Primer on Analysis of Experimental Data and Design of Experiments

Lecture 4. Model Selection and Goodness of Fit

Muhammad A. Alam
alam@purdue.edu

Outline

1. The problem of matching data with theoretical distribution

2. Parameter extractions: Moments, linear regression, maximum likelihood

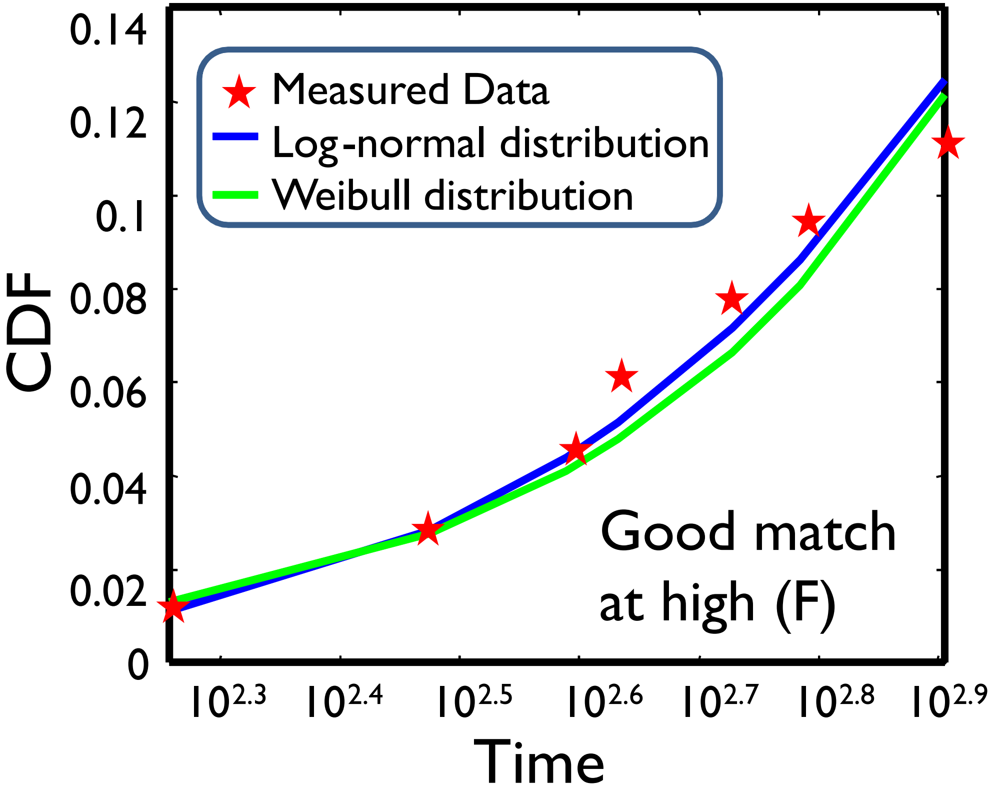
3. Goodness of fit: Residual, Pearson, Cox, Akika

4. Conclusion

Matching moments to distributions

Of 60 oxides, 7 failed in 1000 hrs

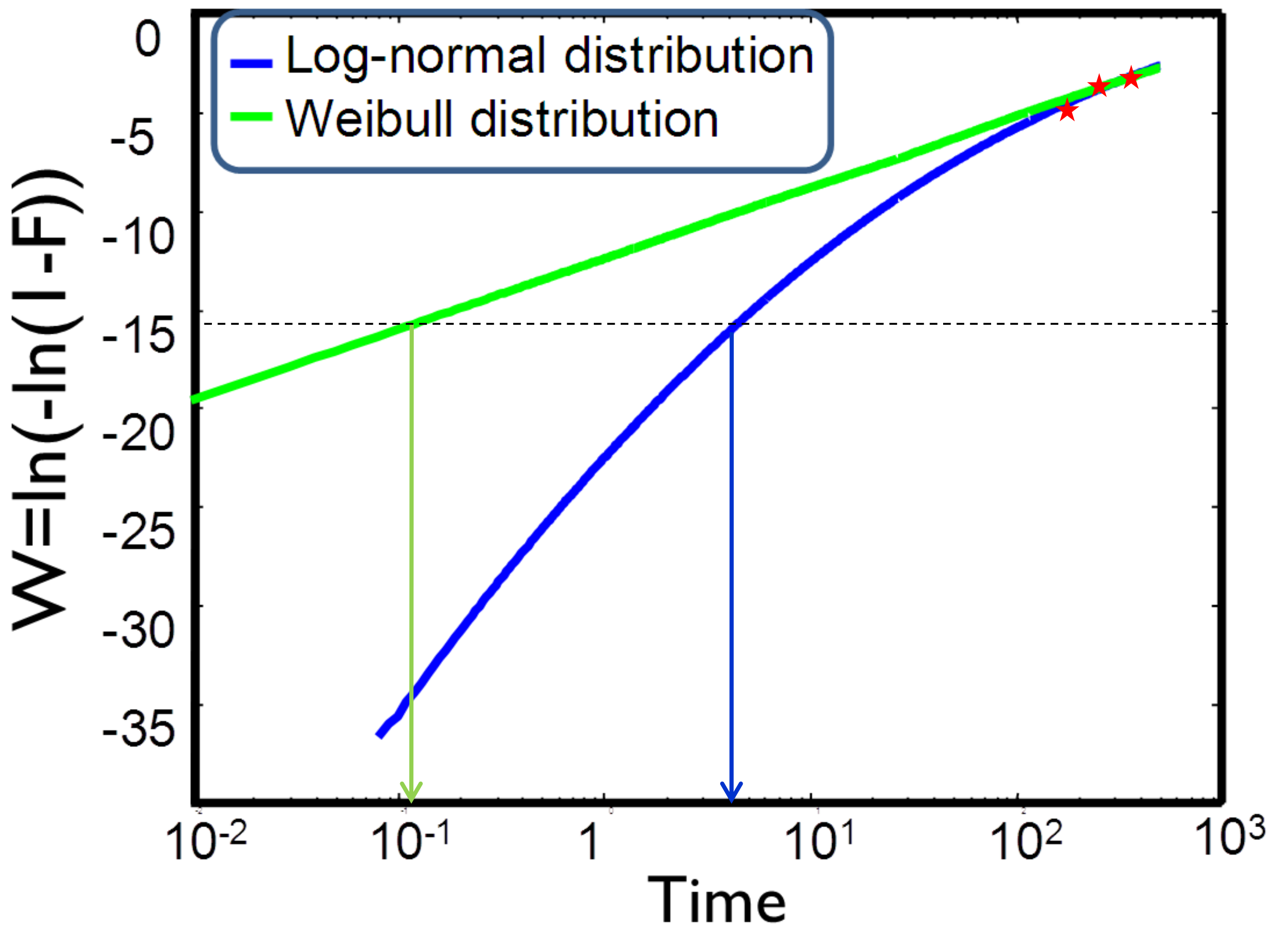
Rank	Lifetime	$F_i = (i - 0.3)/(n + 0.4)$
1	181	0.012
2	299	0.028
3	389	0.045
4	430	0.061
5	535	0.078
6	610	0.094
7	805	0.111



Weibull Distribution Parameters
When $t=\alpha$, $\ln(1-F(t))=-1$, $F(t)=0.632$, $a=2990$
β estimated using parameter fitting as 1.56

Log-Normal Distribution Parameters
$s=\ln(T_{50\%}/T_{15.9\%})$, $\sigma=\ln(3600/980)=1.30$
$\mu=\ln(T_{50\%})=\ln(3600)=8.19$

