 Cytometry 39:300–305 (2000)

# Mean and Variance of Ratio Estimators Used in Fluorescence Ratio Imaging

**G.M.P. van Kempen[1]* and L.J. van Vliet[2]**

[1]Central Analytical Sciences, Unilever Research Vlaardingen, Vlaardingen, The Netherlands
[2]Pattern Recognition Group, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

Received 7 April 1999; Revision Received 14 October 1999; Accepted 26 October 1999

**Background:** The ratio of two measured fluorescence signals (called $x$ and $y$) is used in different applications in fluorescence microscopy. Multiple instances of both signals can be combined in different ways to construct different ratio estimators.
**Methods:** The mean and variance of three estimators for the ratio between two random variables, $x$ and $y$, are discussed. Given $n$ samples of $x$ and $y$, we can intuitively construct two different estimators: the mean of the ratio of each $x$ and $y$ and the ratio between the mean of $x$ and the mean of $y$. The former is biased and the latter is only asymptotically unbiased. Using the statistical characteristics of this estimator, a third, unbiased estimator can be constructed.
**Results:** We tested the three estimators on simulated data, real-world fluorescence test images, and comparative genome hybridization (CGH) data. The results on the simulated and real-world test images confirm the presented theory. The CGH experiments show that our new estimator performs better than the existing estimators.
**Conclusions:** We have derived an unbiased ratio estimator that outperforms intuitive ratio estimators. Cytometry 39:300–305, 2000. © 2000 Wiley-Liss, Inc.

**Key terms:** ratio imaging; CGH; ratio estimation; mean; variance

In fluorescence microscopy, ratio imaging is applied to a number of applications. In ratio labeling, the ratio between the intensities of different fluorophores is used to expand the number of labels for an in situ hybridization procedure (1). The number of fluorophores that can be spectrally separated by fluorescence microscopy normally restricts the total number of labels.

Fluorescence ratio imaging is also used to measure spatial and temporal differences in ion concentrations within a single cell. This is achieved by using fluorophores whose excitation or emission spectrum changes as a function of the $Ca^{2+}$ or pH concentration (2).

In a third application of ratio imaging, known as comparative genome hybridization (CGH) (3,4), one tries to estimate the DNA sequence copy number as a function of the chromosomal location. This is achieved by measuring the hybridization between "tumor" DNA and "normal" DNA to detect gene amplifications and deletions (5).

In fluorescence ratio imaging, one is interested in the ratio $R$ between two random variables $X$ and $Y$,

$$R = \frac{X}{Y}. \tag{1}$$

In practice, one cannot measure $X$ and $Y$, but only "noisy" realizations of $X$ and $Y$, termed $x$ and $y$ in the present article. We will assume $x$ and $y$ to be stochastic variables with $E\{x\} = \mu_x = X$ and $E\{y\} = \mu_y = Y$. In the second section we study the statistical properties of two estimators of the ratio $R$. Based on these results we show in the third section that, under certain conditions (known variance, covariance), a third, unbiased estimator can be constructed. The simulations and experiments in the fourth section support the presented theory. In the fifth section, we discuss the application of the presented theory on CGH ratio imaging. We conclude in the sixth section.

## MEAN AND VARIANCE OF TWO RATIO ESTIMATORS

Given $n$ samples of $x$ and $y$, we can obtain two obvious estimators for the ratio $R = X/Y$:

$$r_1 = \left(\frac{\bar{x}}{\bar{y}}\right) \qquad r_2 = \frac{\bar{x}}{\bar{y}} \tag{2}$$

This research was presented at ASCI'95, First Annual Conference of the Advanced School for Computing and Imaging, Heijen, The Netherlands, 16–18 May 1995.

The research was performed while G.M.P. van Kempen was a member of the Pattern Recognition Group.

*Correspondence to: Dr. Ir. Geert M.P. van Kempen, Unilever Research Vlaardingen, Olivier van Noortlaan 120, 3133 AT Vlaardingen, P.O. Box 114, 3130 AC Vlaardingen, The Netherlands.

