

# ECE695: Reliability Physics of Nano-Transistors

## Lecture 33: Model Selection/Goodness of Fit

Muhammad Ashraful Alam  
alam@purdue.edu

# copyright 2013

This material is copyrighted by M. Alam under the following Creative Commons license:



**Attribution-NonCommercial-ShareAlike 2.5 Generic (CC BY-NC-SA 2.5)**

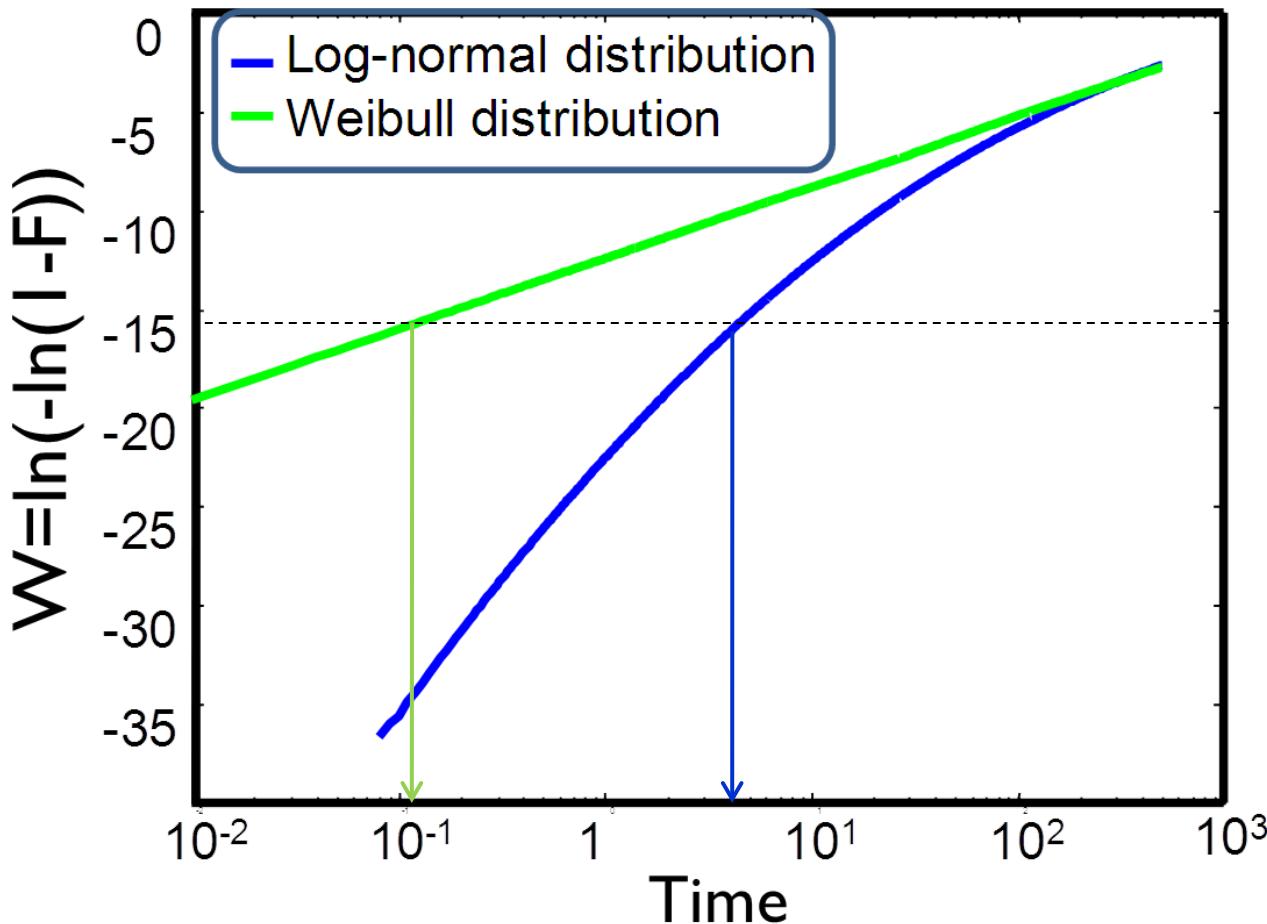
Conditions for using these materials is described at

<http://creativecommons.org/licenses/by-nc-sa/2.5/>

# Outline

- I. The problem of matching data with theoretical distribution
2. Parameter extractions: Moments, linear regression, maximum likelihood
3. Goodness of fit: Residual, Pearson, Cox, Akika
4. Conclusion

# Problem of matching the moments



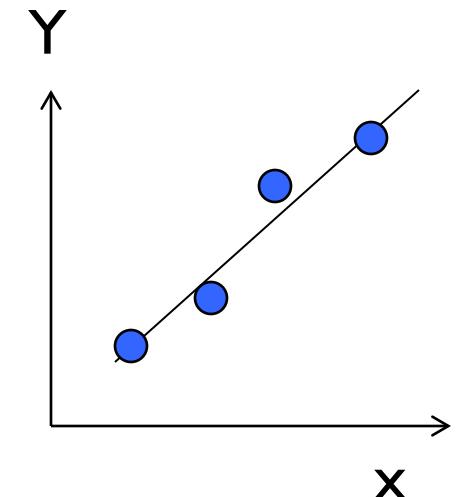
Log-normal distribution is considerably optimistic

# (1) Linear regression: balanced errors

$$W \equiv \ln(-\ln(1 - F)) = \beta \ln t + c$$

Theory:  $y = ax + b$

Data:  $y_i = ax_i + b$



Mimimize  $SSR = \sum_i (y - y_i)^2$

$$a = \left( \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \right) \times \textcolor{red}{D}^{-1}$$

$$b = \left( n \sum x_i y_i - \sum x_i \sum y_i \right) \times \textcolor{red}{D}^{-1}$$

$$\textcolor{red}{D} \equiv n \sum x_i^2 - \left( \sum x_i \right)^2$$

# Uncertainty in regression coefficients

Dependent variable subject to random Gaussian Error of same magnitude at each data point

Theory:  $y = ax + b$

$$\sigma_a^2 = s \left( \sum x_i^2 \right) \times \textcolor{red}{D}^{-1}$$

$$\sigma_b^2 = s \sqrt{n} \times \textcolor{red}{D}^{-1}$$

$$\textcolor{red}{D} \equiv n \sum x_i^2 - \left( \sum x_i \right)^2$$

$$s^2 = \sum (y - y_i)^2 / (n - 2)$$

t-distribution with (n-2) degree of freedom

$$a \pm t_{95\%, (n-2)} \sigma_a$$

$$b \pm t_{95\%, (n-2)} \sigma_b$$

\* Note s and (n-2) ... So called Bessel correction, Because we needed data to calculate a and b.

Use Excel LINEST() function.

# Methods of least squares for weibull

$T_i$  obtained from measurement,

$F_i$  obtained from Hazen or Kaplan-Meier formula.

