Other PDFs

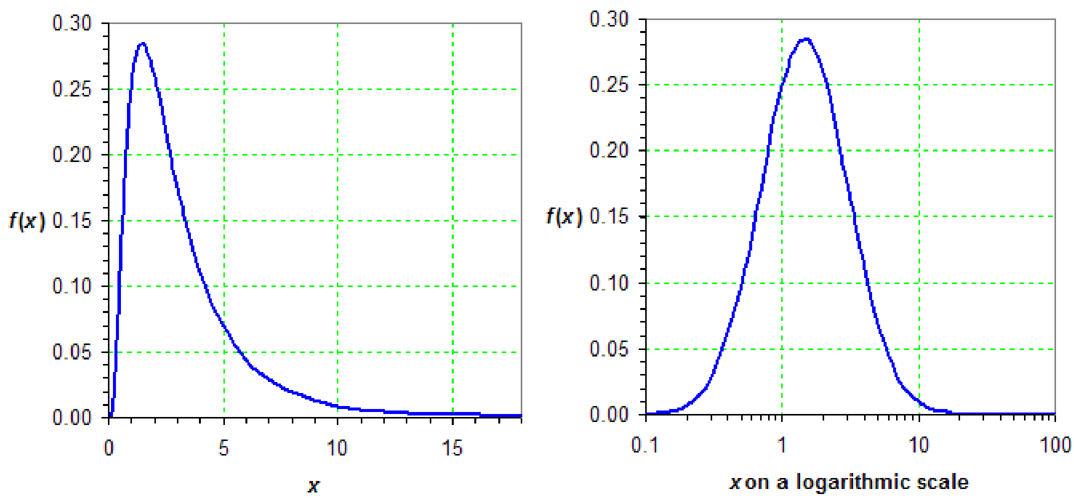
Author: John M. Cimbala, Penn State University
Latest revision: 23 January 2013

Introduction

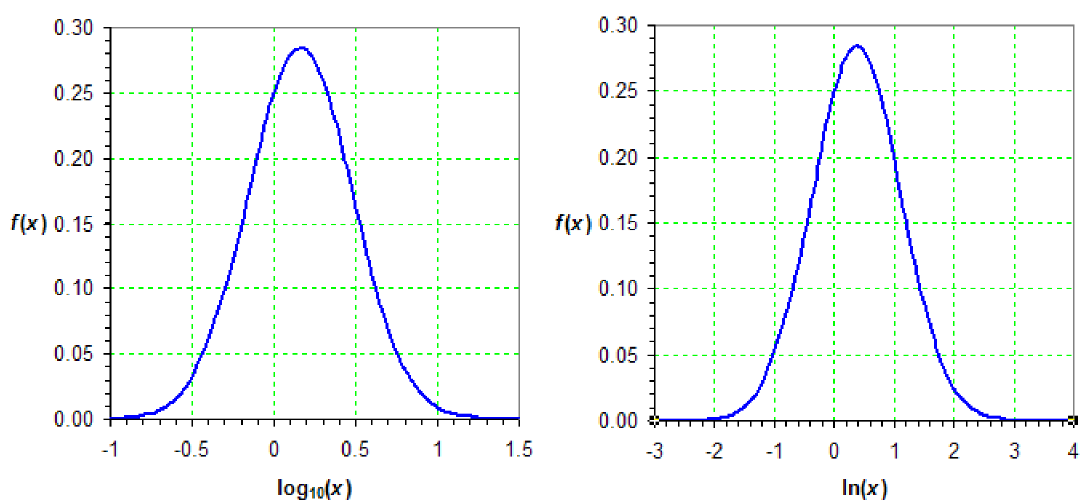
- There are *other* useful standard distributions and PDFs besides the Gaussian PDF. These include the binomial, chi-squared, exponential, gamma, lognormal, Poisson, student's *t*, uniform, and Weibull PDFs.
- We discuss some of these in this learning module, although not in as much detail as for the Gaussian (normal) distribution.

Lognormal PDF

- A ***lognormal PDF*** is defined as ***a PDF that becomes Gaussian when the x-axis is plotted as a log scale***.
  - Lognormal PDFs often appear in air quality measurements, e.g., the size distribution of particles. It is also useful for some life and durability analyses of components and equipment or instruments.
  - When the PDF is plotted as usual (linear *x* scale), it is skewed towards the left (lower values), and has a very long tail to the right (higher values). This is shown on the first plot below.
  - However, when the PDF is plotted with a logarithmic *x* scale, all else being equal, it is no longer skewed, but becomes symmetric. In fact, it's bell shape is identical to that of a Gaussian or normal PDF. This is shown on the second plot below.



