Unbiased Estimator for Fourth Moments

Synge Todo
Department of Physics, University of Tokyo, Tokyo 113-0033, Japan

February 9, 2017

1 Hybrids

Before considering unbiased estimator for the fourth moments, we introduce more hybrids moments in addition to Online Appendix M of Ref. [1].

$$
\begin{aligned}
E[x_1 x_2 \bar{\mathbf{x}}^2] =& \frac{1}{n^2} E[x_1 x_2 (\sum_{i=1}^n x_i)^2] \\
=& \frac{1}{n^2} E[x_1 x_2 (x_1^2 + x_2^2 + \sum_{i=3}^n x_i^2 + 2x_1 x_2 + 2x_1 (\sum_{i=3}^n x_i) + 2x_2 (\sum_{i=3}^n x_i) + \sum_{i=3}^n \sum_{j=i+1}^n x_i x_j)] \\
=& \frac{2}{n^2} E[x_1^3 x_2] + \frac{n-2}{n^2} E[x_1 x_2 x_3^2] + \frac{2}{n^2} E[x_1^2 x_2^2] + \frac{4(n-2)}{n^2} E[x_1^2 x_2 x_3] \\
&+ \frac{(n-2)(n-3)}{n^2} E[x_1 x_2 x_3 x_4] \\
=& \frac{2}{n^2} (\mathcal{S}(\mathbf{x}) + 3\sigma^2(\mathbf{x})\mu + \mu^3)\mu + \frac{5(n-2)}{n^2} (\sigma^2(\mathbf{x}) + \mu^2)\mu^2 + \frac{2}{n^2} (\sigma^2(\mathbf{x})^2 + \mu^2)^2 \\
&+ \frac{(n-2)(n-3)}{n^2} \mu^4 \\
=& \frac{2}{n^2} \mathcal{S}(\mathbf{x})\mu + \frac{5}{n} \sigma^2(\mathbf{x})\mu^2 + \frac{2}{n^2} \sigma^4 + \mu^4
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
E[x_1^2 x_2 \bar{\mathbf{x}}] =& \frac{1}{n} E[x_1^2 x_2 (\sum_{i=1}^n x_i)] \\
=& \frac{1}{n} E[x_1^3 x_2 + x_1^2 x_2^2 + (n-2) x_1^2 x_2 x_3] \\
=& \frac{1}{n} (\mathcal{S}(\mathbf{x}) + 3\sigma^2(\mathbf{x})\mu + \mu^3)\mu + \frac{1}{n} (\sigma^2(\mathbf{x}) + \mu^2)^2 + \frac{n-2}{n} (\sigma^2(\mathbf{x}) + \mu^2)\mu^2 \\
=& \frac{1}{n} \mathcal{S}(\mathbf{x})\mu + \frac{(n+3)}{n} \sigma^2(\mathbf{x})\mu^2 + \frac{1}{n} \sigma^4 + \mu^4.
\end{aligned}
\tag{2}
$$

2 Square of variance, $\sigma^4_{\bar{\mathbf{x}}}(\mathbf{x})$

Expression for square of variance, $\sigma^4_{\bar{\mathbf{x}}}(\mathbf{x})$, given on p. 9 in Online Appendix M of Ref. [1] is not correct. The correct expression is

