The purpose of MCMC is to
provide a fair sampling of some probability distribution

What does acceptance fraction of 0 or 1 mean?

(Integrated) Autocorrelation Time

Autocorrelation time The autocorrelation time is a direct measure of the number of evaluations of the posterior PDF required to produce independent samples of the target density. GW10 show that the stretch-move algorithm has a significantly shorter autocorrelation time on several non-trivial densities. This means that fewer PDF computations

(Integrated) Autocorrelation Time

The autocovariance function of a time series $X(t)$ is

$$C_f(T) = \lim_{t\to\infty} \mathrm{cov}\left[f\left(X(t+T)\right), f\left(X(t)\right)\right]. \tag{11}$$

This measures the covariances between samples at a time lag T. The value of T where $C_f(T) \to 0$ measures the number of samples that must be taken in order to ensure independence. In particular, the relevant measure of sampler efficiency is the integrated autocorrelation time

$$\tau_f = \sum_{T=-\infty}^{\infty} \frac{C_f(T)}{C_f(0)} = 1 + 2\sum_{T=1}^{\infty} \frac{C_f(T)}{C_f(0)}. \tag{12}$$

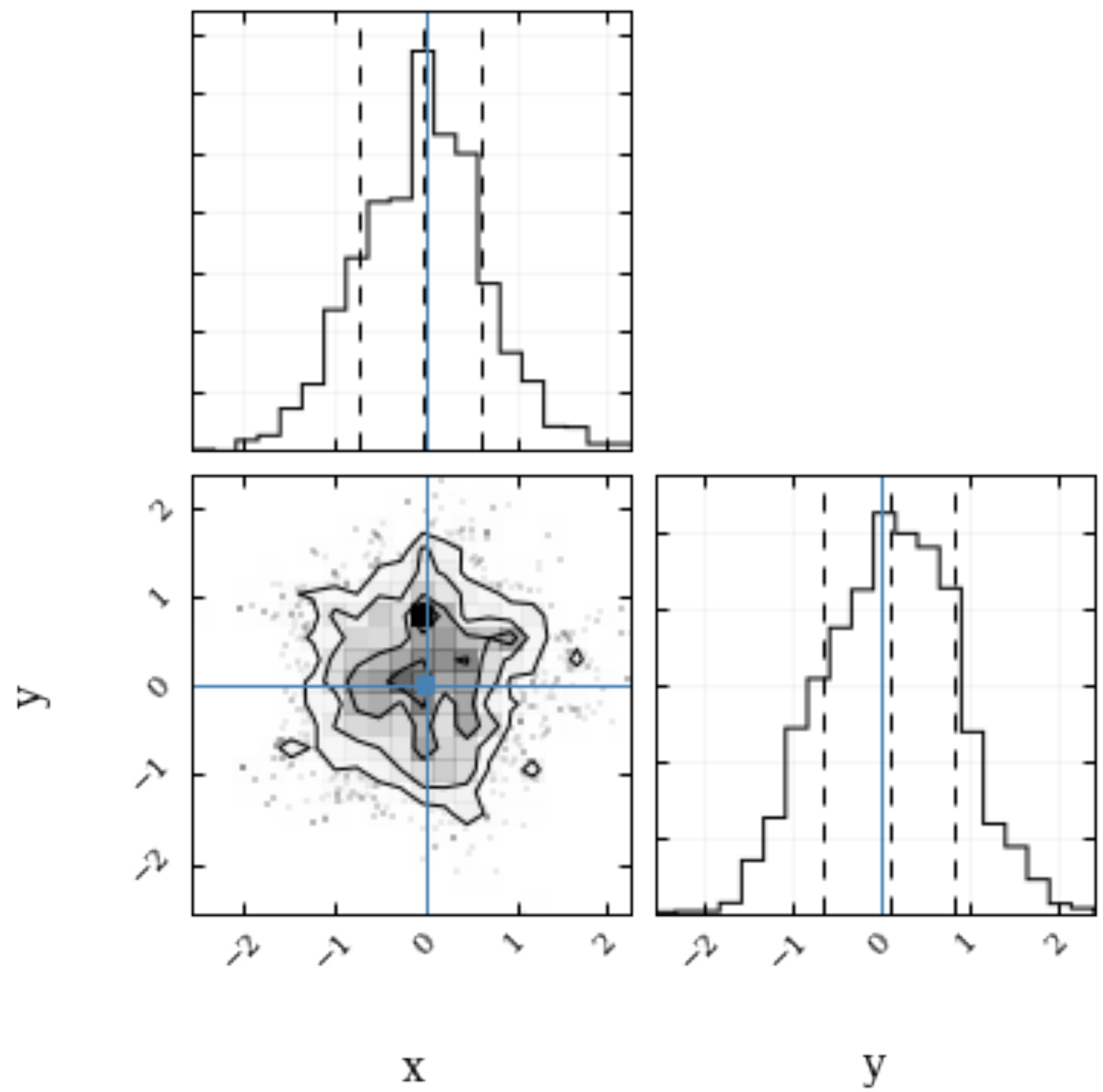
In practice, one can estimate $C_f(T)$ for a Markov chain of M samples as