E-mail: geert-van.kempen@unilever.com.

with $\bar{x}$ the average of $x$. To find an approximate expression for the expectation of $r_1$ and $r_2$, we have used a Taylor series expansion of $x/y$ around $\mu_x$, $\mu_y$:

$$\begin{aligned}\frac{x}{y} \approx \frac{x}{y}\Big|_{\mu_x,\mu_y} &+ (x-\mu_x)\frac{\partial}{\partial x}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} + (y-\mu_y)\frac{\partial}{\partial y}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} \\ &+ \frac{1}{2}(x-\mu_x)^2\frac{\partial^2}{\partial x^2}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} \\ &+ \frac{1}{2}(y-\mu_y)^2\frac{\partial^2}{\partial y^2}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} \\ &+ (x-\mu_x)(y-\mu_y)\frac{\partial^2}{\partial x\partial y}\left(\frac{x}{y}\right)\Big|_{\mu_x,\mu_y} \\ &+ O\left(\left((x-\mu_x)\frac{\partial}{\partial x} + (y-\mu_y)\frac{\partial}{\partial y}\right)^3\left(\frac{x}{y}\right)\right). \end{aligned} \tag{3}$$

A Taylor series expansion of $\bar{x}/\bar{y}$ around $\mu_x$, $\mu_y$ is similar to equation (3). The mean of $r_1$ and $r_2$ can be found by applying the expectation operator to the individual terms (ignoring all terms higher than two),

$$E\{r_1\} = E\left\{\left(\frac{\bar{x}}{\bar{y}}\right)\right\} = E\left\{\frac{x}{y}\right\} \approx \frac{\mu_x}{\mu_y} + \operatorname{var}(y)\frac{\mu_x}{\mu_y^3} - \frac{\operatorname{cov}(x,y)}{\mu_y^2} \tag{4}$$

and

$$\begin{aligned}E\{r_2\} = E\left\{\frac{\bar{x}}{\bar{y}}\right\} = \frac{\mu_x}{\mu_y} + \operatorname{var}(\bar{y})\frac{\mu_x}{\mu_y^3} - \frac{\operatorname{cov}(\bar{x},\bar{y})}{\mu_y^2} &\approx \frac{\mu_x}{\mu_y} \\ + \frac{1}{n}\left(\operatorname{var}(y)\frac{\mu_x}{\mu_y^3} - \frac{\operatorname{cov}(x,y)}{\mu_y^2}\right)&\end{aligned} \tag{5}$$

with $E\{r\}$ the expectation of $r$, $\operatorname{var}(y)$ the variance of $y$, and $\operatorname{cov}(x, y)$ the covariance of $x$ and $y$. It is clear from these two expressions that $r_2$ is asymptotically unbiased ($\lim n \to \infty\ E\{r_2\} = \mu_x/\mu_y$) and that $r_1$ is a biased estimator of $R$. An approximation of the variance of $r_1$ and $r_2$ is obtained by using the first-order terms of the Taylor series expansion:

$$\begin{aligned}\operatorname{var}(r_1) = \operatorname{var}\left(\left(\frac{\bar{x}}{\bar{y}}\right)\right) &= E\left\{\left(\left(\frac{\bar{x}}{\bar{y}}\right) - E\left\{\left(\frac{\bar{x}}{\bar{y}}\right)\right\}\right)^2\right\} \\ &\approx E\left\{\left(\left(\frac{\bar{x}}{\bar{y}}\right) - \frac{\mu_x}{\mu_y}\right)^2\right\} \\ &\approx \frac{1}{n}\left(\frac{\operatorname{var}(x)}{\mu_y^2} + \frac{\mu_x^2\operatorname{var}(y)}{\mu_y^4} - \frac{2\mu_x\operatorname{cov}(x,y)}{\mu_y^3}\right)\end{aligned} \tag{6}$$

