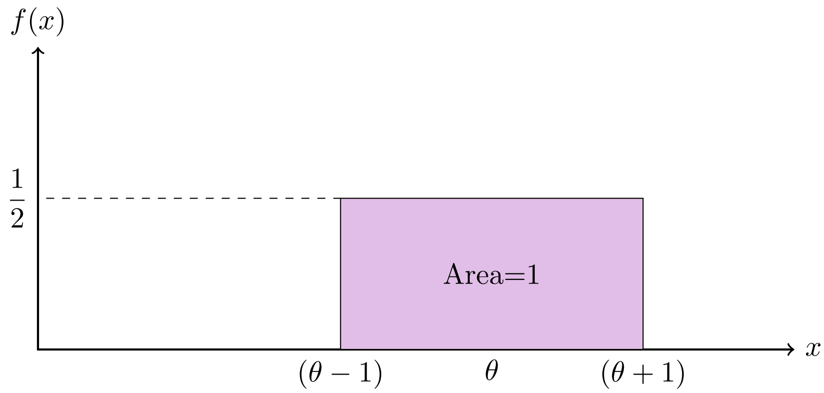
Contents

1 CI Formula Evolution

1.1 Example 1: Uniform Random Variable with 100% CI

1.1.1 Initial Setup

Random Variable x having uniform probability density function $f(x)$.



This simply means, the converge probability,

$$Pr(\theta - 1 \leq x \leq \theta + 1) = 1 \tag{1}$$

That is, the probability that x could be within $\theta \pm 1$ is 1.

1.1.2 CI construction using Pivotal Quantity

In equation 1, by adding $-\theta$ to the inequalities, we get,

$$\begin{aligned} Pr(-\theta + \theta - 1 \leq -\theta + x \leq -\theta + \theta + 1) = 1 \\ Pr(-1 \leq x - \theta \leq 1) = 1 \end{aligned} \tag{2}$$

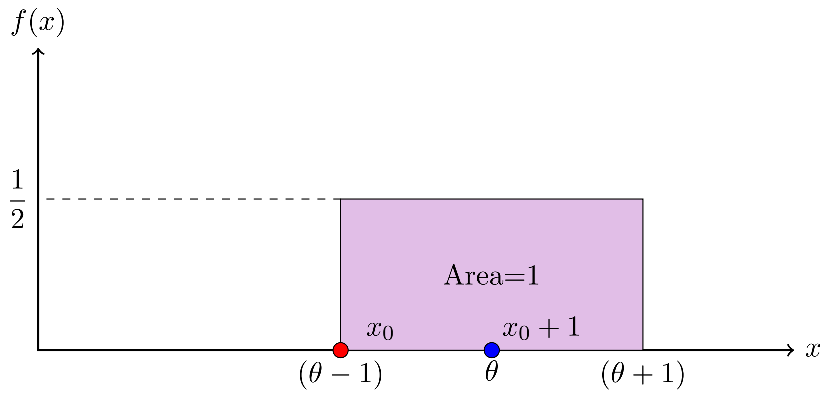
Multiplying by -1, and adding x

$$\begin{aligned} Pr(1 \geq -x + \theta \geq -1) = 1 \\ Pr(x + 1 \geq \theta \geq x - 1) = 1 \\ Pr(x - 1 \leq \theta \leq x + 1) = 1 \end{aligned} \tag{3}$$

Thus while x could take value only between $\theta \pm 1$ for given probability, Equation 3 states, θ could also be only within $x \pm 1$ for same probability
.

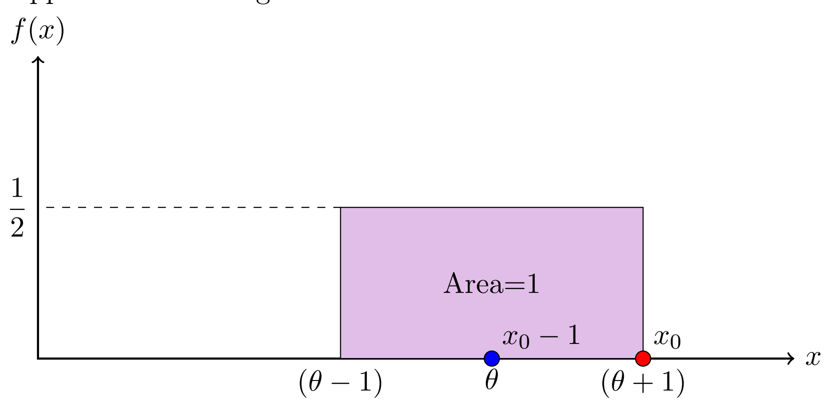
1.1.3 Intuitive Proof

Suppose x takes a left extreme value as below within bounds $\theta \pm 1$.



Then, we could already see, θ is at $x_0 + 1$ still respecting the bounds $x \pm 1$.

Suppose x takes a right extreme value as below within bounds $\theta \pm 1$.



Then, we could already see, θ is at $x_0 - 1$ still respecting the bounds $x \pm 1$.