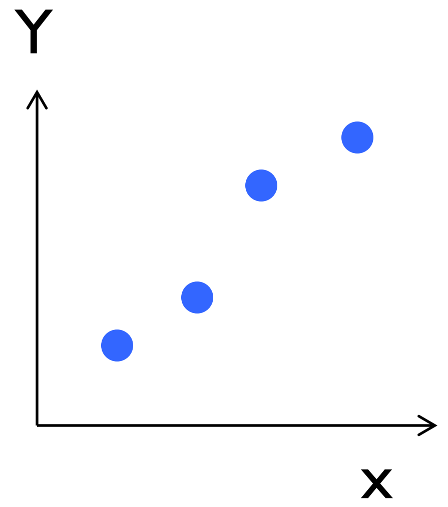
Problem of matching the moments



Log-normal distribution is considerably optimistic

(1) Linear regression: balanced errors

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



$$\text{Theory:} \quad y = ax + b$$

$$\text{Data:} \quad y_i = ax_i + b$$

$$\text{Miminize} \quad SSR = \sum\nolimits_i (y - y_i)^2$$

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

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

$$D \equiv n \sum x_i^2 - \left(\sum x_i\right)^2$$

Uncertainty in regression coefficients

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

Theory: $y = ax + b$

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

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

$$D \equiv n\sum x_i^2 - \left(\sum x_i\right)^2$$

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

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(y,x,c, stat) function.

Methods of least squares for weibull

T_i obtained from measurement,
F_i obtained from Hazen or Kaplan-Meier formula.

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

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

Error and Residual …

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

Fitting of physical models: challenges

Is the error in W Gaussian distributed ?

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

Inverse fitting is more appropriate … $\quad x = a^* + b^* y$

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

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

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

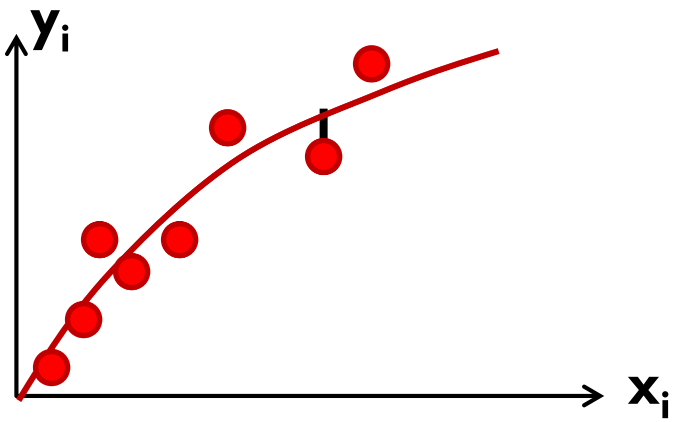
$$\sigma_c^2 = \left(\frac{\delta c}{\delta a}\right)^2 \sigma_a^2 + \left(\frac{\delta c}{\delta b}\right)^2 \sigma_b^2$$

Method of correlation coefficient

$$r = \sqrt{b \times b^*} \qquad y = a + bx \qquad x = a^* + b^* y$$

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

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



Prob. of r when x-y are uncorrelated

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

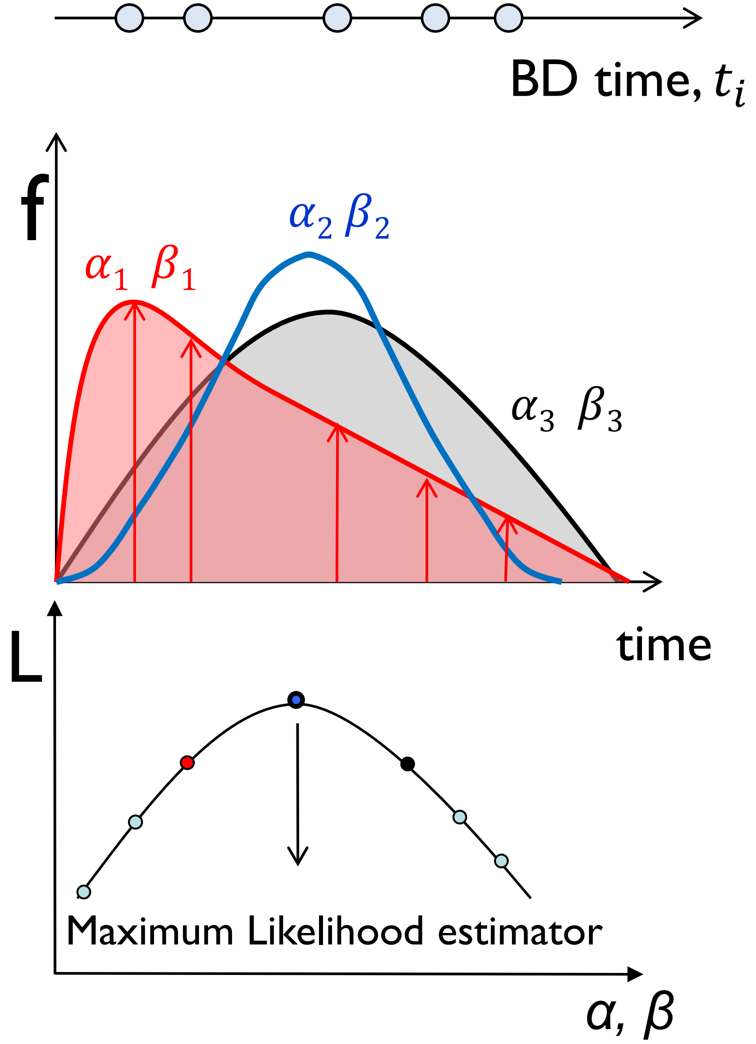
Towering figure, showed that Mendel manipulated data; MALAB fitdist functions

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

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

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

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



Example: origin of least square method

Let the error around each data point be distributed Normally …

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

Then the Likelihood function for this problem is :

$$L=\prod_{i=1}^{N} f(y_i,\mu)=\prod_{i=1}^{N}\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{(y_i-y(x_i))^2}{2\sigma_i^2}}$$

$$=\left[\frac{1}{\sqrt{2\pi}}\right]^n\frac{1}{\prod\sigma_i}e^{-\frac{(y_1-y(x_1))^2}{2\sigma_1^2}}e^{-\frac{(y_2-y(x_2))^2}{2\sigma_1^2}}\ldots e^{-\frac{(y_3-y(x_3))^2}{2\sigma_3^2}}$$

$$=\left[\frac{1}{\sqrt{2\pi}}\right]^n\frac{1}{\prod\sigma_i}e^{-\sum\frac{(y_i-y(x_i))^2}{2\sigma_i^2}}$$

Example (continued)

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

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

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

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

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

$$D \equiv n\sum x_i^2 - \left(\sum x_i\right)^2$$

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

Example: MLE estimator for one-parameter distribution

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

$$L = \left(Kt_1 e^{-Kt_1^2/2}\right) \times \left(Kt_2 e^{-Kt_2^2/2}\right) \times \left(Kt_3 e^{-Kt_3^2/2}\right) \times \left(Kt_4 e^{-Kt_4^2/2}\right) \times \ldots$$

$$= K^n \left(\prod_{i=1}^{n} t_i\right) \exp\left(-\frac{K}{2}\sum_{i=1}^{n} t_i^2\right)$$

$$\ln L = n \ln K + \sum_{i=1}^{n} \ln t_i - [\text{3rd term ?}]$$

(A) $\frac{K}{2}\sum_{i=1}^{n} \ln t_i^{\,2}$ (B) $\frac{K}{2}\sum_{i=1}^{n} t_i^{\,2}$ (C) $\mathrm{K}\sum_{i=1}^{n} t_i^{\,2}/2$

