

4

Monte Carlo and Non-Normal Data



We extend the basic methods to address also non-normal data, because using the normal approximation will often lead to severe over- or underdesign for circuits. Distribution-free estimations are also possible, but usually lead to much wider confidence intervals. One example of an advanced non-normal yield analysis is the application of the new generalized process capability index C_{GPK} .

If we have no normal distribution, what else can we assume? And how accurate can our estimates, e.g., on yield be with a limited number of samples?

Indeed, having a good guess on what type the distribution is (such as lognormal, uniform, and Gaussian mix) always helps to improve estimation accuracy. If we assume “nothing”, then we can use distribution-free estimates like the sample yield and have to live with its wide confidence interval and there is a need for large n [Schmid]!

If we assume no specific shape but have a good estimate for the standard deviation, we can use the Chebyshev theorem, so actually even for non-normal distributions, there is a clear theoretical foundation of the spec distance method

for yield estimation. However, unfortunately, the Chebyshev method leads to wide confidence limits (see Figure 4.6), whereas the Gaussian fit may lead to severe systematic error. So you can try to fit the data to another model (instead of a normal Gaussian one).

For Further Reading:

Older and basic statistical literature focuses (too) often on normal distributions, but nowadays also non-normal analysis has found a huge interest. Also, the topic of which model to choose is a hot one, and new techniques like model selection or model averaging (or fusion) have been created.

- Lange, C. Sohrmann, R. Jancke, J. Haase, B. Cheng, A. Asenov, U. Schlichtmann, *Multivariate Modeling of Variability Supporting Non-Gaussian and Correlated Parameters*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 35, No. 2, pp. 197–210, Feb. 2016.
- Yield Prediction with a New Generalized Process Capability Index Applicable to Non-Normal Data, Weber, S.; Ressurreicao, T.; Duarte, C., IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 35, No. 6, June 2016, p. 931ff.

4.1 Examples of Non-Normal Distributions

A simple non-normal case is the uniform distribution (e.g., fitting often well for many discrete components), and if this is the case, we can get even $1/n$ convergence instead of $1/\sqrt{n}$. If we add samples from multiple uniform distributions, the result is *not* again a uniform distribution, but actually a good approximation of the normal Gaussian distribution. This is due to the central limit theorem, so sometimes nature helps us to apply well-known Gaussian approximations! However, circuits not only do summations or differences! The sum of normal variates is again normal, but actually the sum of two uniform variables is giving a triangular distribution, and the sum of two lognormal variables is not lognormal!

Of course, MC results can look more difficult, e.g., having two modes (“peaks”)—here we may better assume a mix of two normal distributions (having already five parameters in total). In such cases, we may need $10\times$ more samples compared to the fully normal case and in extreme cases may be even $1000\times$ more (depending on yield, mix ratio, etc.). Gaussian mixes can be highly non-normal (in opposite to summing normal variates, which still gives normal Gaussian distributions), and in circuit design, they can occur if our circuit has different modes (or states) of operation, like a multiplexer

giving Gaussian outputs in both cases, but overall providing a mix of both due to duty cycle. In subsection 4.8.2 we give some more examples.

Non-normal methods are needed because circuits really create such non-normal distributions more or less all the time! And *whatever* we assume is almost never the true distribution type! With statistics only, you can select a type which gives a good fit and good *prediction*. The only thing we can have “confidence” in is that at least the data might come from our model, and based on that, we do estimations with a certain confidence level. Unfortunately, even if the data pass a normality test, it might be still non-normal or a non-normal model might fit even better. In a MC analysis, many of the technology parameter distributions might be indeed normal or lognormal or uniform, but not your output—for many reasons:

- You have a measurement in dB
- You look for filter passband ripple or settling time
- You have a circuit with 2 modes, often giving a mix of 2 distributions
- DNL of a flash-ADC is defined by $\max(V_{\text{offset}})$
- Delay of a CMOS gate $\sim 1/(V_{\text{GS}} - V_{\text{TO}})$
- Looking to leakage current
- Using $|x|$ in a specification (e.g., $|V_o|$ is half-normal)

This list shows quite nicely that if something “special” happens in your design or just only in your test bench setup, easily non-normal performance data will be generated! However, unfortunately, it is not always easy to understand which of the causes have actually taken place. Non-normal data are a frequent case, and only sometimes strange distributions indeed clearly indicate design weaknesses.

Experience shows that roughly 35% of all analog measures are so non-normal (look at Figures 4.12 to 4.14 for examples), and even for a moderate yield (like 99.8%), estimations based on the normal assumption may become significantly biased. Only sometimes, the user can easily avoid non-normal data (e.g., by not using a spec in dB). Often, it is the circuit *itself* creating a certain nonlinearity leading to non-normal data. Of course, even circuits regarded as linear (like passive filters) can create non-normal MC data, if you inspect performances such as overshoot, phase margin, and poles and zeroes. It does not mean that it will always happen and you can never trust specific methods and you would always need to stick to pure random MC and sample yield: often indeed, the Gaussian approximation is not so bad, and usually, there is a fix part and the statistical variation is something small on top, like $1 + \Delta$, with $\Delta \ll 1$. If we would apply a nonlinear operation like $f = 1/x$, we would end up in $f = 1 - \Delta + \Delta^2 + \dots$ and still in a very mild form

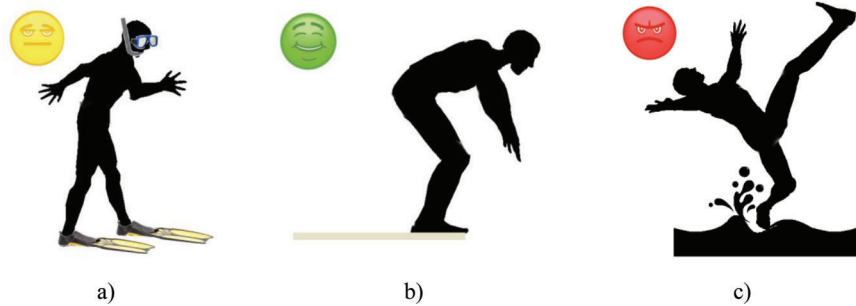


Figure 4.1 (a) Overdesign, (b) optimum design, and (c) design fail.

of nonlinearity due to $\Delta \gg \Delta^2$! Unfortunately, there is no guarantee for this smooth behavior, because there might be stronger nonlinearities, many variables, many correlations, etc.—or you have to go more in the direction of not so small Δ due to circuit specification and technology limitations.

Actually, there is no clear limit, protecting you from making a non-robust design, to be out-of-spec and to have no need for non-normal techniques—there is a smooth transition, a slippery gray area (Figure 4.1)!

What is an outlier? This is a sample that would destroy your fit to the model you “assume”, so actually you have to *decide*! It is best to inspect simulation results of the related MC point manually! The usual rules, like remove points “beyond 6σ ”, are only useful if you can assume near-normal data. The sample yield is quite insensitive to outliers, and they just lower the overall yield a bit, whereas the C_{PK} could heavily impact even by one outlier. For the generalized C_{PK} , you can use the method recommended in [Weber]. The decision if a certain sample is an outlier is not always easy to made, and it could happen that just this “outlier” is truly an indicator that the currently used model is too simple, actually even wrong! In physics, a new model can be a real revolution, and some examples are the quantum Hall effect found in 1980 and the prediction of outer planets.

So if you feel your MC run contains an outlier, then debug it and try to understand and to “repair” the circuit! However, it could be that it is too much effort. If you keep all data including outliers you may overdesign this way. If you completely ignore an outlier you may underdesign, so another method called winsorization is sometimes also usefull: If e.g., the outlier is too extreme (like 1E100) just cut it back at least to the 2nd most extreme sample. This way you can also make sure that e.g., the mean calculation becomes more robust against outliers.

In Chapter 7, we will address the problem of extreme samples to some degree again and in an accurate way by introducing worst-case distances—generation of corner samples with dedicated yield level!

4.2 Identification of Non-Normal Distributions

With the normal quantile plot, we can identify a normal distribution, in less critical cases even with a histogram. Also, numerical tests are available, but any test based on sampled data has its uncertainty. Already small deviations to normal behavior can lead to significant yield errors, so you should get a feeling when the normal assumption might be still applied with acceptable errors and when not.