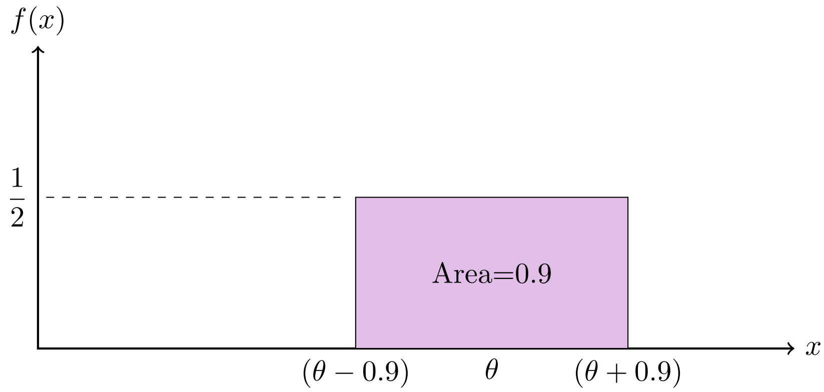
Thus, while x could be only within $\theta \pm 1$, it is also valid to say, θ could vary only within $x \pm 1$.

$$Pr(\theta - 1 \leq x \leq \theta + 1) = Pr(x - 1 \leq \theta \leq x + 1) = 1 \tag{4}$$

1.2 Example 1b: Uniform Random Variable with 90% CI

1.2.1 Initial Setup

Random Variable x having uniform probability density function $f(x)$.



This simply means, the converge probability,

$$Pr(\theta - 0.9 \leq x \leq \theta + 0.9) = 0.9 \tag{5}$$

That is, the probability that x could be within $\theta \pm 0.9$ is 0.9 or 90%.

1.2.2 CI construction using Pivotal Quantity

In equation 5, by adding $-\theta$ to the inequalities, we get,

$$\begin{aligned} Pr(-\theta + \theta - 0.9 \leq -\theta + x \leq -\theta + \theta + 0.9) = 0.9 \\ Pr(-0.9 \leq x - \theta \leq 0.9) = 0.9 \end{aligned} \tag{6}$$

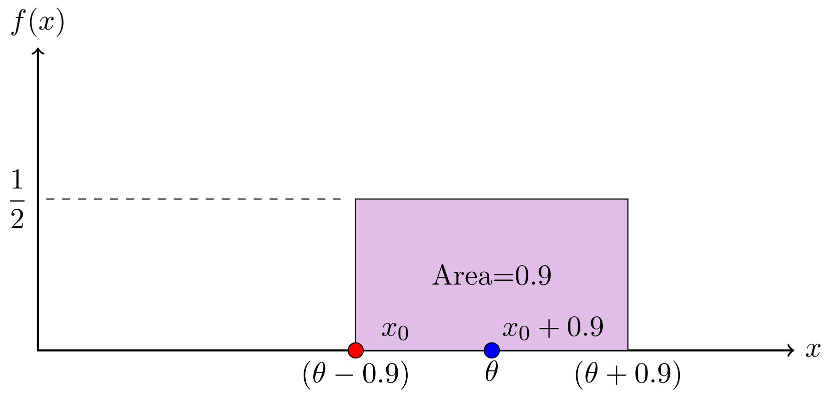
Multiplying by -1, and adding x

$$\begin{aligned} Pr(0.9 \geq -x + \theta \geq -0.9) = 0.9 \\ Pr(x + 0.9 \geq \theta \geq x - 0.9) = 0.9 \\ Pr(x - 0.9 \leq \theta \leq x + 0.9) = 0.9 \end{aligned} \tag{7}$$

Thus, while x could take value only between $\theta \pm 0.9$ for given probability 0.9, above equation states, θ could also be only within $x \pm 0.9$ for same probability.

1.2.3 Intuitive Proof

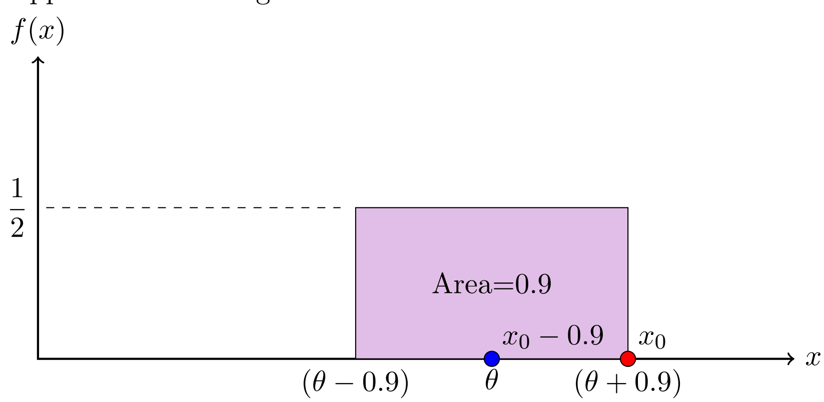
Suppose x takes a left extreme value as below within bounds $\theta \pm 0.9$.