Example: MLE estimator for one-parameter distribution

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

$$L = \left(Kt_1 e^{-Kt_1^2/2}\right) \times \left(Kt_2 e^{-Kt_2^2/2}\right) \times \left(Kt_3 e^{-Kt_3^2/2}\right) \times \left(Kt_4 e^{-Kt_4^2/2}\right) \times \ldots$$

$$= K^n \left(\prod_{i=1}^{n} t_i\right) \exp\left(-\frac{K}{2}\sum_{i=1}^{n} t_i^2\right)$$

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

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

Example: MLE estimator for Weibull

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

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

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

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

Solve for unknowns α, β

HW: MLE for Log-Normal

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

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

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

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

Outline

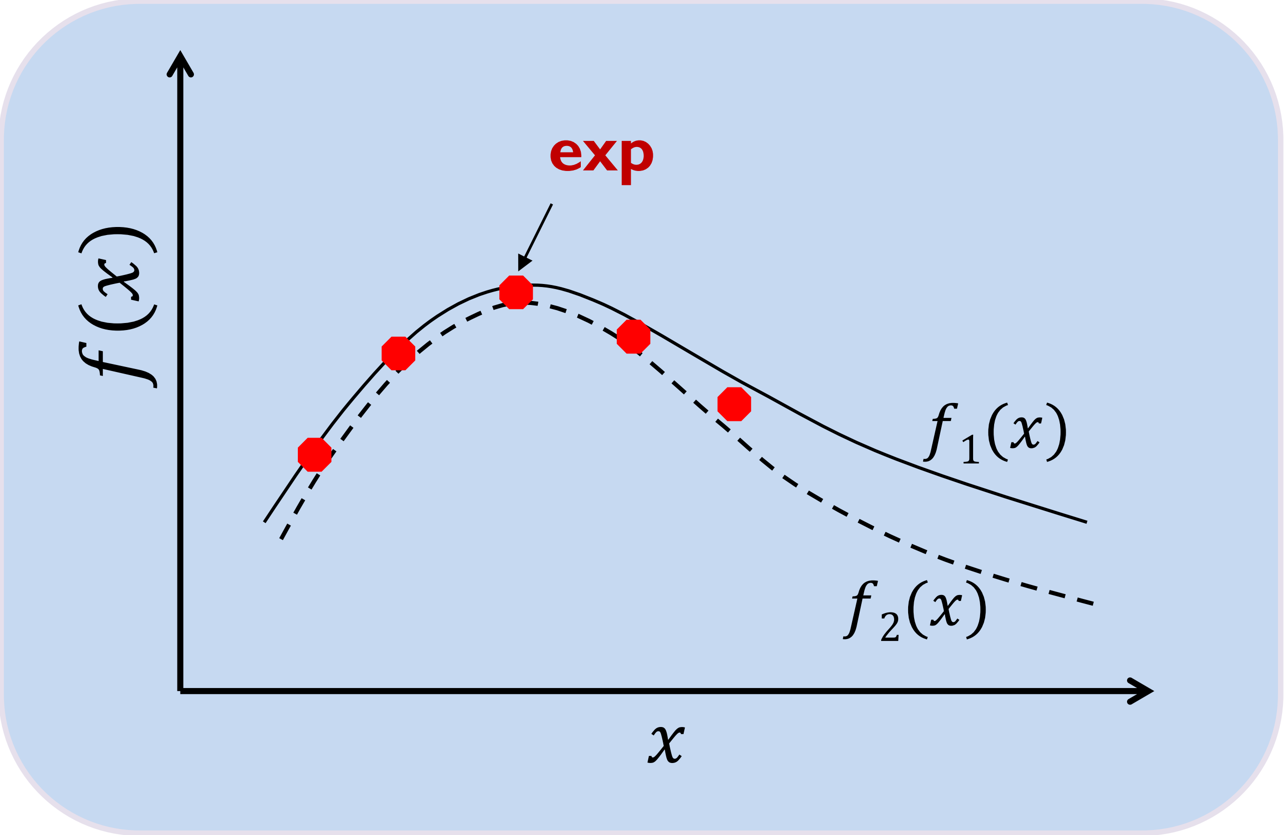
1. Introduction: The problem of matching data with theoretical distribution