Check for Normality? In addition to the normal quantile plot, also pure numerical tests for normality are available, e.g., the Jarque–Bera test (JB based on skew s and kurtosis k).

$$JB = n/6 \cdot (s^2 + (k - 3)^2/4)$$

JB is quite a powerful test, and it combines two measures which can also be easily interpreted by themselves: skew s is a measure of asymmetry, and kurtosis k is a measure of the relationship between inner and outer samples. Symmetric distributions have a skew close to zero, and the normal Gaussian distribution has a kurtosis of 3. If JB is large (like beyond 7), we can usually assume that the data are significantly non-normal. In such cases, we do not use the C_{PK} . Please also inspect the spreadsheet example Figure 4.10 for JB calculation.

Identification also for other distributions can be useful, because often the circuit performances do not follow a normal distribution. Leakage current follows often an exponential law, so assuming here a lognormal behavior is much more meaningful than assuming normal data. So a good analysis for this special case is making a fit to a lognormal distribution and using it for yield estimation. Such approach will typically lead to more accurate estimations in terms of variance and systematic errors. However, in the general case one problem is that it is often hard to say which law we should assume, e.g., multiple effects can have an impact, or already the transistor models use very complex functions. Another interesting problem is how can we treat such mixed cases, e.g., a weighted sum of normal and lognormal variates that can vary smoothly from a perfect normal distribution to a full lognormal behavior.

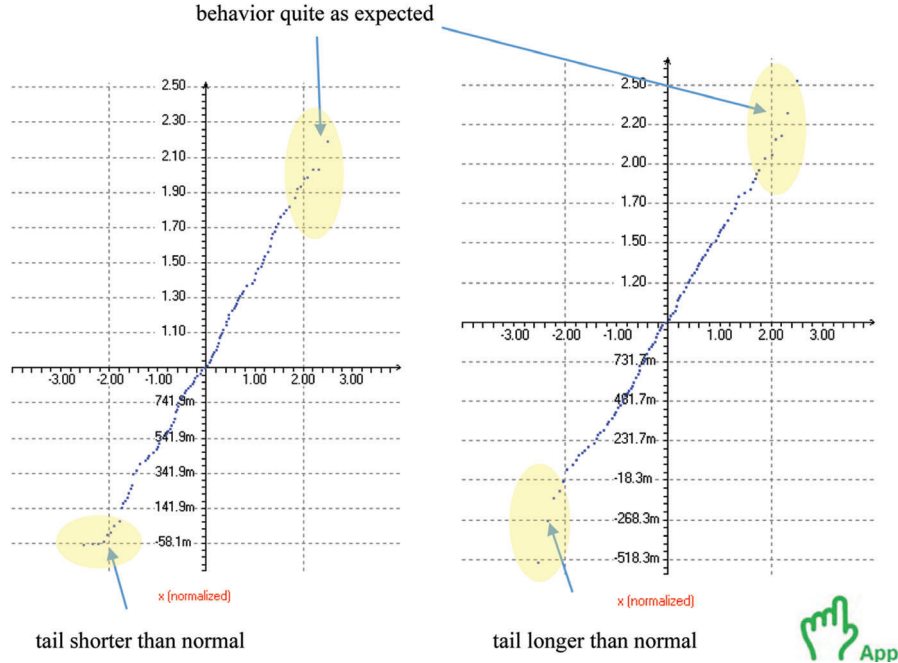


Figure 4.2 Two typical normal quantile plots taken from a normal Gaussian distribution ($n = 256$).

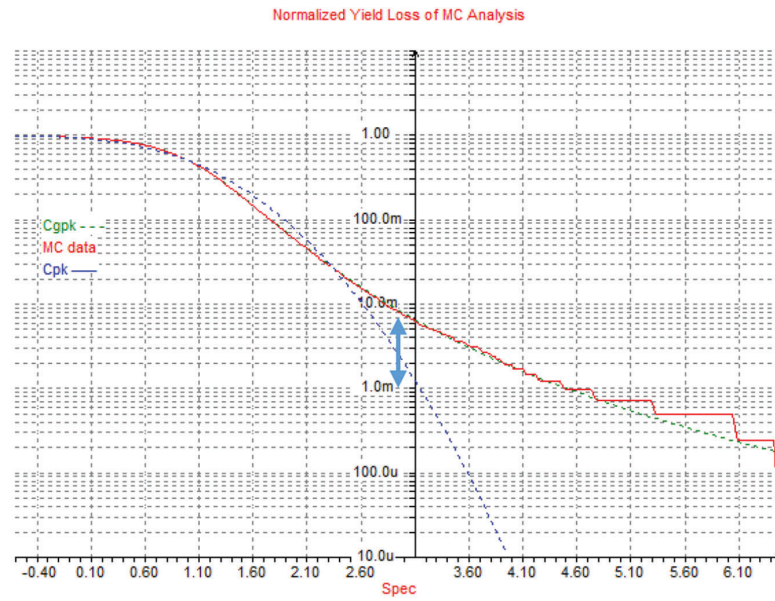
For statistical problems with one variable, we always have a 1-to-1 relation from x_S to f , but unfortunately that would not be the case for multiple variables $\mathbf{x}_S = (x_{S1}, x_{S2}, \dots)^T$. The problem of treating multiple statistical variables will be covered in this Chapter 4.

Figure 4.2 shows how difficult it can be to identify a normal distribution via quantile plot; with 256 points, it can be still hard to decide whether the behavior at $\pm 2.5\sigma$ is Gaussian or not. Figure 4.3(b) shows the quantile plot for a Student-4 distribution (look also to subsection 4.8.1); the yield error in sigma (indicated by the blue arrow) is already roughly 0.5σ , but the quantile plot is just starting to become distinct.

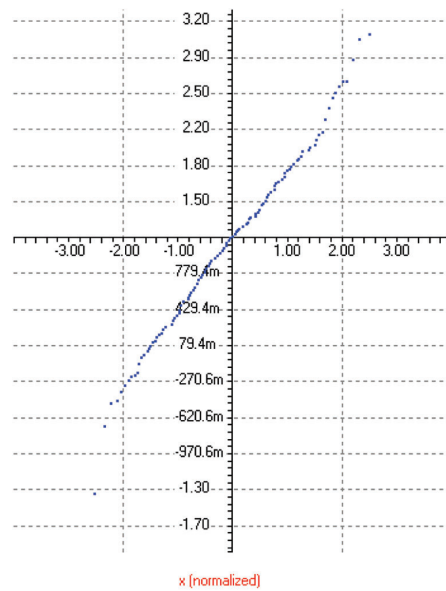
Let us now investigate how we can improve our estimations when dealing directly with an *arbitrary* output distribution.

4.3 Non-Normal Data Analysis via Generalized C_{PK}

We have already inspected two different yield estimation methods, but only the sample yield has no systematic error for non-normal data. On the other



(a)



(b)



Figure 4.3 (a) $\text{Log}(1 - Y)$ for Student-4 and normal distribution fit (averaged MC run with $n = 4$ K, $C_{PK} = 1$, but true $C_{PK} = 0.82$) and (b) quantile plot for Student-4 ($n = 256$, not averaged).

hand, the C_{PK} allows us to interpret also small MC data sets efficiently, and a lower MC count gives the designer a speed-up in making design decisions (Figure 4.4).

The question is: can we obtain a similar speed-up also in the general non-normal case, e.g., by making a more detailed result evaluation?

The old state of the art on process capability index is to make a Gaussian fit, thus extracting the two distribution parameters μ and s , and the normalized spec distance $(USL - \mu)/\sigma$ gives us a yield estimation. This is a distance method, and the good thing is that intuitively the yield is indeed well correlated with the spec distance—although it is not the only measure!

To address this problem, a new generalized C_{PK} has been developed [Weber], which features one more parameter “ t ” to describe also the tail behavior. This way, the generalized C_{PK} is quite accurate also for non-normal data, whereas the C_{PK} can be easily systematically wrong by 50%, in cases where the bias of new is only 5%.

Instead of fitting a normal Gaussian pdf, we fit a “generalized Gaussian” pdf. Also, the fit is not done on the whole data, but only to the spec-sided part—starting at the distribution mode (instead of the mean).