and

$$\begin{aligned}\operatorname{var}(r_2) = \operatorname{var}\left(\frac{\bar{x}}{\bar{y}}\right) &= E\left\{\left(\frac{\bar{x}}{\bar{y}} - E\left\{\frac{\bar{x}}{\bar{y}}\right\}\right)^2\right\} \\ &\approx E\left\{\left(\frac{\bar{x}}{\bar{y}} - \frac{\mu_x}{\mu_y}\right)^2\right\} \\ &\approx \frac{\operatorname{var}(\bar{x})}{\mu_y^2} + \frac{\mu_x^2\operatorname{var}(\bar{y})}{\mu_y^4} - \frac{2\mu_x\operatorname{cov}(\bar{x},\bar{y})}{\mu_y^3} \\ &\approx \frac{1}{n}\left(\frac{\operatorname{var}(x)}{\mu_y^2} + \frac{\mu_x^2\operatorname{var}(y)}{\mu_y^4} - \frac{2\mu_x\operatorname{cov}(x,y)}{\mu_y^3}\right).\end{aligned} \tag{7}$$

Estimators $r_1$ and $r_2$ have, only in first-order Taylor series expansion, an equal variance, which diminishes for an infinite number of samples ($n \to \infty$). Therefore both estimators $r_1$ and $r_2$ are consistent. An expression similar to equations (6) and (7) for the variance of $x/y$ was found by Kendall and Stuart (6).

## AN UNBIASED RATIO ESTIMATOR

In the previous section, we found that the estimator $r_2$ is only asymptotically unbiased. However, having found an analytical expression of the bias, we can derive an unbiased estimator for $R$ based on $r_2$:

$$r_3 = \frac{\bar{x}}{\bar{y}} - \frac{1}{n}\left(\frac{\mu_x}{\mu_y^3}\operatorname{var}(y) - \frac{\operatorname{cov}(x,y)}{\mu_y^2}\right). \tag{8}$$

The expectation for this estimator is, of course,

$$E\{r_3\} = E\left\{\frac{\bar{x}}{\bar{y}} - \frac{1}{n}\left(\frac{\mu_x}{\mu_y^3}\operatorname{var}(y) - \frac{\operatorname{cov}(x,y)}{\mu_y^2}\right)\right\} \approx \frac{\mu_x}{\mu_y}. \tag{9}$$

The variance of this estimator yields

$$\operatorname{var}(r_3) = E\left\{\left(\frac{\bar{x}}{\bar{y}} - E\left\{\frac{\bar{x}}{\bar{y}}\right\}\right)^2\right\} = \operatorname{var}(r_2). \tag{10}$$

## EXPERIMENTS

Some simulations were performed to support the expressions for the mean and variance of $r_1$, $r_2$, and $r_3$. In the first experiment we used computer-generated noise; in the second experiment we used noise images generated in fluorescence image acquisition.

### Simulations

An image of size $n \times m$ was filled with $m$ realizations of $x$, $N(\mu_x, \sigma_x)$, each containing $n$ samples (numerical recipes) (7) and another image was filled with realizations of $y$, $N(\mu_y, \sigma_y)$. Figure 1 shows the mean and variance of $r_1$, $r_2$, and $r_3$ as a function of the number of samples $n$. Note that $r_2$ asymptotically converges to $\mu_x/\mu_y$, $r_1$ remains biased even for large values of $n$, whereas $r_3$ is unbiased even for small number of samples. The means of $r_2$ and $r_3$ are in agreement with the predictions, whereas the mean of $r_1$ converges to a slightly higher ratio than predicted when using a second-order Taylor approximation, as in equation (5). The estimated variances of $r_1$ and $r_3$ are in agreement with the theoretical variances as

