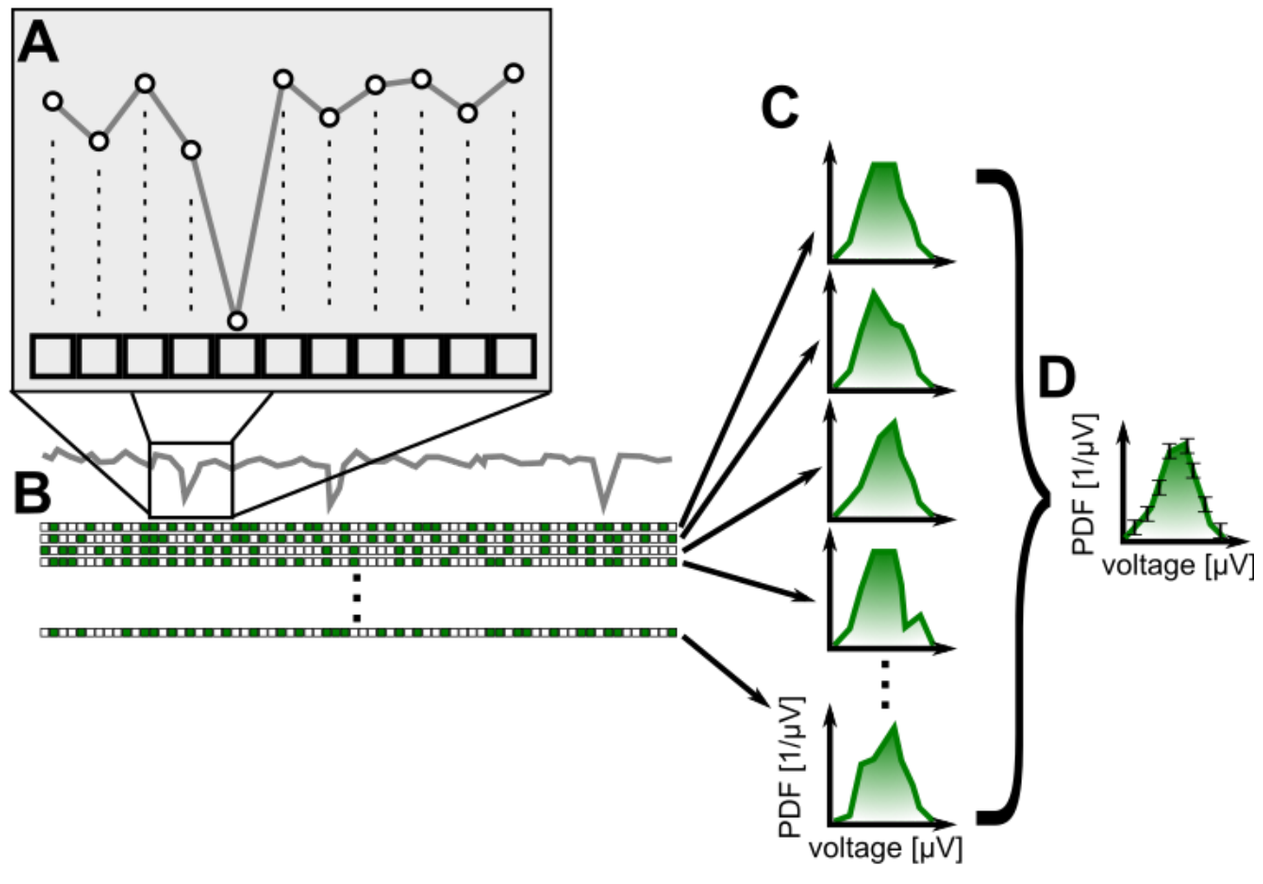
First resampling attempt at error bar estimation

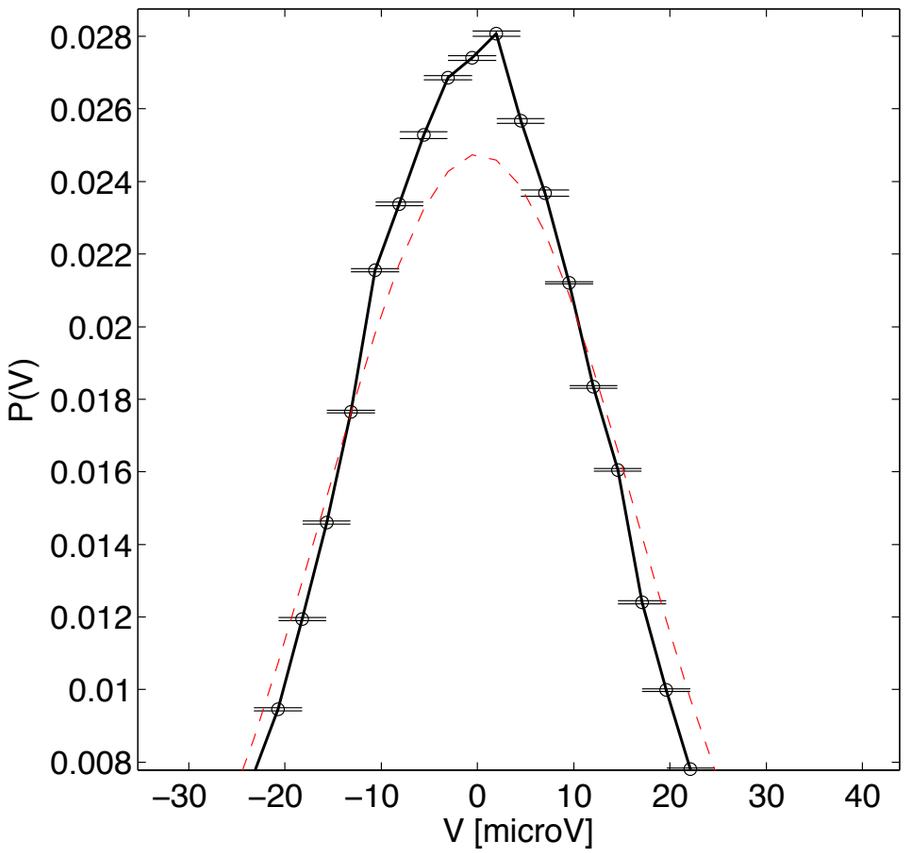
- That's all good, but in normalizing the histogram, we lost the information about errors implicit in raw counts…

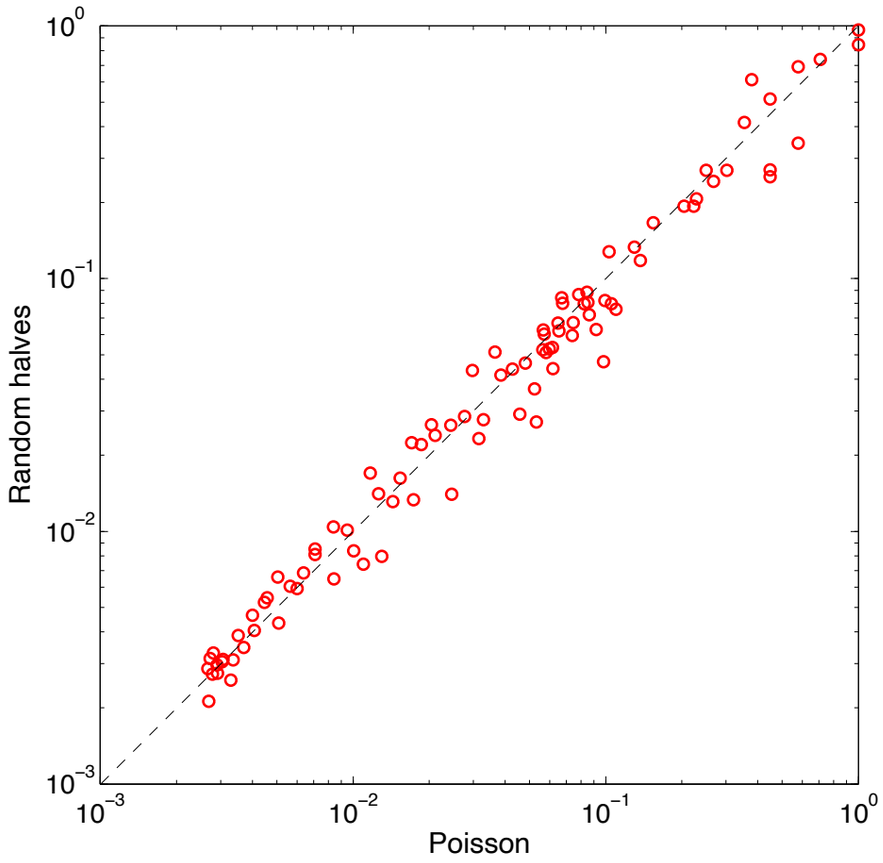
- Rough estimate: evaluate the histograms over random halves of the samples and take std over halves (over the same x-bins!)