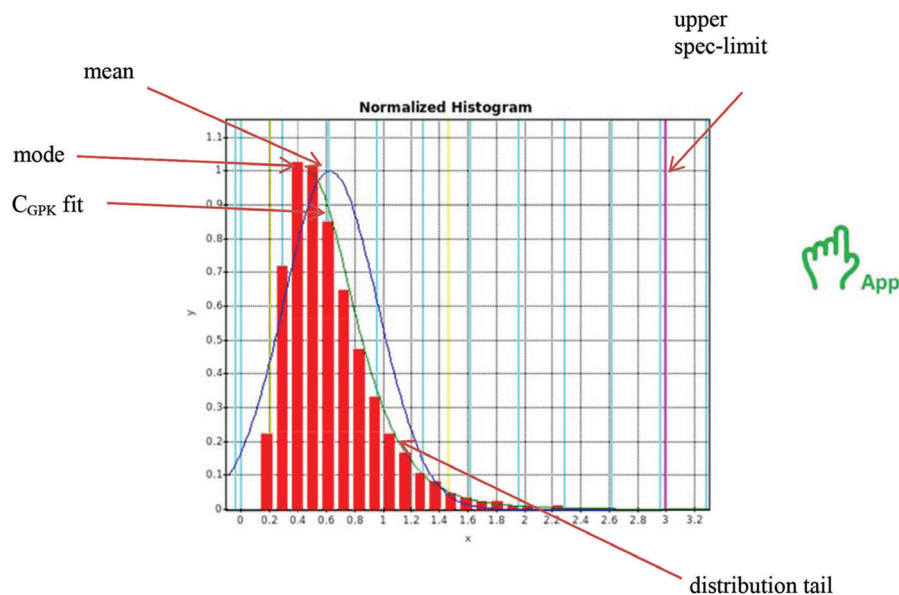


Figure 4.4 Method of the generalized C_{PK} (Gaussian fit is shown as blue curve; it is obviously bad).

Some distributions have multiple modes; here, the C_{GPK} would start at the spec-sided mode. Doing the parametric fit on spec-sided samples has several advantages; for example, non-spec-sided outliers will have no impact, and our fitting function can be formulated much easier, so that indeed, one more parameter (namely t) compared to the “old” C_{PK} gives a dramatic improvement (Figures 4.5 and 4.6).

With the tail parameter t , we can model a much wider range of shapes (Table 4.1).

Note: Some of these distributions are exactly included in the model, and others are only approximated. The parameter t is actually a parameter of the model cdf (see [Weber]); and there is no simple formula as e.g., for the standard deviation, but we can apply MLE or moment fitting. However, Table 4.1 shows that t can be still easily interpreted, just as a normalized tail parameter; complementing the location and scale parameters.

Of course, there is no free lunch: as the C_{GPK} has one more parameter to fit, the statistical variance becomes larger in near-normal cases compared to

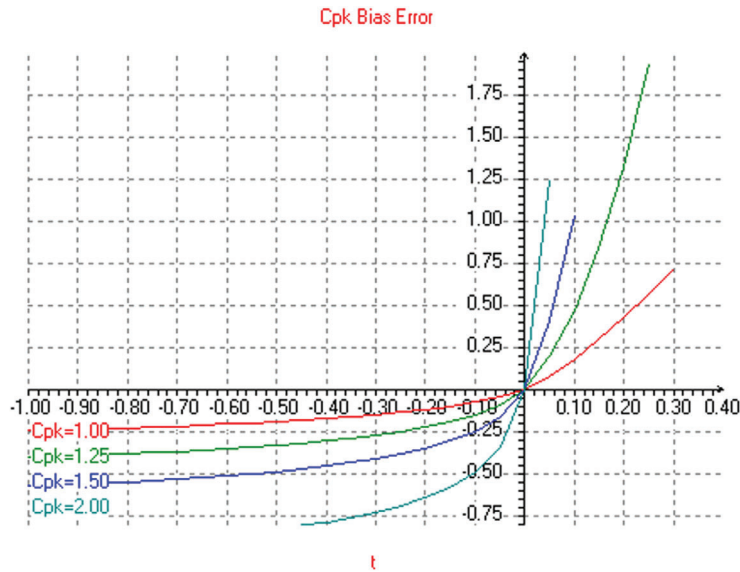


Figure 4.5 Comparison of C_{PK} versus $C_{GPK} - C_{PK}$ bias error for symmetrical cases (C_{GPK} error is zero): as the C_{GPK} is also a distance method, we can also calculate the spec limit for a given target yield.

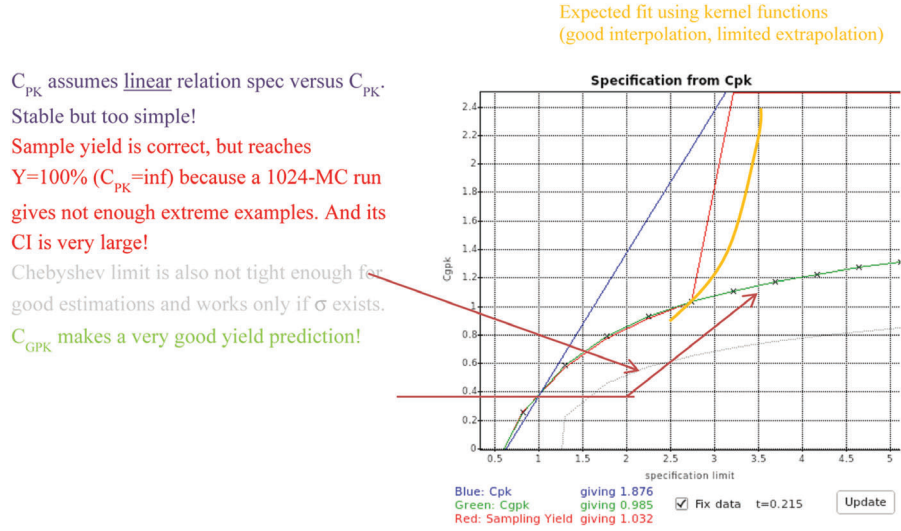


Figure 4.6 Calculating back to spec with different methods (lognormal data used as example).

Table 4.1 Distributions and tail parameter

Distribution	Tail Parameter t	Comment
Uniform	-1	short-tail, low kurtosis
Triangular	$-1 \leq t < 0$	
Parabolic		
Typical bimodal distributions like staggered Gaussian		
Gaussian	0	e^{-x^2} tail, $k = 3$
Peaky distributions like stacked Gaussian	$1 \geq t > 0$	
Student-t, logistic, etc.		
Cauchy	+1	$1/x^2$ tail, infinite k

the C_{PK} . This can be nicely seen if we look to the correlation between yield, C_{PK} and C_{GPK} (look at the scatter plot, Figure 4.7).

As expected, there is a strong correlation with the yield, but the C_{PK} is more stable than the C_{GPK} . So the C_{PK} is still preferable for clearly normal distributions, but for already small deviations, the C_{PK} advantage of lower variance is compensated by its much larger bias error. This bias–variance trade-off is very typical and almost impossible to avoid. Actually, the C_{GPK} is a clever mix of parametric and nonparametric methods, and in opposite to a pure nonparametric modeling (e.g., using the empirical cdf),

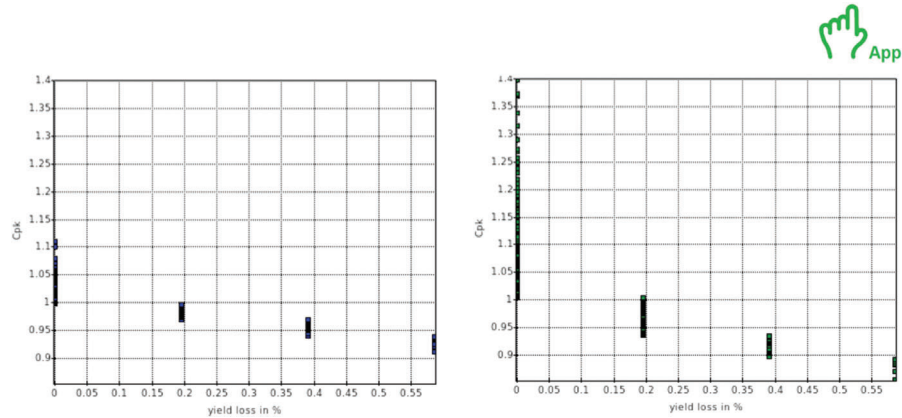


Figure 4.7 Correlation between yield, C_{PK} and C_{GPK} (normal data, $n = 512$, $C_{PK} = 1.0$, 256 MC runs).

nonparametric and tail modeling [MacDonald], or just using a more complex model [Lange], we can efficiently model many difficult distributions like bi- or multimodal distributions and we still fully include the normal Gaussian distribution.