  - Another way to plot lognormal PDFs is to first convert the *x* values to $\log_{10}(x)$ or $\ln(x)$, and then plot using a *linear* abscissa scale. Either way, the PDF again looks like a standard Gaussian PDF, as illustrated below.

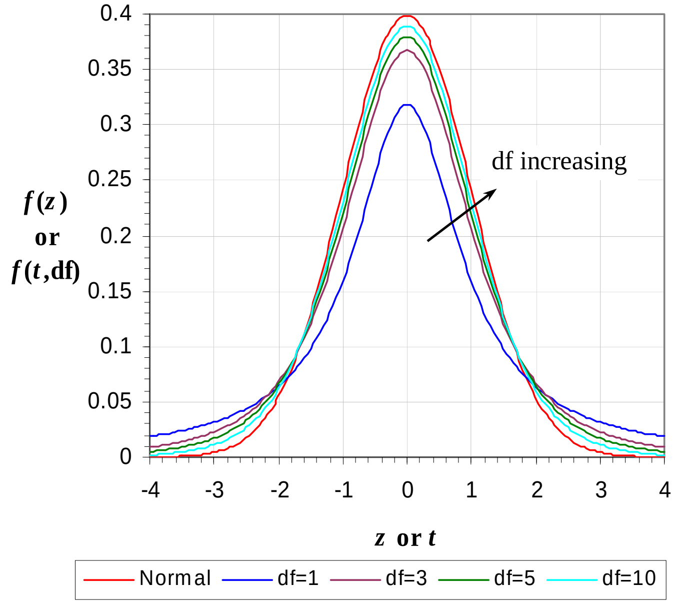


  - To calculate statistics with a lognormal PDF, we substitute either $\log_{10}(x)$ or $\ln(x)$ as our variable instead of *x* itself. For example, if the data are for particle diameter D_p in units of microns (μm), we let our statistics variable be $x = \ln[D_p/(1\ \mu\text{m})]$ instead of D_p itself. All statistics are then based on *x* as usual.

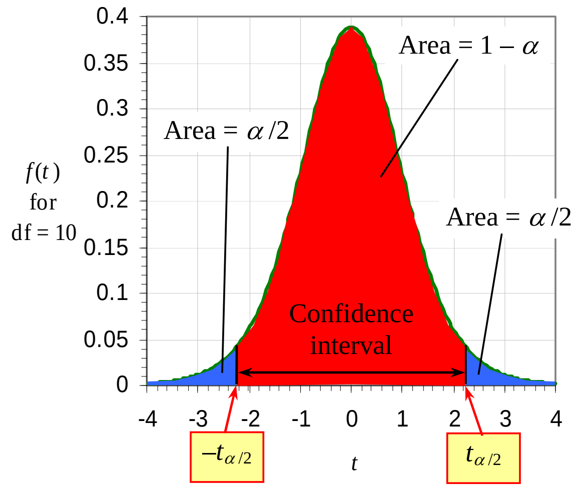
Student's *t* PDF

- A ***student's t PDF*** (also sometimes called simply the ***t PDF***) is ***similar to the Gaussian (normal) PDF, but is used for small sample sizes*** (typically when $n < 30$, where n is the number of data points in the sample).
  - In simple terms, when n is small, the sample mean and sample standard deviation may differ from the population mean and population standard deviation by some unknown amount. So, for a specified confidence level (typically 95%), the student's t PDF is expected to be *wider* than the normal (Gaussian) PDF.
  - Mathematically, the statistic called ***student's t*** is defined as $t = \dfrac{\bar{x} - \mu}{S / \sqrt{n}}$, where $\bar{x}$ is the sample mean, S is the sample standard deviation, and n is the number of data points in the sample. μ is the population mean or expected value, as defined previously, but μ is not necessarily known. (This is the whole point of the student's t analysis in the first place – we want to establish some confidence level in predicting μ.)
  - Statisticians use a parameter called ***degrees of freedom***, with notation df (we do not use italics here so as to not confuse df with *df*, the differential of some variable *f*). Note that some authors use *f* or *v* (lower case italic V) as their notation for degrees of freedom.
  - We define ***degrees of freedom*** as ***the number of measurements minus the minimum number of measurements necessary to estimate a statistic***. For example, it takes only one measurement to estimate the mean value of some quantity x. So, df = n – 1 in this case.
  - Example: If we take n = 10 measurements of pipe diameter, and calculate the sample mean, then df = n – 1 = 10 – 1 = 9. In other words, there are 9 degrees of freedom "left over" after we estimate the mean. [By the way, this is the reason we define standard deviation with n – 1 in the denominator instead of n itself, because we have already "used up" one degree of freedom to calculate the mean; there are thus n – 1 degrees of freedom left over to calculate the standard deviation.]
  - Mathematically, the ***student's t PDF*** is defined as $f(t,\mathrm{df}) = \dfrac{\Gamma\left(\dfrac{\mathrm{df}+1}{2}\right)}{\sqrt{\pi \mathrm{df}} \cdot \Gamma\left(\dfrac{\mathrm{df}}{2}\right)} \left(1 + \dfrac{t^2}{\mathrm{df}}\right)^{-\left(\frac{\mathrm{df}+1}{2}\right)}$,

