Dataset 1

Part A

1. What is the mean value of this data?

The mean is 1.1857

2. What is the standard error in the estimate of the mean?

The standard error is 0.0133

3. What is the standard deviation of the data?

The standard deviation is 0.4012.

Part B

1. How does the standard error in the estimate of the mean change?

The standard error changed from 0.0133 to 0.0181.

2. How does standard deviation change?

The standard deviation changed from 0.4012 to 0.4037.

3. How would you expect these quantities to change as you gradually increase the initial cutoff from 0 to 500?

Standard Error will gradually decrease, as is shown in the definition, while standard error will fluctuate around a certain number and gradually stabilize at that point.

Dataset 2

1. What is the autocorrelation time of this series?

The autocorrelation time is 19.74.

2. What would be the standard error in the estimate of the mean without considering autocorrelation?

The standard error without considering autocorrelation is 0.00326.

3. Now what would be the error in the estimate of the mean with correlation?

The standard error is 0.0145.

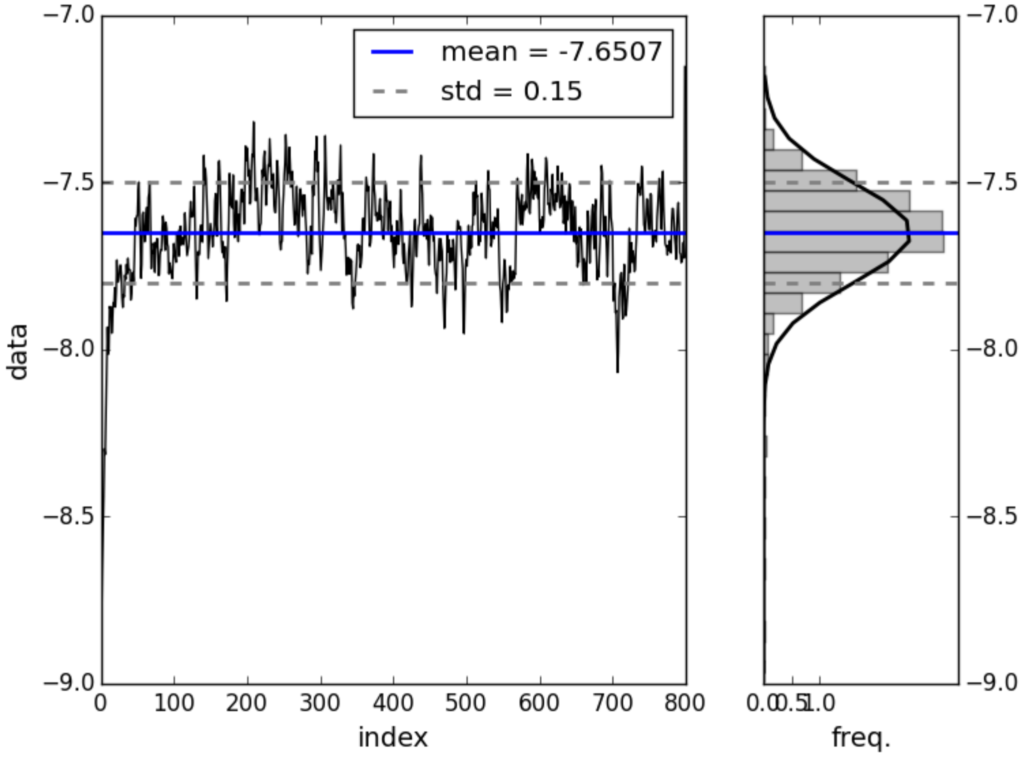
Dataset 3

1. What is the mean (and its error) with the initial cutoff set to zero?

The mean is -7.6507 and the error is 0.0281.

2. Where should the initial cutoff be set?

By checking the plot of trace, the initial cutoff should be set at about 100.



3. What is the mean (and its error) with the new cutoff?

The mean is -7.6293 and its error is 0.0175.

4. Is the difference between the mean in part 1 and part 3 significant?

	Mean	Error
Part1	-7.6507	0.0281
Part3	-7.6293	0.0175

No, they are not. Because two error bars overlap.

Dataset 4

$$P(x) = \frac{b}{|x|^a + c}$$

1. Based on this analytic expression, what do you expect for the mean? the variance? (Hint: for the variance, it's the behavior at large x that matters, so one can use an approximation for the denominator of the distribution function.)

The mean should be zero, since the integrand is an odd function.

For variance, because mean is zero, the equation becomes: $s^2 = \int_{-\infty}^{+\infty} x^2 P(x)\,dx$. When x becomes greater, c can be omitted, so the final integration is $s^2 = \int_{-\infty}^{+\infty} x^2 \frac{b}{|x|^{2.2}}\,dx$, which leads to infinity.

2. Look at the convergence of the mean and sigma by computing these values for five "end cutoffs" from 1000 to 5000 (i.e., 0-999, 0-1999, 0-2999, etc). Do the same for data set 1 or 2 and compare the convergence behavior?