Figure 4.8 compares the sample yield, C_{PK} and C_{GPK} sample count for verification as a function of the true guaranteed C_{PK} (by 95% confidence

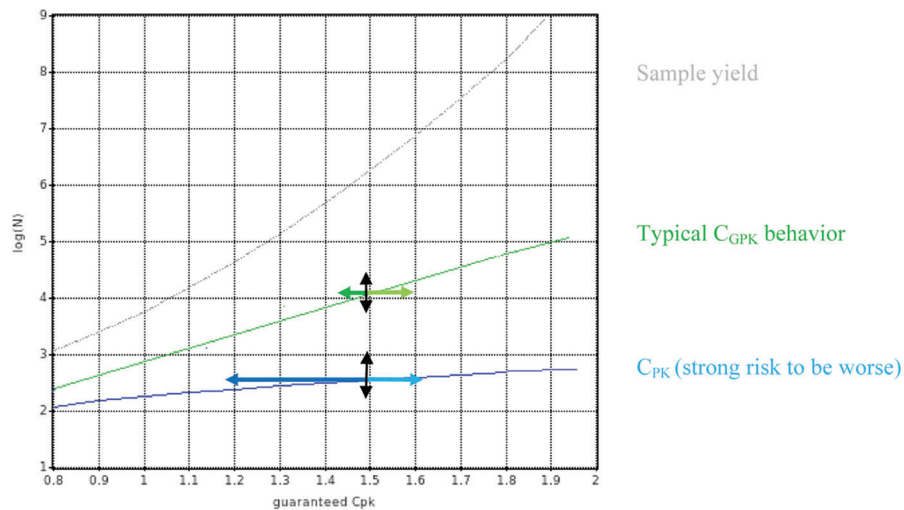


Figure 4.8 Random MC yield verification count for different estimators (design margin 0.375σ).

interval). It shows that the sample yield is highly inefficient especially for high yields, but as mentioned, using it requires no extra-margin for any bias errors due to model limitations (look at the horizontal arrows in Figure 4.8). It can easily happen that the C_{PK} is wrong by more than 1σ , whereas the C_{GPK} bias is usually $5\text{--}10\times$ lower. Note that the green curve for the C_{GPK} is only an average, as it depends also slightly on the distribution type (vertical arrows). In the succeeding subchapters, we will give some concrete examples (for measured production data and for certain mathematical distributions).

Alternative Distribution Fitting Methods. There are indeed many ways to fit data to a certain distribution (we mentioned MLE and moment fitting). Instead of using the C_{GPK} concept we could also do it a bit differently [Weber]. We could also try to fit over the entire data (like the C_{PK} does), or we could also only model the tail (Table 3.3). For instance, we could assume a triple Gaussian mix to model distributions with up to three modes, but obviously a high flexibility comes with the price of many parameters. The advantage of modeling only the tail is some more flexibility and potentially high accuracy in this region of interest! But one big question is where to “start” the fit and where the tail begins? Having enough data, it is indeed possible to answer that question, but to some degree, it results in a model which depends on the fit for quite few data points, just the tail points. So the price for low-bias errors is typically having quite large confidence intervals. For the tail modeling, typically the generalized Pareto or generalized extreme value distribution is assumed. Also when using the C_{GPK} concept, we could plug-in different distributions, or we may extend the concept with one more parameter to be able to model the shape of the mode and the one for tail independently. Also almost completely nonparametric fits are possible, e.g., based on KDE, but typically they are not well suited for high-yield estimation.

4.4 Analyzing Real Production Data

Of course many statistical methods are not only applicable to MC results, but also applicable to real data measured in production. The data in this example (Figure 4.9) have been taken from [Shinde] and come from a USB2 squelch circuit: trip point has been measured on $n = 3,999$ silicon samples.

Let us do two analyses: according to [Shinde], let us inspect what we can estimate if we do not use the full measured data, but only a subset of 25 samples. This is often regarded as minimum requirement for a normal

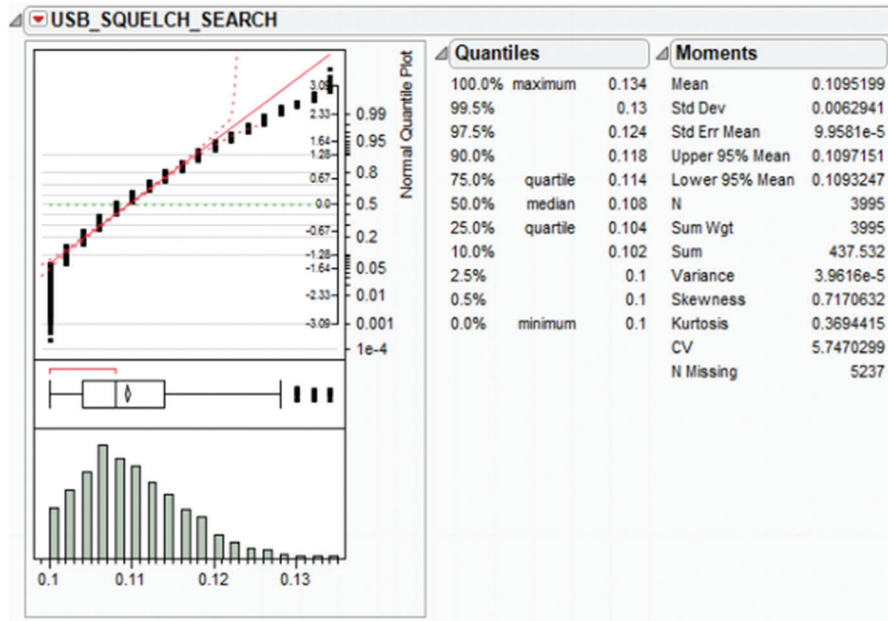


Figure 4.9 Data from a fabricated USB interface [Shinde].

data analysis. In many such cases, designers can expect near-normal data for good reasons, because usually the trip point of a comparator is dominated by device mismatch—and mismatch can usually modeled well with normal distributions. Later, let us check this analysis against an analysis taking the full statistical data into account (Figure 4.10).

The original Intel conference paper exemplified already a yield estimation on a subset of $n = 25$ samples. The authors obtain a sample C_{PK} of 2.04 (6.12σ). By visual inspection of the histogram, the authors regarded the data as normally distributed and they predict 4σ as lower yield 95% confidence limit. This means although the sample C_{PK} is 2.04—indicating a very good design—a statistical analysis based on the assumption of normality can only guarantee a C_{PK} of $4/3 = 1.333$. The difference between 4σ lower CI limit and 6.12σ sample C_{PK} would go to zero for $n \rightarrow \infty$.

This difference looks like a good “safety margin”, but this simple analysis does not take some important aspects into account: sample skew is $s = 0.71$, and the critical specification limit is at the long-tail side. This leads to a too optimistic C_{PK} yield prediction! If we apply the Jarque–Bera normality

188 Monte Carlo and Non-Normal Data

C107: = KURT (C2:C102) - a measure of tail behavior
 C108: = SKEW (C2:C102) - a measure of unsymmetry
 C109: = 100/6*(C108*C108 + C107*C107/4)

C109 : ✕ ✓ f_x =100/6*(C108*C108+C107*C107/4)						
	A	B	C	D	E	F
97		0,3951461	-0,265931241	0		
98		0,9646656	1,807599664	0		
99		0,9106683	1,344881804	0		
100		0,2933183	-0,543716415	0		
101		0,0116437	-2,268687314	0		
102						
103	Avg	0,504315	0,02401982			
104	Stddev	0,2847205	0,96873295			
105	Yield in %			99,0000		
106	Cpk			-0,851965164		
107	Kurtosis		-0,123452811			
108	Skew		0,130741339			
109	JB		0,348390781			



Figure 4.10 Spreadsheet example for the calculation of the Jarque–Bera normality test.

test for the whole data set, we can obtain a value for JB beyond 200, which clearly indicates non-normal data, but for $n = 25$ JB is only 2.2 (indicating only very mild non-normality). A large JB value indicates that we should not apply a Gaussian fit, but better apply the new generalized C_{PK} . This predicts a true C_{PK} of 1.30—instead of 2.04! And of course also the CI for the generalized C_{PK} would be even lower, being at 1.20. Overall, the designer's conclusion should be that the design definitely *needs* significant improvements. And it is unfortunately not enough just to measure more samples to tighten the confidence interval!