$$C_f(T) \approx \frac{1}{M-T} \sum_{m=1}^{M-T} \left[f(X(T+m)) - \langle f\rangle\right] \left[f(X(m)) - \langle f\rangle\right]. \tag{13}$$

(Integrated) Autocorrelation Time

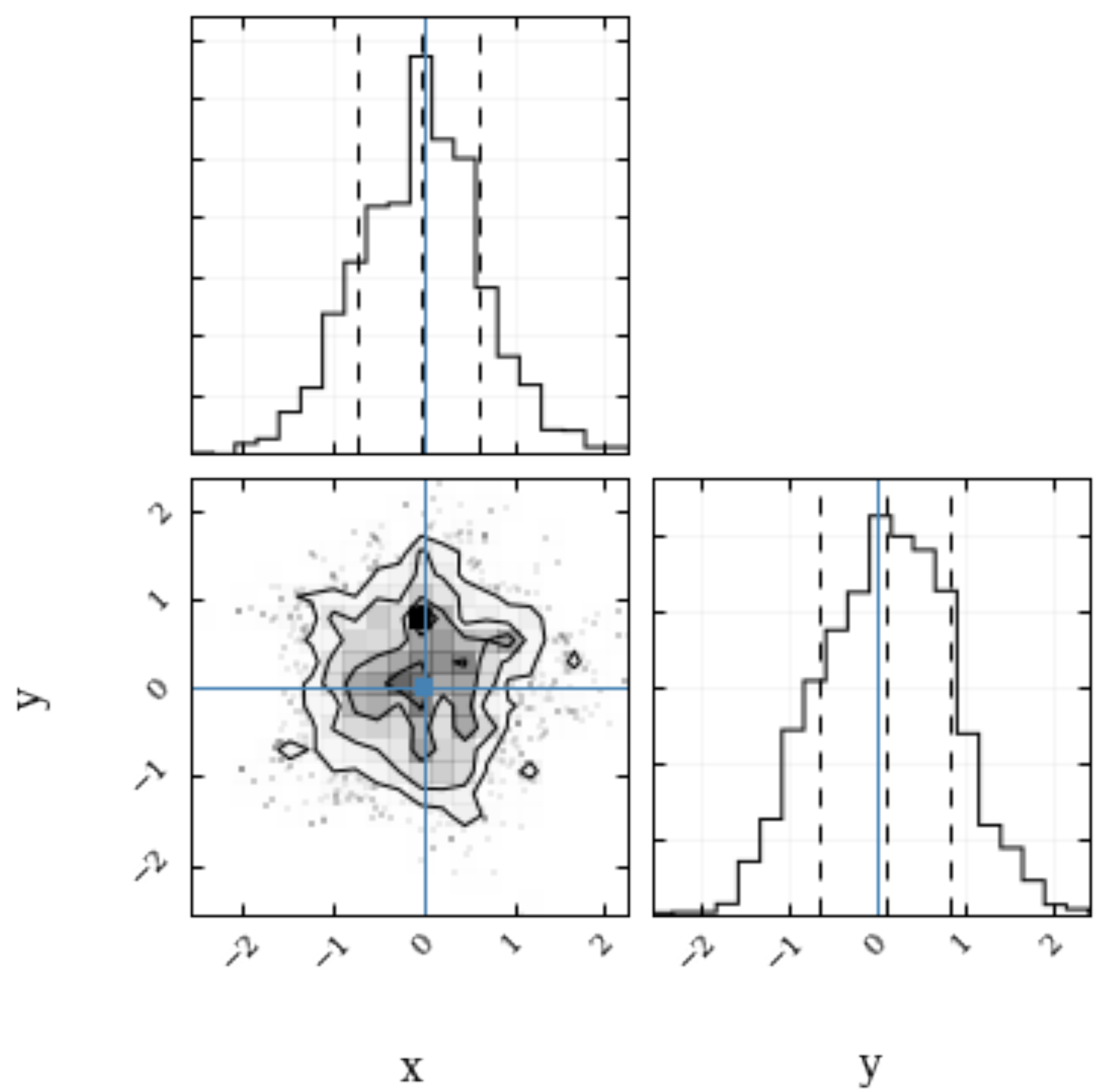
the $E_{p(\theta)}[q(\theta)]$ inside the sum is that estimator too. It is slightly off-topic here, but related to this, if you have a long chain of length K and with autocorrelation time τ and you thin the chain down to every τth sample (instead of using all K samples), you will get just as good sampling-based estimates as if you had used all K samples.