First resampling attempt at error bar estimation

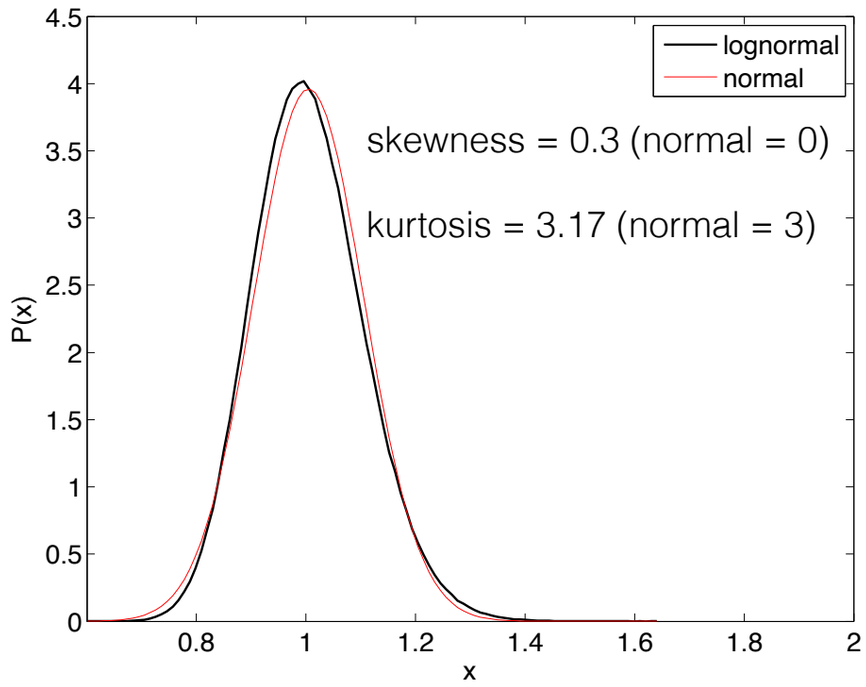
- Looks like the PDF is indeed significantly different from a Gaussian!
- What do you expect for the error bar as a function of the counts? Counting / Poisson errors: **Var[N_i] = N_i**
- Let's say the error bar is ~sqrt(counts), does this match with the rough bootstrap estimate?





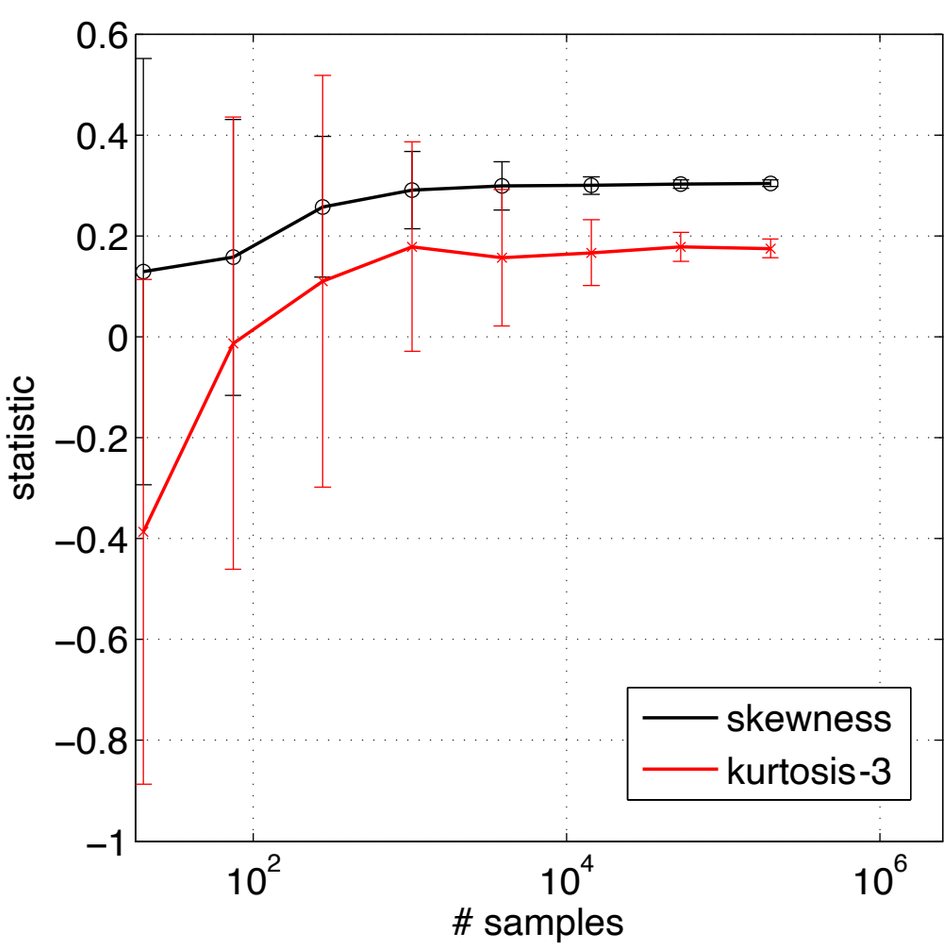
Do you think this procedure for error bar estimation is actually correct for our data? (why Poisson? Why halves? Why independent errors? IID assumption?)

- Higher-order statistics can be very sensitive to sampling and outliers



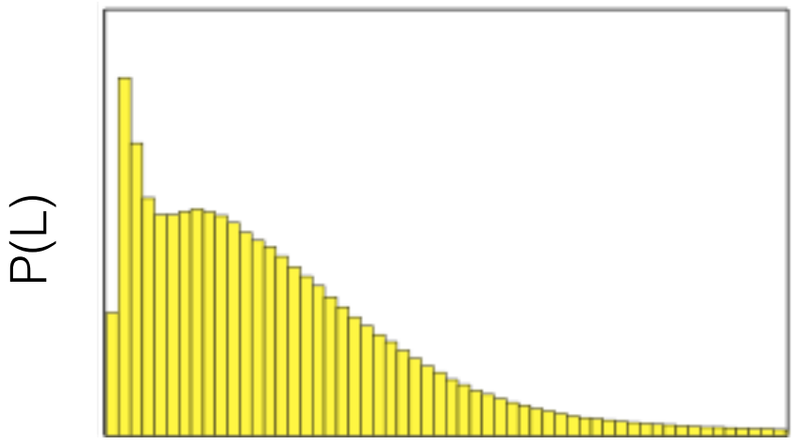
Quantile statistics are much more *robust*
(decide whether to use moments or quantiles
based on the amount of data + field)