Note: In this example, the sample yield is still 100%, because there is no fail even in 3999 points! The sample yield confidence limit is approximately 99.88% and is equivalent to a true C_{PK} of 1.02, which is worse than the CI of the generalized C_{PK} (which is at 1.20). The Clopper–Pearson yield limit (also used in most EDA tools for yield confidence intervals) for 1 fail in 3999 samples would be 98.6% only (equivalent to a true C_{PK} of 0.73!). Also look up: Jarque–Bera is one of many normality tests, neither the best, nor the worst. It is actually quite powerful which means that it needs usually not that many points to make a decision. On the other hand, there is simply no best normality test, because deviations to normal data can be of many kinds.

4.5 Yield Estimation for Non-Normal MC Data via C_{GPK}

In the previous example, we inspected measured production data, and let us now inspect MC data from an operational amplifier. Such amplifier is basically a linear circuit, where most designers would expect to find quite normal histograms. However, we will see that also such classical analog circuit can easily create non-normal data (Figure 4.11).

A complete CMOS op-amp with tricky feed-forward frequency compensation has been designed (but not fully optimized) and verified for almost all common specifications in a big MC run for mismatch only ($n = 1500$). The PDK does not offer global variations, but these would of course lead to wider variations and potentially even more non-normal data.

Many histograms are indeed near-normal (like the one for current consumption and offset voltage), but here are also some interesting non-normal histograms, where we really need the C_{GPK} . In the Figures 4.12 to 4.14 you will find some examples taken from [Weber2016], e.g., Figure 4.12 shows HD_3 data, where the C_{PK} is too pessimistic.

Figure 4.13 shows a second example, looking to the peaking of the closed loop gain; here, the C_{PK} is too optimistic.

A third example is shown in Figure 4.14, the 3-dB-BW (the graph of amplification versus frequency has two peaks due to the feedforward scheme—and the circuit is really functional, no bug).

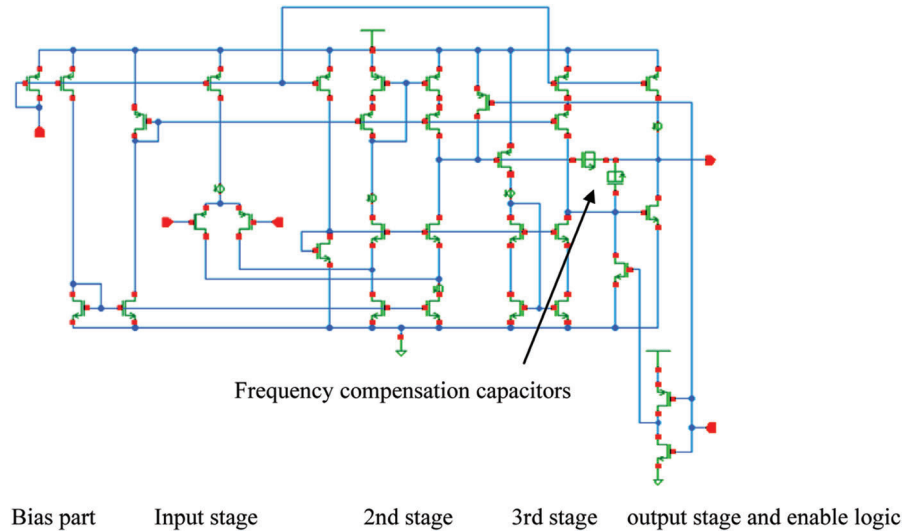


Figure 4.11 Inspected op-amp circuit.

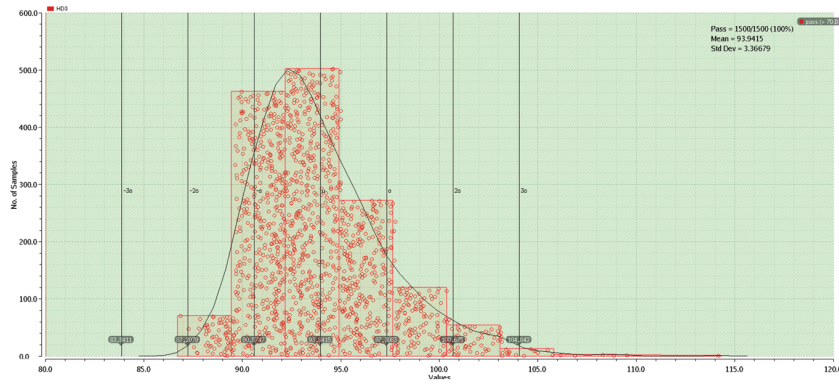


Figure 4.12 Third-order distortion in dB (C_{PK} too pessimistic due to spec at short tail) [Weber2016].

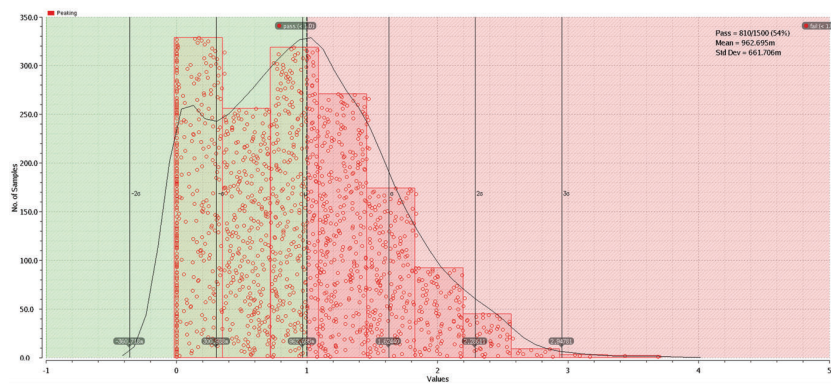


Figure 4.13 Gain peaking in dB (C_{PK} too optimistic, spec on long tail) [Weber2016].

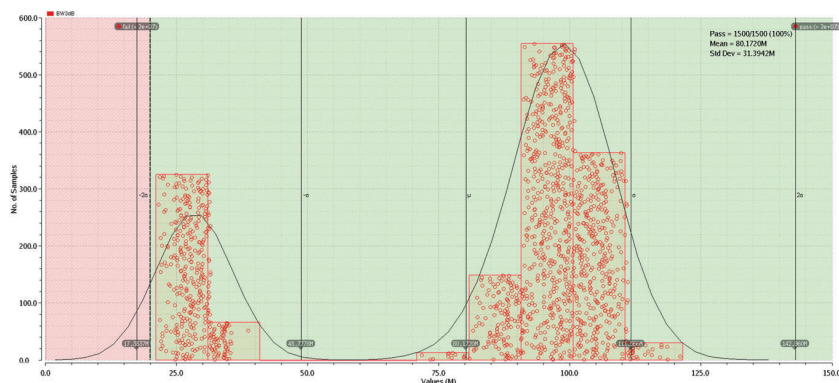


Figure 4.14 3 dB bandwidth (C_{PK} too pessimistic) [Weber2016].

4.6 Questions and Answers

1. How can I check my MC data if it is following a normal Gaussian distribution?

Inspect the normal quantile plot or apply a numerical test like the one according to Jarque–Bera. $JB = n/6(s^2 + (k - 3)^2/4)$.



2. What is the distribution for the output voltage of a two-resistor voltage divider made of discrete resistors?

For discrete elements, it is not realistic to assume a Gaussian distribution, and a uniform distribution is usually more realistic. The sum of two uniform variables gives a triangular distribution! Actually, the output voltage is a nonlinear function of the two resistors, but the nonlinearity is not that large, so indeed we can expect a near-triangular distribution.

3. Is the generalized C_{PK} having a systematic error?

The C_{GPK} includes a much wider class of distributions without such bias error, but if the data are not part of the model pdf, we get some bias. Usually, it is only 10% of the normal C_{PK} bias, like for a lognormal distribution. In such cases, and also on a uniform distribution, the C_{GPK} bias makes the yield estimation a bit too pessimistic—so you are on the safe side, but overdesign a bit.



4. Is the C_{GPK} method an interpolation or extrapolation method?

It is both, depending on the yield level! In opposite to many other extrapolation schemes, the C_{GPK} makes a very meaningful extrapolation. Also the C_{GPK} method might be combined with WCD methods to get rid of the extrapolation risk.

5. $\pm 3\sigma$ around mean is equivalent to capturing approximately 99.73% of the distribution if the data are normal, but how much is it for a Student-5 (which looks very similar to a Gaussian distribution, look at subsection 4.8.1)?