Summary Statistics



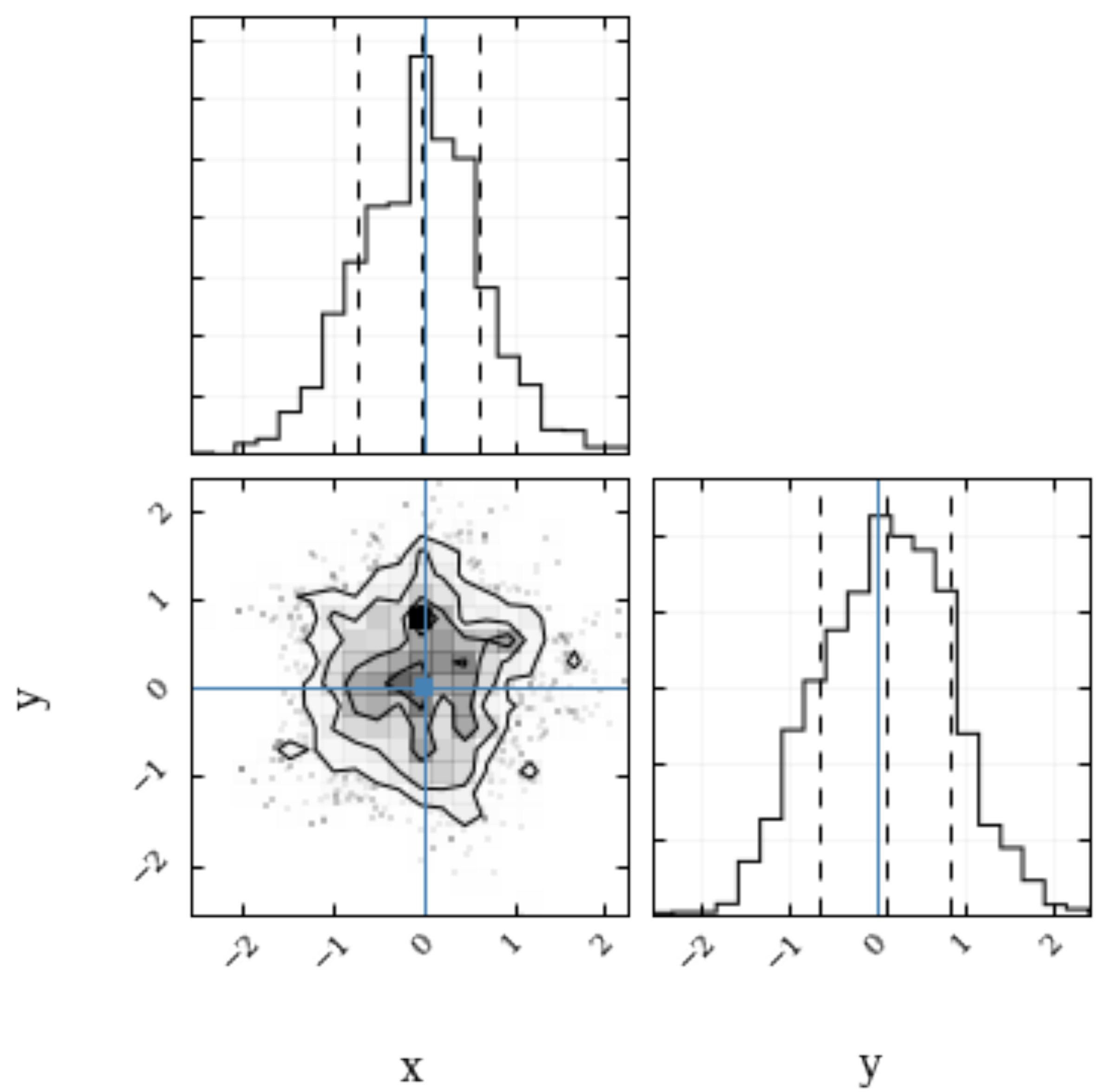
How to report the results of probabilistic inference?