2. Parameter extractions: Moments, linear regression, maximum likelihood

3. Goodness of fit: Residual, Pearson, Cox, Akika

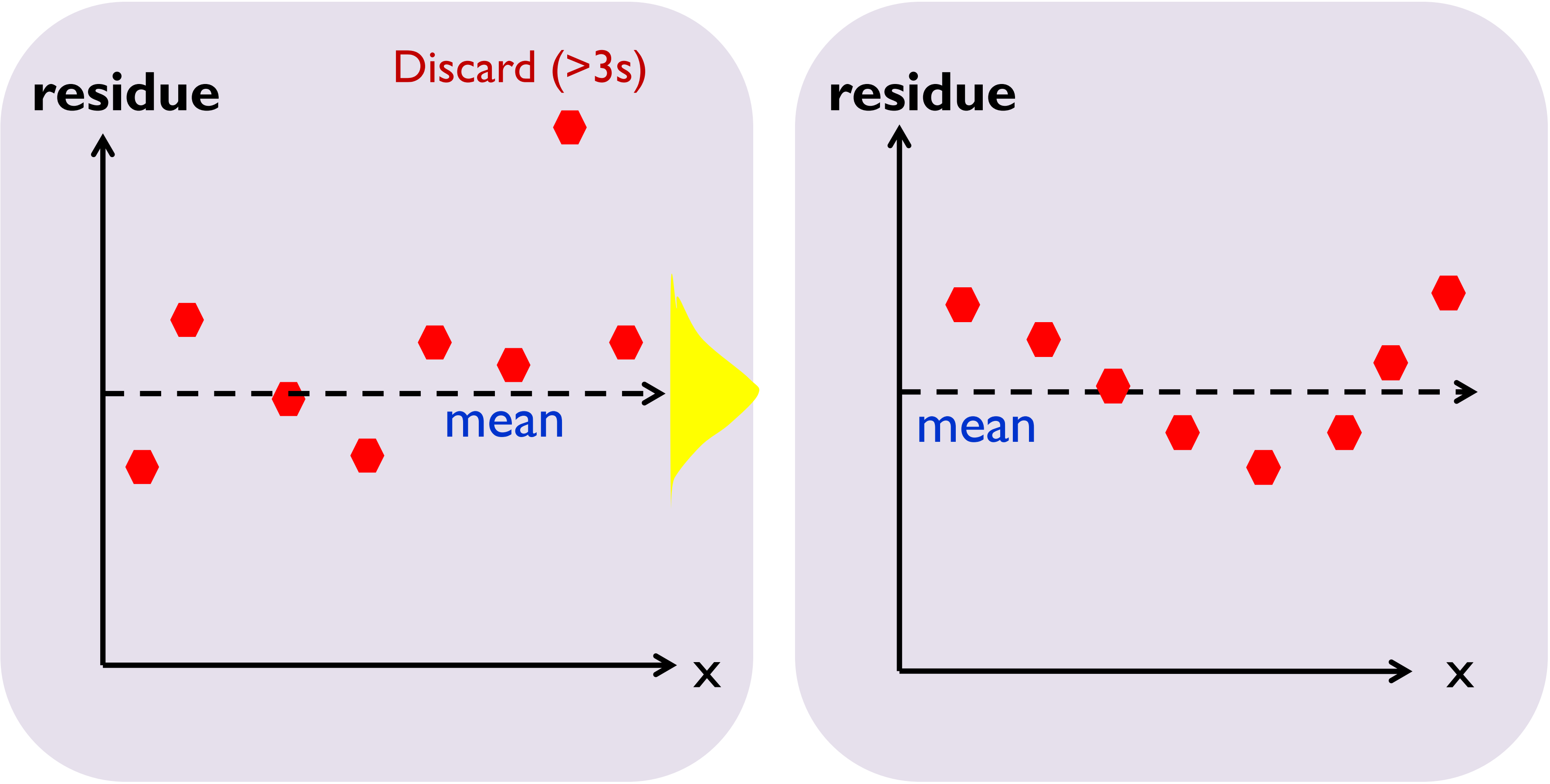
4. Conclusion

(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) Q-Q Method: An example

Data: {3, 6, 7, 8, 8, 10, 13, 15, 16, 20}

What is the first quartile point? (A) 3 (B) 7 (C) 8 (D) 10

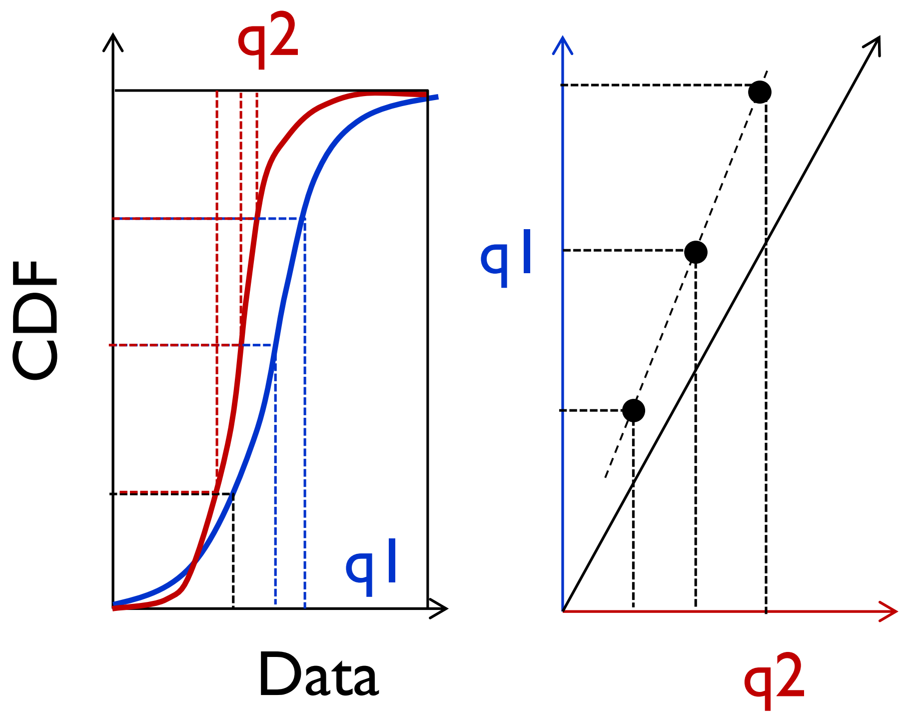
What is the median point? (A) 3 (B) 8 (C) 10 (D) 16

What is the third quartile point? (A) 13 (B) 15 (C) 16 (D) 20

Exponential distribution $Q_2(p, K) = -\ln(1-p)/K$

What is the 2nd quartile point? (A) ln(1.33)/k, (B) ln(2)/K, (C) ln 4/k

(3) Goodness of fit: Q-Q Method



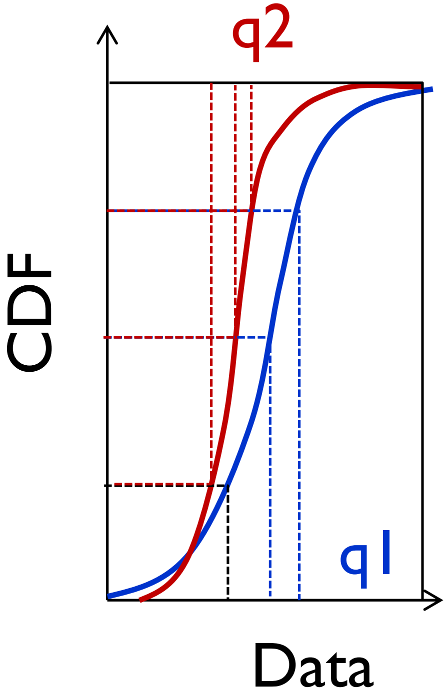
q-Quantile and quartile are different things. Median is 2-quantile, Quartile is 4-quantile, decile is a 10-quantile, percentile is 100-quantile, etc.

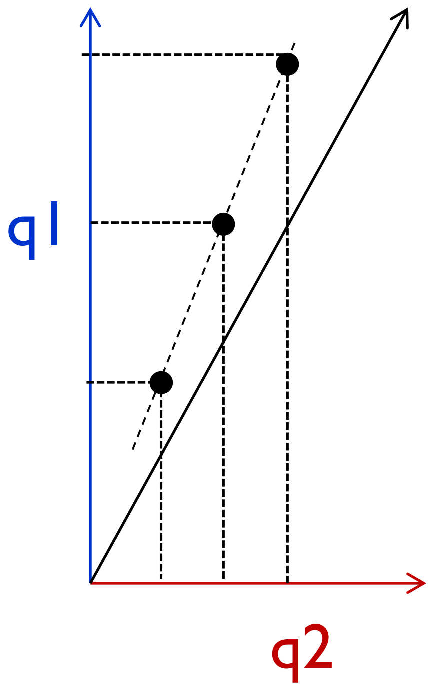
Take the q-quantile values of the original data and plot in the y-axis.

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





median

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

10x1/4~3 10x2/4~5 10x3/4~8

Q1 points=(7,9,15)