    where Γ is a standard mathematical function called the ***gamma function***, defined for integers and half-integers as follows:
    - If y is a whole integer (e.g., 12, 17, 25), $\Gamma(y = \text{integer}) = (y-1)! = (y-1)(y-2)...(3)(2)(1)$ You should recall from math class the ***factorial operation***, represented by "!".
    - if y is a half-integer (e.g., 12.5, 16.5), $\Gamma(y = \text{half-integer}) = (y-1)(y-2)...(3/2)(1/2)\sqrt{\pi}$
    - if y = 1/2, $\Gamma(y = 1/2) = \sqrt{\pi}$
    - Examples: Γ(5) = 4! = 4×3×2×1 = **24**. Γ(5.5) = 4.5×3.5×2.5×1.5×0.5× $\sqrt{\pi}$ = **52.3428**, where the answer is given to 6 significant digits.
  - Like the normal PDF, the student's t PDF is ***symmetric*** about zero, except now we plot the PDF with t instead of z as the abscissa.
  - The above equation for the student's t PDF depends on both t and df, so there is actually a whole *family* of curves representing the student's t PDF.
  - On the next page, the student's t PDF $f(t,\mathrm{df})$ is plotted as a function of t for several values of df. On the same plot is shown the normal (Gaussian) PDF, $f(z)$ as a function of z. As you can see, as df increases, the student's t PDF approaches the normal PDF. In fact, as df $\to \infty$, the student's t PDF becomes *identical* to the normal PDF.



- ***Confidence level*** and ***level of significance*** – In order to use the student's *t* PDF, we must first review the concepts of confidence level and level of significance.
  - ***Confidence level c*** is defined as ***the probability that a random variable lies <u>within</u> a specified range of values***. The range of values itself is called the ***confidence interval***. For example, as discussed previously, we are 95.44% confident that a purely random variable lies within ± two standard deviations from the mean (using the normal PDF). We state this as a confidence level of 95.44%, which we usually round off to 95% for practical engineering statistical analysis.
  - ***Level of significance*** α is defined as ***the probability that a random variable lies <u>outside of</u> a specified range of values***. In the above example, we are 100 – 95.44 = 4.56% confident that a purely random variable lies either *below* or *above* two standard deviations from the mean. (We usually round this off to 5% for practical engineering statistical analysis.)



  - Mathematically, confidence level and level of significance must add to 1 (or in terms of percentage, to 100%) since they are complementary, i.e., $\boxed{\alpha + c = 1}$ or $\boxed{c = 1 - \alpha}$.
  - Both α and confidence level represent ***probabilities***, or areas under the PDF, as sketched above for the student's *t* PDF with df = 10 (ten degrees of freedom).

- ***Estimating the population mean with the student's t PDF*** – Here is the procedure for how to *use* the student's *t* PDF to estimate the population mean to a desired confidence level:
  - For a specified confidence level, we define the significance level α, i.e., $\boxed{\alpha = 1 - c}$.
  - The range of *t* corresponding to this confidence level is $-t_{\alpha/2} < t \le t_{\alpha/2}$, as illustrated on the above plot.
  - The probability that *t* lies within the desired confidence level is $P\left(-t_{\alpha/2} < t \le t_{\alpha/2}\right) = 1 - \alpha$.
  - From previously, the *t* statistic *t* is defined as $t = \dfrac{\overline{x} - \mu}{S/\sqrt{n}}$. We substitute this into the above probability equation, yielding $P\left(-t_{\alpha/2} < \dfrac{\overline{x} - \mu}{S/\sqrt{n}} \le t_{\alpha/2}\right) = 1 - \alpha$.

