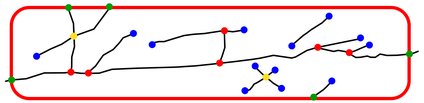
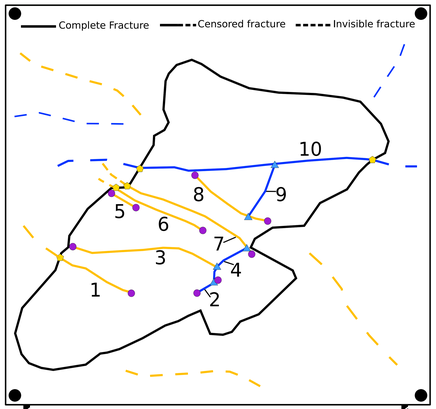
Fracability



A python toolbox for survival analysis in fractured rock systems



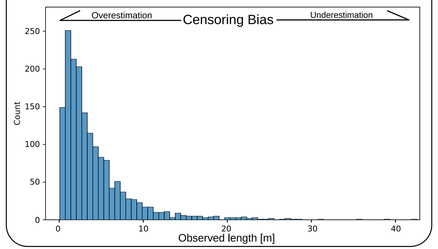
Questions and problems

Fractures can be censored, especially the long ones!

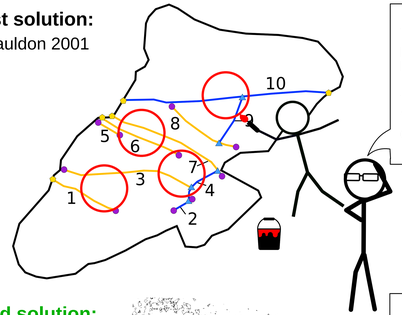
How should we treat censored data?

How does censoring affect estimation?

Which is the "right" model describing the data?



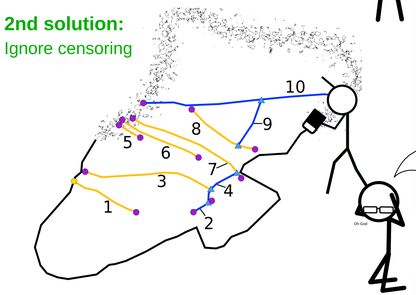
1st solution:
Mauldon 2001

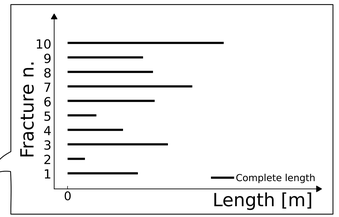


The problem with Mauldon is that it is a non parametric method for estimating the mean length.

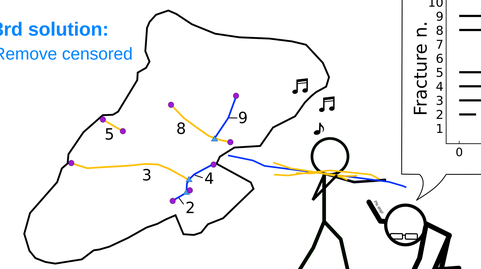
This is an important detail and limitation if the focus is to **model the statistical length distribution** of an outcrop!

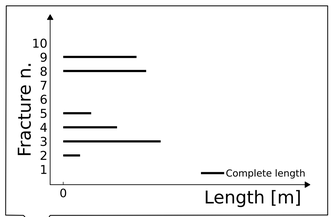
2nd solution:
Ignore censoring



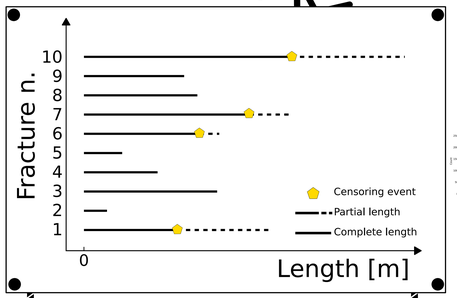


3rd solution:
Remove censored

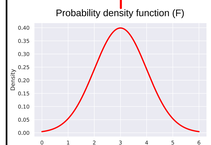


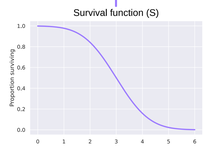


4th solution:
Survival Analysis



$$L(\theta|X_i) = \prod_{i=1}^{n} F(\theta|X_i)^{\delta_i} S(\theta|X_i)^{1-\delta_i}$$





Complete measurements δ_i=1

We are used to this

Censored measurements δ_i=0

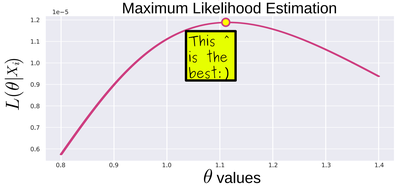
This is new!!

X_i are the known measurements

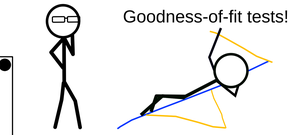
θ are the unknown parameter(s)

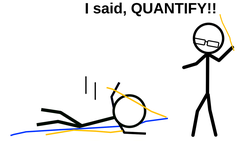
L is function of θ

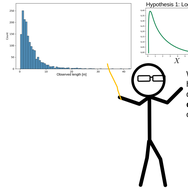
Maximum Likelihood Estimation



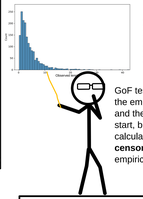
Can we **quantify** which statistical models (hypothesis) are representative of the data?



