Sample Mean/Variance and Confidence Regions

Keith M. Chugg – November 20, 2015– version 0.5

In your homework, you generated 500 realizations of a Gaussian random variable with zero mean and unit variance. Specifically, you generated z_1, z_2, $z_3 \dots z_{500}$, all drawn from the sampling distribution $f_{Z(u)}(z) = \mathcal{N}(z;0;1)$. In this handout, we use these 500 realizations to explore the concepts of sample mean, sample variance, and confidence regions.

1 Sample Mean and Variance

Let us first consider the more general case where n independent realizations are drawn from the sampling distribution $f_{Z(u)}(z)$ – *i.e.,* this means that we generate a realization of $Z_i(u) = z_i$ for $i = 1, 2, \dots n$ where the distribution of each $Z_i(u)$ is $f_{Z(u)}(z)$ and these random variables are independent. In particular, each $Z_i(u)$ has mean m and variance σ^2.

First, the sample mean of n realizations is

$$\hat{m}_n = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{1}$$

and the sample variance is

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \hat{m}_n)^2 \tag{2a}$$

$$= \left(\frac{1}{n-1}\sum_{i=1}^{n} z_i^2\right) - \left(\frac{n}{n-1}\right)\hat{m}_n^2 \tag{2b}$$

Note that the definition of the sample variance includes the sample mean. Also note that these are realizations of random variables themselves. Specifically, the sample mean random variable is

$$\hat{M}_n(u) = \frac{1}{n}\sum_{i=1}^{n} Z_i(u) \tag{3}$$

$$\mathbb{E}\left\{\hat{M}_n(u)\right\} = m \tag{4}$$

$$\text{var}\left[\hat{M}_n(u)\right] = \frac{\sigma^2}{n} \tag{5}$$

The sample variance random variable is[1]

$$S_n^2(u) = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i(u) - \hat{M}_n(u))^2 \tag{6}$$

$$\mathbb{E}\left\{S_n^2(u)\right\} = \sigma^2 \tag{7}$$

[1] The reason that we divide by $(n-1)$ instead of n in the sample variance equation is so that $\mathbb{E}\left\{S_n^2(u)\right\} = \sigma^2$ – this is called an unbiased estimate. Specifically, the sample variance is an unbiased estimate of the variance since they have the same expected value. If the mean is known, but the variance is unknown, we can replace $\hat{m}_n$ by m in (2a) and divide by n. This is known as the population variance.

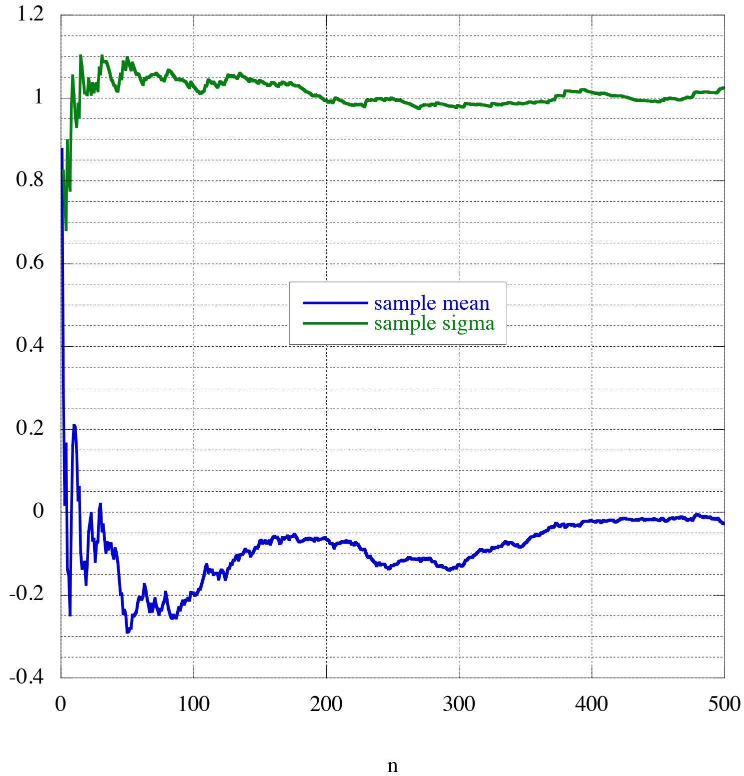

Figure 1: Sample mean and variance as a function of n for realizations of a standard Gaussian random variable.

For the 500 realizations of the standard normal random variable from homework, the sample mean and sample variance are plotted in Fig. 1. Notice that the sample mean varies significantly for small values of n, but stabilizes near the value of $m = 0$ for large values of n as is predicted by the Law of Large Numbers. The sample standard deviation (the square root of the sample variance) exhibits similar behavior – *i.e.,* eventually settling down around the value of $\sigma = 1$.