Define  $E(\alpha, \beta) = \sum_i (F_{i,\text{exp}}(t_i) - F_{i,\text{theroy}}(t_i, \alpha, \beta))^2$

Minimize  $\frac{dE}{d\alpha} = 0, \quad \frac{dE}{d\beta} = 0$

Error and Residual ...

$E(\alpha_0, \beta_0) = \sum_i (F_{i,\text{exp}}(t_i) - F_{i,\text{theroy}}(t_i, \alpha_0, \beta_0))^2$

# Fitting of physical models: challenges

Is the error in  $W$  Gaussian distributed ?

$$W \equiv \beta \ln t + c \quad \ln t \equiv \beta^{-1}W - \beta^{-1}c = a^*W + b^*$$

Inverse fitting is more appropriate ...  $x = a^* + b^*y$

$$a^* = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - (\sum y_i)^2}$$

$$b^* = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\sigma_{\beta}^2 = \left( \frac{\delta \beta}{\delta a} \right)^2 \sigma_a^2 + \left( \frac{\delta \beta}{\delta b} \right)^2 \sigma_b^2$$

$$\sigma_c^2 = \left( \frac{\delta c}{\delta a} \right)^2 \sigma_a^2 + \left( \frac{\delta c}{\delta b} \right)^2 \sigma_b^2$$

# Method of correlation coefficient

$$r = \sqrt{b \times b^*}$$

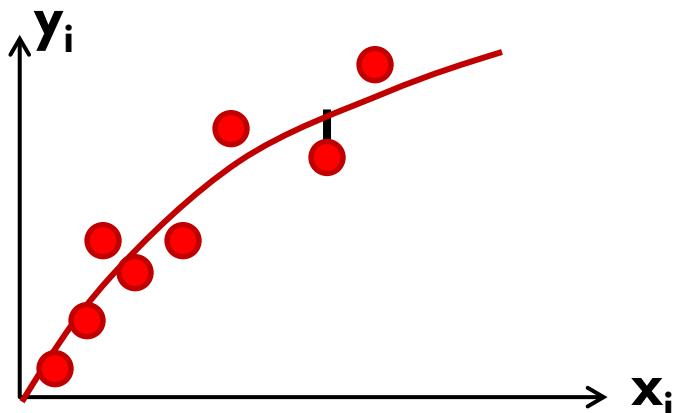
$$y = a + bx$$

$$x = a^* + b^* y$$

Prob. of  $r$  when  $x$ - $y$  are uncorrelated

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - (\sum y_i)^2}$$

$$b^* = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$



$n/r$	0.5	0.7	0.9
3	0.667	0.506	0.287
4			0.1
6			
7			
10	0.141	0.024	

Example. If  $r=0.9$  for  $n=4$ , there is only 10% chance (0.1 value) that this is accidental. If however  $r=0.5$  with  $n=10$ , there is 14.1% chance that it is accidental.

## (2) Fisher's Maximum Likelihood Method

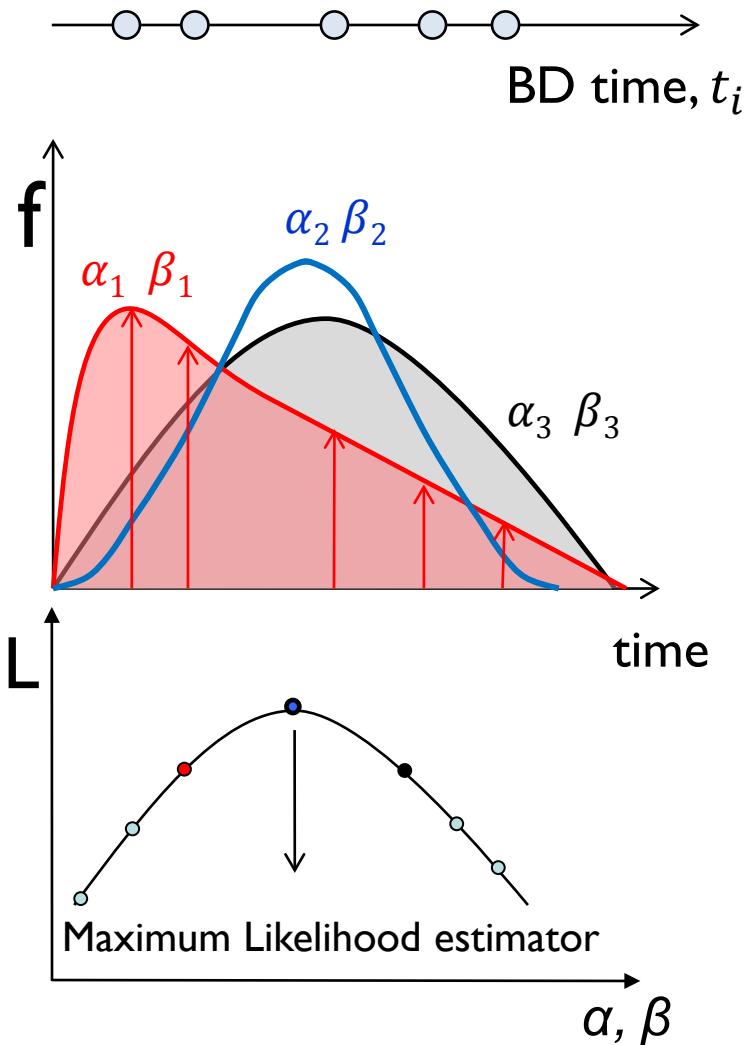
Towering figure, showed that Mendel manipulated data; MATLAB **fitdist** functions

$$f(t_i, \alpha, \beta)$$

$$L = \prod_{i=1}^n f(t_i, \alpha, \beta)$$

$$\ln L = \sum_{i=1}^n \ln f(t_i, \alpha, \beta)$$

$$\frac{d \ln L}{d \alpha} = 0 \quad \frac{d \ln L}{d \beta} = 0$$



# Example: origin of least square method

Let the error around each data point be distributed Normally ...

$$f(y_i, \mu) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - \mu)^2}{2\sigma_i^2}}$$

Then the Likelihood function for this problem is :

$$\begin{aligned} L &= \prod_{i=1}^N f(y_i, \mu) = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - \mu)^2}{2\sigma_i^2}} \\ &= \left[ \frac{1}{\sqrt{2\pi}} \right]^n \frac{1}{\prod \sigma_i} e^{-\frac{(y_1 - \mu)^2}{2\sigma_1^2} - \frac{(y_2 - \mu)^2}{2\sigma_2^2} - \dots - \frac{(y_n - \mu)^2}{2\sigma_n^2}} \\ &= \left[ \frac{1}{\sqrt{2\pi}} \right]^n \frac{1}{\prod \sigma_i} e^{-\sum \frac{(y_i - \mu)^2}{2\sigma_i^2}} \end{aligned}$$