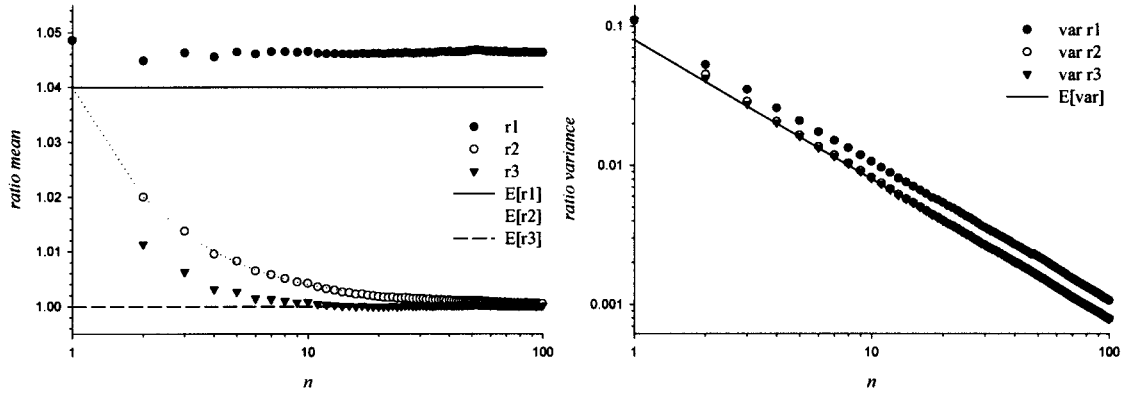


FIG. 1. The mean and variance of $r_1$, $r_2$, and $r_3$ measured as a function of $n$, with $\mu_x = \mu_y = 40.0$ and $\sigma_x = \sigma_y = 8.0$. Each point is an average of 10,000 realizations.

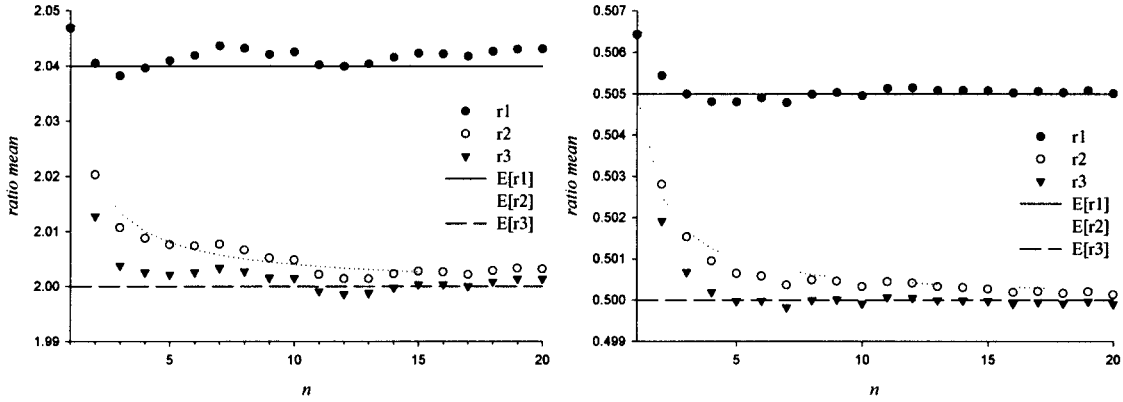


FIG. 2. The mean of $r_1$, $r_2$, and $r_3$ measured as a function of $n$ using Poisson noise, with $\mu_x = 200$ and $\mu_y = 100.0$ on the left and vice versa on the right. In all cases, $\sigma_x = \sqrt{\mu_x}$ and $\sigma_y = \sqrt{\mu_y}$. Each point is an average of 10,000 realizations.

derived in the second section. The estimated variance of $r_1$ is larger than we have expected from our calculations when using only a first-order Taylor approximation (indicated by the continuous line in the right graph).

We performed a second simulation experiment with Poisson noise to test ratios unequal to one. Figure 2 shows that for a ratio of 0.5 and 2.0 the results are in good agreement with the expected ratio values.