Although the kurtosis k is still moderate for the Student5 distribution, and the normal quantile plot indicates no strong non-normality, the yield is pretty much less, approximately 98.8%—so only approximately $\pm 2.5\sigma$. So the error in terms of sigma is approximately 20%, thus often much larger than the confidence interval reports! Remember, confidence intervals do not quantify such systematic errors.

6. Can we extend the C_{PK} and C_{GPK} concept also to multiple performance?

Yes, this is possible and for multivariate pure normal distributions, several solutions exist. These are usually good enough for process monitoring, but in circuit design you have very often to deal with

non-normal distributions. In Chapter 5, we present a good approximated solution.

7. Can it happen that running one more point in MC, which gives a *pass*, leads still to a decrease C_{PK} for a spec like $CMRR > 40\text{dB}$?

Yes, it is possible, and assume we have a short 50-point MC run with sample standard deviation of 2 dB and mean 60 dB, so C_{PK} was 10. Imagine the next MC point is at 100 dB—it would shift the mean down, but the standard deviation would increase even more, so overall the C_{PK} could become worse! In pure Gaussian outputs, such event would be extremely rare, but non-normal data can give such surprises.

8. Imagine you have a Gaussian output y for a certain performance in a MC run. Now you take this in dB, which kind of distribution will you get?

It will be a new distribution, and it is not the lognormal distribution!

9. If we add many samples from independent uniform distributions, we end up by central limit theorem with a normal distribution. Would this also be the case for other distributions?

For instance, even when adding exponential distributions, we would lose the asymmetry; and the left side short would become longer, whereas the longer right tail would become shorter (e^{-x^2} instead of e^{-x})! However, e.g. the Pareto distribution is asymmetric too, but has infinite variance, so here the CLT would not work.

10. Discuss which kind of problems can be solved with pure random MC and sample yield?

Check runtimes, accuracy, design improvements, inclusion of corners, which additional analysis should be done, etc.

11. Imagine you have a Gaussian distribution with mean = 0 V, so we get in a nominal simulation usually also 0 V. Now we apply the exponential function to this output, leading to $\exp(0) = 1$ at the new output. However, what happens in MC? Will the mean be also at 1?

No! The median will be there; and it is now different from the mean (average) value! If our nonlinear function is non-monotonic even the median will usually not be preserved.

4.7 Rules You Have to Know for Monte Carlo

When setting up a MC analysis, one big general question is how to decide on required number of points for certain target accuracy? Actually this problem is not only related to MC but to taking statistical samples in general, like

Table 4.2 Overview on basic normal and non-normal MC techniques

Analysis	#points	Comment
MC for checking mean and standard deviation of performances	e.g., 100	Some mild non-normality allowed. For extreme distributions, bad or no convergence
MC for checking sample yield	app. 2K for 3σ	See Table 3.2
MC for yield via C_{PK}	app. 200	Data should be highly normal, especially for high-yield targets
MC for yield via generalized C_{PK}	app. 500	The number of points depends slightly on yield level and distribution type
MC for checking distribution type	>50	Depend on how accurate you want to model modes and tails; to differentiate between similar distributions, you may need >1K points
MC for correlation analysis	100 – >1000	Dependent on number of variables involved

for production data inspections. It essentially depends also on what kind of estimate you are interested, e.g., 1% yield accuracy is good if the design has a yield of 50%, but it is not good enough if the target is 99.7%! Many measures also depend on distribution shape, and this can never be fully known upfront (Table 4.2).

So best know some basic rules and their prerequisites and make a MC test run, check histograms, and look to confidence intervals. If the CI is $2\times$ too wide, then increase the number of points by approximately $4\times$.

Rules for any kind of data:

1. You can trust the sample yield $Y = n_{\text{pass}}/n$ (because it is a distribution-free estimate).
2. But CI of Y is large. If there are no fails, then CI limit is approx. given as $3/n$.
3. In random MC, there is no dependency on number of statistical variables for estimates like Y , μ , or σ , but of course it might be the case for correlations (Chapter 5).

Basic rules for near-normal case:

1. Most frequently used: 95% confidence interval
(so 5% risk of false decision)
+ assuming a normal Gaussian distribution
+ assuming $n \gg 1$ (like 50)

2. Then, e.g., the 95%-CI of the mean μ becomes roughly $\pm 2\sigma/\sqrt{n}$. This is two times the standard error $SE = \sigma/\sqrt{n}$.
3. So to know variance on mean μ , you need to know σ .
4. Also σ has a variance: $\sim 1/\sqrt{2n}$, e.g.,
 $n = 50$ gives 10%, so if $\sigma V_{\text{offset}} = 10 \text{ mV}$, it is typically within 8 . . . 12 mV with 95% confidence => not so bad
 $n = 200$ gives 5% => often good enough
5. Other measures (such as correlations or the mode) may need more points (like 1000).

Rules for significant non-normality:

1. Apply tests, e.g., via Jarque–Bera and normal quantile plot.
2. σ might not converge for long-tail distributions!
3. σ variance is usually (roughly) proportional to $\sqrt{\text{kurtosis}}$ (4th moment).
4. Do not trust the C_{PK} ! Use C_{GPK} , sample yield or the methods from Chapter 7.
5. If data is not very non-normal, then CI width and sigma follow still often follow the $1/\sqrt{n}$ law. But bias errors can follow any law, and even for infinite n they might be an error (e.g., using the CPK for yield estimation on non-normal data).

4.8 Design with Pictures Part Two

The non-normal distribution examples in this Chapter 4 were quite obvious; that is, by careful visual inspection, it was quite clear that the data are not normal and that a data analysis based on normality would lead to bad results. However, sometimes it is not so easy to decide whether data are normal or not. And in high-yield cases, even small deviations can lead to significant errors, because a Gaussian fit and the C_{PK} based on that are kinds of extrapolation method. An interesting question is how large is the risk that a deviation to the normal distribution can be seen “late”?

In our USB fab data example, the Jarque–Bera JB value was huge for $n = 3,999$, but for the same skew s and kurtosis k with lower MC count, JB will drop and there is a “gray” area (like $JB = 1.4$) where it is quite likely that the data is indeed normal, but the confidence is still quite low, or vice versa the data is non-normal, but too close to normal, so that most people would apply methods based on the normal assumptions, mistakenly.

One further problem with JB and many other normality tests is that they do not take the yield level into account! So the risk of “seeing” non-normality

late and the error in yield prediction depend on how big the non-normality is but also on how much you “extrapolate”. For instance, to “see” the difference between a Student-100 and a normal distribution, you may need 100,000 samples, but between Gaussian and uniform maybe only 50–100 samples. There is even no such thing like a confidence level for such investigation, but at least we can provide guidelines and train ourselves with examples. In statistics and yield verification just multiplying the sigma is risky, actually even when using normality tests.

Note: With the RealTime MC program that complements the book, you can do such investigations in a very short time, because a much faster simulator is built-in, than just plain SPICE.

4.8.1 Normal versus Student-t versus IH Distribution

The normal distribution has a kurtosis of 3 (note Excel gives 0, because the 3 is subtracted internally), whereas the other two distributions which we have chosen as example, the Student-t and the Irwin–Hall IH distribution, feature a parameter to adjust the kurtosis. In our RealTime app, we can tweak this parameter to obtain a kurtosis of 3.3 for the Student-t and 2.7 for the Irwin–Hall distribution.

Actually, the Student-t is a very important distribution, e.g., for confidence interval calculations, but also the IH has a relationship to the normal distribution. It is composed of the sum of uniform distributions, and if we would add up an infinite number, we would end up in the normal distribution (Figure 4.15). Note that many distributions (even discontinuous ones) can be connected to the Gaussian distribution this way, just due to central limit theorem (CLT)! You can also extend the normal distribution, by adding parameters to adjust the shape and to introduce an asymmetry. Actually, there is no single “best” way to do this, and many such generalized Gaussian distributions exist. There are also distribution families with no tight connection to the normal distribution, but in spite of that, they can be still very similar (e.g., the logistic distribution).

Looking to the (smoothed) histograms of all our three examples, you can hardly see any difference (Figure 4.16), just because even for $n = 1024$ the tails are very hard to inspect visually. Even in the normal quantile plots (Figure 4.17) you really need a huge number of points to identify the distributions.

The JB value for checking normality is approx. 4.5 (pretty close to the gray zone), and if we set the spec for a C_{PK} at 1.0, we get the standard deviation of only 2.5%, i.e., n is large enough to really get a stable C_{PK} .