# Example (continued)

$$\left| \frac{\partial \ln L(a,b)}{\partial a} \right| = \frac{\partial}{\partial a} \left[ n \ln(n \ln \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \sum \frac{(y_i - y(x_i))^2}{2\sigma_i^2} \right] = 0$$

$$\frac{\partial}{\partial a} \left[ \sum_{i=1}^n \frac{(y_i - y(x_i))^2}{2\sigma^2} \right] = 0$$
$$\frac{\partial}{\partial b} \left[ \sum_{i=1}^n \frac{(y_i - y(x_i))^2}{2\sigma_i^2} \right] = 0$$

$$a = \left( \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \right) \times \textcolor{red}{D}^{-1}$$
$$b = \left( n \sum x_i y_i - \sum x_i \sum y_i \right) \times \textcolor{red}{D}^{-1}$$
$$\textcolor{red}{D} \equiv n \sum x_i^2 - \left( \sum x_i \right)^2$$

Linear fit is a special case of MLE with requirement that error is distributed normally

# Example: MLE estimator for one-parameter distribution

$$f(t; K) = K \times t \times e^{-Kt^2/2}$$

$$\begin{aligned} L &= \left( Kt_1 e^{-Kt_1^2/2} \right) \times \left( Kt_2 e^{-Kt_2^2/2} \right) \times \left( Kt_3 e^{-Kt_3^2/2} \right) \times \left( Kt_4 e^{-Kt_4^2/2} \right) \times \dots \\ &= K^n \left( \prod_{i=1}^n t_i \right) \exp \left( -\frac{K}{2} \sum_{i=1}^n t_i^2 \right) \end{aligned}$$

$$\ln L = n \ln K + \sum_{i=1}^n \ln t_i - \frac{K}{2} \sum_{i=1}^n \ln t_i^2$$

$$\frac{d \ln L}{dK} = 0 \quad \Rightarrow \quad K = 2n \sqrt{\sum_{i=1}^n t_i^2}$$

# Example: MLE estimator for Weibull

Recall  $f(t; \alpha, \beta) = \frac{\beta}{\alpha^\beta} \cdot t^{\beta-1} \cdot e^{-\left(\frac{t}{\alpha}\right)^\beta}$

$$\begin{aligned}\ln L &= \sum_{i=1}^n \ln f(t_i, \alpha, \beta) \\ &= n \ln \beta - n \ln \alpha + (\beta - 1) \sum_{i=1}^n \ln t_i / \alpha - \sum_{i=1}^n (t_i / \alpha)^\beta\end{aligned}$$

$$\frac{d \ln L}{d \alpha} = 0 \quad \frac{d \ln L}{d \beta} = 0$$

$$\left( \sum_{i=1}^n t_i^{\alpha} \ln(t_i)^{\beta} \middle/ \sum_{i=1}^n t_i^{\beta} \right) - \frac{1}{n} \sum_{i=1}^n \ln(t_i)^{\beta} = 1 \quad \alpha = \left[ \frac{1}{n} \sum_{i=1}^n t_i^{\beta} \right]^{\frac{1}{\beta}}$$

Solve for unknowns  $\alpha, \beta$

# HW: MLE for Log-Normal

$$f(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{\{\ln(t) - \ln(\mu)\}^2}{2\sigma^2}\right]$$

$$\ln L = \sum_{i=1}^n \ln f(t_i, \alpha, \beta)$$

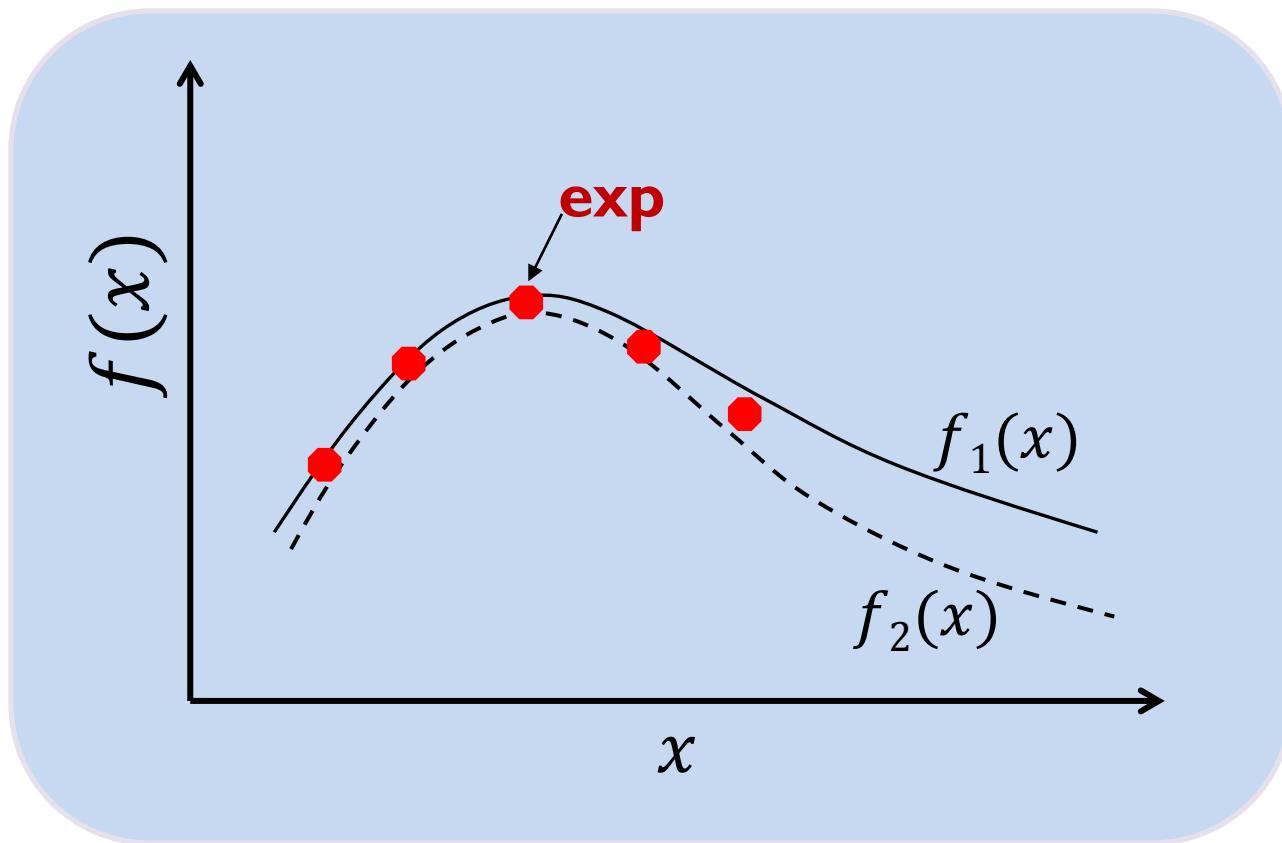
$$\frac{d \ln L}{d\alpha} = 0 \quad \frac{d \ln L}{d\beta} = 0$$

$$\frac{\sum_{i=1}^n t_i^\beta \ln(t_i)^\beta}{\sum_{i=1}^n t_i^\beta} - \frac{1}{n} \sum_{i=1}^n \ln(t_i)^\beta = 1 \quad \alpha = \left[ \frac{1}{n} \sum_{i=1}^n t_i^\beta \right]^{\frac{1}{\beta}}$$