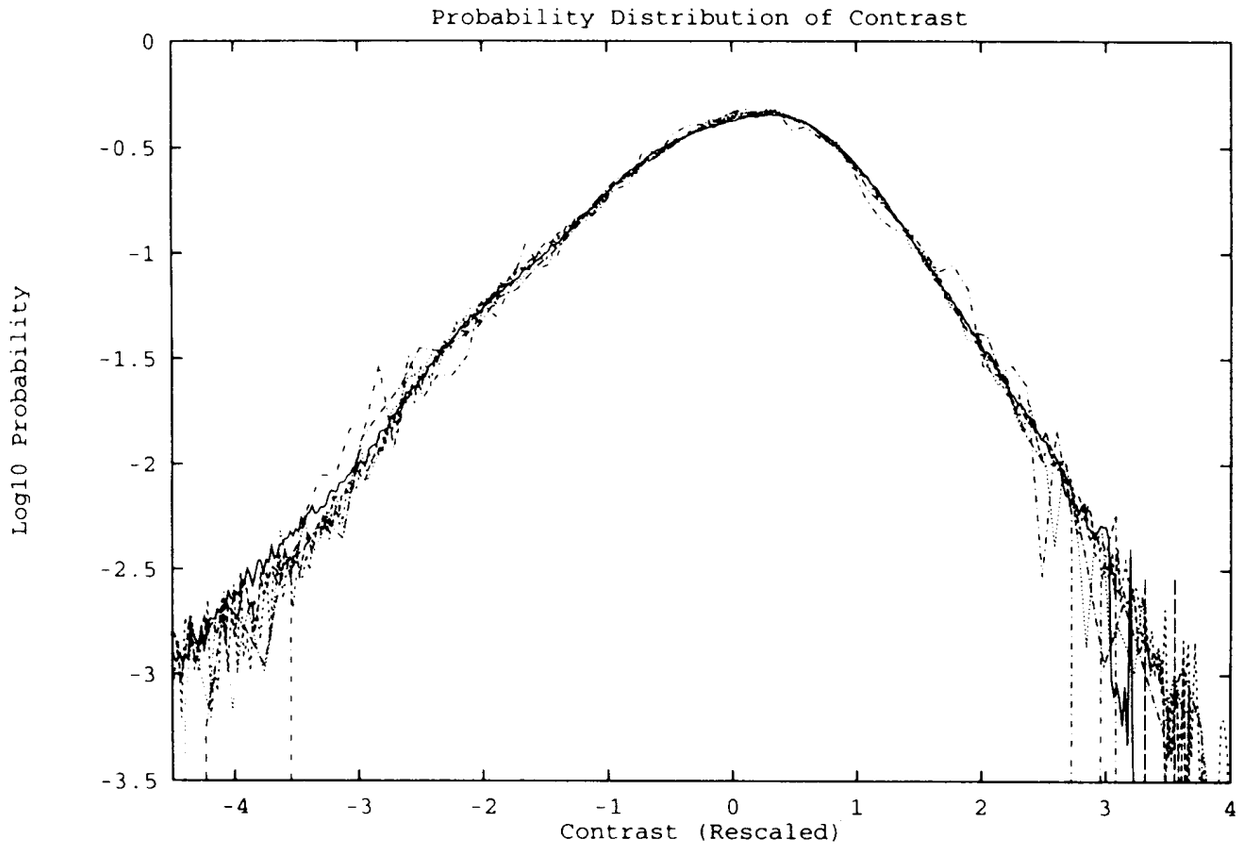
… or try nonlinearly transforming the data?



Excursion: data representation

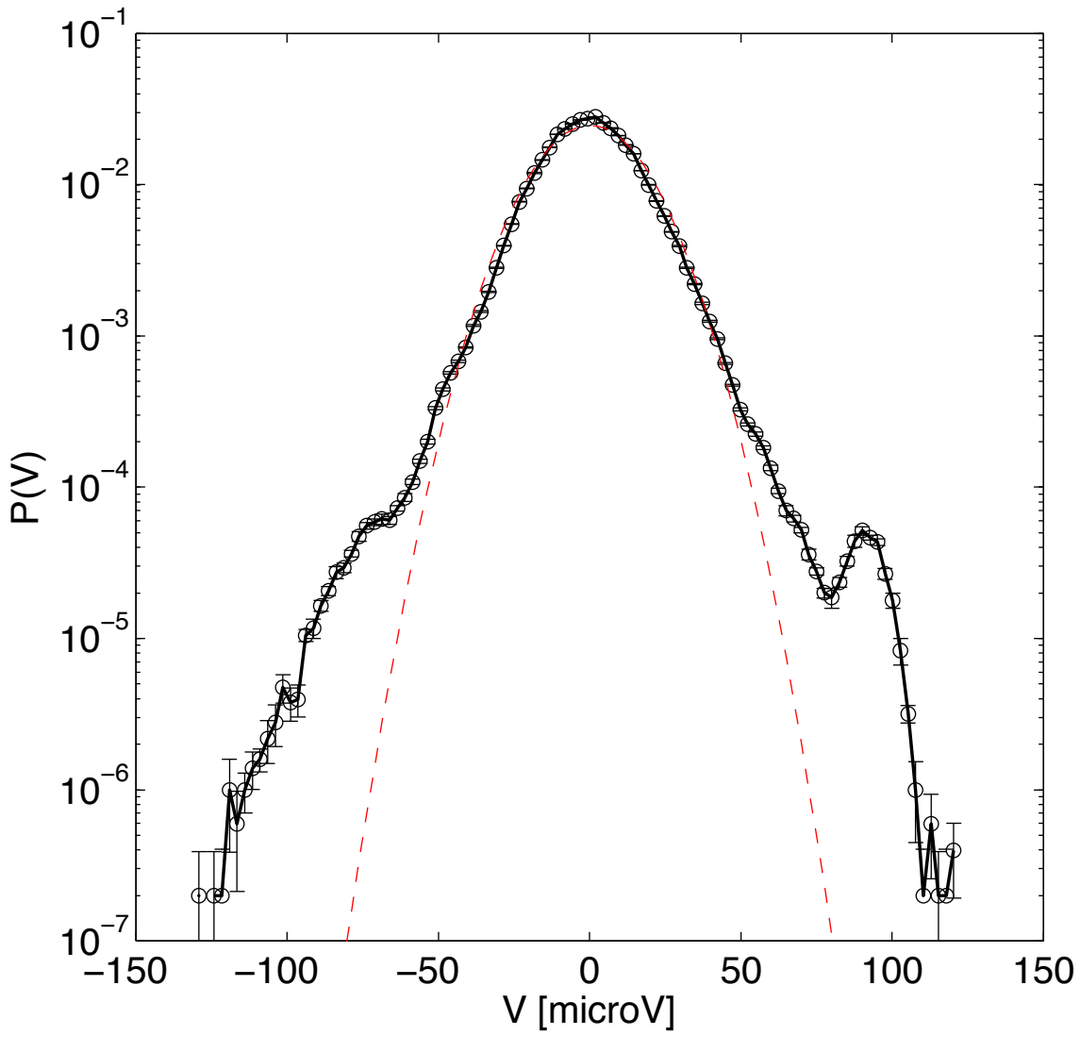
- If you measure x, do you need to look at the histogram of x? Why (not) f(x)?
- e.g. distribution of luminance over natural images (Ruderman et al, PRL 73 (1994)) which is very broad…





- if you define **$C = \log(I / I_0)$** where I_0 is the mean over each image: **(i)** the data has a simpler (bi-exponential) form, better described by central moments (mean, std, skew, kurtosis); and **(ii)** the data collapses
- log-units for the measured variable sometimes make more sense in the context of a particular theory (e.g., log concentration instead of concentration), so it makes sense to reparametrize
- copula: transforming the marginals into the desired (simple) form to study correlations