We tested the three ratio estimators on signals having a covariance. We simulated the covariance by generating a Poisson noise image with a constant mean and used this image as the mean for the Poisson processes that generated the $X$ and $Y$ image. Setting the mean to 100.0, we obtained images with a variance of 200.0 and a covariance of 100.0. The results of this experiment are shown in Figure 3. The results show a good correspondence with the expected ratio values.

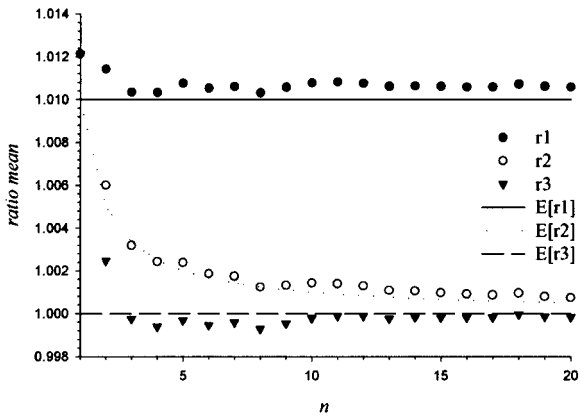


FIG. 3. The mean of $r_1$, $r_2$, and $r_3$ measured as a function of $n$ using Poisson noise, with $\mu_x = \mu_y = 100.0$. The variance is equal to 200.0, and the covariance is 100.0. Each point is an average of 10,000 realizations.

## Experiments on Real Data

To test the presented theory in practice, we performed the following experiment. We used a Nikon inverted microscope equipped with a Photometrics camera (with a KAF 1400 CCD chip; Photometrics, Tucson, AZ) to acquire a fluorescence image of a homogenous fluorescence sample. The acquired data are distorted by several noise sources (8):

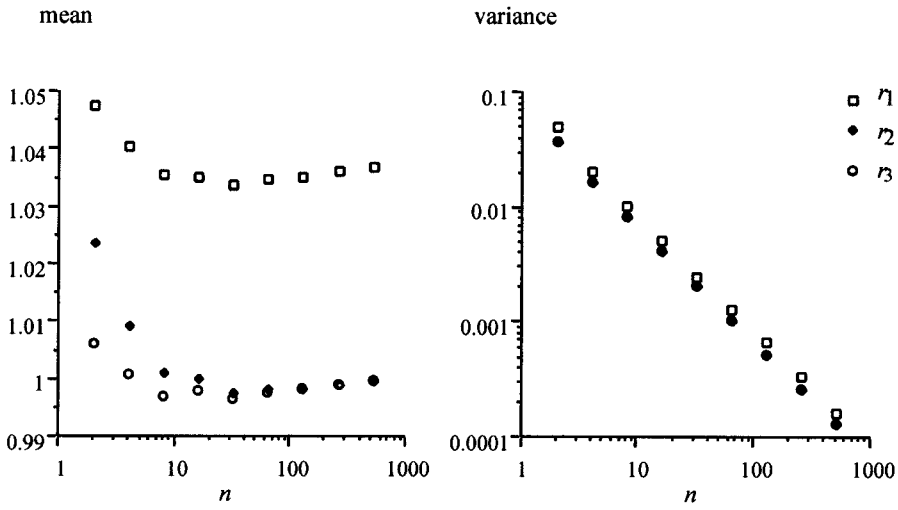


FIG. 4. The mean and variance of $r_1$, $r_2$, and $r_3$ measured as a function of $n$. The mean value of the two noise images was set to 100.0. The variances of the two images are 328.84 and 329.37.

- photon shot noise (Poisson noise due to the counting of electrons induced by single photons)
- read-out noise (Gaussian noise caused by the preamplifier)
- quantitization noise (uniform noise caused by analog-to-digital conversion)
- dark current (very small, Poisson distributed signal)
- bias, space-dependent offset signal.