# Outline

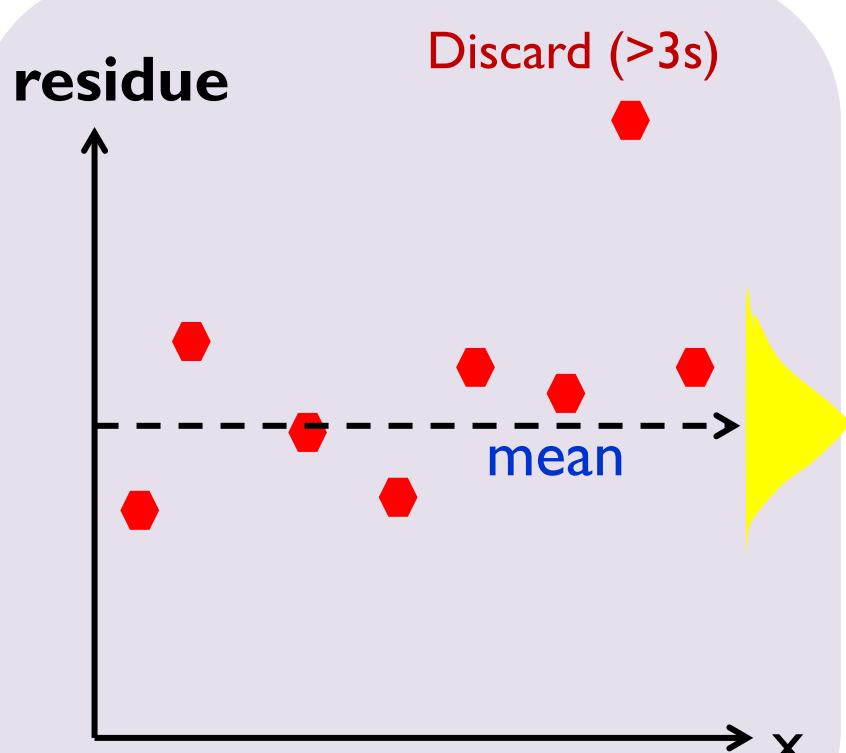
- I. Introduction: The problem of matching data with theoretical distribution
2. Parameter extractions: Moments, linear regression, maximum likelihood
3. Goodness of fit: Residual, Pearson, Cox, Akika
4. Conclusion

# (1) Goodness of Fit: First check visually

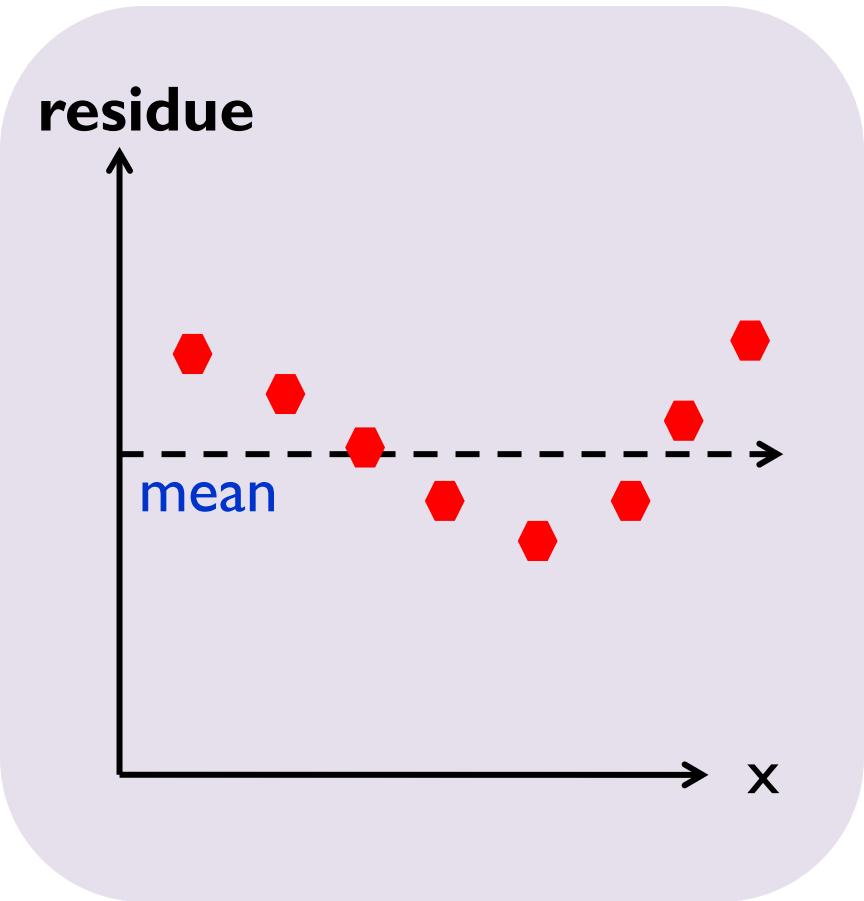


Statistical analysis is helpful only when there is  
a intuitive feel that the fit looks good ...