$$\begin{aligned}
\sigma^4_{\bar{\mathbf{x}}}(\mathbf{x}) =& E[(x_i-\bar{\mathbf{x}})^2(x_j-\bar{\mathbf{x}})^2] \\
=& E[\frac{1}{n}\sum(x_i-\bar{\mathbf{x}})^2\cdot\frac{1}{n}\sum(x_j-\bar{\mathbf{x}})^2] \\
=& \frac{1}{n^2}E[\sum_i\sum_j(x_i-\bar{\mathbf{x}})^2(x_j-\bar{\mathbf{x}})^2] \\
=& \frac{1}{n^2}E[\sum_i(x_i-\bar{\mathbf{x}})^4]+\frac{1}{n^2}E[\sum_i\sum_{j\neq i}(x_i-\bar{\mathbf{x}})^2(x_j-\bar{\mathbf{x}})^2] \\
=& \frac{1}{n}E[(x_1-\bar{\mathbf{x}})^4]+\frac{n-1}{n}E[(x_1-\bar{\mathbf{x}})^2(x_2-\bar{\mathbf{x}})^2] \\
=& \frac{1}{n}E[(x_1-\bar{\mathbf{x}})^4]+\frac{n-1}{n}E[x_1^2x_2^2-4x_1^2x_2\bar{\mathbf{x}}+2x_1^2\bar{\mathbf{x}}^2+4x_1x_2\bar{\mathbf{x}}^2-4x_1\bar{\mathbf{x}}^3+\bar{\mathbf{x}}^4] \\
=& \frac{1}{n}\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})+\frac{n-1}{n}(\sigma^2(\mathbf{x})+\mu^2)^2 \\
&-\frac{4(n-1)}{n}\Big[\frac{1}{n}\mathcal{S}(\mathbf{x})\mu+\frac{(n+3)}{n}\sigma^2(\mathbf{x})\mu^2+\frac{1}{n}\sigma^4+\mu^4\Big] \\
&+\frac{2(n-1)}{n}\Big[\frac{1}{n^2}\mathcal{K}(\mathbf{x})+\frac{2(n+1)}{n^2}\mathcal{S}(\mathbf{x})\mu+\frac{n+5}{n}\sigma^2(\mathbf{x})\mu^2+\frac{n-1}{n^2}\sigma^4+\mu^4\Big] \\
&+\frac{4(n-1)}{n}\Big[\frac{2}{n^2}\mathcal{S}(\mathbf{x})\mu+\frac{5}{n}\sigma^2(\mathbf{x})\mu^2+\frac{2}{n^2}\sigma^4+\mu^4\Big] \\
&-\frac{4(n-1)}{n}\Big[\frac{1}{n^3}\mathcal{K}(\mathbf{x})+\frac{4}{n^2}\mathcal{S}(\mathbf{x})\mu+\frac{6}{n}\sigma^2(\mathbf{x})\mu^2+\frac{3(n-1)}{n^3}\sigma^4+\mu^4\Big] \\
&+\frac{n-1}{n}\Big[\frac{1}{n^3}\mathcal{K}(\mathbf{x})+\frac{4}{n^2}\mathcal{S}(\mathbf{x})\mu+\frac{6}{n}\sigma^2(\mathbf{x})\mu^2+\frac{3(n-1)}{n^3}\sigma^4+\mu^4\Big] \\
=& \frac{1}{n}\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})+\frac{(n-1)(2n-3)}{n^4}\mathcal{K}(\mathbf{x})+\frac{(n-1)(n^3-2n^2-3n+9)}{n^4}\sigma^4(\mathbf{x}) \\
=& \frac{n-1}{n^3}\Big[(n-1)\mathcal{K}(\mathbf{x})+(n^2-2n+3)\sigma^4(\mathbf{x})\Big].
\end{aligned} \tag{3}$$

In the last line in Eq. (3), we use the expression for the fourth sample central moment:

$$\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x}) = E[(x_i-\bar{\mathbf{x}})^4] = \frac{n-1}{n^3}\Big[(n^2-3n+3)\mathcal{K}(\mathbf{x})+(6n-9)\sigma^4(\mathbf{x})\Big], \tag{4}$$

which is given on p. 8 in Online Appendix M of Ref. [1].

3 Unbiased estimator for $\sigma^4(\mathbf{x})$, $\mathcal{K}(\mathbf{x})$, and fourth cumulant

By solving Eqs. (3) and (4), we obtain the unbiased estimators for the square of the variance, the fourth central moment, and the fourth cumulant as

$$\sigma^4(\mathbf{x}) = \frac{n}{(n-1)(n-2)(n-3)}\Big[(n^2-3n+3)\sigma^4_{\bar{\mathbf{x}}}(\mathbf{x})-(n-1)\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})\Big] \tag{5}$$

$$\mathcal{K}(\mathbf{x}) = \frac{n}{(n-1)(n-2)(n-3)}\Big[(n^2-2n+3)\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})-(6n-9)\sigma^4_{\bar{\mathbf{x}}}(\mathbf{x})\Big] \tag{6}$$

$$\kappa_4(\mathbf{x}) = \mathcal{K}(\mathbf{x})-3\sigma^4(\mathbf{x}) = \frac{n^2}{(n-1)(n-2)(n-3)}\Big[(n+1)\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})-3(n-1)\sigma^4_{\bar{\mathbf{x}}}(\mathbf{x})\Big], \tag{7}$$

respectively.