- Lets circle back to error bars in real data…

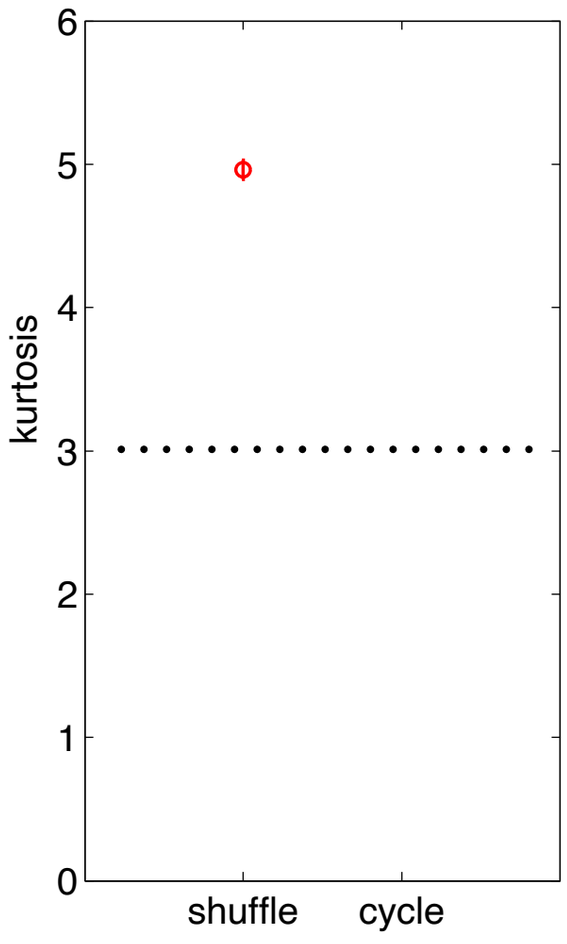


skewness = 0.11 (normal = 0)

kurtosis = 4.97 (normal = 3)

- To see if this is significantly different from Gaussian, let's estimate over 1/20 (or 1/10) "fractions" of the data and get error bars

- What do we mean by fractions? Random shuffle draws, or contiguous blocks (e.g., by cycle-permuting data)?



- A **huge difference in estimated error bar!**

- Non-IID samples due to temporal correlation?

- Estimate the effect: var(cycle)/var(shuffle) ~ 100, 2M samples ~ 20k IID samples?

  Still not completely OK… depends on choice of data fractions etc.

Bootstrap re/sub-sampling for error estimation

When there is NO correlation in the data (IID samples)

- **Basic idea:** given limited data, resampling from empirical distribution approximates sampling from a true (but unknown) underlying distribution

1. Take the empirical distribution based on N samples
2. Resample (with replacement) / noise from the same distribution many times N samples
3. Evaluate the statistic of interest and its error bar as std over resampling draws

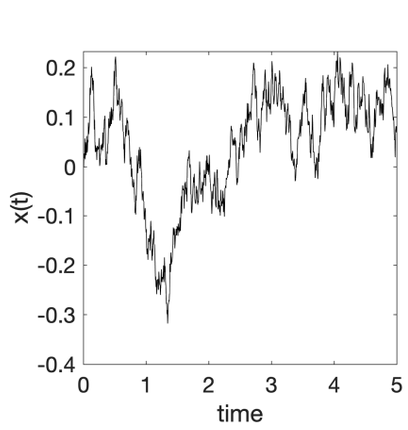
When there are serial, short range correlations in the data

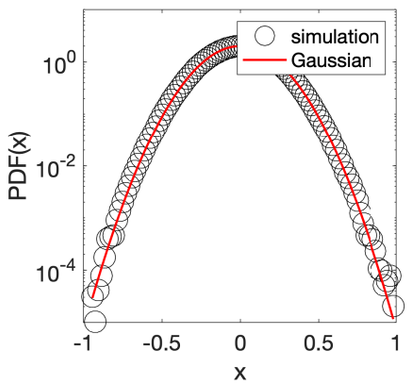
- **Basic idea:** subsample without replacement from *contiguous subsets* of data to preserve the correlation (assumptions?)

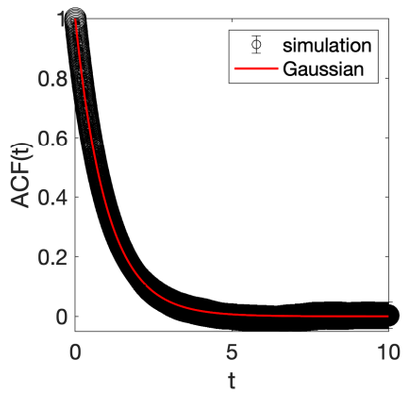
1. Split the data into blocks of different sizes (many times)
2. Evaluate the statistic on each block and get its std across blocks
3. Extrapolate the std in the statistic from different subset sizes to the whole data size

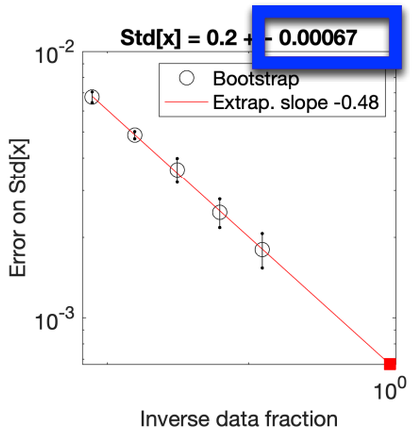
Need stationary data (second “I” in “IID”): statistics do not change with time!

