Notes on Error Bars and Log-Scale Error Bars

Jungtaek Kim

University of Pittsburgh
Pennsylvania, USA
jungtaek.kim@pitt.edu

January 4, 2023

These notes describe an error bar and, in particular, log-scale error bars, where an error bar shows the error or uncertainty of repeated experiments.

1 Sample Mean, Sample Variance, Standard Error of the Sample Mean, and Confidence Interval

Suppose that $X_1, \ldots, X_n$ are n observations of a variable X. The sample mean of $X_1, \ldots, X_n$ is defined as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{1}$$

Along with the sample mean $\bar{X}$, the sample variance of $X_1, \ldots, X_n$ is defined as

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2. \tag{2}$$

The standard error of the sample mean is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}. \tag{3}$$

A confidence interval for an unknown parameter μ satisfies the following inequalities with a probability p:

$$P(\bar{X} - c\sigma_{\bar{X}} \leq \mu \leq \bar{X} + c\sigma_{\bar{X}}) = p, \tag{4}$$

where $c \geq 0$ is a coefficient for controlling the confidence interval. To determine the confidence interval, one of the following coefficients can be chosen; see Table 1. For example, if we choose c as 1.960, we can obtain a 95% confidence interval for μ. See textbooks on probabilities, e.g., a book by Papoulis [1991], for the details.

2 Error Bars

As shown in Figure 1, given 50 observations sampled from $\mathcal{N}(0, 4^2)$, we can compute the sample mean of the observations, the standard deviation of the observations, the standard error of the sample mean, and 90.0%/95.0%/99.9%/99.9% confidence intervals, by (1), (2), (3), (4), and Table 1.

Table 1: Coefficients for confidence intervals.

p	0.000	0.500	0.600	0.700	0.800	0.900	0.950	0.980	0.990	0.999
c	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

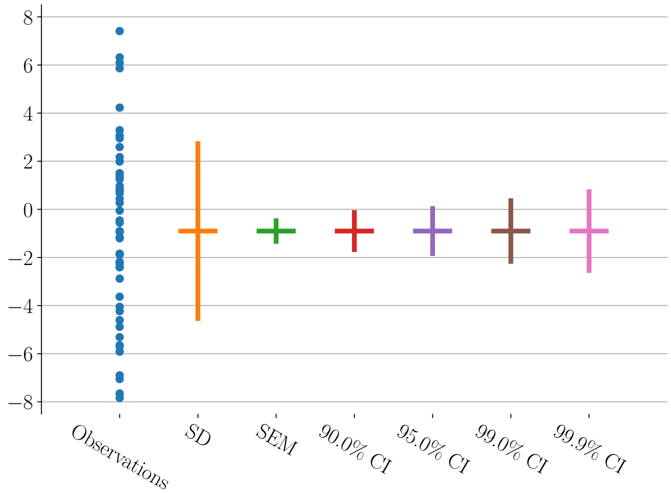

Figure 1: Error bars for the standard deviation of observations (denoted as SD), the standard error of the sample mean (denoted as SEM), and 90.0%/95.0%/99.0%/99.9% confidence intervals (denoted as CIs). Horizontal bars indicate the sample mean of observations. Observations are sampled from $\mathcal{N}(0, 4^2)$.

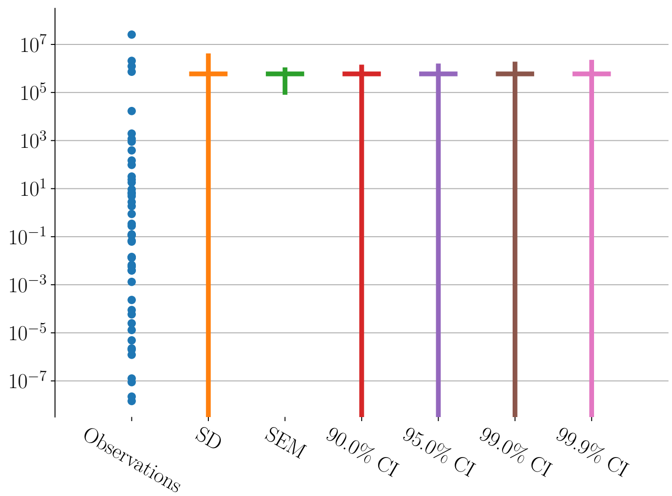

Figure 2: Log-scale error bars for the standard deviation of observations (denoted as SD), the standard error of the sample mean (denoted as SEM), and 90.0%/95.0%/99.0%/99.9% confidence intervals (denoted as CIs), computing the statistics of Gaussian distributions. Horizontal bars indicate the sample mean of observations. Observations are sampled from $\text{Lognormal}(0, 4^2)$. Precisely, the exactly same points shown in Figure 1 are sampled, and then an exponential function is applied to those points.