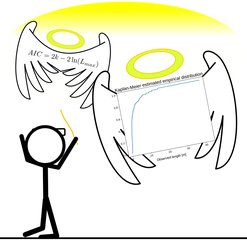




We have some data and a set of **sensible** hypothesis (statistical models), which one describes the data the best? Can we **quantify and rank from** best to worst our hypothesis?

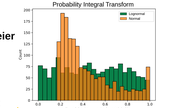


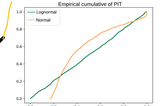
GoF tests are based on **distances** between the empirical distribution of the observations and the theoretical distributions. This is a good start, but the data are censored!! How can we calculate the **empirical distribution of censored data**? Also, can we completely avoid empirical distributions?

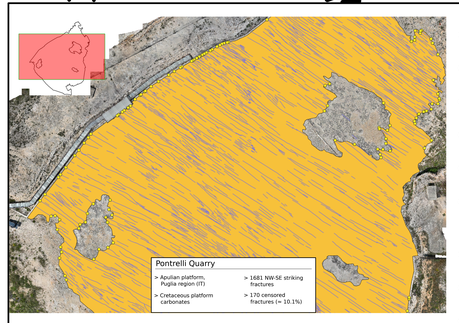


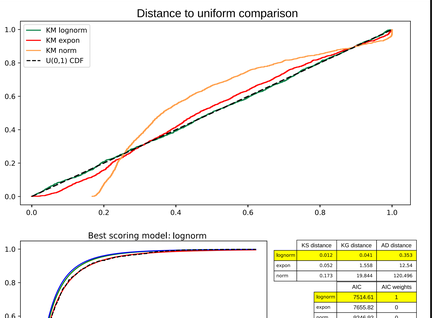
Models can be **ranked** according to **their distance to the Kaplan-Meier** distribution and Akaike value!
The smaller, the better.

Moreover, we can **visualize** different fits using the Probability Integral Transform to also have a visual comparison.
(the best fit is the closest to the uniform distribution)
Kim (2019)

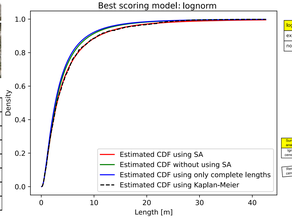








	Lognorm estimated parameters ignoring censoring	Lognorm estimated parameters removing censored measures	Lognorm estimated parameters using survival
Mean	4.33	4.12	4.94
Median	2.76	2.63	2.98
Mode	1.12	1.07	1.08
B5	0.57	0.55	0.57
B95	13.17	12.50	15.58
Std	5.24	4.96	6.53
Var	27.51	24.69	42.69



	KS distance	KG distance	AD distance
lognorm	0.012	0.041	0.353
expon	0.052	1.558	12.54
norm	0.173	19.844	120.496

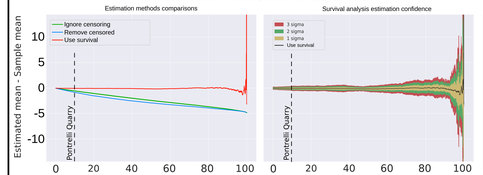
	AIC	AIC weights
lognorm	7514.61	1
expon	7655.82	0
norm	9246.92	0

	AIC	AIC weights
Natural fractures	7514.61	1
Artificial fractures	8018.40	0

Natural fractures	7054.78
Artificial fractures	

We cannot compare different datasets!

If we remove censored fractures we are fitting a completely different model to different data!



Concluding remarks

> Ignoring censoring does not help:
2nd solution underestimates the mean by -0.52 ± 0.10 cm
3rd solution by -0.77± 0.15 cm

> By using survival analysis (4th sol): -0.023 ± 0.10 cm. Moreover, it stays quite stable around the sample mean up to 80% censoring

> Paradigm shift to select and **quantify** the best model from a list of hypothetical models.

Gabriele Benedetti[1], Stefano Casiraghi[1], Andrea Bistacchi[1] & Daniela Bertacchi[2]

[1] Università degli Studi di Milano-Bicocca, Dipartimento di Scienze dell'Ambiente e della Terra, Milano, Italy

[2] Università degli Studi di Milano-Bicocca,Dipartimento di Matematica e Applicazioni, Milano, Italy

Anake, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53(282), 457–481. https://doi.org/10.2307/2281868

Kim, N. (2019). Tests based on EDF statistics for randomly censored normal distributions when parameters are unknown. Communications for Statistical Applications and Methods, 26(5), 431–441. https://doi.org/10.29220/CSAM.2019.26.5.431

Mauldon, M., Dunne, W. M., & Rohrbaugh, M. B. (2001). Circular scanlines and circular windows: New tools for characterizing the geometry of fracture traces. Journal of Structural Geology, 23(2–3), 247–258. https://doi.org/10.1016/S0191-8141(00)00094-9