- We multiply by the denominator, yielding $P\left(-t_{\alpha/2}\frac{S}{\sqrt{n}} < \bar{x}-\mu \le t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha$.
- Finally, we subtract $\bar{x}$ from all three terms, and multiply all three terms by –1, yielding $\boxed{P\left(\left(\bar{x}-t_{\alpha/2}\frac{S}{\sqrt{n}}\right) \le \mu < \left(\bar{x}+t_{\alpha/2}\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha}$.
- We state our final result as follows: $\boxed{\mu = \bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}}$ with a confidence level of 1 – α.
- In other words, we are confident to a level of c = 1 – α that the true mean (population mean) μ lies within the calculated range.
- Unlike the normal PDF, we cannot create one single table for the student's t PDF, since the PDF is a function of both t and df, i.e., $f = f(t,\text{df})$. Instead, for a given confidence level (or for a given level of significance α), and for a given number of degrees of freedom, df, we need to calculate $t_{\alpha/2}$. This value of $t_{\alpha/2}$ is called the ***critical value***.
- Excel provides a built-in function called **TINV(α,df)** that calculates the critical value $t_{\alpha/2}$. For example, if df = 10 and α = 0.05 (95% confidence level), the critical value is $t_{\alpha/2}$ = TINV(0.05,10) = 2.2281 (to 5 significant digits), as indicated on the above plot. This also agrees with the value in the table.

- Example:

  Given: We randomly grab pipes off an assembly line, and measure their diameter. We use only 6 pipes in our first sample (n = 6). The sample mean is 1.372 cm, and the sample standard deviation is 0.114 cm.

  To do: Estimate the population mean pipe diameter, along with its confidence interval for standard engineering (95%) confidence.

  Solution:
  - Since only *one* measurement is necessary to estimate the mean, df = n – 1 = 6 – 1 = 5.
  - For 95% confidence, α = 1 – 0.95 = 0.05.
  - To calculate $t_{\alpha/2}$, we need to calculate the value of t such that the area under the PDF between t and ∞ is equal to α/2 = 0.025. We use Excel's TINV(α,df) function, i.e., $t_{\alpha/2}$ = TINV(0.05,5) = 2.5706.
  - Our estimate for the population mean then becomes
    $$\mu = \bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 1.372 \pm (2.5706)\frac{0.114}{\sqrt{6}} = 1.372 \pm 0.120 \text{ cm.}$$
  - Our final answer is **population mean pipe diameter = 1.372 ± 0.120 cm**.

  Discussion: Do not confuse this notation for the final answer with the standard engineering notation for a mean with ± 2σ random error (approximately 95% confidence level), as discussed previously. In the present problem, we are predicting the *population mean value of pipe diameter and its confidence interval*, not the level of random fluctuations.

- Example:

  Given: From the same pipe assembly line as in the previous example, we grab 6 more pipes at random, and measure their diameters. For the combined sample, n = 12. The new $\bar{x}$ is 1.367 cm, and S is 0.109 cm.

  To do: Estimate the population mean pipe diameter, along with its confidence interval for standard engineering (95%) confidence.

  Solution:
  - The procedure is identical to the previous problem, except that n, $\bar{x}$, and S have changed. For α = 0.05 and df = n – 1 = 12 – 1 = 11, the table gives $t_{\alpha/2}$ = 2.2010. Or, $t_{\alpha/2}$ = TINV(0.05,11) = 2.2010.
  - Our estimate for the population mean then becomes
    $$\mu = \bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 1.367 \pm (2.2010)\frac{0.109}{\sqrt{12}} = 1.367 \pm 0.069 \text{ cm.}$$
  - Our final answer is **population mean pipe diameter = 1.367 ± 0.069 cm**.

  Discussion: The confidence interval has decreased significantly (from ±0.120 to ±0.069) by taking twice as many data points in the sample. Obviously, the bigger n is, the smaller the confidence interval (the closer we are to the *real* population mean).