4 Numerical tests

We test numerically the above expressions for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$. We evaluate Eqs. (1)–(7) for the sample size $n = 4, 8, 16, 32, \cdots, 1024$. For each sample size, we take the average over 65536 sample sets and estimate the error bar. The results are shown in Figs. 1–7.

References

[1] Klements, B. *Modeling with Data: Tools and Techniques for Scientific Computing* (Princeton University Press, Princeton, 2009). URL `http://modelingwithdata.org`.

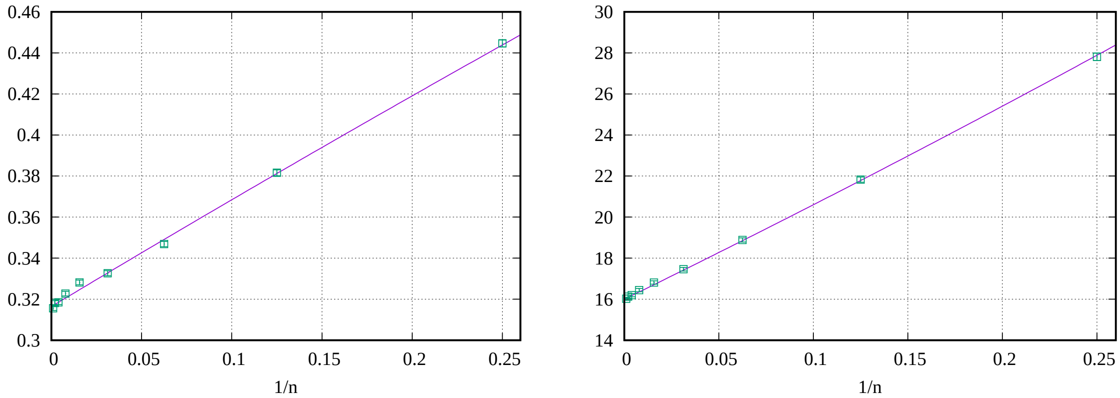

Figure 1: Sample-size dependence of $E[x_1 x_2 \bar{\mathbf{x}}^2]$ for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes Eq. (1) calculated by using the exact moments.

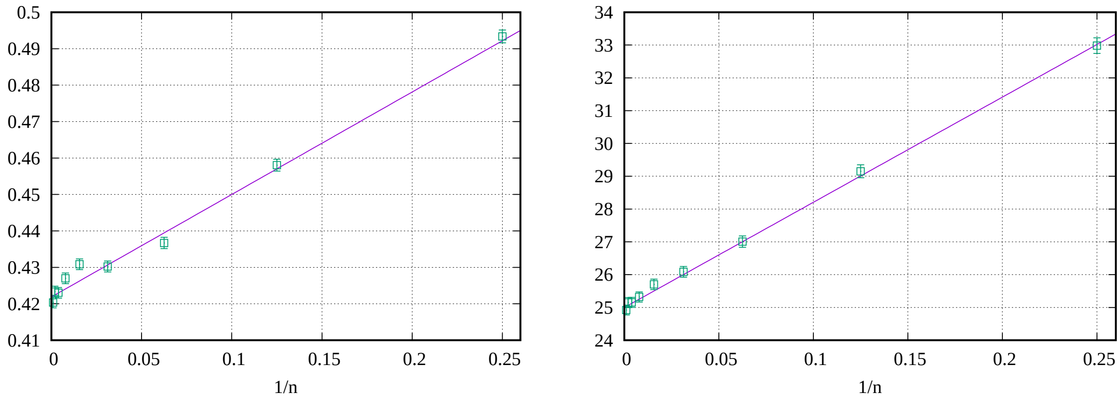

Figure 2: Sample-size dependence of $E[x_1^2 x_2 \bar{\mathbf{x}}]$ for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes Eq. (2) calculated by using the exact moments.