2 Confidence Regions

Suppose we would like to characterize the qualitative notion of "settling" to the true value or, equivalently, put some quantitative measure on the Law of Large Numbers. For example, we can seek a region of the real line $B_{1-\alpha}$ such that

$$\text{Pr}\left\{X(u) \in B_{1-\alpha}\right\} = 1 - \alpha \tag{8}$$

where $\alpha \in (0, 1)$. The region $B_{1-\alpha}$ is called an "$100(1-\alpha)$ percent confidence region" for $X(u)$. In other words, we known with confidence $1 - \alpha$ that $X(u)$ will lie in the region $B_{1-\alpha}$.

Common choices for α are 0.1, 0.05, and 0.01 corresponding to 90%, 95%, and 99% confidence regions, respectively.

The most common choice for the confidence region is a symmetric interval around the mean

$$\text{Pr}\{m_X - v \leq X(u) \leq m_X + v\} = \text{Pr}\{|X(u) - m_X| \leq v\} = 1 - \alpha \tag{9}$$

An important example of this is when $X(u)$ is Gaussian with mean m and variance σ^2 for which

$$\begin{aligned} \text{Pr}\left\{m - z_{\alpha/2}\sigma \leq X(u) \leq m + z_{\alpha/2}\sigma\right\} &= \text{Pr}\left\{\tfrac{|X(u)-m_X|}{\sigma} \leq z_{\alpha/2}\right\} && (10)\\ &= 1 - 2\text{Q}(z_{\alpha/2}) && (11)\\ &= 1 - \alpha && (12) \end{aligned}$$

where we have implicitly defined $z_{\alpha/2}$ via

$$2\text{Q}(z_{\alpha/2}) = \alpha \qquad \Longleftrightarrow \qquad z_{\alpha/2} = \text{Q}^{-1}(\alpha/2) \tag{13}$$

This is standard notation in the statistics literature: $z_{\alpha/2}$ is specifically associated with a Gaussian distribution as defined above. The values of $z_{\alpha/2}$ for some common values of $1 - \alpha$ are shown in Fig. 2.

We can apply this notion to the sample mean $\hat{M}_n(u)$ or, more precisely, to a standardized version thereof:

$$\text{Pr}\left\{-z_{\alpha/2} \leq \tfrac{\hat{M}_n(u)-m}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\} = 1 - \alpha \tag{14}$$

Note that, because $\frac{\hat{M}_n(u)-m}{\sigma/\sqrt{n}}$ is normal with mean zero and unit variance, the values of $z_{\alpha/2}$ (*i.e.*, inverse Q-function) define the confidence region. In our specific example, $m = 0$, $\sigma = 1$, so the resulting confidence region for $\hat{M}_n(u)$ is $(-z_{\alpha/2}/\sqrt{n}, +z_{\alpha/2}/\sqrt{n})$. For $\alpha = 0.05$, we have the 95% confidence region for the sample mean

$$\left(\frac{-1.96}{\sqrt{n}}, \frac{+1.96}{\sqrt{n}}\right) \quad 95\% \text{ confidence for sample mean } (m = 0,\ \sigma^2 = 1) \tag{15}$$

which uses the fact that $z_{0.025} = 1.96$. This concept is illustrated in Fig. 3 where the sample mean is plotted as a function of n along with the 95% confidence region. Note that some realizations of the sample mean our outside this region – in fact that is to be expected. Specifically, for this 95% confidence region, we would expect to see approximately 5% of the realizations outside of the confidence region for large sample sizes.

2.1 Confidence Region: Unknown Mean, Known Variance

The problem considered above is really a probability problem as opposed to a statistics problem. Specifically, we knew the mean, variance and pdf of a random variable and we simply computed the probability that it was within some range around the mean. A related statistics problem is: given realizations of the sample mean, find a confidence region without knowledge of the mean.

This is really a matter of interpreting the above mathematics. Specifically, the following still holds

$$\text{Pr}\left\{-z_{\alpha/2} \leq \tfrac{\hat{M}_n(u)-m}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right\} = 1 - \alpha \tag{16}$$

but now we should interpret this with m being an unknown parameter. This can also be written as

$$\text{Pr}\left\{\hat{M}_n(u) - z_{\alpha/2}\tfrac{\sigma}{\sqrt{n}} \leq m \leq \hat{M}_n(u) + z_{\alpha/2}\tfrac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha \tag{17}$$

Table of t_(alpha/2), such that P(T_v>t_(alpha/2))=alpha/2 when T_v is standard T-distributed

confidence=1-alpha	0.5	0.8	0.9	0.95	0.98	0.99
deg. of freedom v = (n-1)	alpha/2 0.25	0.1	0.05	0.025	0.01	0.005
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
31	0.682	1.309	1.696	2.040	2.453	2.744
32	0.682	1.309	1.694	2.037	2.449	2.738
33	0.682	1.308	1.692	2.035	2.445	2.733
34	0.682	1.307	1.691	2.032	2.441	2.728
35	0.682	1.306	1.690	2.030	2.438	2.724
36	0.681	1.306	1.688	2.028	2.434	2.719
37	0.681	1.305	1.687	2.026	2.431	2.715
38	0.681	1.304	1.686	2.024	2.429	2.712
39	0.681	1.304	1.685	2.023	2.426	2.708
40	0.681	1.303	1.684	2.021	2.423	2.704
infinity	0.674	1.282	1.645	1.960	2.326	2.576