Then, we could already see, θ is at $x_0 + 0.9$ still respecting the bounds $x \pm 0.9$.

Simply put, when x is at $x_0 = \theta - 0.9$, then it automatically implies, $\theta = x_0 + 0.9$

Suppose x takes a right extreme value as below within bounds $\theta \pm 1$.



Then, we could already see, θ is at $x_0 - 0.9$ still respecting the bounds $x \pm 0.9$.

Thus, while x could be only within $\theta \pm 0.9$, it is also valid to say, θ could vary only within $x \pm 0.9$.

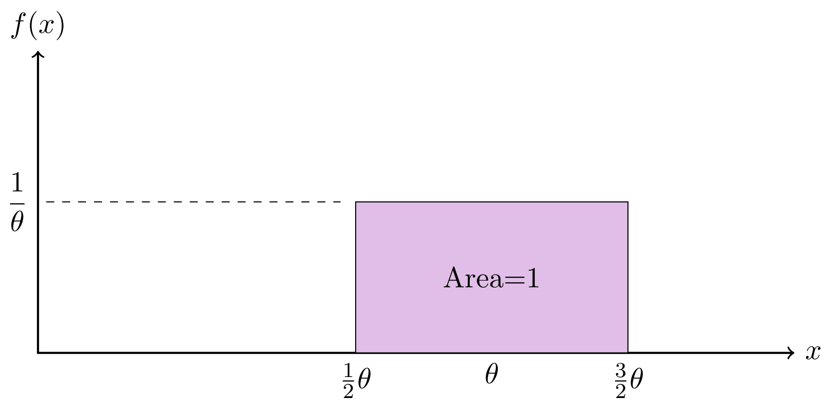
$$Pr(\theta - 0.9 \leq x \leq \theta + 0.9) = Pr(x - 0.9 \leq \theta \leq x + 0.9) = 0.9 \qquad (8)$$

1.3 Example 2: Uniform Random Variable with 100% CI

1.3.1 Initial Setup

Random Variable x having uniform probability density function

$$f(x) = \frac{1}{\theta} \quad \text{for} \quad \frac{1}{2}\theta \le x \le \frac{3}{2}\theta \tag{9}$$



This simply means, the converge probability,

$$Pr\Big(\frac{1}{2}\theta \le x \le \frac{3}{2}\theta\Big) = 1 \tag{10}$$

That is, the probability that x could be within $\theta \pm \frac{1}{2}\theta$ is 1

1.3.2 CI construction using Pivotal Quantity

Multiplying by 2 in the inequalities,

$$Pr\Big(\theta \le 2x \le 3\theta\Big) = 1$$

Dividing by θ,..

$$Pr\Big(1 \le \frac{2x}{\theta} \le 3\Big) = 1$$

Dividing by x and inversing the inequalities, and again multiplying by 2..

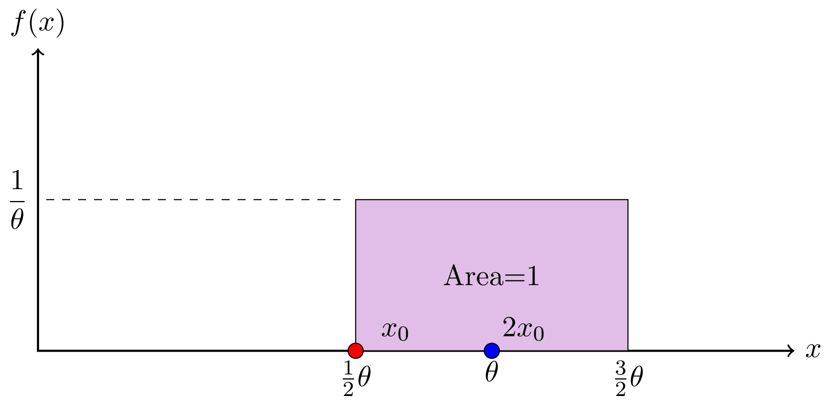
$$Pr\Big(\frac{1}{x} \leq \frac{2}{\theta} \leq \frac{3}{x}\Big) = 1$$
$$Pr\Big(x \geq \frac{\theta}{2} \geq \frac{x}{3}\Big) = 1$$
$$Pr\Big(2x \geq \theta \geq \frac{2x}{3}\Big) = 1$$

which is same as

$$Pr\Big(\frac{2x}{3} \leq \theta \leq 2x\Big) = 1 \tag{11}$$

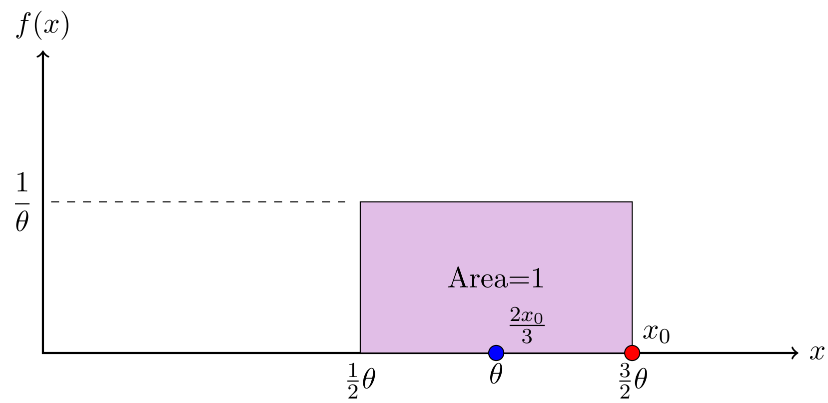
1.3.3 Intuitive Proof

Suppose x takes a left extreme value as below within bounds $\theta \pm \dfrac{\theta}{2}$.