Q2: Exponential distribution

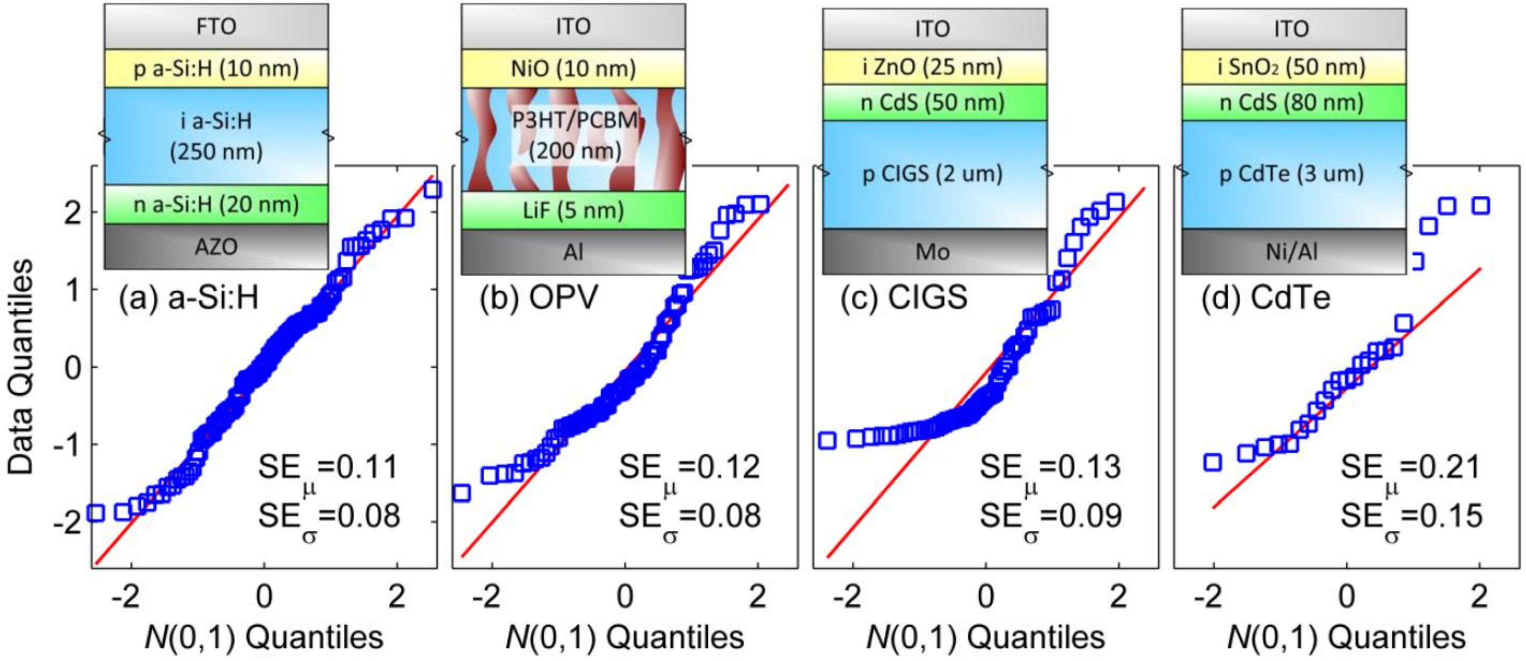
$$Q_2(p,K) = -\ln(1-p)/K$$

1st q: $\ln(4/2)/K$

2nd q: $\ln(4/3)/K$ Given K

3rd q: $\ln(4/1)/K$

Q-Q method: an example



Data against log-normal plot: Optimize (μ, σ)

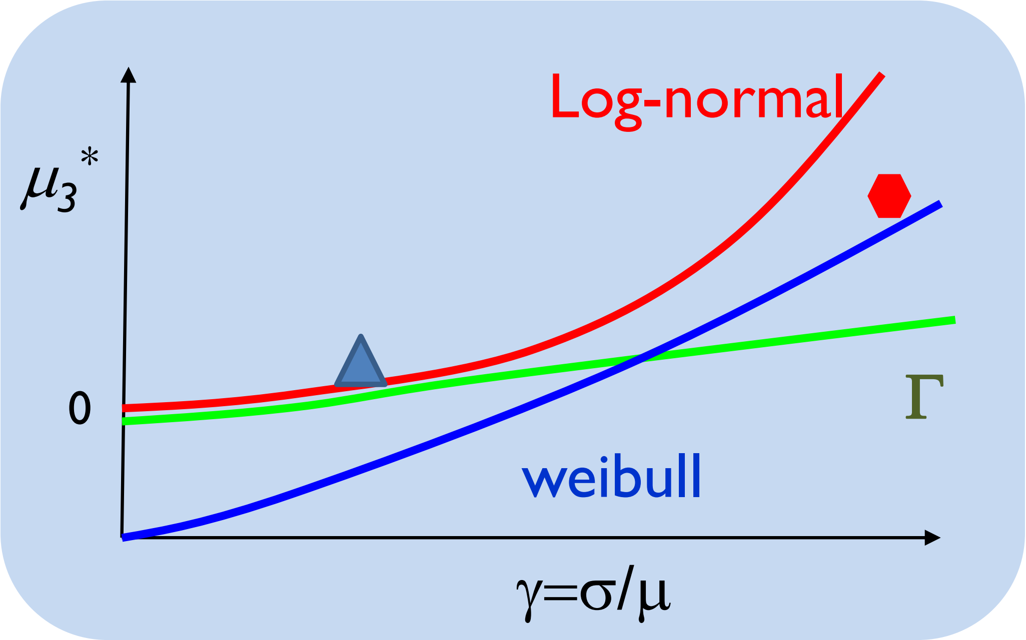
(4) Goodness of Fit: Cox-Oakes measure

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

Solid lines are known for various distributions.

Example. For a given α, β, Weibull has a specific μ,σ,μ_3 (blue triangle)

Logic: Every distribution has different shape.

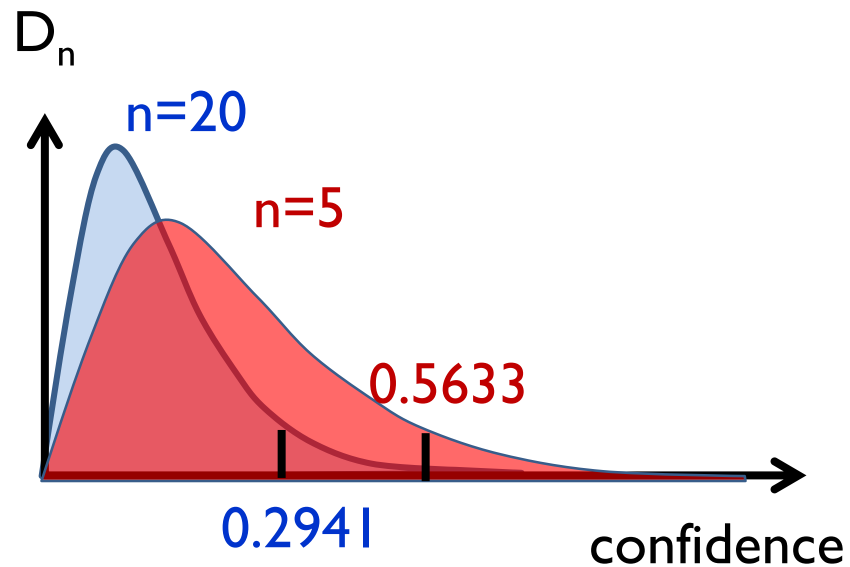


(5) Kolmogorov-Smirnov algorithm

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

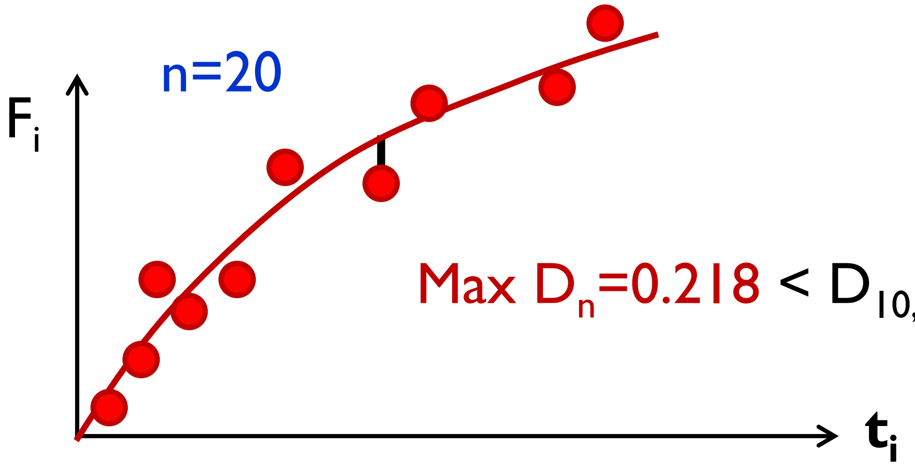
Compute … $D_n = \max \left| F_{obs}(t_i) - F_{theory}(t_i) \right|$

Sample size

If $D_n > D_n^{crit}$, fit is poor …

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

t_i	3	20	40	52	53	54	85	318	429	553
$F_i = (i - .3)/(n + .4)$										



A good fit …

Max D_n=0.218 < $D_{10,crit}$=0.4092

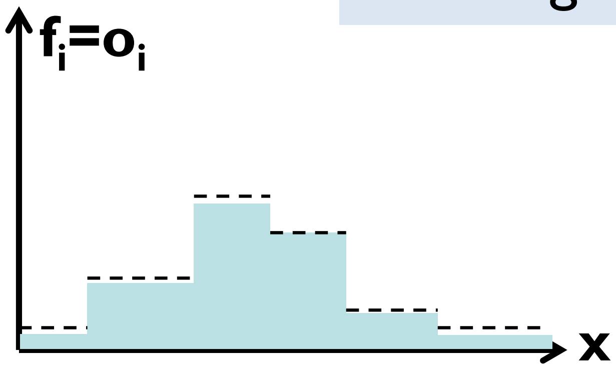
```
[h,p] = kstest(x,'CDF',test_cdf,'Alpha',0.01)
```

(6) Pearson χ^2 – test algorithm

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

o … Observed
e … expected
n* … datapoints
p_i …. probability
v … deg. of freedom

ν	**5% (χ^2)**
2	5.99
4	9.49
10	18.307
20	27.68



If the value observed χ^2 value exceeds critical value, the fit is poor.