Bootstrap on synthetic data - Ornstein-Uhlenbeck



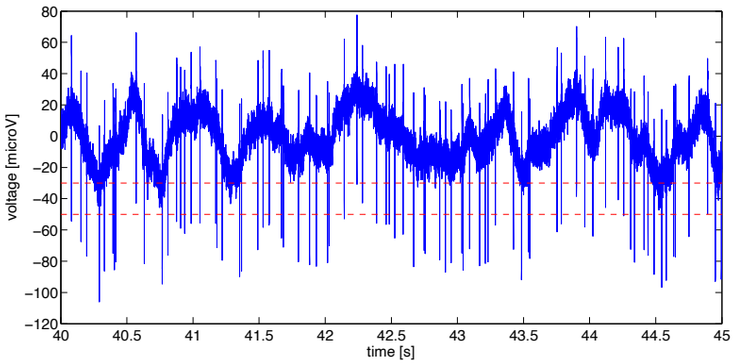


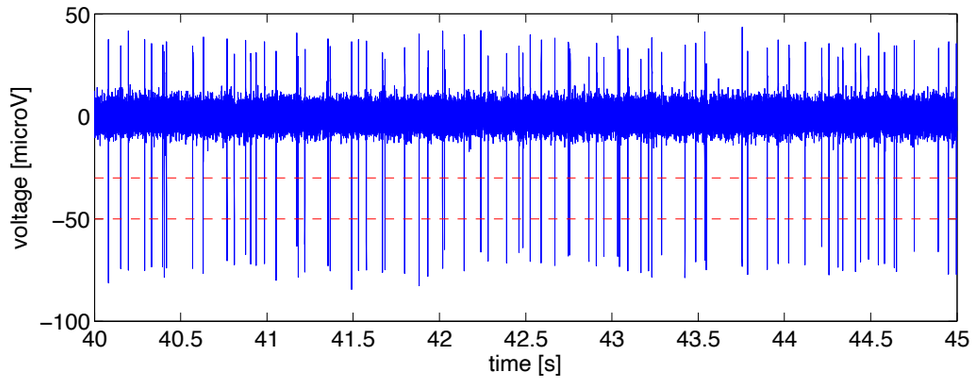


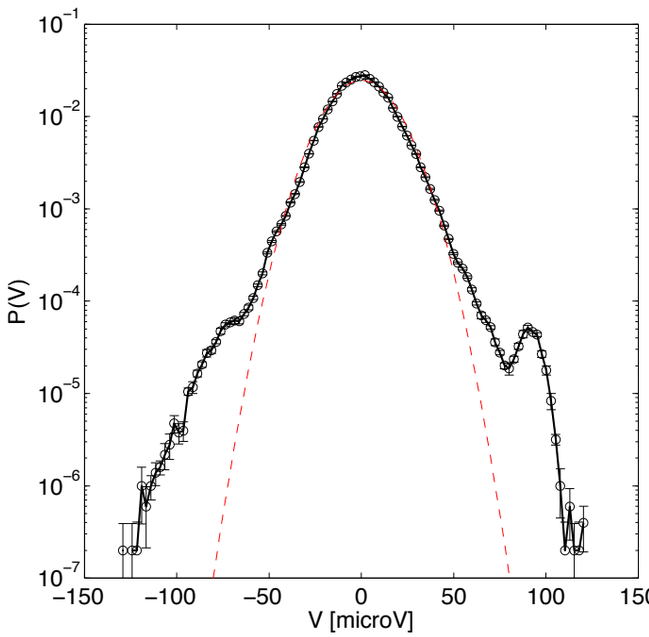


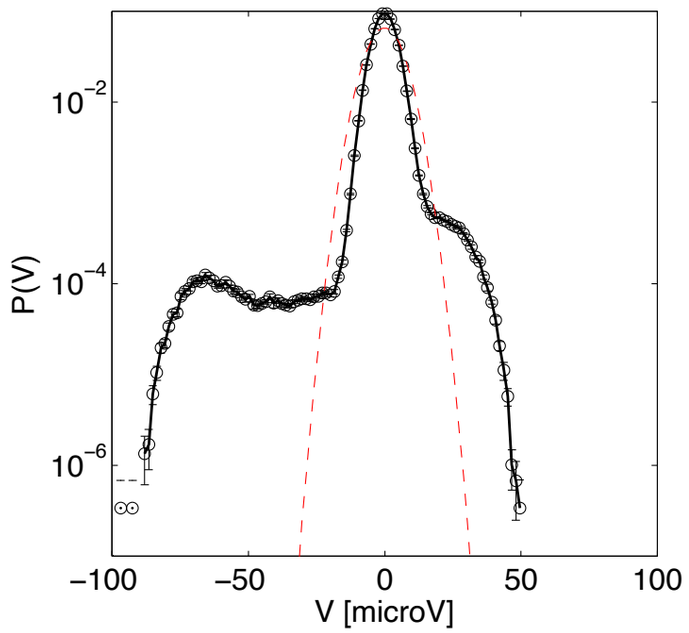
10 x longer correlation

10 x shorter correlation









Original distribution

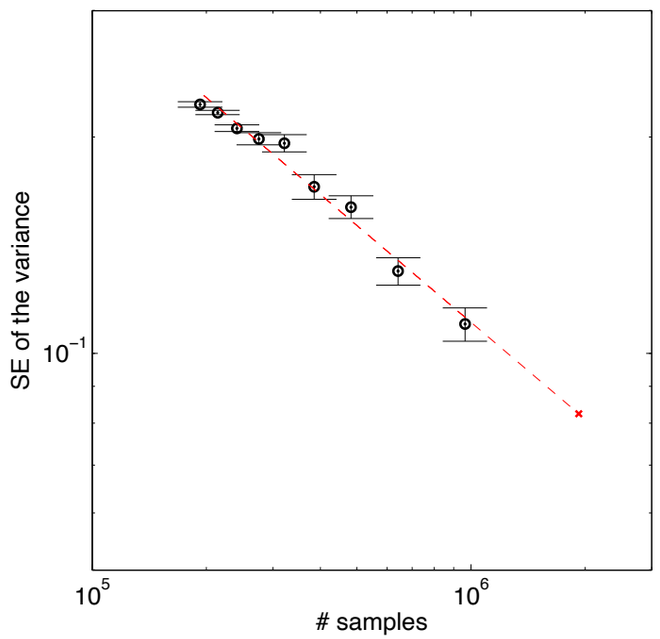
Baseline subtracted

Spikes subtracted

Let's use bootstrapping to estimate the variance
+- SE of the spike-subtracted signal…

Variance = 16.6 +- 0.09 uV²

(+- 0.01 w/o correlations)

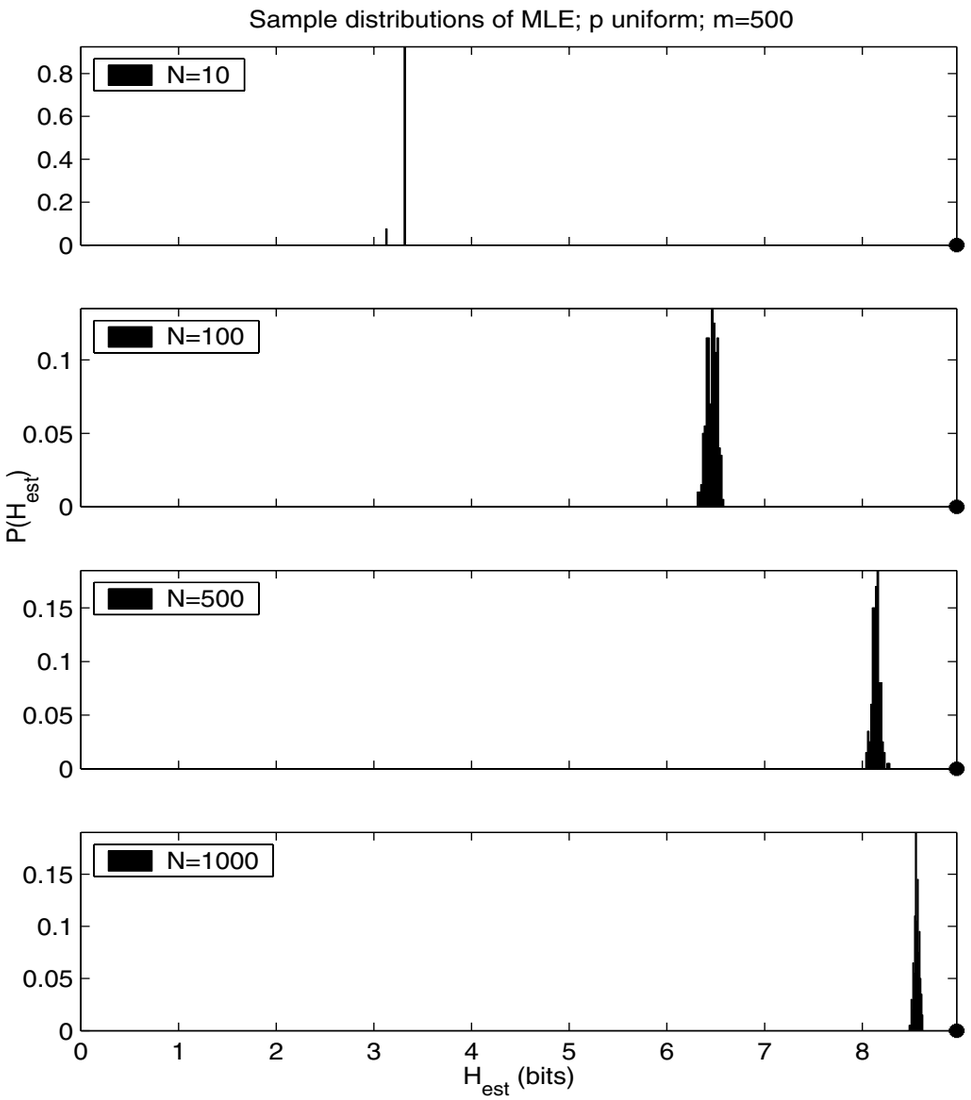


Excursion: using bootstrap for de-biasing

- **Task:** estimate the entropy of a discrete distribution given a finite number of samples