Table of z_(alpha/2), such that P(Z>z_(alpha/2))=alpha/2 when Z is Gaussian, m=0,sigma=1

confidence=1-alpha	0.5	0.8	0.9	0.95	0.98	0.99
alpha/2=	0.25	0.1	0.05	0.025	0.01	0.005
z_(alpha/2)=	0.674	1.282	1.645	1.960	2.326	2.576

Figure 2: Table of values for $t_{\alpha/2}(n-1)$ and $z_{\alpha/2}$ for common values of confidence $1-\alpha$.

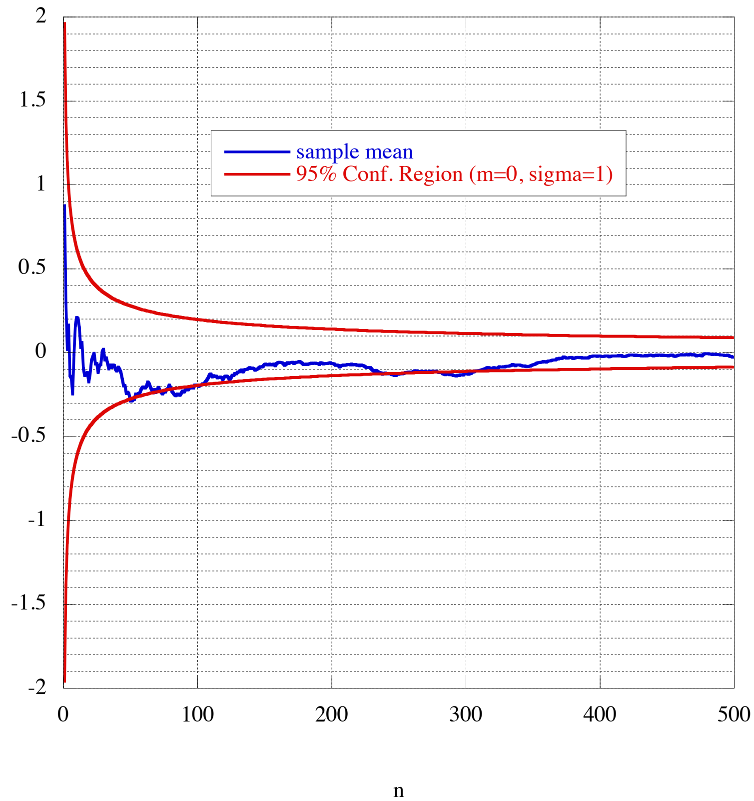

Figure 3: Sample mean with 95% confidence region ($m = 0$, $\sigma = 1$ are known).

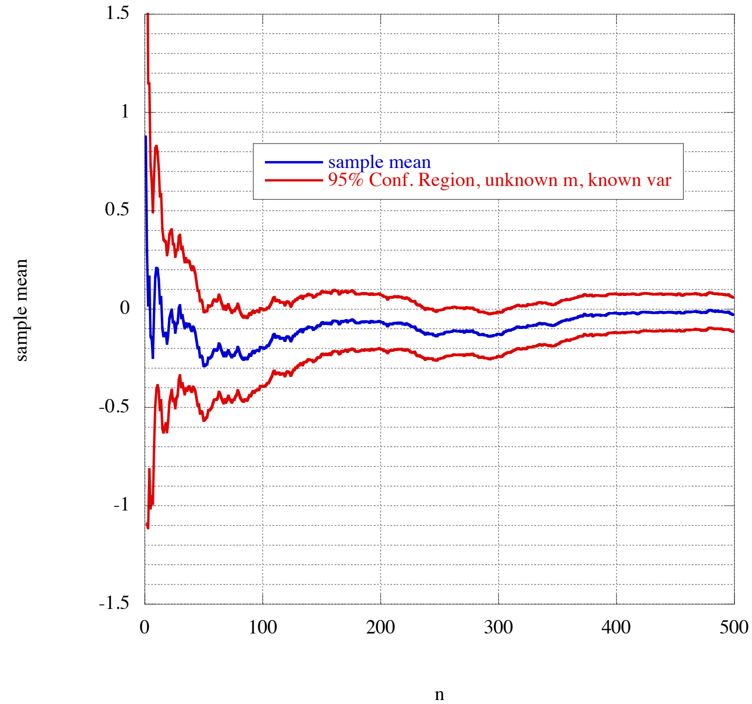

Figure 4: The 95% confidence region for m from the observed realizations ($\sigma = 1$ known).