A famous example: Schon story

Degree of freedom=8-2-1=5

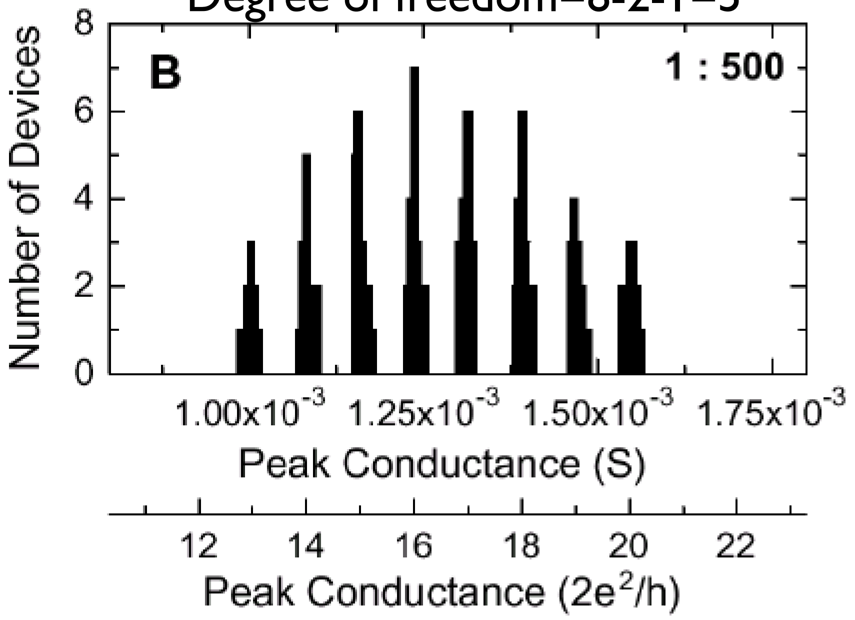


Figure 46. Figure 3(B) from "SingleMolecule" Paper (XIII), showing a histogram of conductances from diluted SAMFETs,

χ^2=0.125! Real data χ^2=1

Chance of this happening is 1.2%

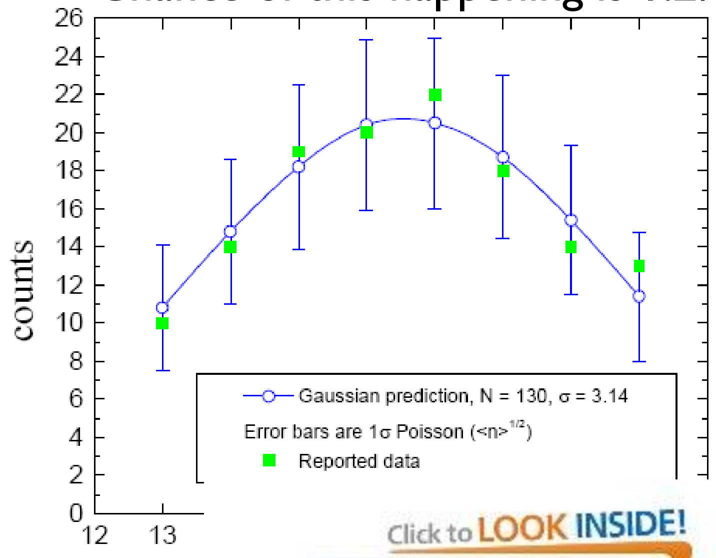


Figure 47.
devices in ea

The data indicating conductance quantization did not arise from an objec

measurement process. At a minimum, the assignment of conductance values was colc

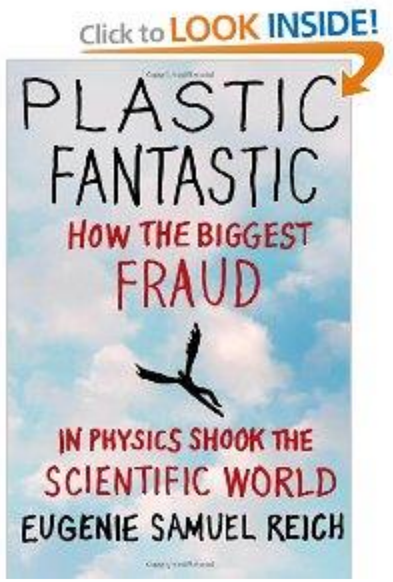
by the expected shape of the final distribution. Such a biased process cannot prov

convincing evidence for quantization. The response to this concern appears to

deliberately deceptive, suggesting that this misrepresentation was intentional.

The preponderance of evidence indicates that Hendrik Schön committed scient

misconduct, specifically data fabrication, in this case.



Parameter number vs. goodness of fit

n = number of samples, M=number of parameters

1) Method of adjusted residual …

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

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

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

1. 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 Einstein'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.

Review Questions

1. 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 χ^2 test?

7. How does the methods affected by censored data (e.g. , TDDB test yet to finish?)

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. Seindenfeld, 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.
- "Formalas 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 Critetia is found in "Making Sense out of Akaike's Information Critetion", Mark. J. Mazerolle, Appendix 1: http://www.theses.ulaval.ca/2004/21842/apa.html

Appendix: Variability by Bootstrap method

Ref. courses.washington.edu/matlab2/Lesson_6.html

Uncertainty in parameters: Least Square

Is the error in W Gaussian distributed ?

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

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

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

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

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

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

MLE estimator for Weibull: Did not discuss variability

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

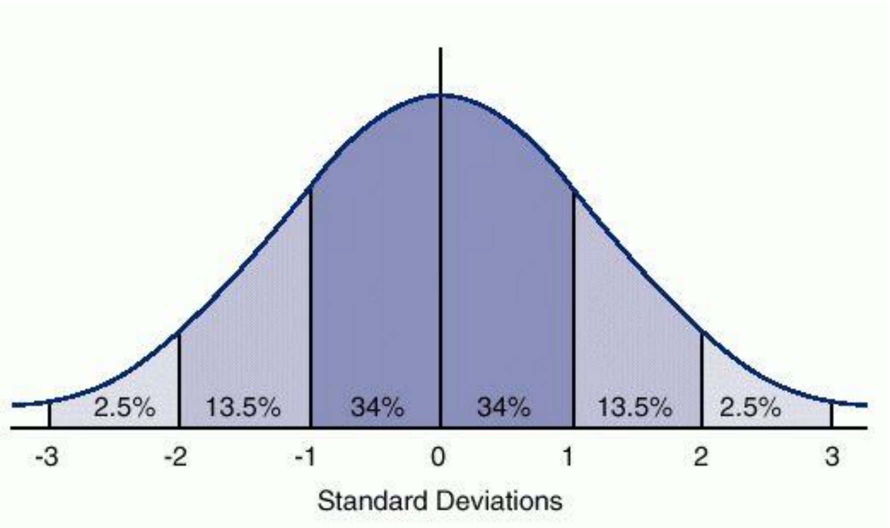
$$\ln L=\sum_{i=1}^{n}\ln f(t_i,\alpha,\beta)$$
$$=n\ln\beta-n\ln\alpha+(\beta-1)\sum_{i=1}^{n}\ln t_i/\alpha-\sum_{i=1}^{n}\left(t_i/\alpha\right)^{\beta}$$