Dataset 4	1000	2000	3000	4000	5000
Mean	0.0905	0.1560	0.2722	0.2366	0.2399
Sigma	1.9759	1.9547	4.2772	3.8666	3.7107

Dataset 1	200	400	600	800	1000
Mean	1.1908	1.1801	1.1929	1.1874	1.1857
Sigma	0.4022	0.3995	0.4077	0.4049	0.4012

Dataset 2	200	400	600	800	1000
Mean	-7.3201	-7.2852	-7.2822	-7.2716	-7.2617
Sigma	0.1056	0.1041	0.1031	0.1032	0.1031

Compared to Dataset 1 or 2, the fluctuation of mean and sigma of Dataset 4 is really huge, which indicates it may not have converged at 5000.

Comparison of Datasets

1. Compute the mean, variance, and the estimate of the error of the mean for A and B separately, assuming each run is **uncorrelated** with the others.

	mean	variance	Error
A	1.253	0.0249	0.0644
B	1.352	0.0272	0.0673

2. Show that the probability that the two runs are (NOT) drawn from the same distribution is ~ 29% (71%).

$$\frac{\mu_{\mathrm{a}} - \mu_b}{\sqrt{\delta_a{}^2 + \delta_b{}^2}} = 1.06$$

$$p(0,1.06) = 0.3554$$

$$p = 2 \times p(0,1.06) = 71\%$$

3. Assuming that all the data is drawn from the same distribution, estimate the mean and the error of the mean of the combined data set.

The mean is 1.3025, and the error is 0.0662.

Bias from Unequilibrated Data

1. Give an expression for m as a function of m1, m2, t1 and t2. Give also an expression of the systematic error, ε = m - m2, that comes from including the bad data, in terms of the

prolongation λ = t2/t1 and Δ = m1 - m2.

$$m = \frac{m_1 t_1 + m_2(t_2 - t_1)}{t_2}$$

$$m - m_2 = \frac{m_1 t_1 + m_2(t_2 - t_1) - \mathrm{m}_2\mathrm{t}_2}{t_2} = \frac{(m_1 - m_2)t_1}{t_2} = \frac{\Delta}{\lambda}$$

2. Give an expression for σ, grouping your terms in inverse powers of λ.

$$\sigma^2 = E(x^2) - E^2(x)$$

$$E^2(x) = m^2 = m_2^2 + \frac{2\Delta m_2}{\lambda} + \frac{\Delta^2}{\lambda^2}$$

$$E(x^2) = \frac{t_1 E(x_1^2) + (t_2 - t_1)E(x_2^2)}{t_2} = \frac{t_1(\sigma_1^2 + \mathrm{m}_1^2) + (t_2 - t_1)(\sigma_2^2 + m_2^2)}{t_2}$$

$$= \frac{\sigma_1^2 + m_1^2}{\lambda} + (\sigma_2^2 + m_2^2 - \frac{\sigma_2^2 + m_2^2}{\lambda})$$

$$\sigma^2 = \frac{\sigma_1^2 + m_1^2}{\lambda} + (\sigma_2^2 + m_2^2 - \frac{\sigma_2^2 + m_2^2}{\lambda}) - (m_2^2 + \frac{2\Delta m_2}{\lambda} + \frac{\Delta^2}{\lambda^2})$$

$$= \sigma_2^2 + \frac{\sigma_1^2 - \sigma_2^2 + \Delta^2}{\lambda} - \frac{\Delta^2}{\lambda^2}$$

$$\sigma = \sqrt{\sigma_2^2 + \frac{\sigma_1^2 - \sigma_2^2 + \Delta^2}{\lambda} - \frac{\Delta^2}{\lambda^2}}$$

3. Consider t2 » t1, i.e. λ » 1. At what time t2 will the statistical error of the mean equal the systematic error ε from the unequilibrated data, utilizing the lowest nontrivial order of approximation. Don't forget the autocorrelation time τ.

When λ » 1, $\sigma \approx \sqrt{\sigma_2^2 + \frac{\sigma_1^2 - \sigma_2^2 + \Delta^2}{\lambda}}$

To get t2, we need to compute: $\frac{\sigma}{\sqrt{t_2 / \tau}} = \varepsilon = \frac{\Delta}{\lambda}$

After simplification $\mathrm{t}_2 = \frac{t_1^2 \Delta^2}{\tau\sigma^2} - \frac{t_1(\sigma_1^2 - \sigma_2^2 + \Delta^2)}{\sigma_2^2}$

4. With τ=1 and t1 = 100, estimate this time by eyeballing the appropriate ratio from the plot.

Under the same scale $m_1 = 54, m_2 = 10.8, \sigma_1 = 10.2, \sigma_2 = 6.2$

After substitution, I get:

$$t_2 = \frac{100^2 \times (54 - 10.8)^2}{6.2^2} - \frac{100 \times (10.2^2 - 6.2^2 + (54 - 10.8)^2}{6.2^2} = 4.8 \times 10^5$$