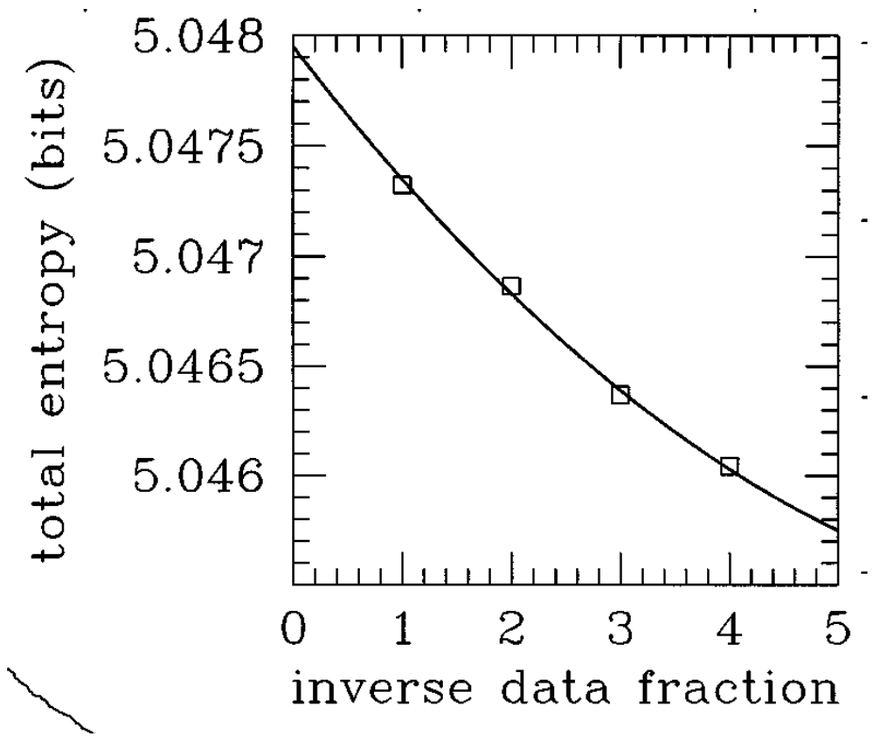
$$S[p] = -\sum_i p_i \log_2 p_i$$

- **Problem:** maximum likelihood estimate ("naive" estimate / histogram statistic) is very biased at finite *N* (bias dominates over variance unless N >> m)



Paninski L, *Neural Computation* 15 (2003)

$$S(N) = S_{\text{true}} + S_1/N + S_2/N^2$$



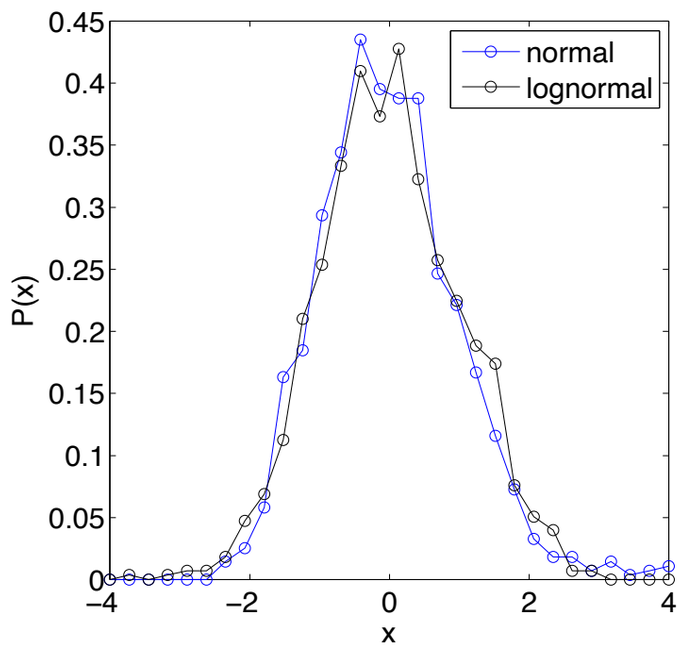
Strong SP et al, *Phys Rev Lett* 80 (1998)

Subsample data and extrapolate the bias to infinite data limit.

Resampling methods summary

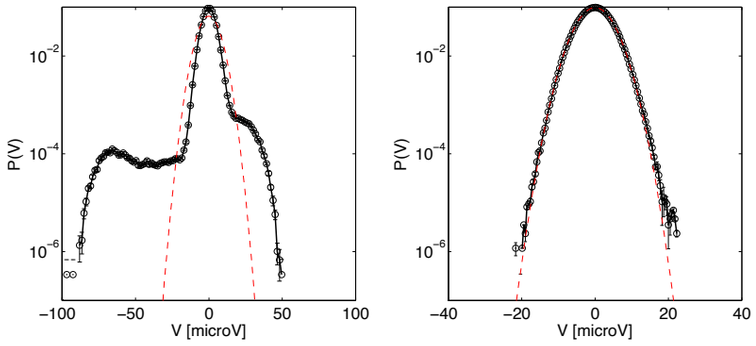
- Bootstrap resampling for IID data
- Bootstrap subsampling for correlated data
- Jackknife
  1. Compute the statistic *f* on all leave-one-out subsamples of the data
  2. $\sigma_f^2 = \frac{N-1}{N}\sum_{t=1}^{N}(f_{\mathcal{D}\setminus t} - f_{\mathcal{D}})^2$
- Debiasing estimators
- Creating null distributions to test for statistical significance (permutation tests)

CPU time is cheap, permute away (carefully)!



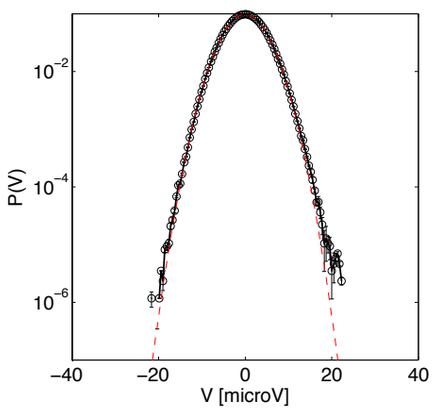
Are two distributions equal? Continuous data

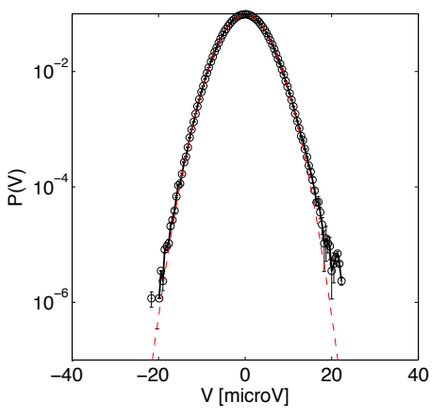
- nonparametric Kolmogorov-Smirnov test for 1D distributions, reparametrization invariant
- P(D>observed | distributions same) = universal function of sqrt(N) and D
- not equally sensitive at all values, but generalizations exist