## (2) Goodness of Fit: Residual method

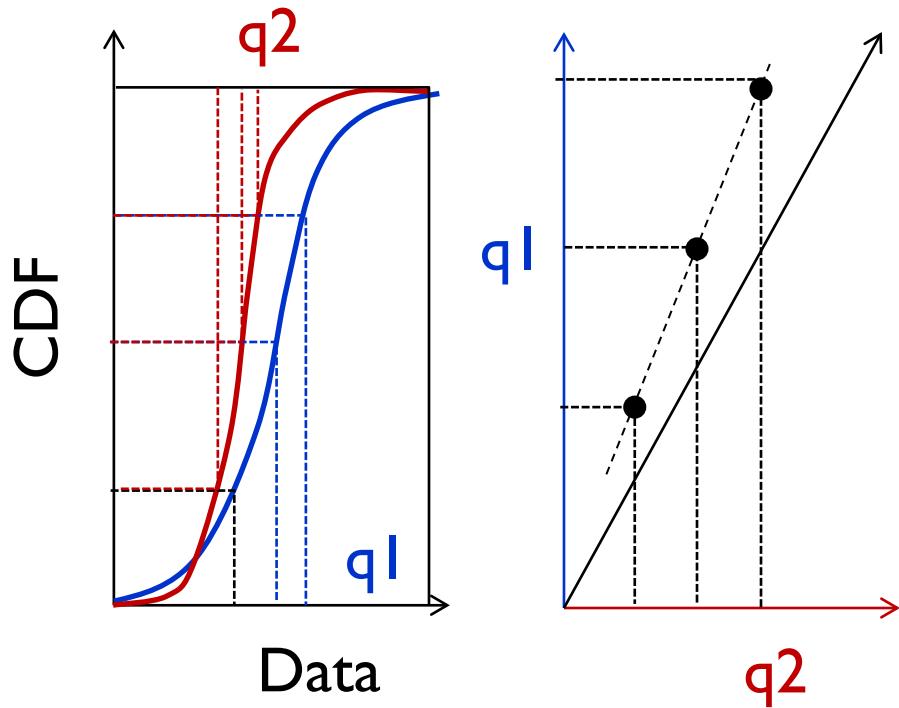


## A good fit (normal distribution of residue)



## A bad fit (systematic distribution in residual)

### (3) Goodness of fit: Q-Q Method



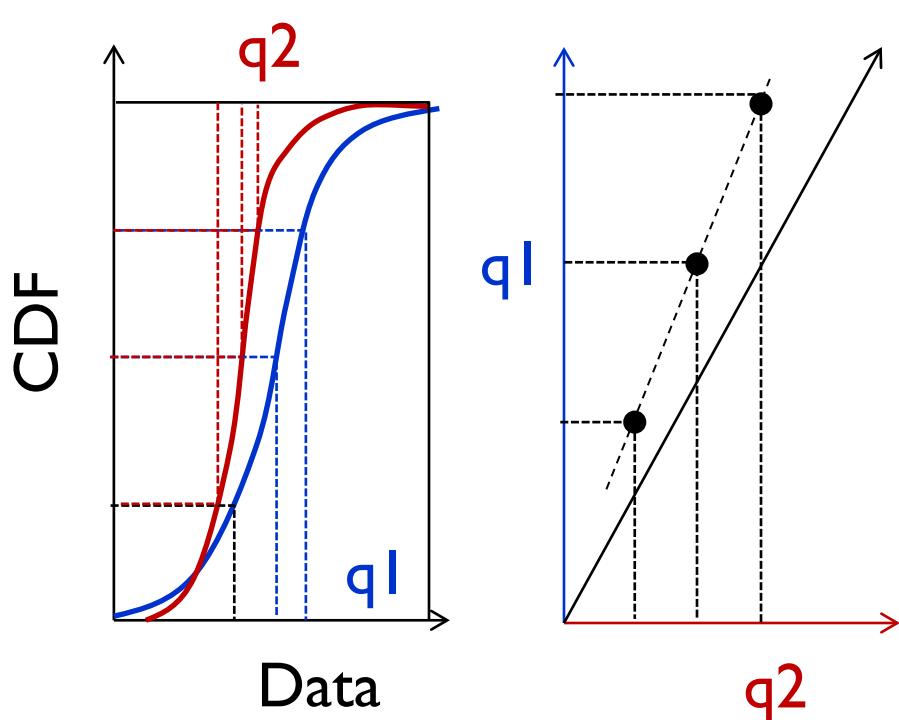
q-Quantile and quartile are different things. Median is 2-quantile, Quartile is 4-quantile, decile is a 10-quantile, percentile is 100-quantile, etc.

Take the q-quantile values of the original data and plot in the y-axis.

Take the q-quantile values of the test-distribution (i.e., calculate  $x=F(-1)(q)$ ) to define the x-axis.

Visually inspect and establish deviation from linearity.

# Q-Q Method: An example



$\{3, 6, 7, 8, 8, 10, 13, 15, 16, 20\}$

$10 \times 1/4 \sim 3$        $10 \times 2/4 \sim 5$        $10 \times 3/4 \sim 8$

$q_1 = (7, 9, 15)$

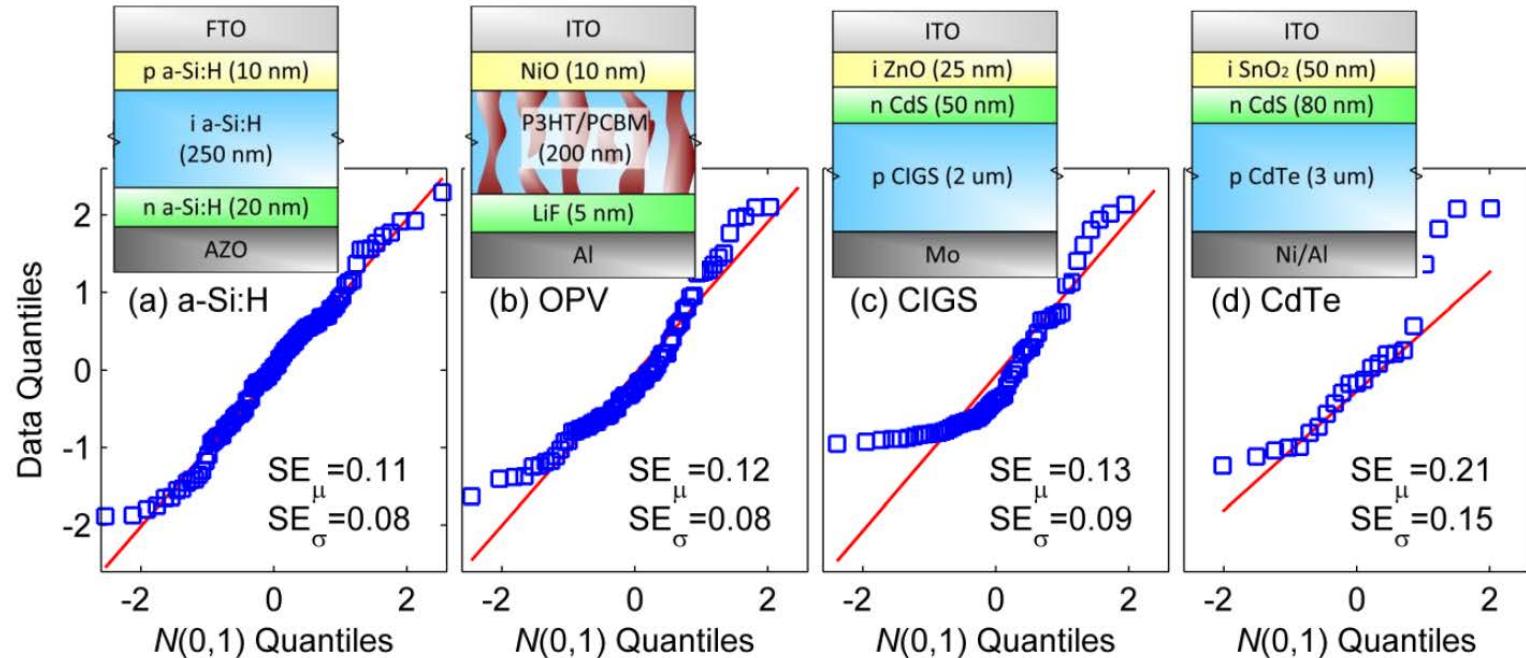
Exponential distribution

$$Q(p; \lambda) = \frac{-\ln(1 - p)}{\lambda},$$

$$\left. \begin{array}{l} \text{1st q: } \ln(4/3)/\lambda \\ \text{2nd q: } \ln(2)/\lambda \\ \text{3rd q: } \ln(4)/\lambda. \end{array} \right\}$$