This may seem odd at first since m is not random, it is a deterministic parameter. Nonetheless, this is the more common notion of a confidence interval in statistics: this provides us with a $100(1-\alpha)\%$ confidence for the mean. The way this is used in practice is that we obtain a set of samples $\{z_i\}_{i=1}^n$, compute the associated sample mean $\hat{m}_n$, and then the associated confidence region for m is

$$\boxed{\left(\hat{m}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{m}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \quad 100(1-\alpha)\% \text{ confidence for } m} \tag{18}$$

Another way that this is often written is

$$m = \hat{m}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{19}$$

and the quantity $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is known as the margin of error (MOE). As a specific example if $\sigma = 1$, then the 95% confidence region is

$$m = \hat{m}_n \pm \frac{1.96}{\sqrt{n}} \tag{20}$$

That is, if we knew the variance was 1, but did not know the mean, then we could conclude the above about the value of m. This concept is illustrated in Fig. 4. Note that, again, there are some points where the true mean, which we know is zero in this case since we generated the data, is outside of the 95% confidence region.

2.2 Confidence Region: Unknown Mean, Unknown Variance

The above example is still usually unrealistic in practice: it is unlikely we would know the value of σ and not know the mean. Instead of assuming the variance is known, we can use the sample variance to find a confidence region on the mean. Specifically, we can consider the random variable

$$T_{n-1}(u) = \frac{\hat{M}_n(u) - m}{S_n(u)/\sqrt{n}} \tag{21}$$

If the underlying sampling distribution is Gaussian (*i.e.,* the pdf of $Z_i(u)$ is Gaussian), then the numerator is Gaussian and $S_n^2(u)$ is a χ^2 random variable with $n-1$ degrees of freedom. Also, the numerator and denominator are, remarkably, independent. It follows that the pdf of $T_{n-1}(u)$ can be computed in closed form and $T_{n-1}(u)$ is called a Student-T or simply a T random variable with $n-1$ degrees of freedom:

$$f_{T_{n-1}(u)}(t) = \frac{\Gamma(n/2)}{\Gamma((n-1)/2)\sqrt{\pi(n-1)}} \left(\frac{1}{\sqrt{1+\frac{t^2}{n-1}}} \right)^{-n/2} \tag{22}$$

which is valid for integer $n \geq 2$.[2] Note that the pdf $f_{T_{n-1}(u)}(t)$ is an even function – *i.e.,* $f_{T_{n-1}(u)}(t) = f_{T_{n-1}(u)}(-t)$ so that the mean of $T_{n-1}(u)$ is zero and

$$\Pr\left\{-t_{\alpha/2}(n-1) \leq T_{n-1}(u) \leq t_{\alpha/2}(n-1)\right\} = 1 - \alpha \tag{23}$$

In other words, since the pdf is symmetric around zero, we can identify $t_{\alpha/2}(n-1)$ with the above property since the integral under the tail to the left and right are the same.

Similarly to how $z_{\alpha/2}$ is related to the inverse cdf of a standard Gaussian, $t_{\alpha/2}(n-1)$ is related to the inverse cdf of a T random variable with $n-1$ degrees of freedom. This inverse cdf has no closed form, but like the inverse Q-function, it is common in statistics and is therefore available in tables and as a built in function in many numerical packages.

In practice, this is applied as follows: given a set of realizations of a Gaussian $\{z_i\}_{i=1}^n$, compute the sample mean and sample variance via (1) and (2a), respectively. Then, the confidence region for the mean

$$\boxed{\left(\hat{m}_n - t_{\alpha/2}(n-1)\frac{s_n}{\sqrt{n}}, \hat{m}_n + t_{\alpha/2}(n-1)\frac{s_n}{\sqrt{n}}\right) \quad 100(1-\alpha)\% \text{ confidence for } m} \tag{24}$$

Again, this is often written as

$$m = \hat{m}_n \pm t_{\alpha/2}(n-1)\frac{s_n}{\sqrt{n}} \tag{25}$$

and the quantity $t_{\alpha/2}(n-1)\frac{s_n}{\sqrt{n}}$ is known as the margin of error (MOE).

Note that $t_{\alpha/2}(n-1)$ changes with n. Thus, if we want a 95% confidence region, we need to compute $t_{\alpha/2}(n-1)$ for for each value of n considered. For example, if $n=4$, then $t_{0.025}(3) = 3.18$ and for $n=20$, $t_{0.025}(19) = 2.09$ so that

$$m = \hat{m}_4 \pm \frac{3.18 s_4}{\sqrt{4}} \tag{26}$$

$$m = \hat{m}_{20} \pm \frac{2.09 s_{20}}{\sqrt{20}} \tag{27}$$

[2]See Leon-Garcia, page 434 for a plot of this pdf and a comparison to the standard Gaussian.