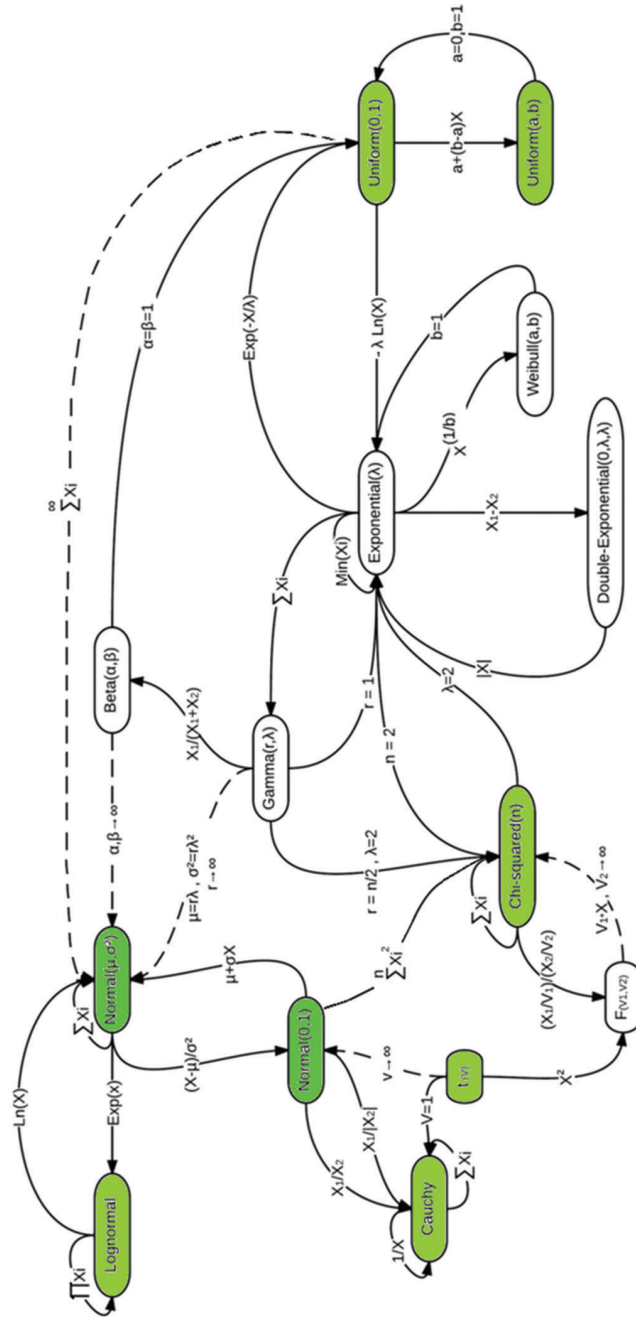


Figure 4.15 Some important distributions and their relationships.

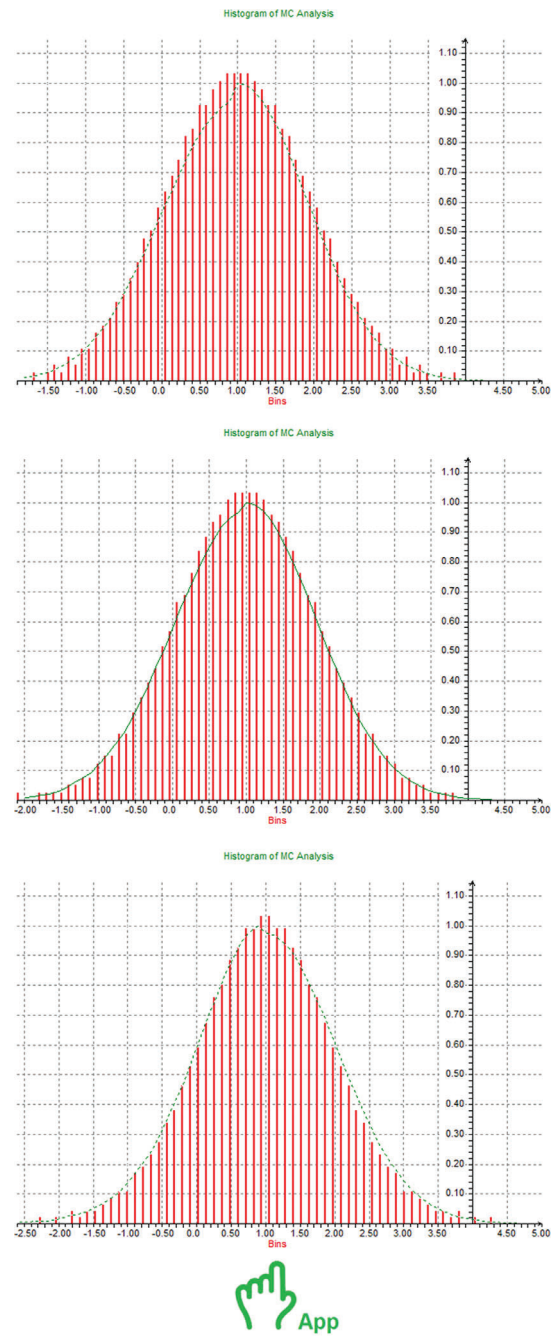


Figure 4.16 Histograms for the three inspected distributions (averaged).

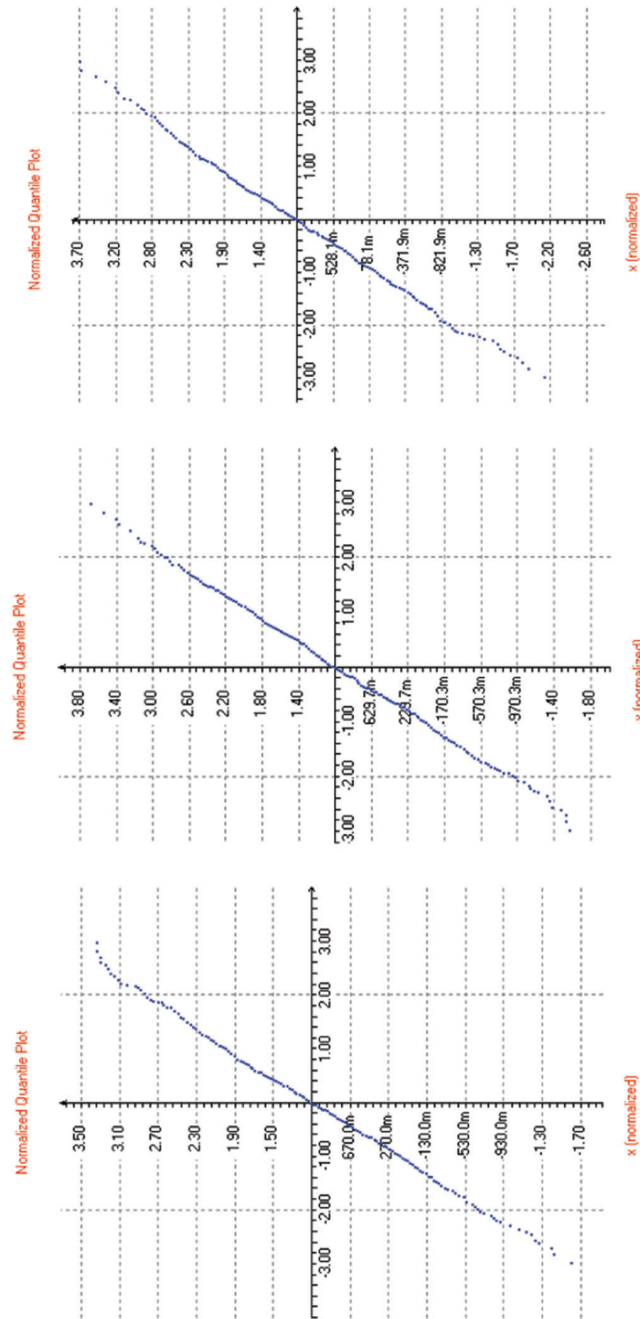


Figure 4.17 Normal quantile plots for $n = 1024$ (non-averaged).

The question is now how large is the systematical C_{PK} error, especially at higher yields. This can e.g., be checked by running a very long MC analysis using 64 K points and using the generalized C_{PK} , which has a highly reduced bias error. We set the spec limit to obtain a C_{PK} of 1.5, and for the Student-t distribution (Figure 4.18), the C_{PK} yield loss prediction is too optimistic by 2.5 orders of magnitude (whereas the C_{GPK} has no bias error in this case)! For the IH distribution, it is vice versa and the C_{PK} estimation is too pessimistic by approx. $30\times$ (Figure 4.19).