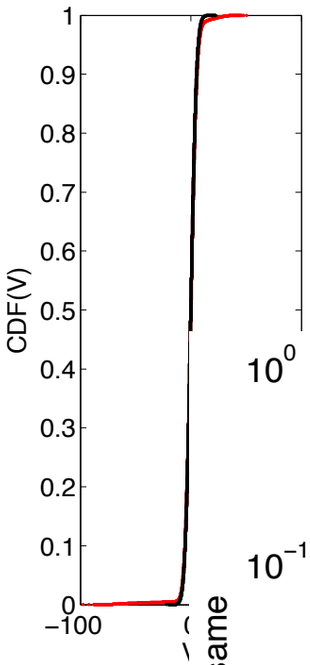


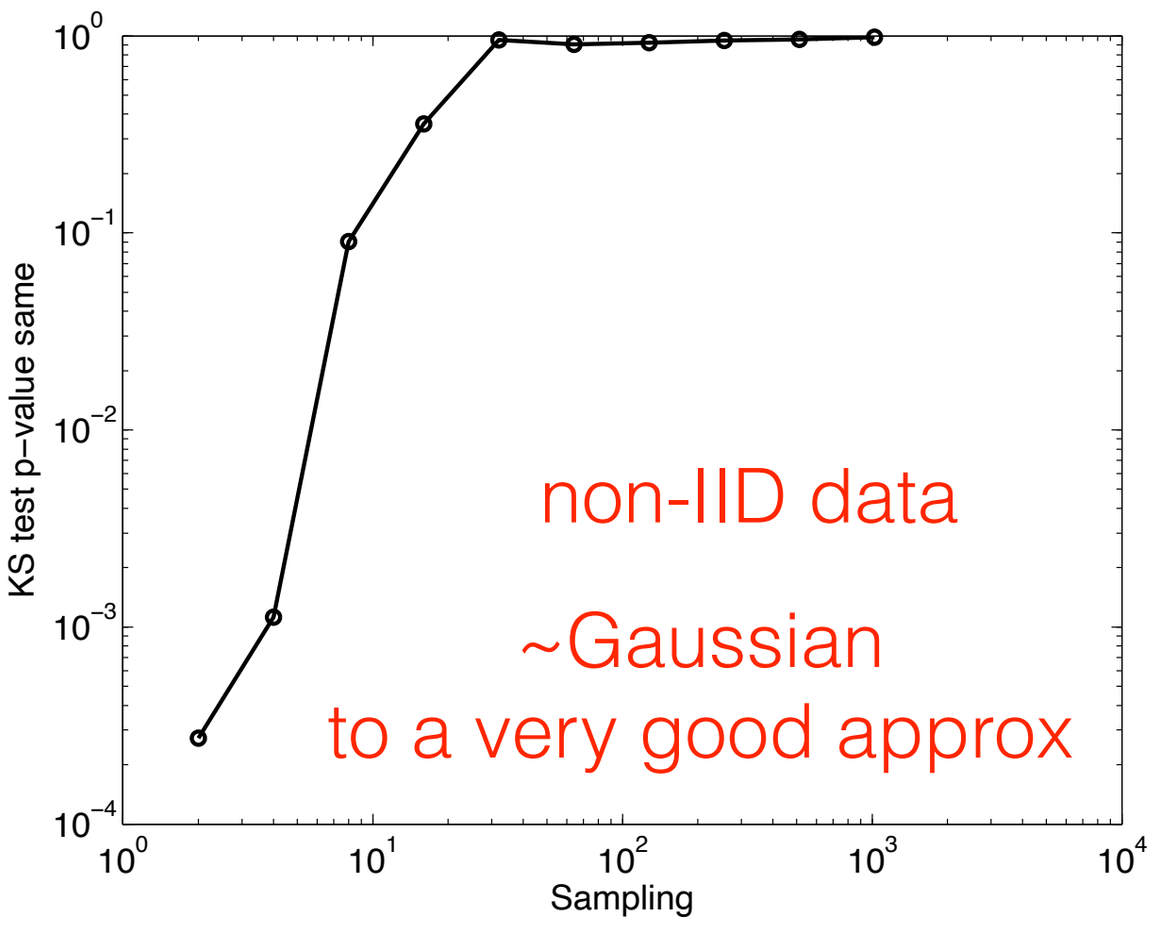
vs

1/2 vs 2/2 of data









Are two distributions equal? Binned data

- Suppose we are given integer counts N_i in a binned histogram. Let's compare to expected number of counts (a model, or a reference distribution), n_i.
- Since the expected errors are ~sqrt(n_i) if the data is IID, we can define the chi2 statistic

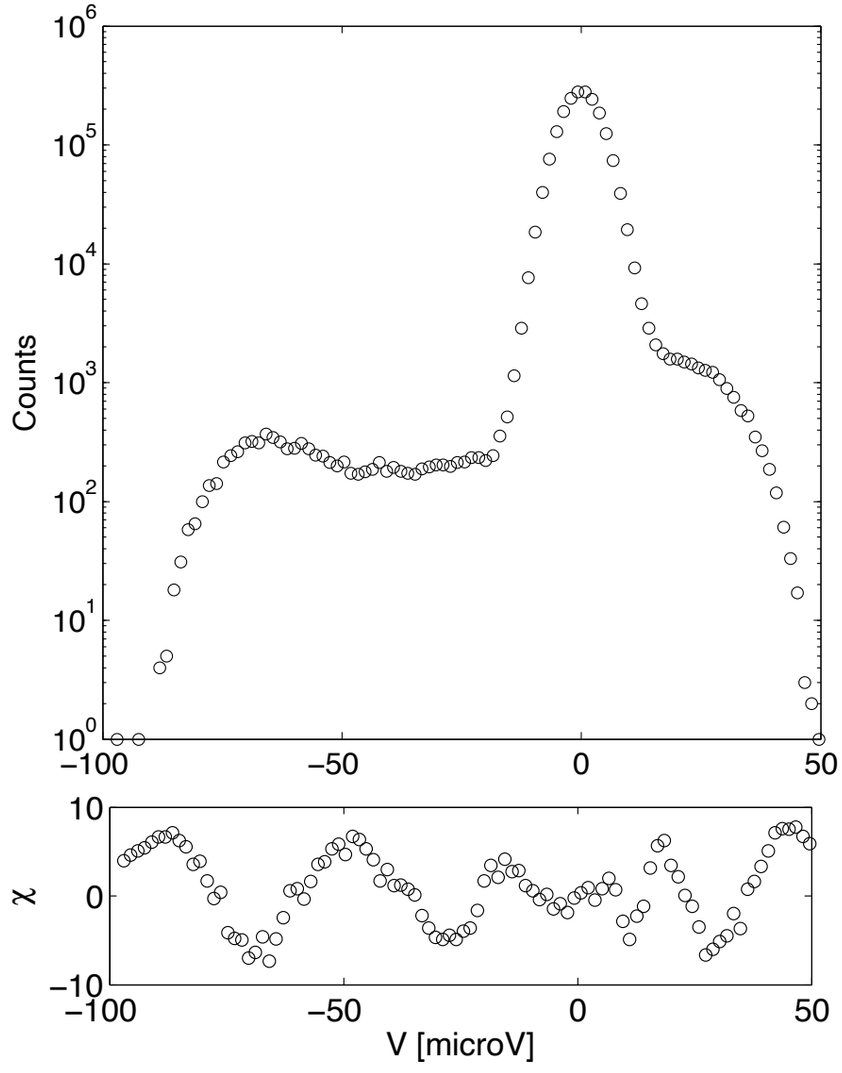
$$\chi^2 = \sum_i \frac{(N_i - n_i)^2}{n_i}$$

- Significance of the deviation is given by looking up getting the p-value from the chi2 distribution with k degrees of freedom

$$P(\chi^2|k) = \frac{1}{2^{k/2}\Gamma(k/2)}\chi^{k-2}e^{-\chi^2/2}$$

We can use this as a error metric to fit distribution models to histograms, e.g., using nonlinear least-squares