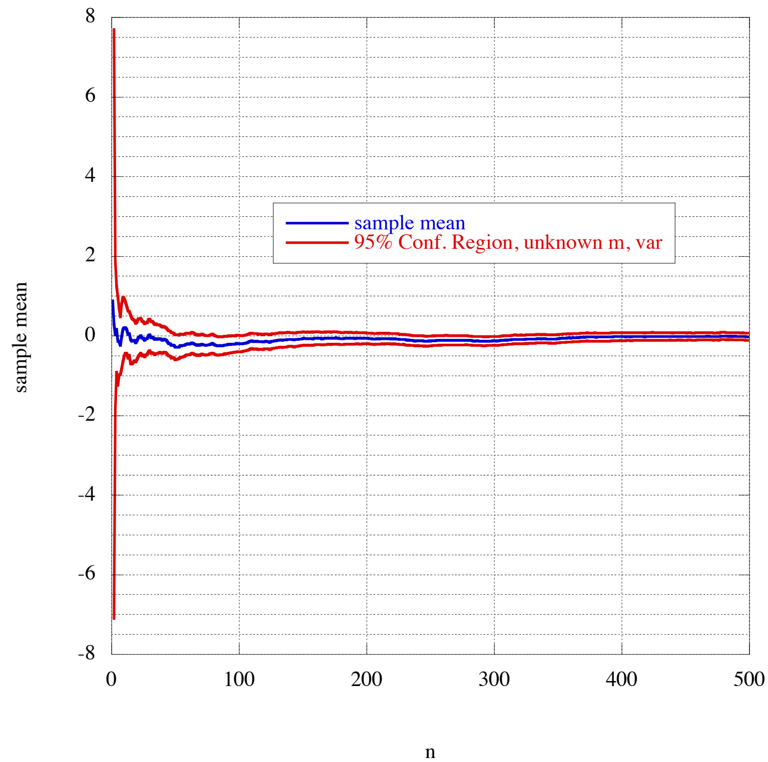

Figure 5: The 95% confidence region for m from the observed realizations (m and σ unknown).

This concept is illustrated in Fig. 5, which shows the confidence region for the mean as a function of n when both the mean and variance are unknown. Note that this looks like a very small MOE as compared to that in Fig. 4, but that is due to the range of the y-axis – *i.e.,* the MOE is large for small sample sizes. Figs. 6-7 compare the confidence regions under the assumptions of known and unknown variance.

Note that the confidence region with unknown variance is always larger than that for known variance. This is beacuse

$$z_{\alpha/2} \leq t_{\alpha/2}(n-1) \tag{28}$$

which can also be viewed as a statement that the tails of the T pdf are heavier (*i.e.,* have more mass) than those of the Gaussian pdf. However, the numerical results shown suggest that for large n, $t_{\alpha/2}(n-1) \approx z_{\alpha/2}$ and this is, in fact, true – *i.e.,* for large values of n, the pdf $T_{n-1}(u)$ is well approximated by $\mathcal{N}(\cdot;0;1)$. Thus, even though in most practical applications we do not know the variance, statisticians still use $z_{\alpha/2}$ for large sample sizes. For example, for $n = 100$, $t_{0.025}(99) = 1.98$ as compared to $z_{0.025} = 1.96$.

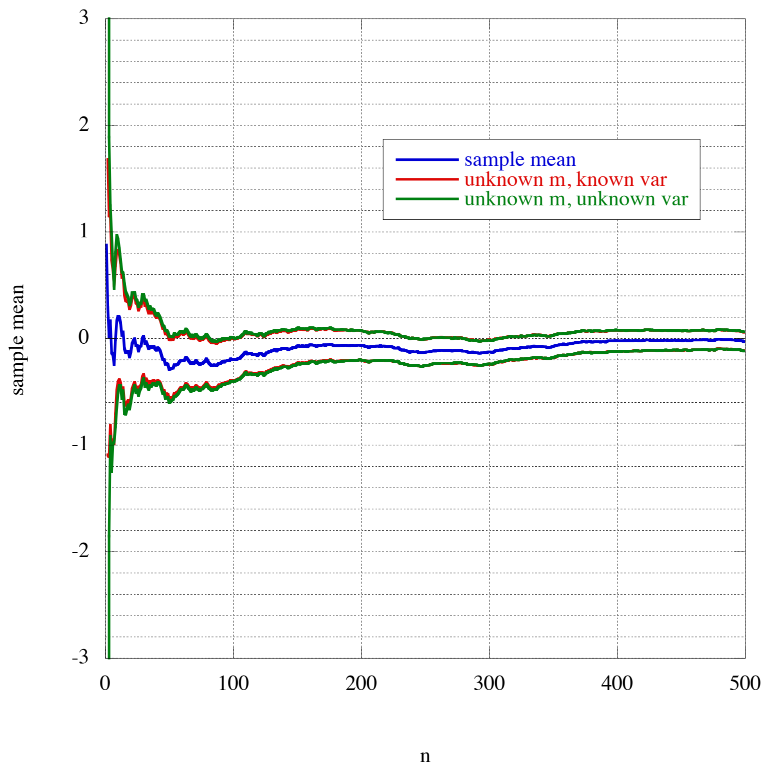

Figure 6: Comparison of 95% confidence region for m obtained when the variance is known and when the variance is unknown.

