björkland, M., and Merilä, J. (1997), "Why Some Measures of Fluctuating Asymmetry are so Sensitive to Measurement Error," *Annales Zoologica Fennici*, 34, 133–137.

Falconer, D. S. (1981), *Introduction to Quantitative Genetics* (2nd ed.), London: Longman.

Hilborn, R., and Mangel, M. (1997), *The Ecological Detective*, Princeton, NJ: Princeton University Press.

Houle, D. (2000), "A Simple Model of the Relationship Between Asymmetry and Developmental Stability," *Journal of Evolutionary Biology*, 13, 720–730.

Hubert, W. A., and Alexander, C. B. (1995), "Observer Variation in Counts of Meristic Traits Affects Fluctuating Asymmetry," *North American Journal of Fisheries Management*, 15, 156–158.

Kristensen, T. N., Pertoldi, C., Andersen, D. H., and Loeschcke, V. (2003), "The Use of Fluctuating Asymmetry and Phenotypic Variability as Indicators of Developmental Instability: A Test of a New Method Employing Clonal Organisms and High Temperature Stress," *Evolutionary Ecology Research*, 5, 53–68.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics* (3rd ed.), New York: McGraw-Hill.

Palmer, A. R. (1994), "Fluctuating Asymmetry Analyses: A Primer," in *Developmental Instability: Its Origins and Evolutionary Implications*, ed. T. A. Markow, Dordrecht, Netherlands: Kluwer, pp. 335–364.

Palmer, A. R., and Strobeck, C. (1992), "Fluctuating Asymmetry as a Measure of Developmental Stability: Implications of Non-normal Distributions and Power of Statistical Tests," *Acta Zoologica Fennica*, 191, 55–70.

——— (2003), "Fluctuating Asymmetry Analyses Revisited," in *Developmental Instability (DI): Causes and Consequences*, ed. M. Polak, New York: Oxford University Press, pp. 279–319.

Polak, M., and Starmer, W. T. (2001), "The Quantitative Genetics of Fluctuating Asymmetry," *Evolution*, 55, 498–511.

Swain, M. W. (1987), "A Problem With the Use of Meristic Characters to Estimate Developmental Stability," *The American Naturalist*, 129, 761–768.

Van Dongen, S. (1999), "Accuracy and Power in the Statistical Analysis of Fluctuating Asymmetry: Effects of Between-Individual Heterogeneity in Developmental Instability," *Annales Zoologica Fennici*, 36, 45–52.

——— (2000), "Unbiased Estimation of Individual Asymmetry," *Journal of Evolutionary Biology*, 13, 107–112.

Van Dongen, S., Molenberghs, G., and Matthysen, E. (1999), "The Statistical Analysis of Fluctuating Asymmetry: REML Estimation of a Mixed Regression Modelm" *Journal of Evolutionary Biology*, 12, 94–102.

Van Valen, L. (1962), "A Study of Fluctuating Asymmetry," *Evolution*, 16, 125–142.

Weiner, J. G., and Rago, P. J. (1987), "A Test of Fluctuating Asymmetry in Bluegills (*Lepomis macrochirus Rafinesque*) as a Measure of pH-Related Stress," *Environmental Pollution*, 44, 27–36.