Notes: At some point, the red curve (showing the sample yield) drops to infinity, because it is still 100% even for $N = 64$ K. The green curve is the result of yield estimation by the generalized C_{PK} , which makes a very meaningful extrapolation.

In conclusion, an MC run with 1024 giving normal data according to standard tests and applying the C_{PK} can give still give big yield errors for $C_{PK} \geq 1.5$. Having no fails, the Clopper–Pearson lower confidence bound LCB would be only at 99.64% (equivalent to $C_{PK} = 0.896$). The C_{GPK} LCB for the Gaussian case is approximately 1.2—and this is (without stronger assumptions) what you can “guarantee” at best. The C_{PK} LCB is 1.43 for normal data, but even this (and not only the point estimate of 1.5) is still too optimistic compared to the true yield and true C_{PK} being at 1.235.

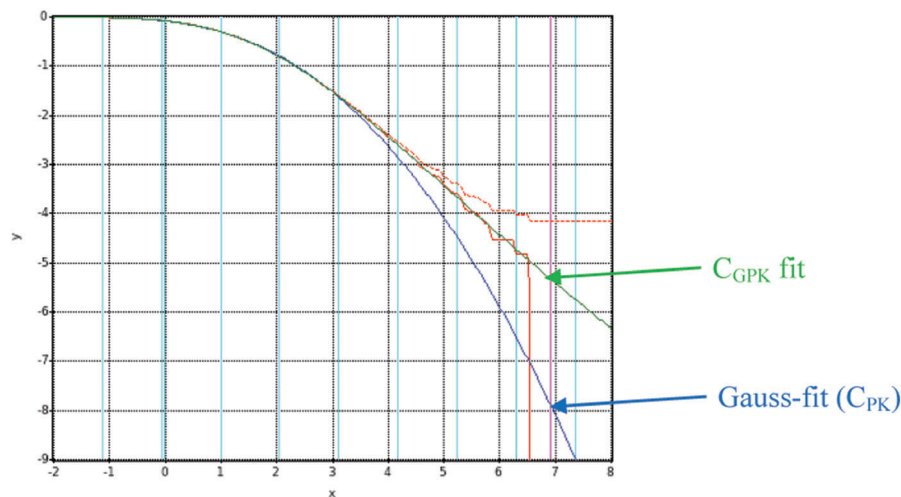


Figure 4.18 Plot of $\log(1 - Y) = f(\text{spec})$ for Student-t distribution.

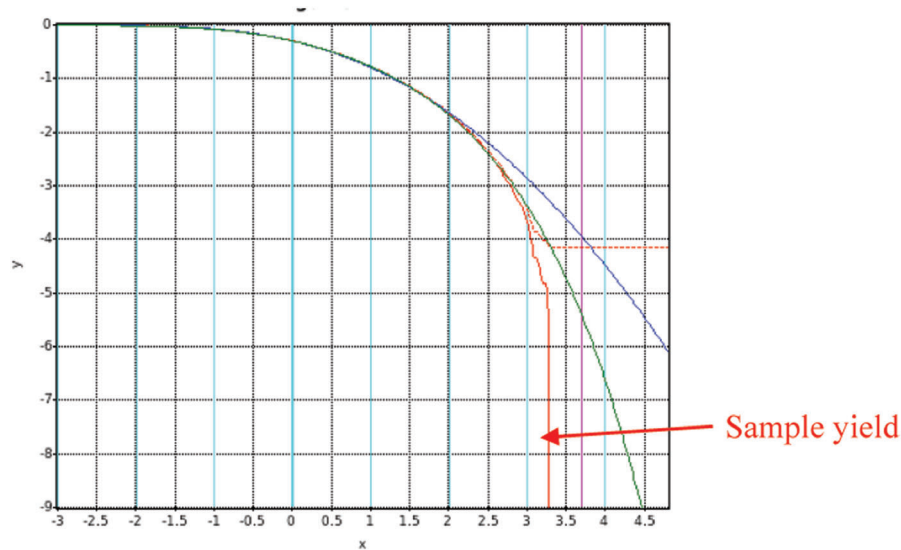


Figure 4.19 Plot of $\log(1 - Y) = f(\text{spec})$ for Irwin-Hall distribution.

4.8.2 Calculations with Random Numbers

Can we calculate with random numbers, like we do with real or complex numbers? Yes, you can, but indeed some rules will *change*, and only few will remain the same. For instance, taking a random variable X with normal distribution and multiplying it by 2 gives a normal distribution with doubled σ . However, adding two *independent* normal random variables (of same σ) gives only $\sqrt{2} \cdot \sigma$, so $X+X$ is not always equal to $2X$. Also taking the difference is special, because $X-X$ gives us the same distribution as $X+X$ for independent *standard normal* variables! Also note that also the rule $X+X = \sqrt{2}X$ would only work for Gaussian variables, not for uniform or lognormal ones (here we would get a *change* in the distribution type); (only) for Cauchy variable we would indeed observe $X+X = 2X$.

Taking $\exp(X)$ gives us the lognormal distribution, but adding two independent lognormal variables gives no lognormal distribution again! However, interestingly adding two Cauchy distributions gives us a Cauchy distribution, so to some degree the normal and the Cauchy distribution are special. Taking the absolute value of a standard normal distribution gives us a *half-normal* distribution, but taking the difference from these would give us again a normal distribution.

What about more difficult operations such as multiplication and division? For instance, dividing two normal independent normal variables gives another distribution, which is the Cauchy distribution! Actually the division spreads the distribution a lot, so the result (the Cauchy distribution) has much stronger tails than the normal distribution. The Cauchy tails are so strong that even mean and sigma does not exist; it looks a bit like a normal distribution with many outliers.

Adding two independent *uniform* variables gives us a triangle distribution; and adding really many independent uniform variables gives us a very good approximation to the *normal* distribution; and this is true for even any infinite sum of random variables; just *finite variances* are required. So even adding e.g., lognormal variables (being quite asymmetric) would end up more and more

Table 4.3 Overview on calculations with independent random variables

X_1	X_2	Operation	Result	Comment
Std-Normal	–	$-X$	Std-Normal	$\mu = 0, \sigma = 1$
Normal	–	$\exp(X)$	Lognormal	$\log(X)$ is not lognormal
Std-Normal	–	$ X $	Half-normal	$\mu = 0, \sigma = 1$
Normal	–	$ X $	Folded-normal	Appear for performances like $ V_{\text{offset}} $
Std-Normal	–	X^2	χ_1^2	Chi-square, important for confidence intervals
Normal	Normal	$X_1 - X_2$	Normal	Mean subtracts, variance adds up still
Uniform	Uniform	$X_1 + X_2$	Triangle	Mean and variance added
Triangle	Uniform	$X_1 + X_2$	Quadratic	Mean and variance added
Cauchy	Cauchy	$X_1 + X_2$	Cauchy	Location and scale add up
Std-Normal	Std-Normal	X_1/X_2	Std-Cauchy	Very wide tails
Half-normal	Half-normal	$X_1 - X_2$	Normal	
Lognormal	Lognormal	$X_1 + X_2$	Not lognormal	New distribution, looking slightly more normal
Std-Normal	Chi	X_1/X_2	Student-T	important for confidence intervals
Std-Uniform	Std-Normal	X_1/X_2	Slash	Similar to Cauchy

in a normal distribution (which is symmetric). Again the Cauchy distribution is special, because it has no finite variance. The central limit theorem CLT will tell us even more, because if we know the mean values and variance of the original distributions, we can calculate the over-all mean and variance just as the sum of the “input” distributions. And the normal distribution with that parameters will often give an excellent fit to the sum distribution. However, this fit is usually only good near the distribution center, not in the tail regions.

Actually on all these things there is quite nice material available in the internet! By creating little testbenches and running MC analysis can find such relationships directly from circuit simulations. For instance, simulating a multiplexer with two inputs driven by normal distributions, you can obtain a Gaussian *mix*. Such mixes are often multimodal, so not normal Gaussian at all. With Verilog-A you can perform almost anything you want, because it also supports random number generation, even for “very” special distributions, and of course it also supports many math functions. Although Verilog-A does not support so many distributions (e.g., no Cauchy distribution), you can often easily create whatever you want with moderate effort by calculations with random variables (see Table 4.3).