When x is at $x_0 = \dfrac{\theta}{2}$, then $\theta = 2x_0$

Suppose x takes a right extreme value as below within bounds $\theta \pm \dfrac{\theta}{2}$.



When x is at $x_0 = \dfrac{3\theta}{2}$, then $\theta = \dfrac{2x_0}{3}$

Thus, while x could be only within $\theta \pm \dfrac{\theta}{2}$, it is also valid to say, θ could vary only within $\left(\dfrac{2x}{3}, 2x\right)$.

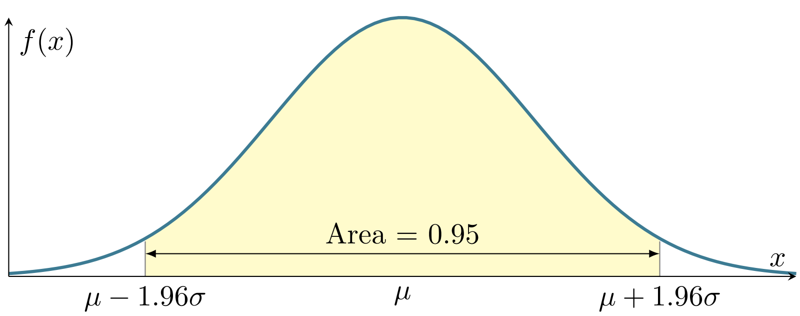
$$Pr\left(\theta - \frac{\theta}{2} \leq x \leq \theta + \frac{\theta}{2}\right) = Pr\left(\frac{2x}{3} \leq \theta \leq 2x\right) = 1 \tag{12}$$

1.4 Example 3: Normal Distribution with 95% CI

1.4.1 Initial Setup

Random Variable x having uniform probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \tag{13}$$



This simply means, the converge probability,

$$Pr(\mu - 1.96\sigma \leq x \leq \mu + 1.96\sigma) = 0.95 \tag{14}$$

That is, the probability that x could be within $\mu \pm 1.96\sigma$ is 0.95 or 95%

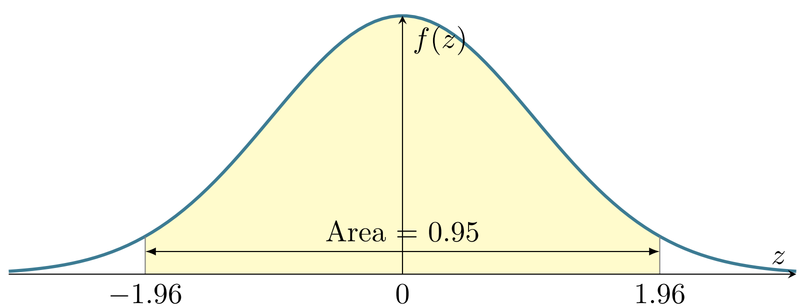
1.4.2 Why 1.96?

Let us standardize the distribution to standard normal distribution, $Z = \dfrac{X - \mu}{\sigma}$.

When $X = \mu + 1.96\sigma$, $Z = \dfrac{\mu + 1.96\sigma - \mu}{\sigma} = 1.96$

When $X = \mu - 1.96\sigma$, $Z = \dfrac{\mu - 1.96\sigma - \mu}{\sigma} = -1.96$

The transformed distribution would look like below.



If we look at the Z table for $Z = 1.96$, we will find value as 0.975

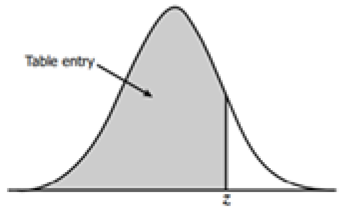


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

If we look at the Z table for $Z = -1.96$, we will find value as 0.025

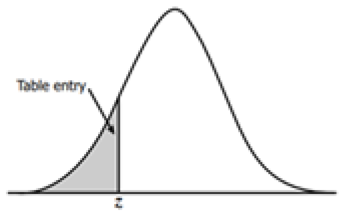


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233

The area between 0.975 and 0.025 is $0.975 - 0.025 = 0.95$ or 95%. Thus, the value 1.96 was born. It depends on the area we are interested. Here, we were interested in 95% area, so we get $Z = \pm 1.96$

Note The Z table might be left tailed as we just saw or also sometimes right tailed due to symmetrical nature of the curve. This realization is important because when we generalize CI, we will often take right tailed. I used the conventional left tailed table above just to state this explicitly as undoubting readers may miss this point.