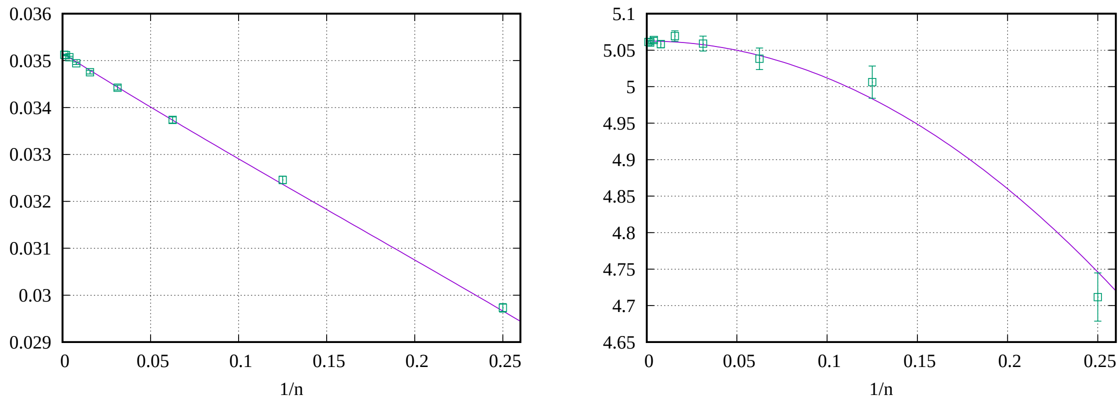

Figure 3: Sample-size dependence of $\sigma_{\bar{\mathbf{x}}}^4(\mathbf{x})$ for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes Eq. (3) calculated by using the exact moments.

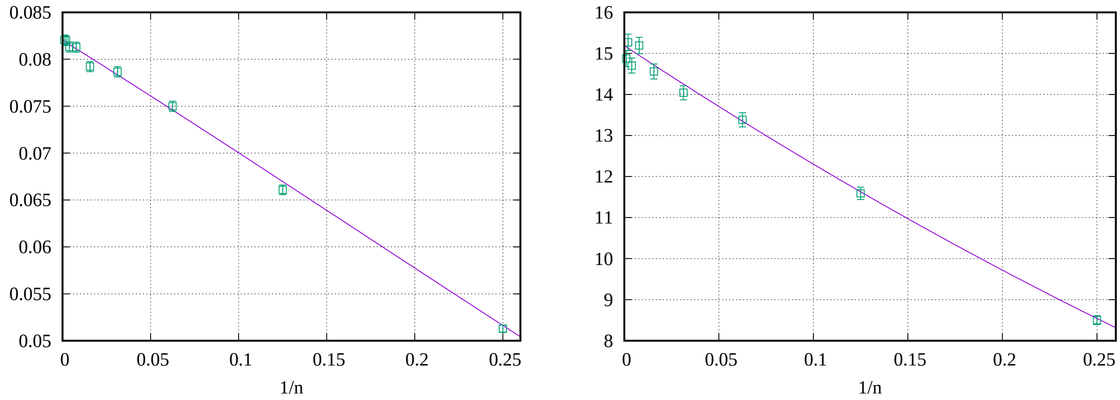

Figure 4: Sample-size dependence of $\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})$ for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes Eq. (4) calculated by using the exact moments.

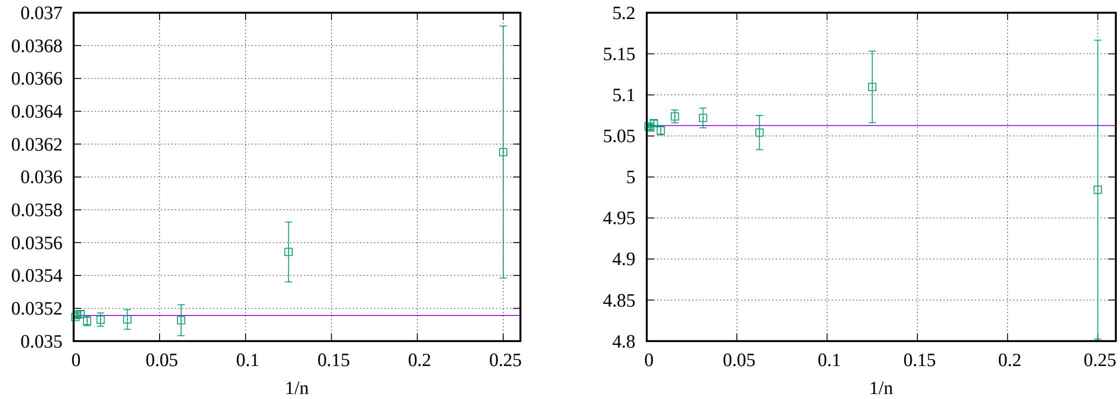

Figure 5: Results for the square of the variance, $\sigma^4(\mathbf{x})$, by using the unbiased estimator (5) for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes the exact value.