Such a noise image will not have a constant mean number of photons over the image due to nonhomogeneity of the illumination. If we acquire a second image of the same scene and subtract the two, we are left with a noise image of zero mean and twice the variance. This can be repeated to produce a second independent noise image. Both noise images can be added to a constant valued image, after which the ratio estimators $r_1$, $r_2$, and $r_3$ can be applied. Noise images acquired this way will contain photon shot noise, read-out noise, and quantization noise. Figure 4 shows the mean and variance of the ratio of the two noise images, as estimated by the three estimators $r_1$, $r_2$, and $r_3$ as a function of the number of samples $n$.

## APPLICATION TO CGH

CGH (3–5) estimates the DNA sequence copy number as function of the chromosomal location. Ratio imaging of "tumor" DNA and "normal" DNA allows detection of gene amplifications and deletions.

To improve the precision and accuracy of such a ratio profile, it is common practice in CGH analysis to average over a number of ratio profiles of the same chromosome (i.e., $r_2$ is used to calculate one profile, and $r_1$ to average these profiles). If we call this ratio estimator $r_{\mathrm{mn}}$, we can write it as,

$$r_{mn} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \Big/ \frac{1}{n}\sum_{i=1}^{n} y_i\right\} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{\bar{x}}{\bar{y}}\right\} \qquad (11)$$

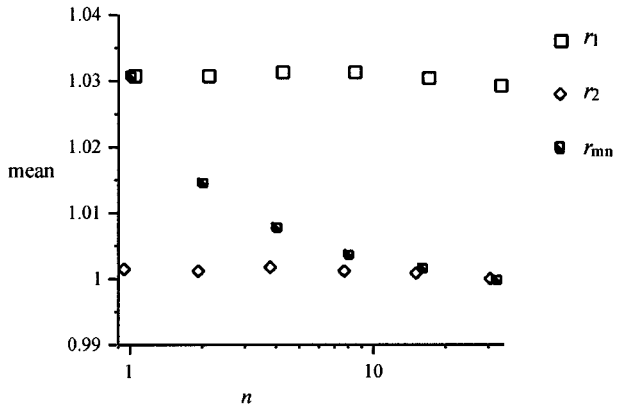


FIG. 5. The mean of $r_{\mathrm{mn}}$, $r_1$, and $r_2$ measured as function of $n$, with a constant number of samples divided over $m$ and $n$ in such a way that $m \times n = 32$, $\mu_x = \mu_y = 36.0$, and $\sigma_x = \sigma_y = 6.0$.

where $m$ is the number of profiles and $n$ is the number of $x$, $y$ values at a certain position on each profile. Equation (11) shows that estimator $r_{\mathrm{nm}}$ is identical to applying estimator $r_2$ $m$ times on $n$ samples followed by averaging over the obtained $m$ ratios. An approximation of the expectation and variance of $r_{\mathrm{mn}}$ can be found with a Taylor series expansion:

$$E\{r_{\mathrm{mn}}\} = E\{r_1\} \qquad \mathrm{var}(r_{\mathrm{mn}}) = \frac{1}{m}\,\mathrm{var}(r_2). \qquad (12)$$

The expressions for the mean and variance of $r_{\mathrm{mn}}$ show that averaging over ratio profiles only improves the precision of the estimation of $R$. It does not improve the accuracy.

A simulation experiment was performed to compare the performance of estimator $r_{\mathrm{mn}}$ with $r_1$ and $r_2$ for a