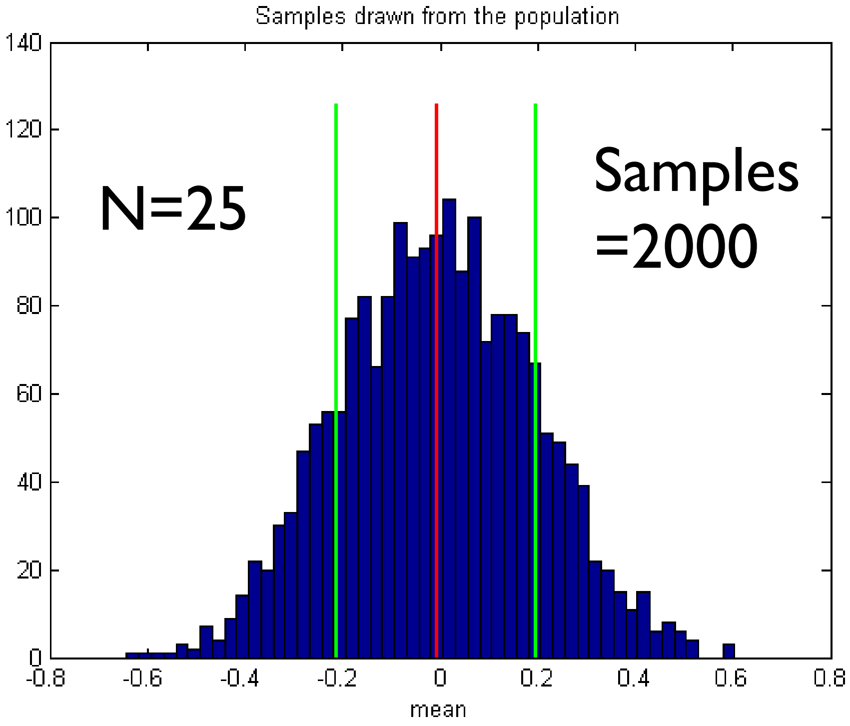
$$\frac{d\ln L}{d\alpha}=0 \qquad \frac{d\ln L}{d\beta}=0$$

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

Solve for unknowns α, β

(1) Bootstrap method: Introduction





68% between +0.2 to -0.2
95% between +0.4 to -0.4

$$s = \sqrt{\frac{\sum_{j=1,N=25} (t_j - \langle t \rangle)^2}{N-1}} \sim \sqrt{\frac{1}{24}} \sim 0.2$$

Ref. courses.washington.edu/matlab2/Lesson_6.html

Working with a single sample

0.2 -0.1 0.5 0.3 -0.6   All you have is a single sample ..

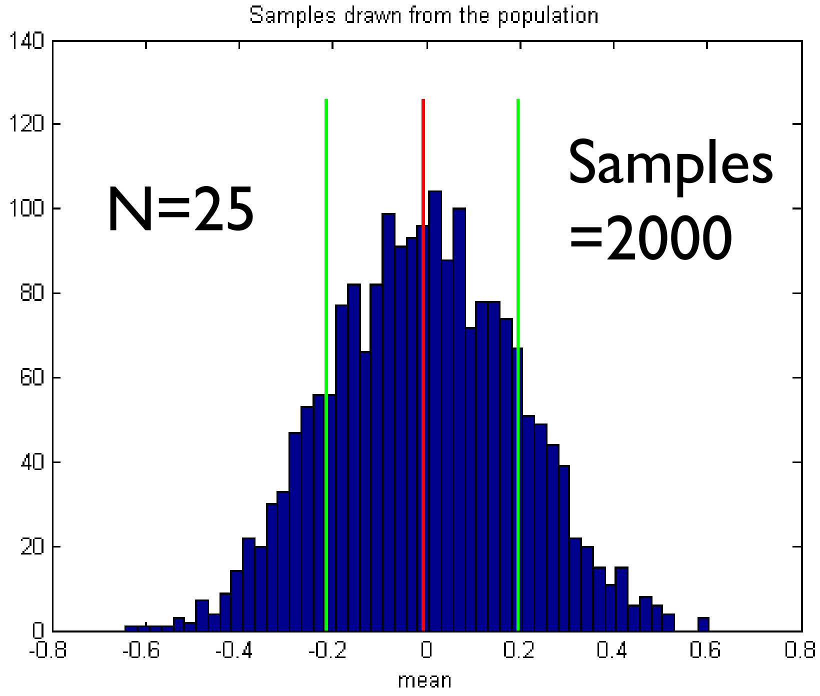
Generate synthetic samples from the original (with replacement)

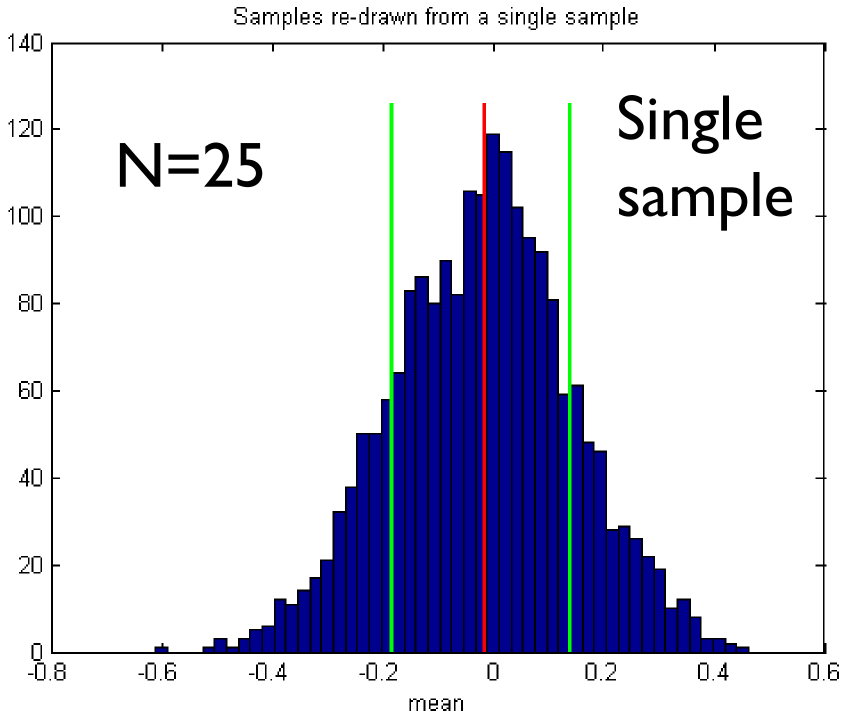
0.2 -0.1 -0.6 -0.1 0.5   Synthetic sample 1

0.3 0.2 -0.6 0.2 0.5   Synthetic sample 2

0.5 -0.1 0.5 0.2 0.3   Synthetic sample 3

Multiple sample vs. single sample





Bootstrap average is not zero!

And yet, the s~0.18, just from a single sample.