1.4.3 CI construction using Pivotal Quantity

From equation 14, adding $-\mu$ on both sides of inequalities, we get,

$$Pr(-\mu + \mu - 1.96\sigma \le x - \mu \le -\mu + \mu + 1.96\sigma) = 0.95$$
$$Pr(-1.96\sigma \le x - \mu \le 1.96\sigma) = 0.95$$

And then adding $-x$

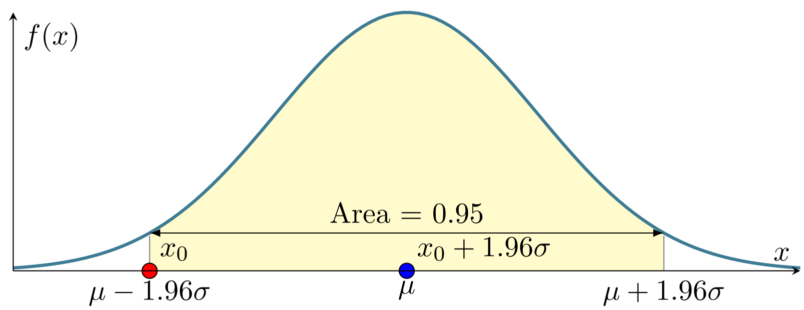
$$Pr(-x - 1.96\sigma \leq -x + x - \mu \leq -x + 1.96\sigma) = 0.95$$
$$Pr(-x - 1.96\sigma \leq -\mu \leq -x + 1.96\sigma) = 0.95$$

Multiplying by -1

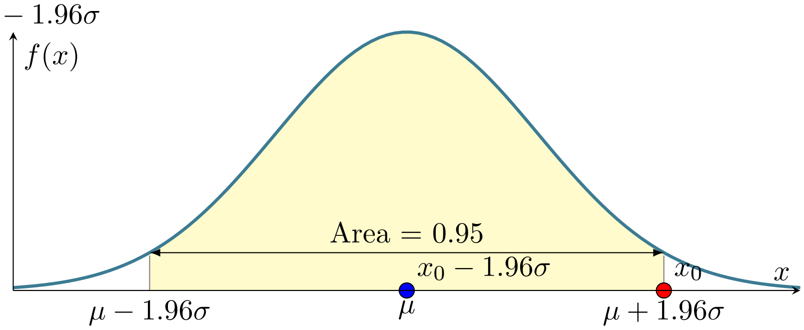
$$Pr(x + 1.96\sigma \geq \mu \geq x - 1.96\sigma) = 0.95$$
$$Pr(x - 1.96\sigma \leq \mu \leq x + 1.96\sigma) = 0.95 \tag{15}$$

1.4.4 Intuitive Proof

Suppose x takes left extreme value within bounds $\mu \pm 1.96\sigma$. That is, $x_0 = \mu - 1.96\sigma$ Then, $\mu = x_0 + 1.96\sigma$



Similarly, when $x_0 = \mu + 1.96\sigma$ then directly we could derive, $\mu = x_0 - 1.96\sigma$



So, as x_0 varies from $\mu - 1.96\sigma$ to $\mu + 1.96\sigma$, implicitly, μ varies from $x_0 + 1.96\sigma$ to $x_0 - 1.96\sigma$. Thus,

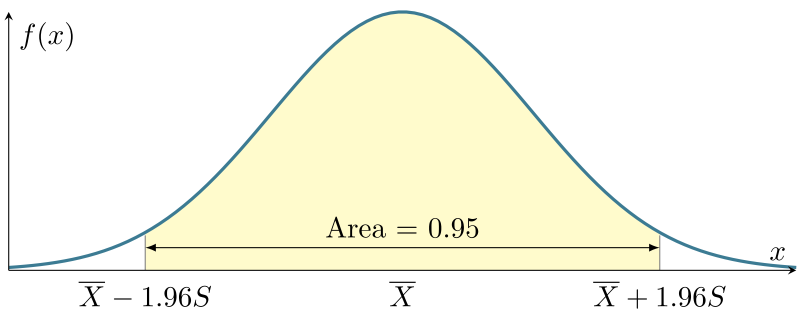
$$Pr(\mu - 1.96\sigma \leq x \leq \mu + 1.96\sigma) = Pr(x - 1.96\sigma \leq \mu \leq x + 1.96\sigma) = 0.95 \tag{16}$$

2 CI for Sampling Distribution

2.1 95% CI as a Corollary

We already have seen, any sampling distribution for sample proportions or sample means, will approach normal distribution, with $\overline{X} \rightarrow \mu$ and $S \rightarrow \dfrac{\sigma}{\sqrt{n}}$, where μ, σ, n are population mean, population standard deviation, and sample size respectively, when respective conditions[1] are met as per Central Limit theorem (CLT). Note each x is a sample mean.

We thus have a normal distribution like below representing sampling distribution.