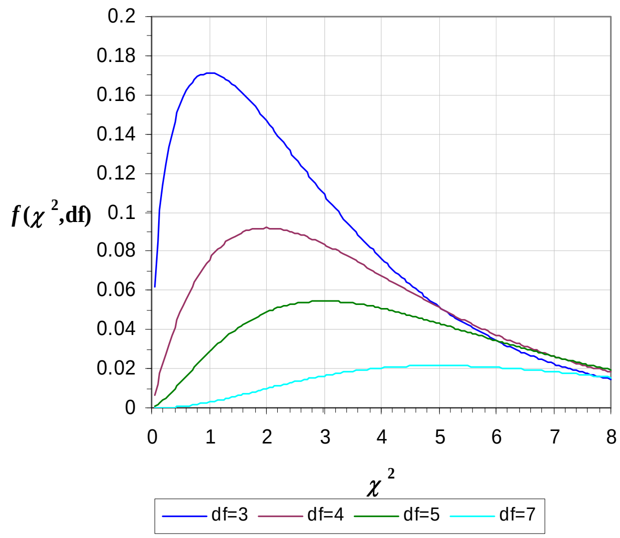
- ***Table of critical values for the student's t PDF*** – Finally, for convenience, and when a calculator or computer or is not readily available, we generate a table of the critical values associated with the *t* PDF.
  - We use Excel's TINV(α,df) function to calculate the critical value $t_{\alpha/2}$ for several values of df and confidence levels. The table is shown below. This kind of table appears in most statistics books.

Values of $t_{\alpha/2}$ (critical values) for the student's t distribution				
	90% confidence	95% confidence	98% confidence	99% confidence
α = →	0.10	0.05	0.02	0.01
df = ↓				
1	6.3137	12.7062	31.8210	63.6559
2	2.9200	4.3027	6.9645	9.9250
3	2.3534	3.1824	4.5407	5.8408
4	2.1318	2.7765	3.7469	4.6041
5	2.0150	2.5706	3.3649	4.0321
6	1.9432	2.4469	3.1427	3.7074
7	1.8946	2.3646	2.9979	3.4995
8	1.8595	2.3060	2.8965	3.3554
9	1.8331	2.2622	2.8214	3.2498
10	1.8125	2.2281	2.7638	3.1693
11	1.7959	2.2010	2.7181	3.1058
12	1.7823	2.1788	2.6810	3.0545
13	1.7709	2.1604	2.6503	3.0123
14	1.7613	2.1448	2.6245	2.9768
15	1.7531	2.1315	2.6025	2.9467
16	1.7459	2.1199	2.5835	2.9208
17	1.7396	2.1098	2.5669	2.8982
18	1.7341	2.1009	2.5524	2.8784
19	1.7291	2.0930	2.5395	2.8609
20	1.7247	2.0860	2.5280	2.8453
21	1.7207	2.0796	2.5176	2.8314
22	1.7171	2.0739	2.5083	2.8188
23	1.7139	2.0687	2.4999	2.8073
24	1.7109	2.0639	2.4922	2.7970
25	1.7081	2.0595	2.4851	2.7874
26	1.7056	2.0555	2.4786	2.7787
27	1.7033	2.0518	2.4727	2.7707
28	1.7011	2.0484	2.4671	2.7633
29	1.6991	2.0452	2.4620	2.7564
30	1.6973	2.0423	2.4573	2.7500
35	1.6896	2.0301	2.4377	2.7238
40	1.6839	2.0211	2.4233	2.7045
50	1.6759	2.0086	2.4033	2.6778
100	1.6602	1.9840	2.3642	2.6259
500	1.6479	1.9647	2.3338	2.5857
1000	1.6464	1.9623	2.3301	2.5807
1.00E+10	1.6448	1.9600	2.3264	2.5758

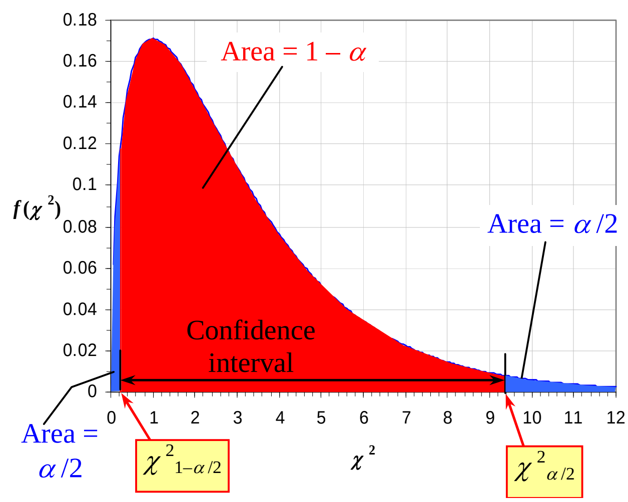
  - The 95% confidence level case (α = 0.05) is highlighted since it is the engineering standard.
  - The last row is for a *very large* value of df, which we approximate as infinity.