Optimize  $\lambda$

# Q-Q method: an example



Data against log-normal plot: Optimize  $(\mu, \sigma)$

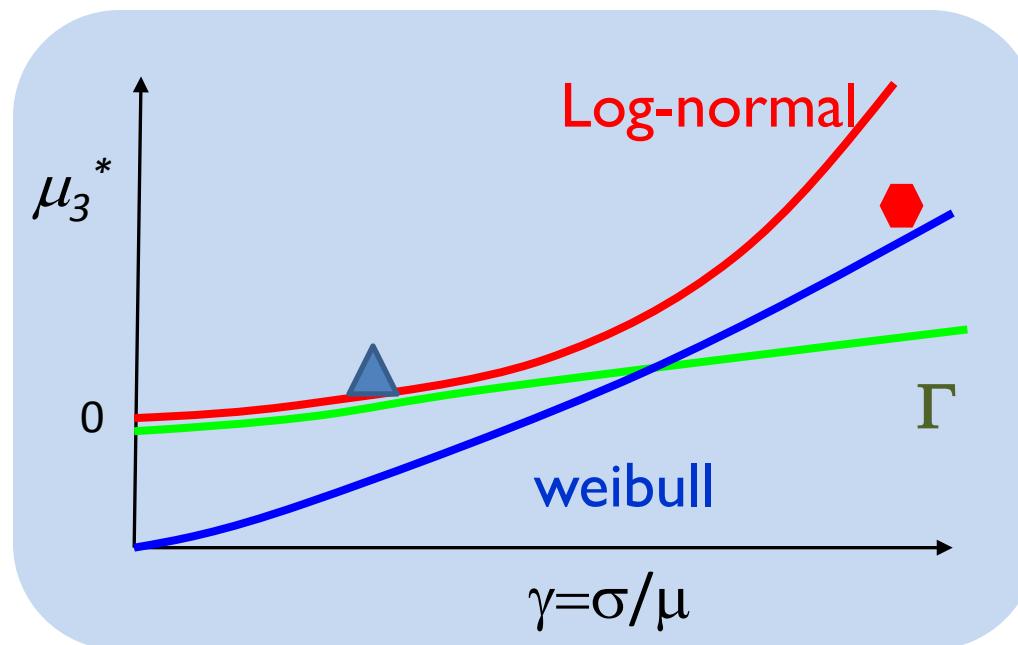
## (4) Goodness of Fit: Cox-Oakes measure

$$\mu_3^* = \frac{\mu_3}{\sigma^3} \quad \mu = \frac{1}{n} \sum_{i=1}^n t_i \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (t_i - \mu)^2$$

*Solid lines are known for various distributions.*

Example. For a given  $\alpha, \beta$ , Weibull has a specific  $\mu, \sigma, \mu_3$  (blue triangle)

Logic: Every distribution has different shape.



## (5) Kolmogorov-smirnov algorithm

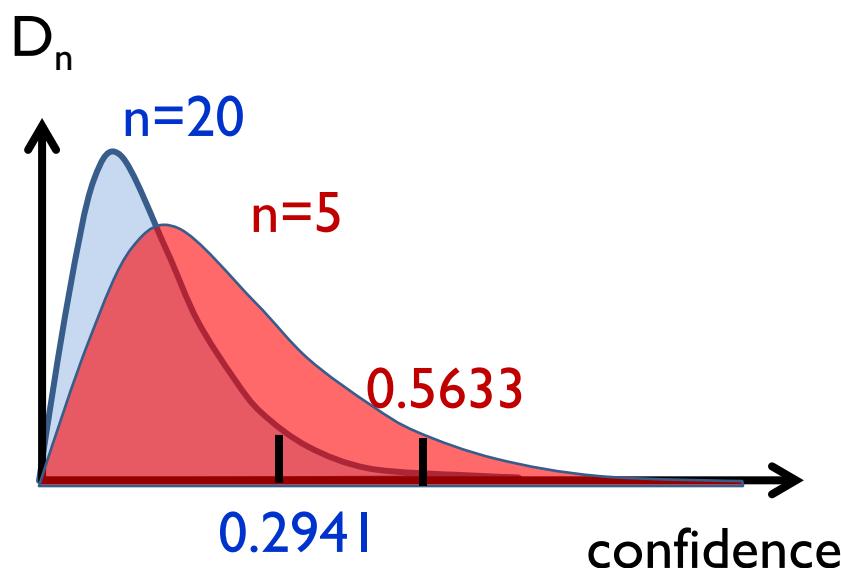
Compute ...  $D_n = \max \left| F_{obs}(t_i) - F_{theory}(t_i) \right|$

Sample size

If  $D_n > D_n^{crit}$ , fit is poor ...

5% significance level

n	$D_{crit}(n)$
5	0.5633
10	0.4092
20	0.2941
50	0.1884



# Example: Kolmogorov-Smirnov Test

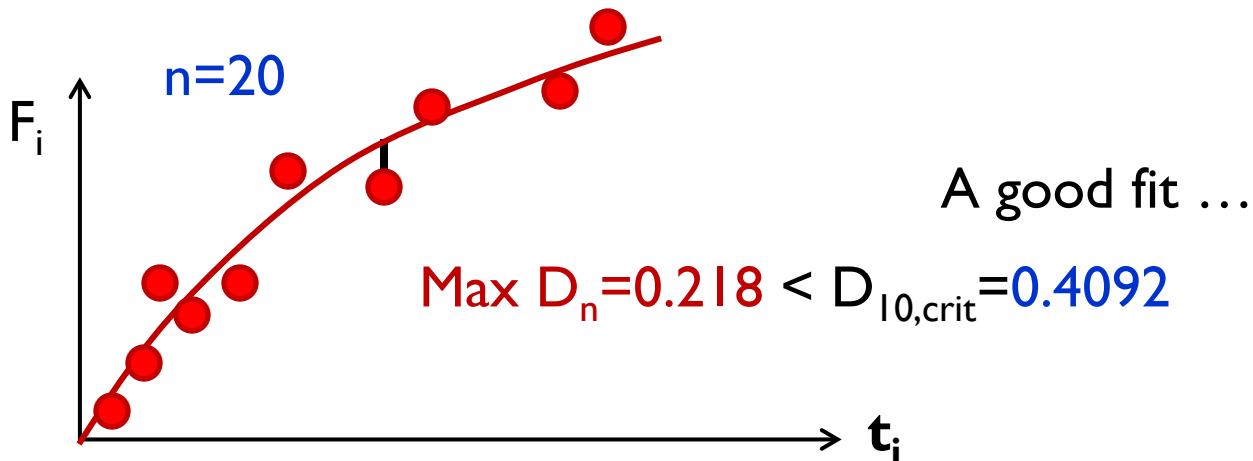
Compute ...  $D_n = \max |F_{obs}(t_i) - F_{theory}(t_i)|$

Sample size

If  $D_n > D_n^{crit}$ , fit is poor ...