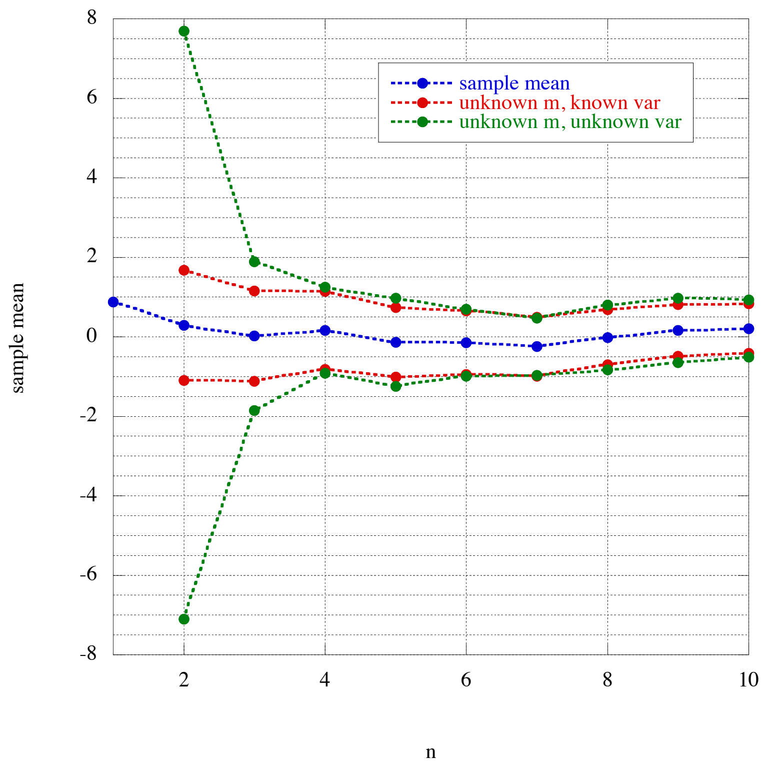

Figure 7: Same comparison as shown in Fig. 6, but with focus on small n.

2.3 Confidence Region: Unknown Mean, Unknown Variance, Unknown Distribution

There is still one potentially unrealistic assumption in the last section: the underlying sampling distribution is assumed to be Gaussian. This assumption can often be justified via the Central Limit Theorem. However, you will see this assumed frequently in the literature whether or not it is justified. If the underlying sampling distribution $f_{Z(u)}(z)$ is unknown, one approach is the "method of batch means". In this method, we form non-overlapping averages from the original data and then process this reduced data set with the Gaussian assumption.

As an example, consider the exponential random samples that were generated in the homework. We know that the sample distribution was exponential with mean 3 ($\lambda = 1/3$) and variance 9. Suppose we did not know this information, we could use the method of batch means to find a confidence interval on the mean. For example, we can use batch size 10 to form 50 batches from 500 original samples. Since each of these is the average of 10 iid random variables, the CLT implies that they are approximately Gaussian. Therefore, we can apply the above approach to this new set of 50 data points. Treating these 50 points as coming from a Gaussian sample with unknown mean and variance yields the confidence region plot in Fig. 8.

As a comparison, let's consider the case where the original samples are (incorrectly) assumed to be Gaussian with unknown mean and variance. Note that the method of batch means will not affect the sample mean (*i.e.,* it is the average of averages), only the MOE. If we assumed that the original 500 samples were Gaussian, we would obtain

$$m = 3.03 \pm 0.25 \tag{29}$$

This is to be compared to the method used, where, based on the 500 samples, we obtained

$$m = 3.03 \pm 0.28 \tag{30}$$

So, by being cautious, we suffered a slightly larger MOE. The incorrect assumption of a Gaussian sample can be more dangerous for small sample sizes.