The χ^2 PDF

- The χ^2 PDF is used somewhat like the student's *t* PDF, except applies to the standard deviation rather than to the mean.
  - The student's *t* PDF is used to estimate the confidence interval for the ***population mean***, μ.
  - The χ^2 PDF is used to estimate the confidence interval for the ***population variance***, σ^2. [Recall, variance is the square of the standard deviation.]
  - Mathematically, the statistic called χ^2 **(chi-squared)** is defined as $\chi^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}$, where x_i is an individual data point, *n* is the number of data points in the sample, μ is the population mean or expected value, as defined previously, and σ is the population standard deviation.
  - Comparing the above definition to that of the sample variance (square of the sample standard deviation) $S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$, we see that there is a relationship between χ^2 and S^2, namely, $\chi^2 \approx (n-1)\frac{S^2}{\sigma^2}$ [assuming $\mu \approx \bar{x}$].
  - Mathematically, the χ^2 **PDF** is defined as $f(\chi^2,\text{df}) = \frac{(\chi^2)^{(\text{df-2})/2}e^{-\chi^2/2}}{2^{\text{df}/2}\cdot\Gamma(\text{df}/2)}$, where Γ is the *gamma function*, defined previously, and df is the degrees of freedom, also defined previously.
  - Shown here is a plot of the χ^2 PDF for several values of df, noting that χ^2 must always be ≥ 0:



  - Unlike the normal (Gaussian) or student's *t* PDF, the χ^2 PDF is *not symmetric* – it is **skewed**.
- We use the χ^2 PDF to estimate a confidence interval for the population variance, much like we use the student's *t* PDF to estimate a confidence interval for the population mean.
  - It turns out that the confidence interval for the variance is $(n-1)\frac{S^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq (n-1)\frac{S^2}{\chi^2_{1-\alpha/2}}$, where
    - $\chi^2_{\alpha/2}$ is the value of χ^2 at which the area under the PDF is equal to $\alpha/2$ on one side of the PDF.
    - $\chi^2_{1-\alpha/2}$ is the value of χ^2 at which the area under the PDF is equal to $\alpha/2$ on the *other side* of the PDF.
  - This is illustrated on the PDF shown here for the case in which df = 3 (3 degrees of freedom).



  - Just as with the *t* PDF, we can generate a table of critical χ^2 values as a function of df and α.
  - The main difference is that the χ^2 PDF is *not symmetric*, so we must list on our table *two* critical values – for $\chi^2_{\alpha/2}$ and for $\chi^2_{1-\alpha/2}$.
  - For example, at 95% confidence level, $\alpha = 0.05$, $\alpha/2 = 0.025$, and $1-\alpha/2 = 0.975$. For df = 3, as in the plot, $\chi^2_{1-\alpha/2} = 0.2158$ and $\chi^2_{\alpha/2} = 9.3484$.
  - Excel provides a built-in function called **CHIINV(probability,df)** that calculates the critical values $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$. For example, if df = 3 and $\alpha = 0.05$ (95% confidence level), the critical values are $\chi^2_{1-\alpha/2}$ = CHIINV(0.975,3) = 0.2158 and $\chi^2_{\alpha/2}$ = CHIINV(0.025,3) = 9.3484, as plotted above.

- ***Table of critical values for the χ^2 PDF*** – Finally, for convenience, and when a computer or calculator is not readily available, we generate a table of the critical values associated with the χ^2 PDF.
    - We use Excel's CHIINV(probability,df) function to calculate critical values $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ for several values of df and confidence levels. The table is shown below. This kind of table appears in most statistics books.