3 Log-Scale Error Bars

Unlike the error bars described in Section 2, we can plot different styles of error bars where an axis is log-scale:

1. error bars with the statistics of a Gaussian distribution (see Figure 2);
2. error bars with the statistics of a log-normal distribution (see Figure 3).

To visualize error bars, 50 observations are sampled from $\text{Lognormal}(0, 4^2)$ as shown in Figures 2 and 3. Note that observations in Figures 1, 2, and 3 are shown identically, since they are converted to match using an exponential function.

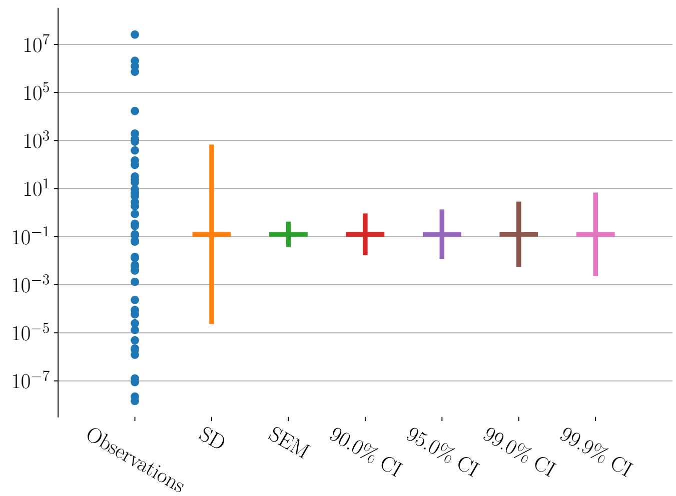

Figure 3: Log-scale error bars for the standard deviation of observations (denoted as SD), the standard error of the sample mean (denoted as SEM), and 90.0%/95.0%/99.0%/99.9% confidence intervals (denoted as CIs), computing the statistics of log-normal distributions. Horizontal bars indicate the sample mean of observations. Observations are the same as the observations presented in Figure 2.

As presented in Figure 2, the sample mean is $10^{5.7740}$, which is mostly affected by a few large observations. Moreover, error bars are asymmetric due to a log-scale axis. Given the results in Figure 2, we need to think of whether an error bar with the statistics of a Gaussian distribution is appropriate or not. Instead of the Gaussian distribution, we can plot error bars with the statistics of a log-normal distribution as shown in Figure 3. The sample mean is computed as follows:

$$\bar{X}_{\text{Lognormal}} = 10^{\frac{1}{n}\sum_{i=1}^{n} \log_{10} X_i}. \tag{5}$$

With (5), the sample standard deviation is computed:

$$\log_{10} \sigma_{\text{Lognormal}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\log_{10} X_i - \log_{10} \bar{X}_{\text{Lognormal}})^2}, \tag{6}$$

$$\sigma_{\text{Lognormal,upper}} = \bar{X}_{\text{Lognormal}}\sigma_{\text{Lognormal}} - \bar{X}_{\text{Lognormal}}, \tag{7}$$

$$\sigma_{\text{Lognormal,lower}} = \bar{X}_{\text{Lognormal}} - \bar{X}_{\text{Lognormal}}/\sigma_{\text{Lognormal}}. \tag{8}$$

Interestingly, the upper and lower standard deviations are different as in (7) and (8). By both (7) and (8), an error bar for the sample standard deviation is symmetric. Similar to (3), the standard error of the sample mean is computed:

$$\log_{10} \sigma_{\bar{X}_{\text{Lognormal}}} = \frac{1}{\sqrt{n}} \log_{10} \sigma_{\text{Lognormal}}, \tag{9}$$

$$\sigma_{\bar{X}_{\text{Lognormal}},\text{upper}} = \bar{X}_{\text{Lognormal}}\sigma_{\bar{X}_{\text{Lognormal}}} - \bar{X}_{\text{Lognormal}}, \tag{10}$$

$$\sigma_{\bar{X}_{\text{Lognormal}},\text{lower}} = \bar{X}_{\text{Lognormal}} - \bar{X}_{\text{Lognormal}}/\sigma_{\bar{X}_{\text{Lognormal}}}. \tag{11}$$

Also, by (10) and (11), an error for the standard error is symmetric. Likewise, a confidence interval is computed with the following:

$$\sigma_{\bar{X}_{\text{Lognormal}},c} = 10^{\frac{c}{\sqrt{n}} \log_{10} \sigma_{\text{Lognormal}}}, \tag{12}$$

where a coefficient for the confidence interval c is given. Eventually, we can readily employ (12) in (10) and (11), for computing the confidence interval.

Acknowledgements

This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided. Specifically, this work used the HTC cluster, which is supported by NIH award number S10OD028483.

References

A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, 1991.