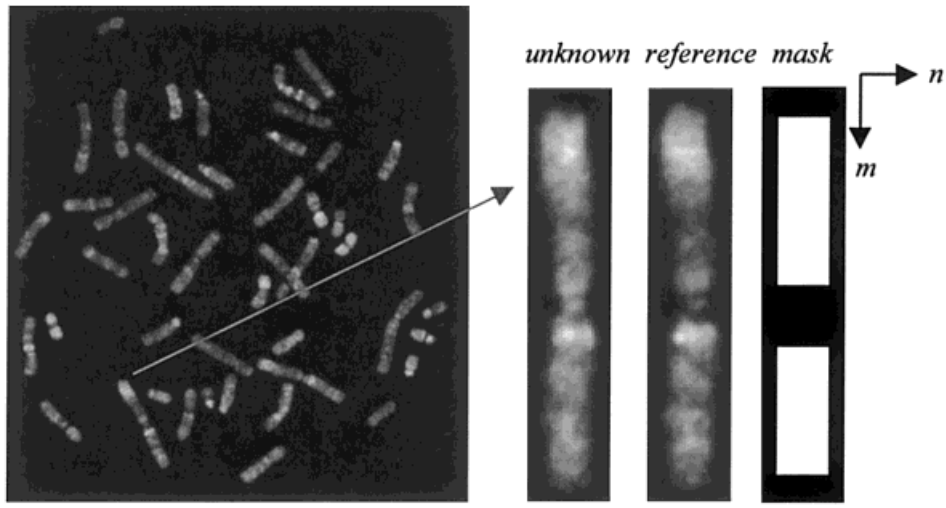


FIG. 6. Image of the metaphase (left) from which a single chromosome has been selected. The right three images show the "normal" or "reference" chromosome. The "tumor" or "unknown" chromosome, and the mask used to calculate the ratio.

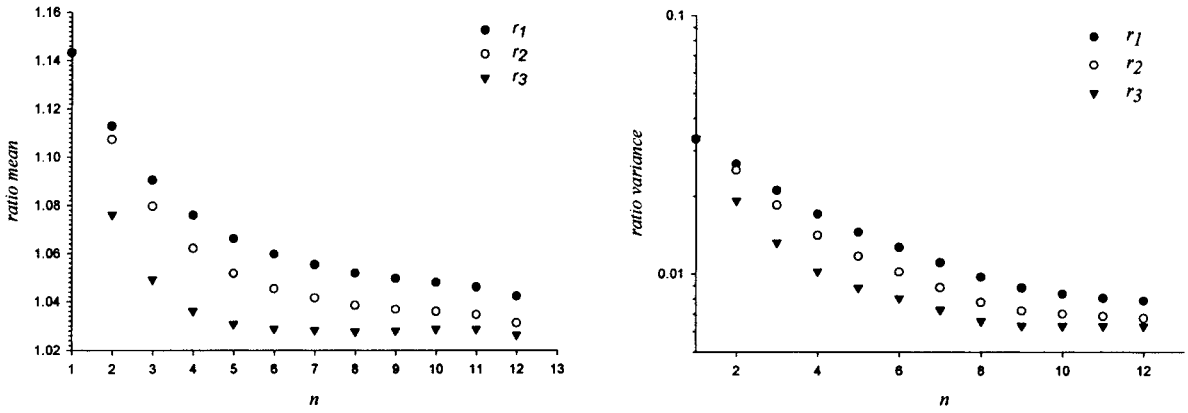


FIG. 7. The mean (left) and variance (right) of $r_1$, $r_2$, and $r_3$ (with and without estimating the covariance) measured as function of $n$.

constant number of $m \times n$ samples. Two images were filled with Gaussian distributed noise, and the ratios between both images were estimated using $r_{mn}$, $r_1$, and $r_2$. The mean and variance of $r_{mn}$ were calculated as function of the ratio between $n$ and $m$; the mean and variance of $r_1$ and $r_2$ were calculated from $m \times n$ samples.

Figure 5 shows the results of an experiment with $r_{mn}$, $r_1$, and $r_2$ where $m \times n$ equals 32. Note that estimator $r_1$ has already converged for this number of samples. It is clear that estimator $r_{mn}$ yields a suboptimal estimation of the ratio for a given number of $m \times n$ samples. This shows that averaging before taking the ratio yields a better ratio estimate.

In a second CGH experiment we used CGH control images to verify whether the differences between estimators $r_1$, $r_2$, and $r_3$ could be found on real data. The CGH control images are made with normal DNA for both test and reference DNA; thus, after calibration a ratio of 1.0 is to be expected. From a set of CGH images, we selected a straight chromosome (Fig. 6) to avoid errors due to the straightening of the chromosome pixel data. We used a normalized Gaussian convolution to estimate the background intensity, which was subtracted from the chromosome intensity values (9). To correct for the different chromatic efficiencies of the color filters and camera used, the intensities of the two CGH signals of the chromosome were scaled such that the average intensities were equal.