The success of the method relies on precision measurement

Parametric vs. non-parametric Bootstrap

0.2 -0.1 0.5 0.3 -0.6

Fit the distribution of your choice by Maximum likelihood estimators (MLE) (obtain parameters, i.e. η_0, β_0)

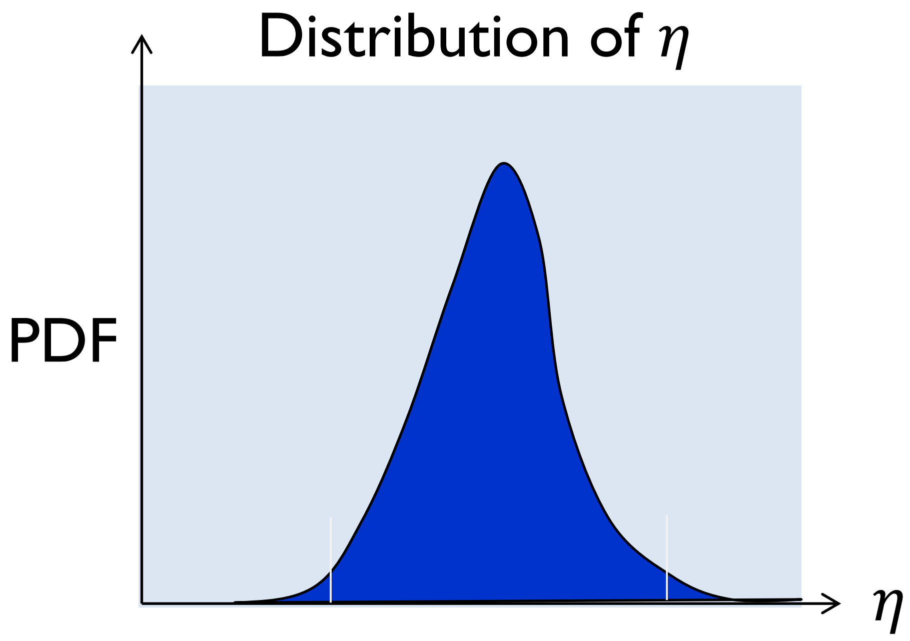
Generate synthetic samples based on the parametric distribution

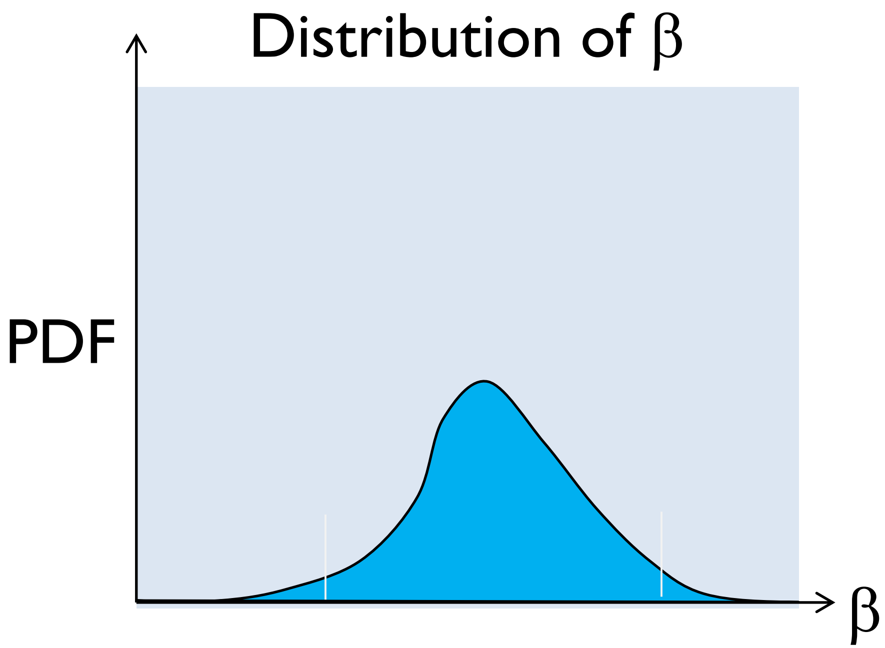
0.12 -0.17 -0.44 -0.71 0.52   Synthetic sample 1 (new η_1, β_1)

0.32 0.21 -0.69 0.23 0.58   Synthetic sample 2 (new η_2, β_2)

Plot distribution of statics η_i, β_i

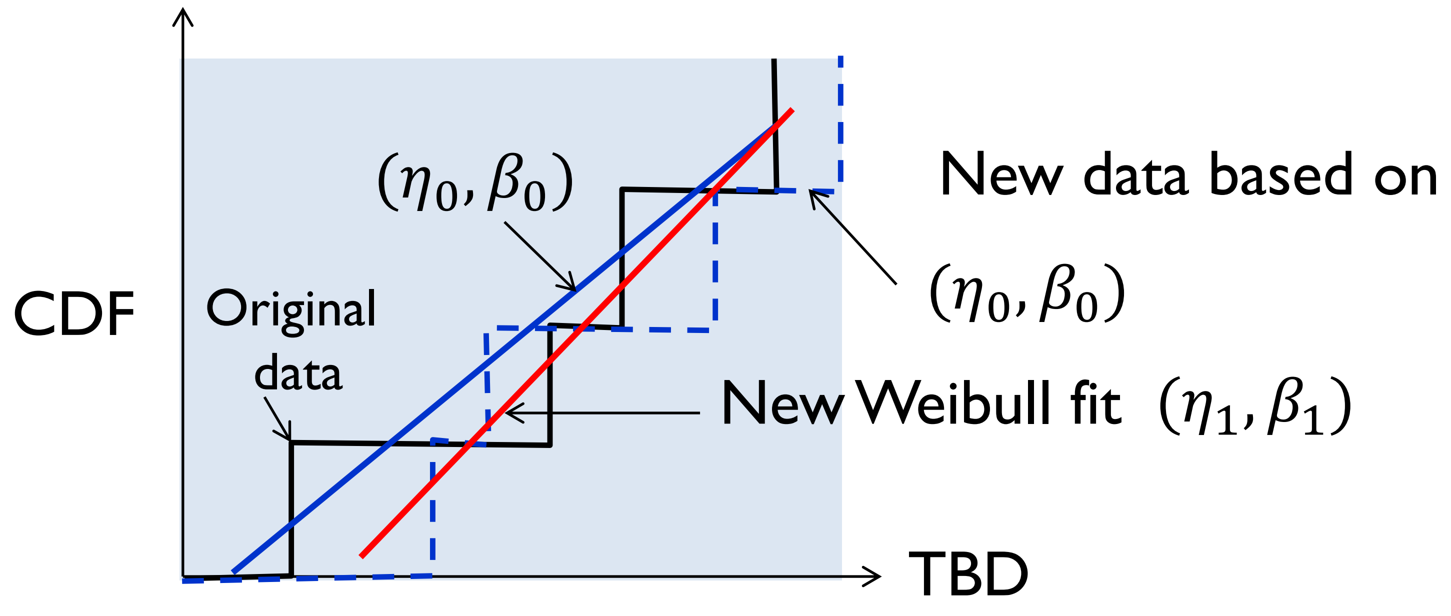
Distribution of α and β





Same technique for polling and tenure rate of faculty!

Why resampling from the same distribution generates new fit parameters



Samples taken from the same distribution (η_0, β_0) generates datapoints that are fitted with new (η_i, β_i)

References

1. "Detecting Novel Associations for large scale dataset", D. Reshef et al. , Science 334, p. 1418, 2011.
2. "Survival Analysis of Faculty Retention in Science and Engineering", D. Kaminski et al., Science, 335, 864, 2012.
3. 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.
4. Modified R^2 is discussed by Neter, Kutner, Nachtheim and Wasserman (1996).