Critical values of $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ for the χ^2 distribution (both tails)								
Two tails →	90% confidence		95% confidence		98% confidence		99% confidence	
α →	0.10		0.05		0.02		0.01	
Single tail →	$1-\alpha/2$	$\alpha/2$	$1-\alpha/2$	$\alpha/2$	$1-\alpha/2$	$\alpha/2$	$1-\alpha/2$	$\alpha/2$
Probability →	0.95	0.05	0.975	0.025	0.99	0.01	0.995	0.005
df = ↓	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$
1	0.0039	3.8415	0.0010	5.0239	0.0002	6.6349	0.0000	7.8794
2	0.1026	5.9915	0.0506	7.3778	0.0201	9.2104	0.0100	10.5965
3	0.3518	7.8147	0.2158	9.3484	0.1148	11.3449	0.0717	12.8381
4	0.7107	9.4877	0.4844	11.1433	0.2971	13.2767	0.2070	14.8602
5	1.1455	11.0705	0.8312	12.8325	0.5543	15.0863	0.4118	16.7496
6	1.6354	12.5916	1.2373	14.4494	0.8721	16.8119	0.6757	18.5475
7	2.1673	14.0671	1.6899	16.0128	1.2390	18.4753	0.9893	20.2777
8	2.7326	15.5073	2.1797	17.5345	1.6465	20.0902	1.3444	21.9549
9	3.3251	16.9190	2.7004	19.0228	2.0879	21.6660	1.7349	23.5893
10	3.9403	18.3070	3.2470	20.4832	2.5582	23.2093	2.1558	25.1881
11	4.5748	19.6752	3.8157	21.9200	3.0535	24.7250	2.6032	26.7569
12	5.2260	21.0261	4.4038	23.3367	3.5706	26.2170	3.0738	28.2997
13	5.8919	22.3620	5.0087	24.7356	4.1069	27.6882	3.5650	29.8193
14	6.5706	23.6848	5.6287	26.1189	4.6604	29.1412	4.0747	31.3194
15	7.2609	24.9958	6.2621	27.4884	5.2294	30.5780	4.6009	32.8015
16	7.9616	26.2962	6.9077	28.8453	5.8122	31.9999	5.1422	34.2671
17	8.6718	27.5871	7.5642	30.1910	6.4077	33.4087	5.6973	35.7184
18	9.3904	28.8693	8.2307	31.5264	7.0149	34.8052	6.2648	37.1564
19	10.1170	30.1435	8.9065	32.8523	7.6327	36.1908	6.8439	38.5821
20	10.8508	31.4104	9.5908	34.1696	8.2604	37.5663	7.4338	39.9969
21	11.5913	32.6706	10.2829	35.4789	8.8972	38.9322	8.0336	41.4009
22	12.3380	33.9245	10.9823	36.7807	9.5425	40.2894	8.6427	42.7957
23	13.0905	35.1725	11.6885	38.0756	10.1957	41.6383	9.2604	44.1814
24	13.8484	36.4150	12.4011	39.3641	10.8563	42.9798	9.8862	45.5584
25	14.6114	37.6525	13.1197	40.6465	11.5240	44.3140	10.5196	46.9280
26	15.3792	38.8851	13.8439	41.9231	12.1982	45.6416	11.1602	48.2898
27	16.1514	40.1133	14.5734	43.1945	12.8785	46.9628	11.8077	49.6450
28	16.9279	41.3372	15.3079	44.4608	13.5647	48.2782	12.4613	50.9936
29	17.7084	42.5569	16.0471	45.7223	14.2564	49.5878	13.1211	52.3355
30	18.4927	43.7730	16.7908	46.9792	14.9535	50.8922	13.7867	53.6719
35	22.4650	49.8018	20.5694	53.2033	18.5089	57.3420	17.1917	60.2746
40	26.5093	55.7585	24.4331	59.3417	22.1642	63.6908	20.7066	66.7660
50	34.7642	67.5048	32.3574	71.4202	29.7067	76.1538	27.9908	79.4898
100	77.9294	124.3421	74.2219	129.5613	70.0650	135.8069	67.3275	140.1697
500	449.1467	553.1269	439.9360	563.8514	429.3874	576.4931	422.3034	585.2060
1000	927.5944	1074.6794	914.2572	1089.5307	898.9124	1106.9690	888.5631	1118.9475
∞	∞	∞	∞	∞	∞	∞	∞	∞

- The 95% confidence level case ($\alpha = 0.05$) is highlighted since it is the engineering standard.
- The last row shows that as df gets large, both values of critical χ^2 approach infinity.