Then, using equation 16, we have,

$$Pr(\overline{X} - 1.96S \leq x \leq \overline{X} + 1.96S) = Pr(x - 1.96S \leq \overline{X} \leq x + 1.96S) = 0.95$$
$$Pr(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq x \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}) = Pr(x - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95 \tag{17}$$

Thus, 95% CI for a Sampling distribution would be $\left(x \pm 1.96\frac{\sigma}{\sqrt{n}}\right)$

[1] $np \geq 10$ and $nq \geq 10$ for sample proportions, $n \geq 30$ for sample means

2.2 Generalized CI

As hinted in 1.4.2, we will use a right tailed Z table for generalization. We already saw, at Z = -1.96, the area spanned would be 0.025. This could be written as

$$z_{0.025} = -1.96$$

Substituting in 17, we get,

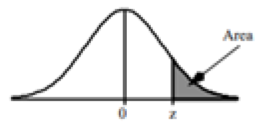
$$Pr(x - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$
$$Pr(x + z_{0.025}\frac{\sigma}{\sqrt{n}} \leq \mu \leq x - z_{0.025}\frac{\sigma}{\sqrt{n}}) = 0.95$$

This is kind of counter intuitive. Additive term comes on the LHS. Though one would later discover, $z_{0.025}$ is negative, it could be better if this is not raising any confusion in first place. This is why we use right tailed Z table.

In case of right tailed Z table as below, note, at Z = 1.96, the area spanned is 0.025. Thus we could write it as

$$z_{0.025} = 1.96$$

Normal Curve Areas
Standard normal probability
in right-hand tail



	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233

Substituting in 17, we get,

$$Pr(x - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$
$$Pr(x - z_{0.025}\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + z_{0.025}\frac{\sigma}{\sqrt{n}}) = 0.95$$

This is good.

Let α be the desired significance level (which we will learn in hypothesis testing). In our case, it is 5% or $\alpha = 0.05$. Thus, $1-\alpha = 0.95$ and $\frac{\alpha}{2} = 0.025$

We could then rewrite above equation as,

$$Pr(x - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha \tag{18}$$

This is the generalized CI equation where $1-\alpha$ is called the **confidence coefficient** and $z_{\frac{\alpha}{2}}$ is called the **critical value**

Note Confidence Interval CI indicates not an interval, where population mean is contained 95% of time, but, if one continues to take many such samples and CI for each sample, then 95% of those CIs would contain population mean. We do not know what those CIs are unless we know the population mean and take many such sample sets and their CIs. Once we have taken enough such sample sets (each sample set of size n) calculating CI each time, we could expect that 95% of those CIs have population mean.

2.3 When σ is known

In 18, we have population standard deviation σ in both end points of the inequalities. Often population parameters are not known in reality. So we have two cases: One when you are lucky enough to known σ and another, you do not know. When you do know, still there are some more parts in play. For example, the more the samples are taken from population, the closer the resulting sampling distribution is to Normal (or Normal approximation is becoming better), so when do you say, sample size n is good enough? This depends on various conditions.

1. If we sample from population whose distribution is itself normal, then even small sample size $\boldsymbol{n \geq 5}$ would suffice because our sampling distribution easily approximates to Normal. Our current CI equation holds good.

$$Pr(x - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

2. If we sample from population whose distribution is not normal but symmetric, unimodal and of the continuous type, then as per Central limit theorem (CLT), sample size $\boldsymbol{n \geq 30}$ should be adequate generally as this would result in sampling distribution becoming almost normal so our equation could still be approximately good. That is,

$$Pr(x - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq x + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) \approx 1 - \alpha \tag{19}$$

3. If distribution is non normal and also highly skewed, even above approximation would not work. In that case, it would be safer to use certain nonparametric methods for finding a CI for the median of the distribution.

2.4 When σ is not known

This is often the case in reality. In this case, depending on certain conditions like above, we could use student's t distribution[2]. The t distribution looks like normal, except the tails are bigger, and also depends on degrees of freedom (which usually is $n-1$). The proof is exhaustive, so we will take at face value for now (and prove in future if time permits)

1. If we sample from population whose distribution is itself normal, and if sample size $\boldsymbol{n} \leq \mathbf{30}$, then our CI equation would be,

$$Pr(x - t_{\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} \leq \mu \leq x + t_{\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}}) = 1 - \alpha \tag{20}$$

   where $t_{\frac{\alpha}{2},(n-1)}$ is the t value for probability area $\frac{\alpha}{2}$, for degrees of freedom $(n-1)$ from corresponding right tailed t table.