2.4 Bernoulli Sampling Distribution (Estimating Probabilities)

Consider the case where $Z_i(u)$ is drawn from a Bernoulli sampling distribution – *i.e.,* $Z_1(u)$, $Z_2(u)$, $Z_3(u)$, $\ldots Z_n(u)$ are i.i.d. and each is 1 with probability p and 0 with probability $q = 1-p$. In this case, $m = p$ and $\sigma^2 = pq$ and the sample mean is a binomial random variable, scaled by $1/n$, that has mean p and variance pq/n. For large values of np (*i.e.,* when then number of successes observed is large), the sample mean can therefore be approximated as Gaussian. If we use the confidence region for a Gaussian sample mean with known variance, we obtain

$$p = \hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{pq}{n}} \tag{31}$$

where $\hat{p}_n = k_n/n$ with k_n being the number of successes observed in the sample of size n. This is not directly applicable because the variance $\sigma^2 = pq$ is not known. However, since the variance is only a function of the mean (*i.e.,* p is the mean and the variance is $p(1-p)$), we may adopt the above confidence interval. Specifically, we are typically interested in cases where the $\hat{p}_n$ is a

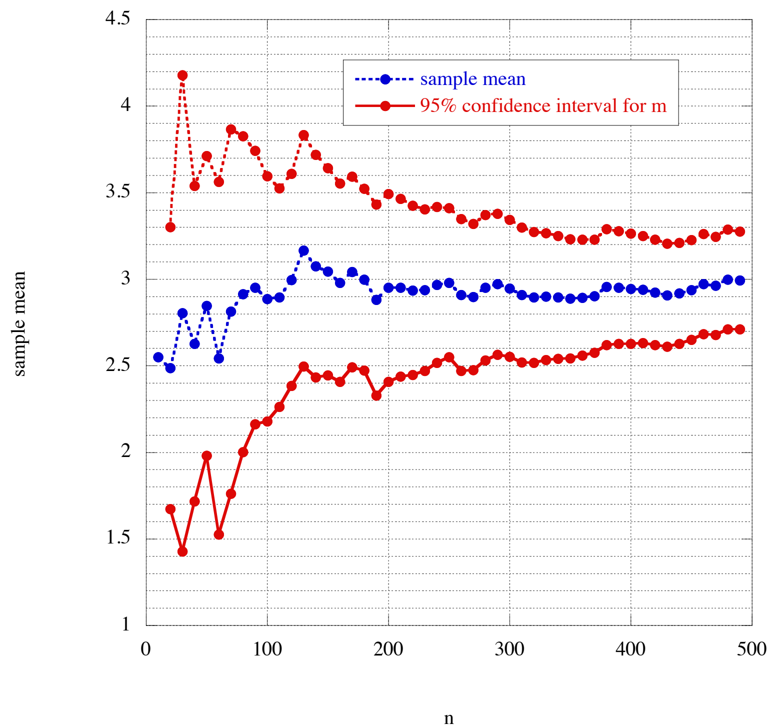

Figure 8: Using batches of size 10 to find the confidence region using Gaussian sample approaches.

reasonably accurate estimate of p. In this case, it is reasonable to replace pq in the MOE expression by the corresponding estimates yielding

$$\boxed{p = \hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_n\hat{q}_n}{n}} = \hat{p}_n\left(1 \pm z_{\alpha/2}\sqrt{\frac{\hat{q}_n}{k_n}}\right)} \tag{32}$$

where we have also shown the MOE as a fraction of the the estimate. This is another nice aspect of the Bernoulli sample – *i.e.,* the MOE can be expressed as a relative error.

2.4.1 Example: Estimating a pdf/pmf

As a specific example, the his histogram generated in homework to approximate the uniform pdf on $[0, 1]$ is actually estimating the probability of being in each bin from a sample. This histogram is shown in Fig. 9 for reference. For example, the first bin contains 20 of the 500 total samples, so the probability estimate is $\hat{p}_{500} = 20/500 = 0.04$. There were 20 bins in the histogram that I generated for the solution, so the actual values of p is $1/20 = 0.05$. Using the above expression we have a 90% confidence region of

$$p = 0.04 \pm (1.645)\sqrt{\frac{(0.04)(0.96)}{500}} = 0.04 \pm 0.014 = 0.4(1 \pm 0.36) \tag{33}$$

so that, with 20 observed samples in the bin, we have a relative MOE of 36%. Notice that over all the bins in this histogram, we have the smallest number of samples to be 13 and the largest to be 34. For $k_{500} = 13$, we have $\hat{p}_{500} = 0.026$ with a relative MOE of 45%. For 34 samples in the bin, we have $\hat{p}_{500} = 0.068$ with a 27% relative MOE (both with 90% confidence). Note that, for the case of 34 samples in a bin, the actual value of $p = 0.05$ is just inside the 90% confidence region, while for 13 samples in a bin, 0.05 is not in the 90% confidence region. Of the 20 bins, all but one have the true value of p within the 90% confidence region.

2.4.2 Example: Estimating $p \approx 0.5$

Consider the case of $p \approx 0.5$ which is typically the case in opinion polling – *e.g.,* political polling in a close race. Here, $pq \lesssim 1/4$ – *i.e.,* $pq < 1/4$ always holds, but $pq \approx 1/4$ for $p \approx 0.5$. It follows that for $p \approx 0.5$ we have

$$p = \hat{p}_n \pm \frac{z_{\alpha/2}}{2\sqrt{n}} \qquad (p \approx 0.5) \tag{34}$$

For example, suppose a political poll is taken with 1000 people polled and the race is 47% to 53%. Then, the MOE for 95% confidence is

$$\frac{1.96}{2\sqrt{1000}} = 0.030 \tag{35}$$

It follows that to have an accuracy of ± 3 percentage points[3] (with 95% confidence) you need to poll about 1000 people.