<b>n</b>	<b>D<sub>crit</sub>(n)</b>
5	0.5633
10	0.4092
20	0.2941
50	0.1884

t <sub>i</sub>	3	20	40	52	53	54	85	318	429	553
F <sub>i</sub> =(i-.3)/(n+.4)										



[h,p] = kstest(x,'CDF',test\_cdf,'Alpha',0.01)

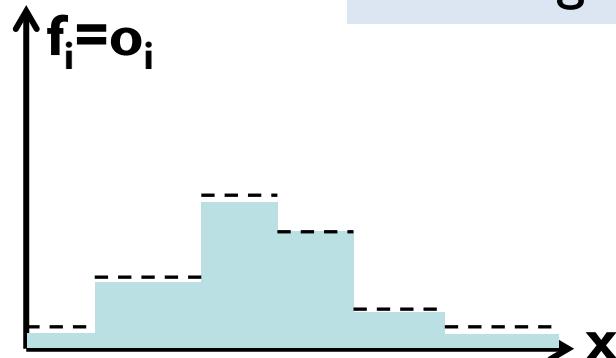
# (6) Pearson $\chi^2$ – test algorithm

Calculate ...

$$\chi_s^2 = \sum_{i=1}^{n^*} \frac{(o_i - e_i)^2}{e_i}$$
$$e_i = n^* \times p_i$$

v	5% ( $\chi^2$ )
2	5.99
4	9.49
10	18.307
20	27.68

- o ... Observed
- e ... expected
- n\* ... datapoints
- p<sub>i</sub> .... probability
- v ... deg. of freedom



If the value observed  $\chi^2$  value exceeds critical value, the fit is poor.

# A famous example: Schon story

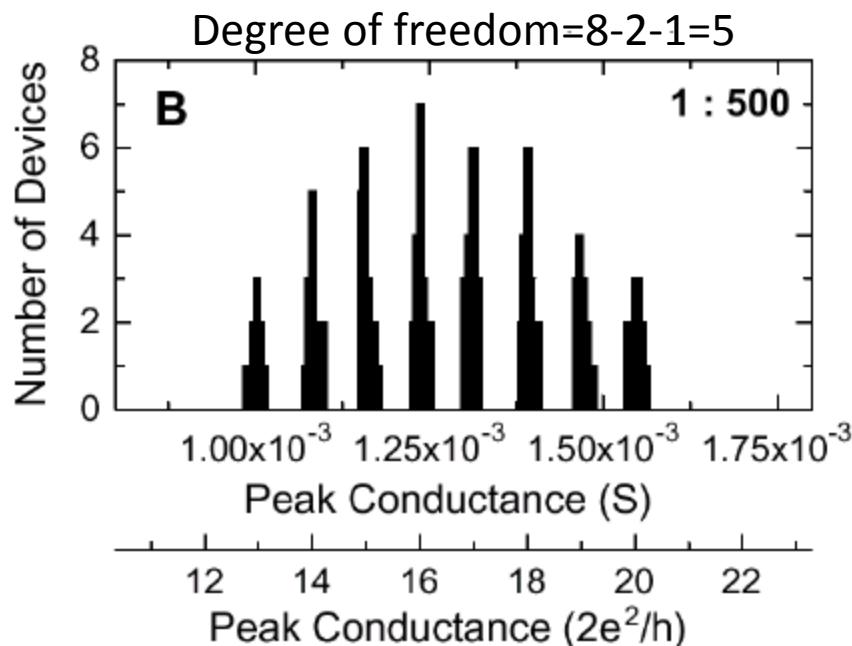
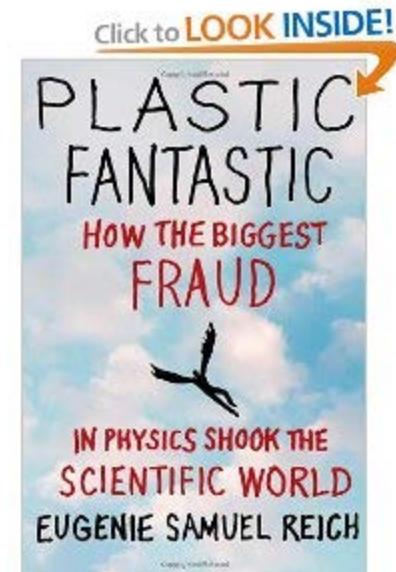
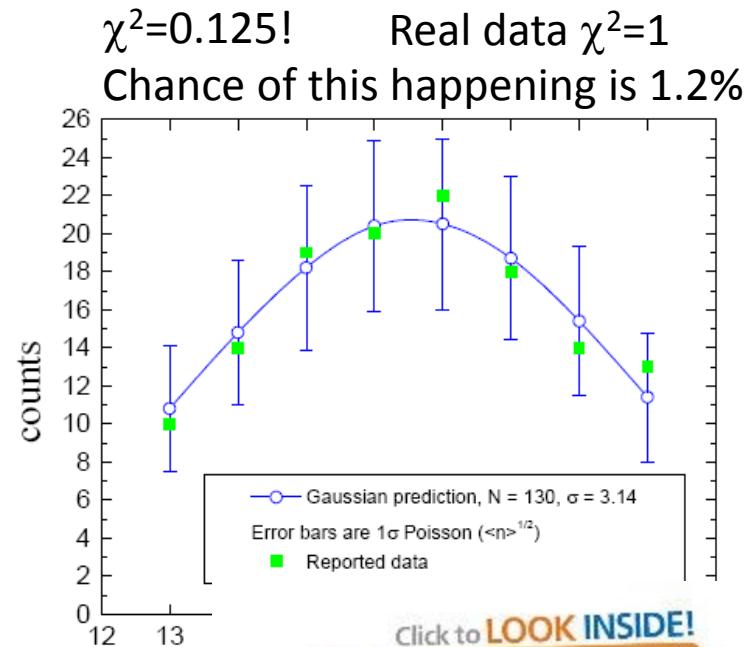