2. If we sample from population whose distribution is itself normal, and if sample size $\boldsymbol{n} > \mathbf{30}$, then our t distribution would already be almost equal to normal (and resulting sampling distribution would be normal) so we could use as below,

$$Pr(x - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq x + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}) = 1 - \alpha \tag{21}$$

3. If we sample from population whose distribution is not normal but symmetric, unimodal and of the continuous type, and sample size $\boldsymbol{n} \leq \mathbf{30}$, we get approximate CI as below.

$$Pr(x - t_{\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}} \leq \mu \leq x + t_{\frac{\alpha}{2},(n-1)}\frac{s}{\sqrt{n}}) \approx 1 - \alpha \tag{22}$$

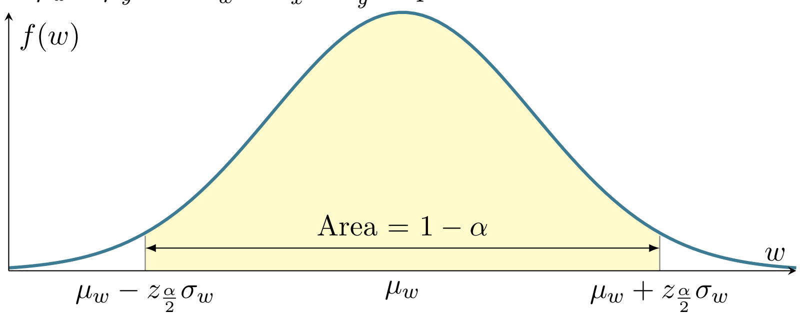
4. If distribution is non normal and also highly skewed, even above approximation would not work. In that case, it would be safer to use certain nonparametric methods for finding a CI for the median of the distribution.

2.5 CI for difference between two means

This section is heavily inspired by Robert et al. [1], and I have tried to articulate in my style to my understanding. Suppose that we are interested in comparing two approximately normal sampling distributions described by random variables $\overline{X} = N(\mu_{\overline{x}}, \sigma^2_{\overline{x}})$ and $\overline{Y} = N(\mu_{\overline{y}}, \sigma^2_{\overline{y}})$, created

[2] http://pages.wustl.edu/montgomery/articles/2757

from population distributions described by random variables $X(\mu_x, \sigma_x^2)$ and $Y(\mu_y, \sigma_y^2)$. Note that $\overline{X}$ represents collection of sample means from sampled sets sampled from X and similarly for $\overline{Y}$. Since both $\overline{X}$ and $\overline{Y}$ are normally distributed, and assuming both are independent to each other, the distribution $W = \overline{X} - \overline{Y}$ would be again a normal distribution $W(\mu_w, \sigma_w^2)$, where $\mu_w = \mu_{\overline{x}} - \mu_{\overline{y}}$ and $\sigma_w^2 = \sigma_{\overline{x}}^2 + \sigma_{\overline{y}}^2$ as proved in **??**



Since W is a normal distribution now, we have the confidence interval as follows directly following equation 14

$$\begin{aligned}
Pr(\mu - z_{\frac{\alpha}{2}}\sigma \leq x \leq \mu + z_{\frac{\alpha}{2}}\sigma) &= 1 - \alpha \\
Pr(\mu_w - z_{\frac{\alpha}{2}}\sigma_w \leq W \leq \mu_w + z_{\frac{\alpha}{2}}\sigma_w) &= 1 - \alpha \\
Pr(-z_{\frac{\alpha}{2}}\sigma_w \leq W - \mu_w \leq z_{\frac{\alpha}{2}}\sigma_w) &= 1 - \alpha \\
Pr(-z_{\frac{\alpha}{2}} \leq \frac{W - \mu_w}{\sigma_w} \leq z_{\frac{\alpha}{2}}) &= 1 - \alpha
\end{aligned}$$

$$\begin{aligned}
Pr(-z_{\frac{\alpha}{2}} \leq \frac{W - \mu_w}{\sigma_w} \leq z_{\frac{\alpha}{2}}) &= 1 - \alpha \\
Pr\Big(-z_{\frac{\alpha}{2}} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_{\overline{x}} - \mu_{\overline{y}})}{\sqrt{\sigma_{\overline{x}}^2 + \sigma_{\overline{y}}^2}} \leq z_{\frac{\alpha}{2}}\Big) &= 1 - \alpha \\
Pr\left(-z_{\frac{\alpha}{2}} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_{\overline{x}} - \mu_{\overline{y}})}{\sqrt{\frac{\sigma_{\overline{x}}^2}{n} + \frac{\sigma_{\overline{y}}^2}{m}}} \leq z_{\frac{\alpha}{2}}\right) &= 1 - \alpha \qquad (23)
\end{aligned}$$

where $Z = \frac{W - \mu_w}{\sigma_w}$ would be the "standardized" normal distribution N(0,1), n and m are sample set sizes of $X(\mu_x, \sigma_x)$ and $Y(\mu_y, \sigma_y)$ respectively.