- **Example**:
  Given: [same pipe example data as previously] We randomly grab pipes off an assembly line, and measure their diameter. We use only 6 pipes in our first sample (n = 6). The sample mean is 1.372 cm, and the sample standard deviation is 0.114 cm.
  To do: Estimate the population standard deviation, along with its confidence interval for standard engineering (95%) confidence.
  Solution:
  - First we calculate df = n – 1 = 6 – 1 = 5. [*Note*: The degrees of freedom is n – 1 for this kind of problem, even though we are dealing with standard deviation, and we may have thought to use n – 2. At the time of this writing, I am not sure why we use n – 1 instead of n – 2.]
  - For 95% confidence, $\alpha = 1 - 0.95 = 0.05$, $\alpha/2 = 0.025$, and $1-\alpha/2 = 0.975$.
  - For df = 5, we read from the above table: $\chi^2_{1-\alpha/2} = 0.8312$ and $\chi^2_{\alpha/2} = 12.8325$. Or, we use Excel's CHIINV function: $\chi^2_{1-\alpha/2}$ = CHIINV(0.975,5) = 0.8312 and $\chi^2_{\alpha/2}$ = CHIINV(0.025,5) = 12.8325.
  - Our estimate for the population variance then becomes
    $(n-1)\dfrac{S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le (n-1)\dfrac{S^2}{\chi^2_{1-\alpha/2}} \quad \rightarrow \quad 5\dfrac{(0.114\text{ cm})^2}{12.8325} \le \sigma^2 \le 5\dfrac{(0.114\text{ cm})^2}{0.8312}$ which yields
    $0.0005064\text{ cm}^2 \le \sigma^2 \le 0.007818\text{ cm}^2$.
  - Taking the square root of all three values yields the final result, the range of estimated population standard deviation, which we write as $\boxed{\mathbf{0.071\text{ cm}} \le \sigma \le \mathbf{0.280\text{ cm}}}$ to 95% confidence. Our best prediction of σ is of course $\sigma = S =$ **0.114 cm**. This is the best we can do based on these few measurements, but now at least we have a *range* and *confidence interval* for σ in addition to this best estimate.

  Discussion: Sample standard deviation S lies within the range calculated for σ as it must – if S were outside of the calculated range, it would surely be an indication of an algebra error. Note that we give our final answer to the same number of decimal places as the original value of S. Any additional digits beyond this would not be significant.

- ***One-tail problems*** – In many situations in practice when using the χ^2 PDF, we are concerned with only one side or one "***tail***" of the PDF. For such cases, we need to keep in mind that the confidence levels on the above table include *both* the left and right tails of the PDF. ***Be careful of factor of 2 errors here.***
  - Consider the above example of pipe diameter measurements. We are 95% confident that σ lies between 0.071 and 0.280 cm. But in a manufacturing situation like this, we are not concerned with a σ that is "too low" – a low σ is actually quite desirable (good quality control)! Instead, we worry if σ is too high.
  - In other words, we are concerned only with the area to the right of the left-most tail on the PDF, for which the probability is $1 - \alpha/2 = 1 - 0.05/2 = 0.975$. In this case, we are 97.5% confident that our population standard deviation is less than 0.280 cm. This is a ***one-tail statistical analysis***.
- **Example**:
  Given: The same six pipe measurements of the previous example. The boss wants to be 99% confident that the population standard deviation is less than 0.2 cm.
  To do: Estimate how many more pipes need to be measured to satisfy the boss's request.
  Solution:
  - For a 99% confidence interval, the probability of *both* tails combined is $\alpha = 1 - 0.99 = 0.01$. But here, for 99% confidence outside of *one tail only*, we set $\alpha/2 = 0.01$, and we are not concerned with the other tail.
  - The right half of the equation for the confidence interval for variance is $\sigma^2 \le (n-1)\dfrac{S^2}{\chi^2_{1-\alpha/2}}$.
  - Setting σ to the desired value, we solve for df = n – 1, yielding $\text{df} = \dfrac{\sigma^2 \chi^2_{1-\alpha/2}}{S^2} = 3.0779\chi^2_{1-\alpha/2}$.
  - There are two unknowns in the above equation (df and $\chi^2_{1-\alpha/2}$). The table of critical values (or the CHINV function) is a second "equation," using 0.99 as the probability ($1-\alpha/2 = 1 - 0.01 = 0.99$). After some iteration, we determine that 13 < df < 14. To be conservative, we choose df = 14.
  - Finally, n = df + 1 = 15, which represents **9 additional measurements** beyond the 6 already available.

  Discussion: This is only an *estimate*, since the actual value of S will change as n is increased. It is critical in these kinds of problems to determine whether we are doing a one-tail or two-tail analysis.