[3] Note that this is not a relative margin of error – it is an absolute measure of error. In relative terms, 3 percentage points is about 6% of a 0.5 probability.

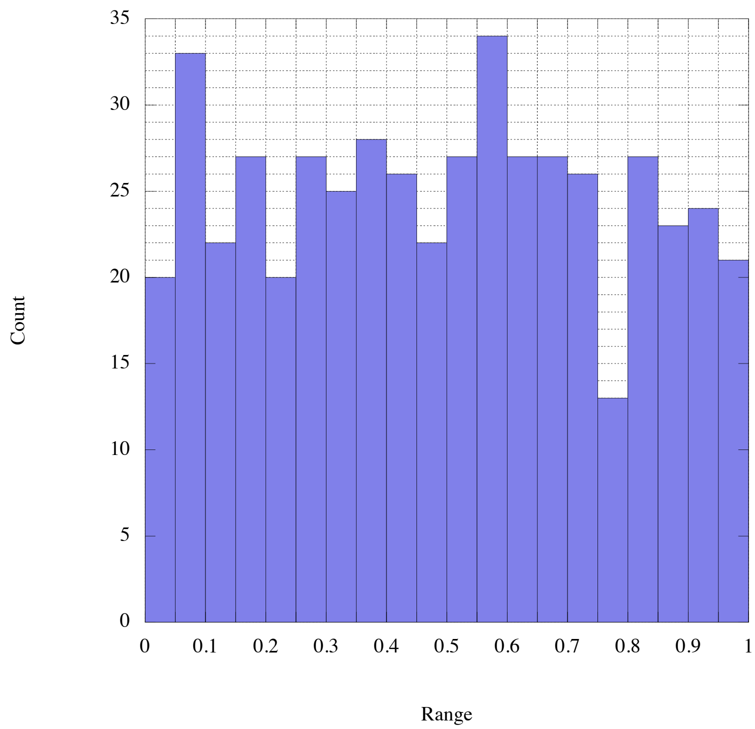

Figure 9: The histogram obtained from 500 samples uniformly distributed on $[0, 1]$.

2.4.3 Example: Estimating $p \ll 1$

Now, consider the case when p is very small. This occurs, for example, in the simulation of communication systems that are designed to have very low error probabilities. Here, we expect p to take values such as 10^{-6} or 10^{-10}. In a simulation, we run Bernoulli trials by simulating a complex system and then we see if an error is made. As a concrete example, suppose that you run 1 million trials and you observe 2 errors,[4] would you feel confident in estimating the $p = 2 \times 10^{-6}$? You should not have high confidence in this estimate because you have not observed many of the rare events (errors). To take this to the extreme, if you observed 0 errors in 1000 trials, would you be confident that your system has an error probability of 0? In fact, no error in 1000 trials in not unlikely if $p = 10^{-9}$.

So, the idea is that when estimating p, you need to observe a large number of rare events to have high confidence. Returning to (32) and noting that for p very small, $\hat{q}_n \approx 1$, we have

$$p = \hat{p}_n \left(1 \pm \frac{z_{\alpha/2}}{\sqrt{k_n}}\right) = \hat{p}_n(1 \pm \epsilon) \qquad (p \ll 1) \tag{36}$$

where ϵ is the relative MOE. It follows that to achieve a relative MOE of ϵ, we must have

$$k_n \geq \left(\frac{z_{\alpha/2}}{\epsilon}\right)^2 \tag{37}$$

which shows explicitly the need to observe a minimum number of rare events to have a small relative error with high confidence. For example, to have a relative error of $\epsilon = 0.2$ with 95% confidence, we need

$$k_n \geq \left(\frac{1.96}{0.2}\right)^2 = 96 \tag{38}$$

In fact, this is a rule of thumb in simulating communication links: collect 100 errors to achieve about 20% relative estimation error with 95% confidence.

It should be noted that all of the above analysis for the Bernoulli sampling distribution assumes that the Gaussian approximation is valid. If this is not the case, confidence regions can still be computed using the exact binomial distribution or the Poisson approximation. For example, suppose you have run $n = 10^6$ trials and observed no errors. It is reasonable to conclude with high confidence, for example, that $p < 10^{-2}$. This can be formalized using the Poisson distribution. For example, although the details are skipped here, if zero errors are observed then with 95% confidence one can conclude that $p \in (0, 3.7/n)$. So, in the example of no observed errors in a million trials, the actual error probability is in $(0, 3.7 \times 10^{-6})$ with confidence 95%.

[4] In this case errors are "successes" in the Bernoulli random variable – *i.e.,* we are trying to estimate p.