Assuming σ unknown

Most of the times in reality, the population paramters are not known. So when the sample sizes n, m are sufficiently large, we could use sample SDs $(s_{\overline{x}}, s_{\overline{y}})$ in place of (σ_x, σ_y).

$$Pr\left(-z_{\frac{\alpha}{2}} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_{\overline{x}} - \mu_{\overline{y}})}{\sqrt{\frac{s_{\overline{x}}^2}{n} + \frac{s_{\overline{y}}^2}{m}}} \leq z_{\frac{\alpha}{2}} \right) \approx 1 - \alpha$$

And also rewriting, to find CI for $(\mu_{\overline{x}} - \mu_{\overline{y}})$, we get,

$$Pr\Big((\overline{X} - \overline{Y}) - z_{\frac{\alpha}{2}} s_w \leq (\mu_{\overline{x}} - \mu_{\overline{y}}) \leq (\overline{X} - \overline{Y}) + z_{\frac{\alpha}{2}} s_w \Big) \approx 1 - \alpha \tag{24}$$

where, $s_w = \sqrt{\frac{s_{\overline{x}}^2}{n} + \frac{s_{\overline{y}}^2}{m}}$, and n, m are large.

When n, m are small

We would then use student's t distribution as suggested by ***Welch and Aspin***. The proof is currently beyond the scope so we take it at face value.

$$Pr\Big((\overline{X} - \overline{Y}) - t_{(\frac{\alpha}{2}, r)} s_w \leq (\mu_{\overline{x}} - \mu_{\overline{y}}) \leq (\overline{X} - \overline{Y}) + t_{(\frac{\alpha}{2}, r)} s_w \Big) \approx 1 - \alpha \tag{25}$$

where r is degrees of freedom. Since two distributions are involved, calculating r is complicated. It is given as follows:

$$r = \frac{\left(\frac{s_{\overline{x}}^2}{n} + \frac{s_{\overline{y}}^2}{m} \right)^2}{\frac{1}{n-1} \left(\frac{s_{\overline{x}}^2}{n} \right)^2 + \frac{1}{m-1} \left(\frac{s_{\overline{y}}^2}{m} \right)^2} \tag{26}$$

Protection when $\sigma_x = \sigma_y$

Since we do not know σ_x, σ_y, it might be that they are also equal. If they happen to be equal, r could be proven as below.

$$r = (n-1) + (m-1) = n + m - 2$$

The equation 26 protects in the sense that, the r value from that is lesser than above equation, so t value is higher, or t distribution of wider variance assumed, thus being conservative. Some texts simply also take $r = min(n-1, m-1)$ as conservative approach.

2.6 CI for difference between two proportions

Suppose that we are interested in comparing two approximately normal sampling distributions described by random variables $\frac{Y_1}{n_1} = N\Big(p_1, \frac{p_1 q_1}{n_1}\Big)$ and $\frac{Y_2}{n_2} = N\Big(p_2, \frac{p_2 q_2}{n_2}\Big)$, created from population distributions which are Bernoulli distributions.

Note that Y_1 represents the sum of *successes* in a sample set, and thus $\frac{Y_1}{n_1}$ represents sample proportions. For example, for any *kth* sample set of $\frac{Y_1}{n_1}$, we calculate sample proportion statistic, $\frac{Y_{1k}}{n_1} = \frac{1}{n}\sum_{i=1}^{n} Y_{1ki}$, where Y_{1ki} is ith sample in kth sample set of sampling distribution described by $\frac{Y_1}{n_1}$. Similarly for $\frac{Y_2}{n_2}$

We could then rewrite 23 as below

$$Pr\Bigg(- z_{\frac{\alpha}{2}} \leq \frac{(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \leq z_{\frac{\alpha}{2}} \Bigg) = 1 - \alpha$$

In case you are wondering about the parameters inside, say $W = \frac{Y_1}{n_1} - \frac{Y_2}{n_2}$, then

$\mu_w = \mu_{y1/n_1} - \mu_{y2/n_2} = p_1 - p_2$
$\sigma_w^2 = \sigma^2_{y_1/n_1} + \sigma^2_{y_2/n_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \therefore \sigma_w = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

Assuming σ unknown

Most of the times in reality, the population paramters are not known. So when the sample sizes n, m are sufficiently large, we could use sample statistics $(\frac{\hat{p_1}\hat{q_1}}{n1}, \frac{\hat{p_2}\hat{q_2}}{n2})$ in place of $(\frac{p_1 q_1}{n1}, \frac{p_2 q_2}{n2})$. This results in further approximation of our confidence intervals. Thus when a sample is observed, we have statistics

$\hat{p_1} = \frac{y_1}{n_1}, \hat{q_1} = 1 - \frac{y_1}{n_1}, \hat{p_2} = \frac{y_2}{n_2}, \hat{q_2} = 1 - \frac{y_2}{n_2},$

Thus we could rewrite further as,

$$Pr\Bigg(- z_{\frac{\alpha}{2}} \leq \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{\hat{p_1}\hat{q_1}}{n_1} + \frac{\hat{p_2}\hat{q_2}}{n_2}}} \leq z_{\frac{\alpha}{2}} \Bigg) \approx 1 - \alpha \tag{27}$$

When n, m are small

Currently I do not have an answer for this question and could not find online. Raised a ticket(?!) here

References

[1] Robert, Elliot, and Dale. *Probability and Statistical Inference*. Pearson, 9th edition, 2015. URL http://www.nylxs.com/docs/thesis/sources/Probability%20and%20Statistical%20Inference%209ed%20%5B2015%5D.pdf.