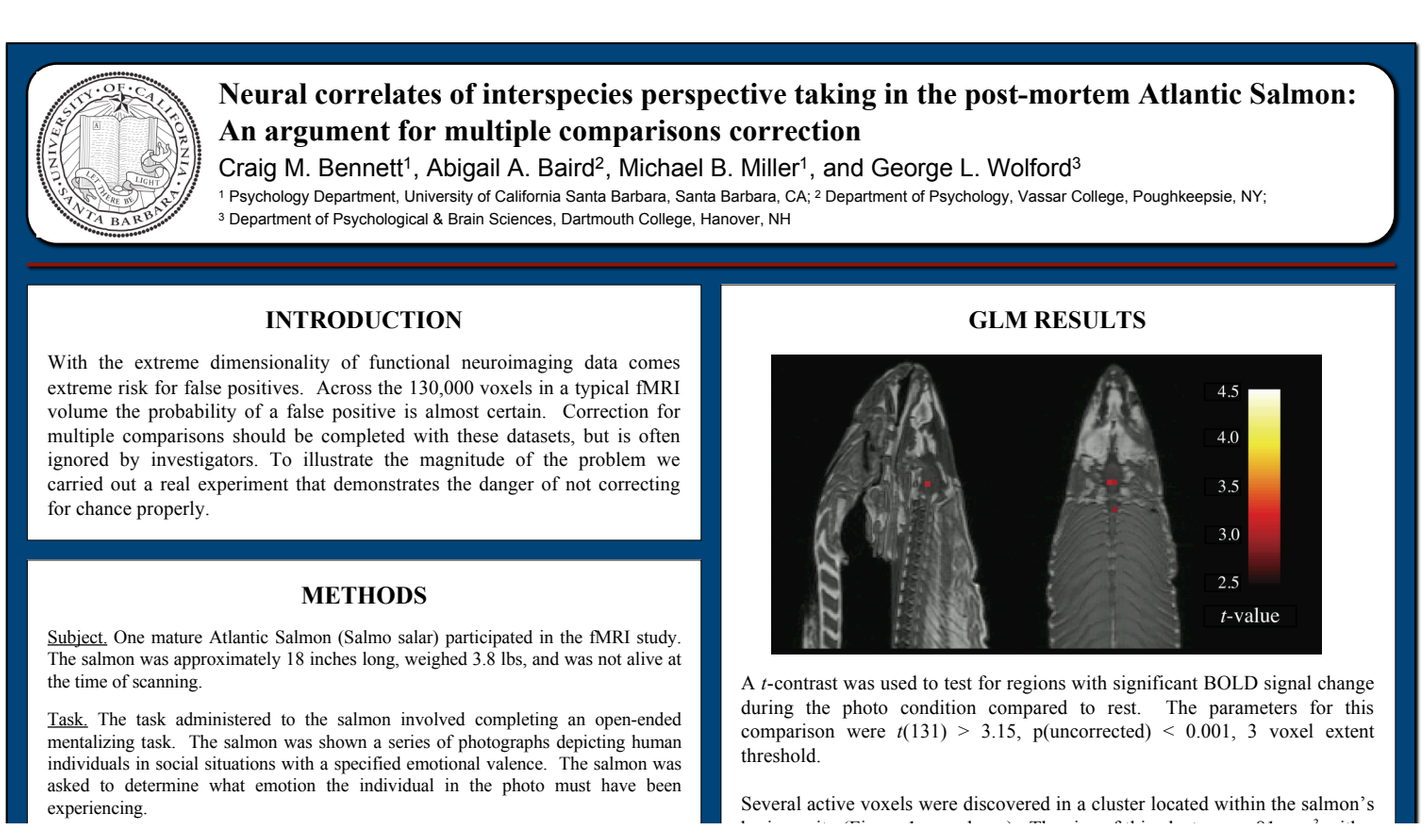
Fitting histograms vs fitting Gaussian mixture models?



Excursion: excess events at LHC and dead salmons

One can view this also as set of hypothesis tests for the localized deviation…

but be careful about significance in multiple comparisons!



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY; [3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we carried out a real experiment that demonstrates the danger of not correcting for chance properly.

METHODS

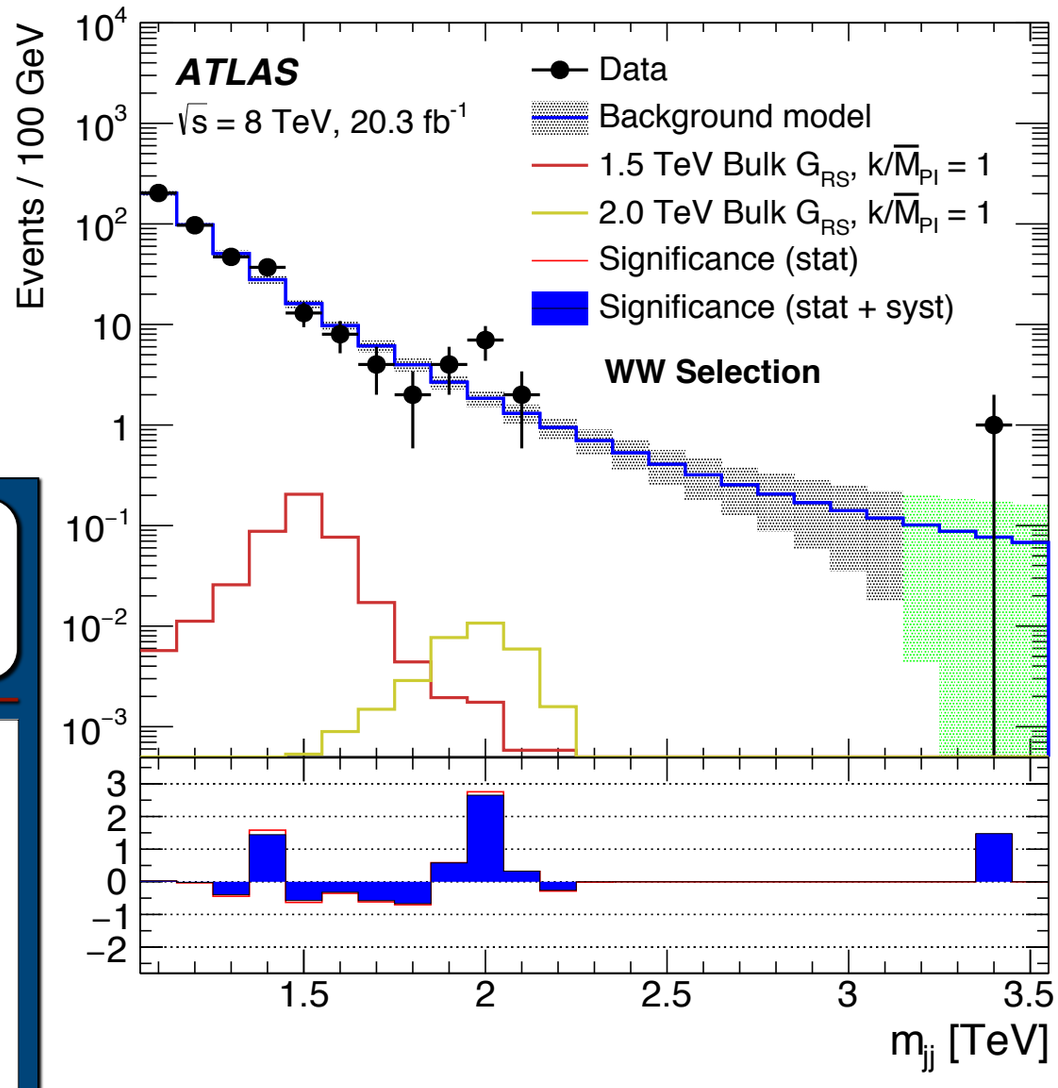
Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

GLM RESULTS

A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

Several active voxels were discovered in a cluster located within the salmon's



ATLAS Collaboration, arXiv.org: 1506.00962 (2016)