We calculated the mean and variance of $r_1$, $r_2$, and $r_3$ from a region of interest (mask in Fig. 6), with size $n$, in the middle of both chromosomes that excluded the centromeric and telomeric regions of the chromosomes. Figure 7 shows the mean and variance values of the three estimators as a function of the number of samples $n$, taken in the horizontal direction of the image. The mean and variance were estimated over 95 realizations (the vertical dimension of the mask is 95 pixels). The experiment shows that the estimator $r_3$ outperforms the other two estimators. All three estimators, however, do decrease their bias as function of the number of samples, which we contribute to the fact, that for this application, our model

for the $X$ and $Y$ signals (constant signals with noise) is quite a crude model for the chromosomes. Nevertheless, the results show that the estimator based on this simple approximation (estimator $r_3$) does improve the estimation of the ratio.

## CONCLUSIONS

In this article we have derived expressions for the mean and variance of two intuitive ways for estimating the ratio between two random variables. We have shown that one of these estimators ($r_1$) is biased and that the other ($r_2$) is only asymptotically unbiased. However, from the derived expression for the bias of $r_2$, we propose an unbiased ratio estimator $r_3$. These results were supported by experiments. The noise data for these experiments were either computer generated or measured camera noise.

For a particular application of ratio imaging, CGH, we have shown that the usual averaging over multiple ratio profiles is not optimal. Measurements on CGH images confirm the differences between the ratio estimators, as found in the other presented experiments.

It should be noted that, when the population of realizations of $X$ and $Y$ is heterogeneous, the estimators $r_2$ and $r_3$ will mask such heterogeneity, and a histogram of primary $x/y$ ratios will be necessary to reveal the heterogeneity. The example shown in Figure 6, therefore, only makes sense because the ratio is known a priori to have an expectation value of 1.0 for all positions. For a true biological sample, averaging along the chromosome would not be a sensible procedure.

## ACKNOWLEDGMENTS

We thank Jim Piper, Joe Gray, and Damir Sudar for all fruitful discussions and for providing us with a CGH dataset. This work was supported in part by the Royal Netherlands Academy of Arts and Sciences (KNAW) and by the Rolling Grants program of the Foundation for Fundamental Research in Matter (FOM). We thank the reviewers and the associate editor for their constructive comments to improve the technical note.

## LITERATURE CITED

1. Nederlof PM. Methods for quantitative and multiple in situ hybridization. Leiden: Leiden University; 1991.
2. Shotton D. Electronic light microscopy. New York: Wiley-Liss; 1993.
3. Kallioniemi A, Kallioniemi O-P, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genome hybridization for molecular cytogenetic analysis of solid tumors. Science 1982;258:818–821.
4. Piper J, Rutovitz D, Sudar D, Kallioniemi A, Kallioniemi OP, Waldman FM, Gray JW, Pinkel D. Computer image analysis of comparative genomic hybridization. Cytometry 1995;19:10–26.
5. Moore DH, Pallavicini M, Cher ML, Gray JW. A t-statistic for objective interpretation of comparative genomic hybridization (cgh) profiles. Cytometry 1997;28:183–190.
6. Kendall M, Stuart A. Advanced theory of statistics. London: Charles Griffin & Company; 1977.
7. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C, 2nd ed. Cambridge: Cambridge University Press; 1992.
8. van Vliet LJ, Sudar D, Young IT. Digital fluorescence imaging using cooled charge-coupled device array cameras. In: Celis JE, editor. Cell biology: a laboratory handbook, 2nd ed, volume 3. London: Academic Press, 1998. p 109–120.
9. Ligtenberg KAW. Analysis of CGH images [master's thesis]. Delft (The Netherlands): Pattern Recognition Group, Delft University of Technology; 1996.