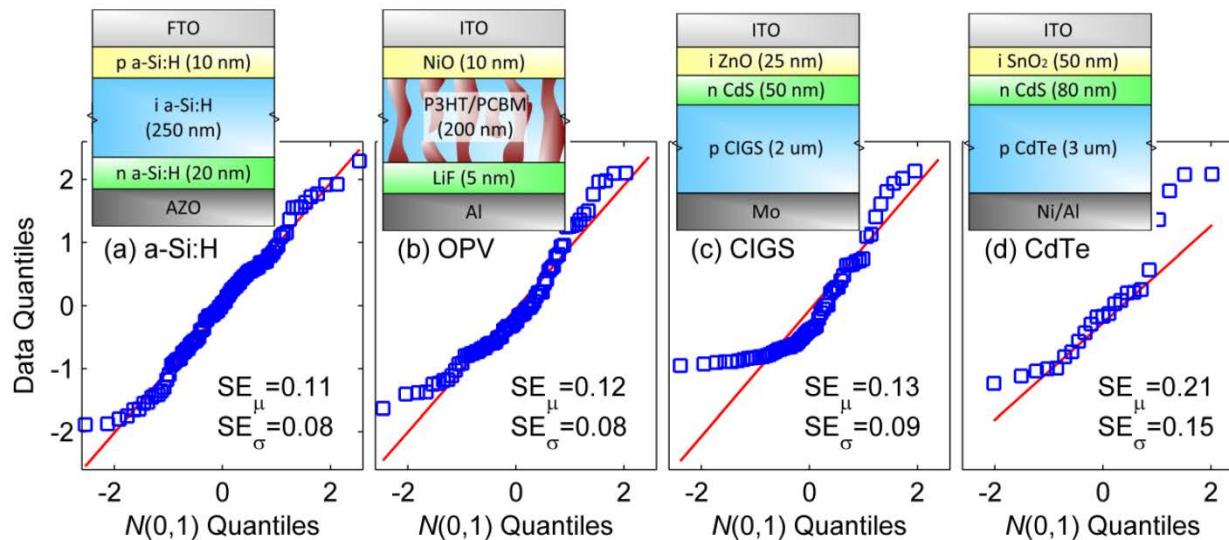
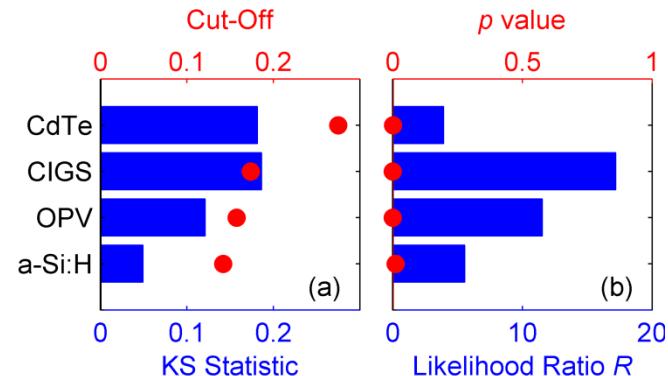
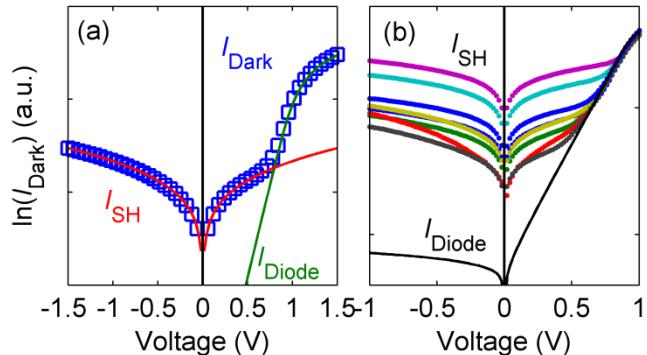
Figure 46. Figure 3(B) from "SingleMolecule" Paper (XIII), showing a histogram of conductances from diluted SAMFETs,

The data indicating conductance quantization did not arise from an objective measurement process. At a minimum, the assignment of conductance values was color-coded by the expected shape of the final distribution. Such a biased process cannot provide convincing evidence for quantization. The response to this concern appears to be deliberately deceptive, suggesting that this misrepresentation was intentional.

The preponderance of evidence indicates that Hendrik Schön committed scientific misconduct, specifically data fabrication, in this case. Alam ECE 695



# Example: Shunt statistics is log-normal



# Parameter number vs. goodness of fit

$n$  = number of samples,  $M$ =number of parameters

1) Method of adjusted residual ...

$$R_{adj}^2 = \frac{(n-1)R^2 - (M-1)}{n-M}$$

$$R \equiv \sum_{n=1}^n (t_i - t_{i,fit})^2$$

2) Akaike Information Criterion

$$AIC = n \times \ln R^2 + 2M$$

Error penalty  
Parameter Penalty

2) Schwarz Information Criterion

$$BIC = n \times \ln R^2 + M \times \ln n$$

# Conclusions

- I. Assuming that we have a correct distribution, it is very important to extract the parameters of the distribution and their error bars as accurately as possible.
2. This must be followed by a rigorous check to see if the data and the fit are statistically significant. Fisher found that Mendel's analysis was incorrect, the analysis that confirmed Einstine's general theory of relativity was flawed, and I found that Schon was cheating (see the book 'Plastic Fantastic').
3. Once you are convinced that fit is reasonable, go back to these checks as more data are collected. It is not uncommon to find that while small dataset supported the possibility of statistical significant, statistical significance is lost as more data are collected.

# References

Linda C. Wolsterholme, "Reliability Modeling – A Statistical Approach, Chapman Hall, CRC, 1999. Chapter 1-7 has excellent summary of 'Goodness of Fit' analysis.

F. Yates, "Sir Ronald Fisher and the Design of Experiments", *Biometrics*, vol. 20, no. 2, In Memoriam: Ronald Aylmer Fisher, 1890-1962., pp. 307-321, (Jun. 1964).

F. J. Massey Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," *SIAM Review*, vol. 51, no. 4, p. 661, Nov. 2009.

F. J. Massey Jr., "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

- Fisher's reservation about Mendel is discussed in "Ending The Mendel-Fisher Controversy", Allan Franklin, A. W. F. Edwards, D. J. Fairbanks, D. L. Hartl, and T. Seidenfeld, Univ. Pittsburgh Press, 2008. Also, see CSI: Mendel, S. M. Stigler, *American Scientist*, Sept. 2008.
- Several other interesting papers, "Making Reliability Estimates When Zero Failures Are Seen in Laboratory Aging", F. R. Nash, AT&T Memo, 52321-900221-11-TM, 1988.
- "Formulas to Describe the Bias and Standard Deviation of ME-estimated Weibull Shape Parameters" R. Ross, IEEE TDEI, 1(2), p. 247, 1994. ibid, TDEI 3(1), 1996.
- The example of KS analysis was taken from "Introduction to Probability and Mathematical Statistics", L. J. Bain, Ex. 13.8.3.
- A nice discussion of Akaike's Information Criteria is found in "Making Sense out of Akaike's Information Criterion", Mark. J. Mazerolle, Appendix I: <http://www.theses.ulaval.ca/2004/21842/apa.html>

# Review Questions

- I. With higher number of model parameters, you can always get a good fit – why should you minimize the number of parameters?
2. Least square method is a subset of maximum likelihood approach to data fitting. Is this statement correct?
3. What aspect of the distribution function does Cox-Oakes method emphasize?
4. Can MLE be used for any distribution function?
5. How would you change the MLE condition if you had 3 independent parameters to estimate?
6. Does increase in model parameters increase chances of passing  $\chi^2$  test?
7. How does the methods affected by censored data (e.g., TDDB test yet to finish?)