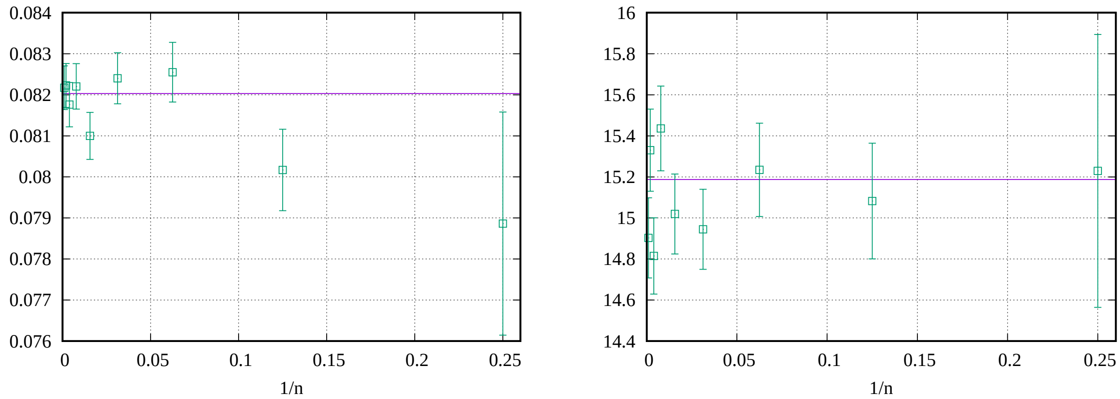

Figure 6: Results for the fourth central moment, $\mathcal{K}(\mathbf{x})$, by using the unbiased estimator (6) for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes the exact value.

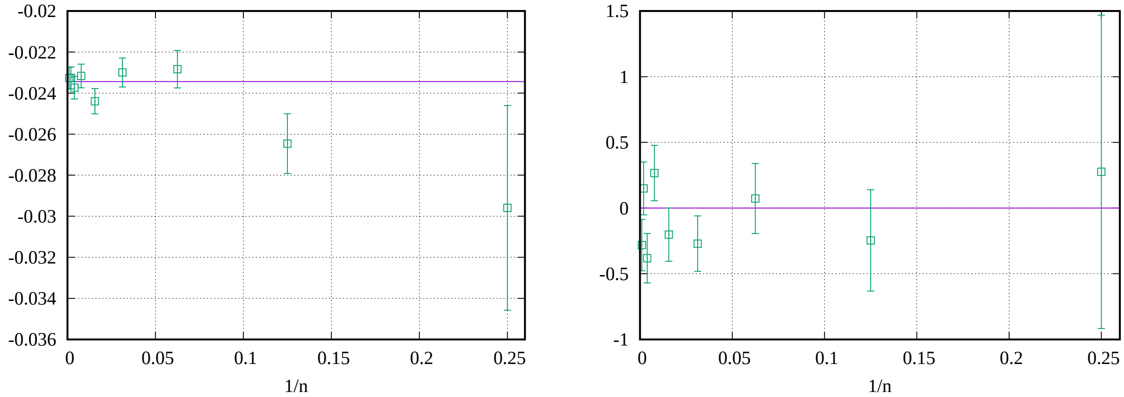

Figure 7: Results for the fourth cumulant, $\kappa_4(\mathbf{x})$, by using the unbiased estimator (6) for the Bernoulli distribution $\mathcal{B}(p = 3/4)$ (left) and the normal distribution $\mathcal{N}(\mu = 2, \sigma = 3/2)$ (right). The green squares denotes the numerical results and the purple line denotes the exact value.