Summary Statistics



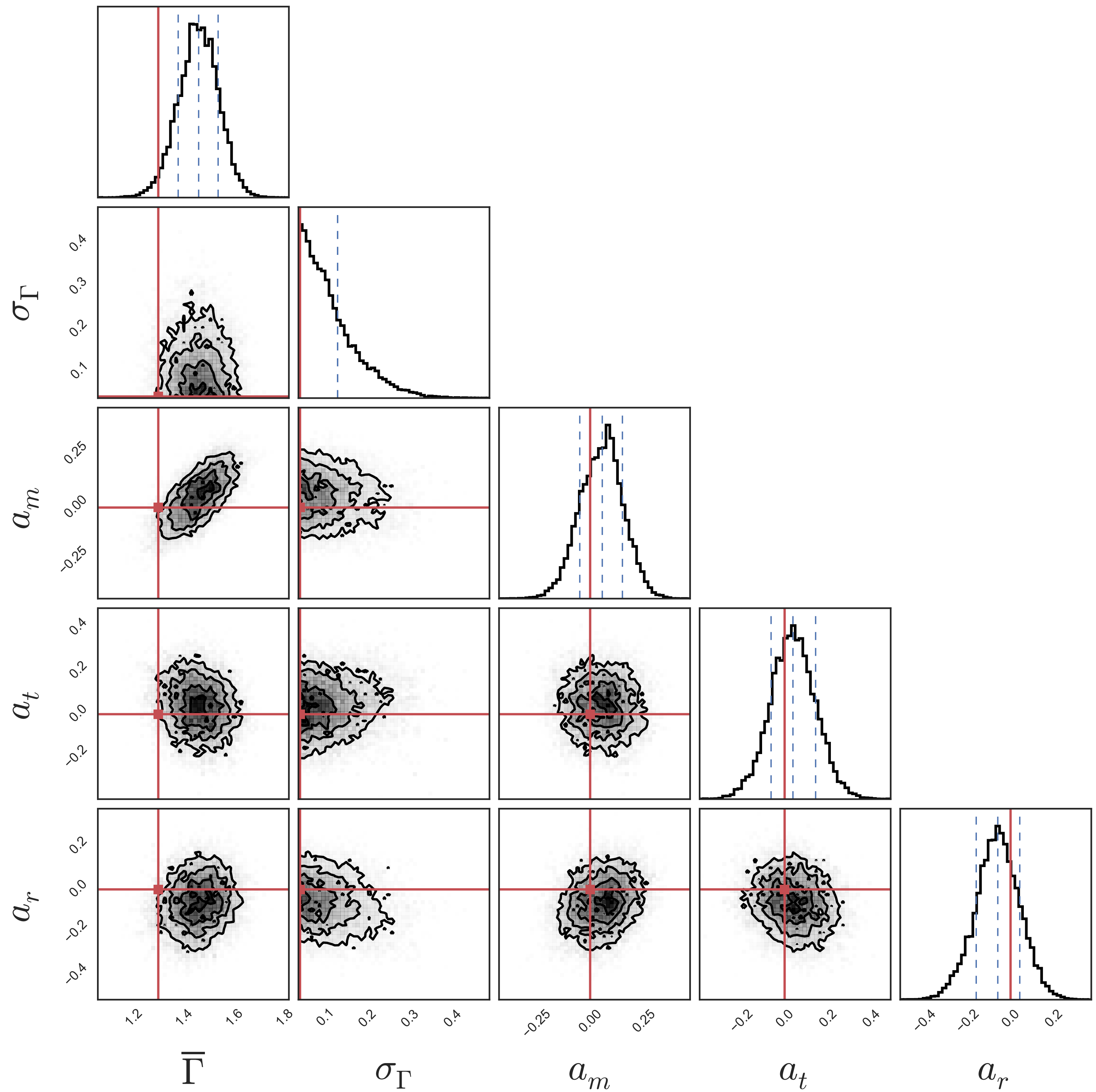
mean, median, mode,
percentiles,
HPD (highest probability density), …

Summary Statistics



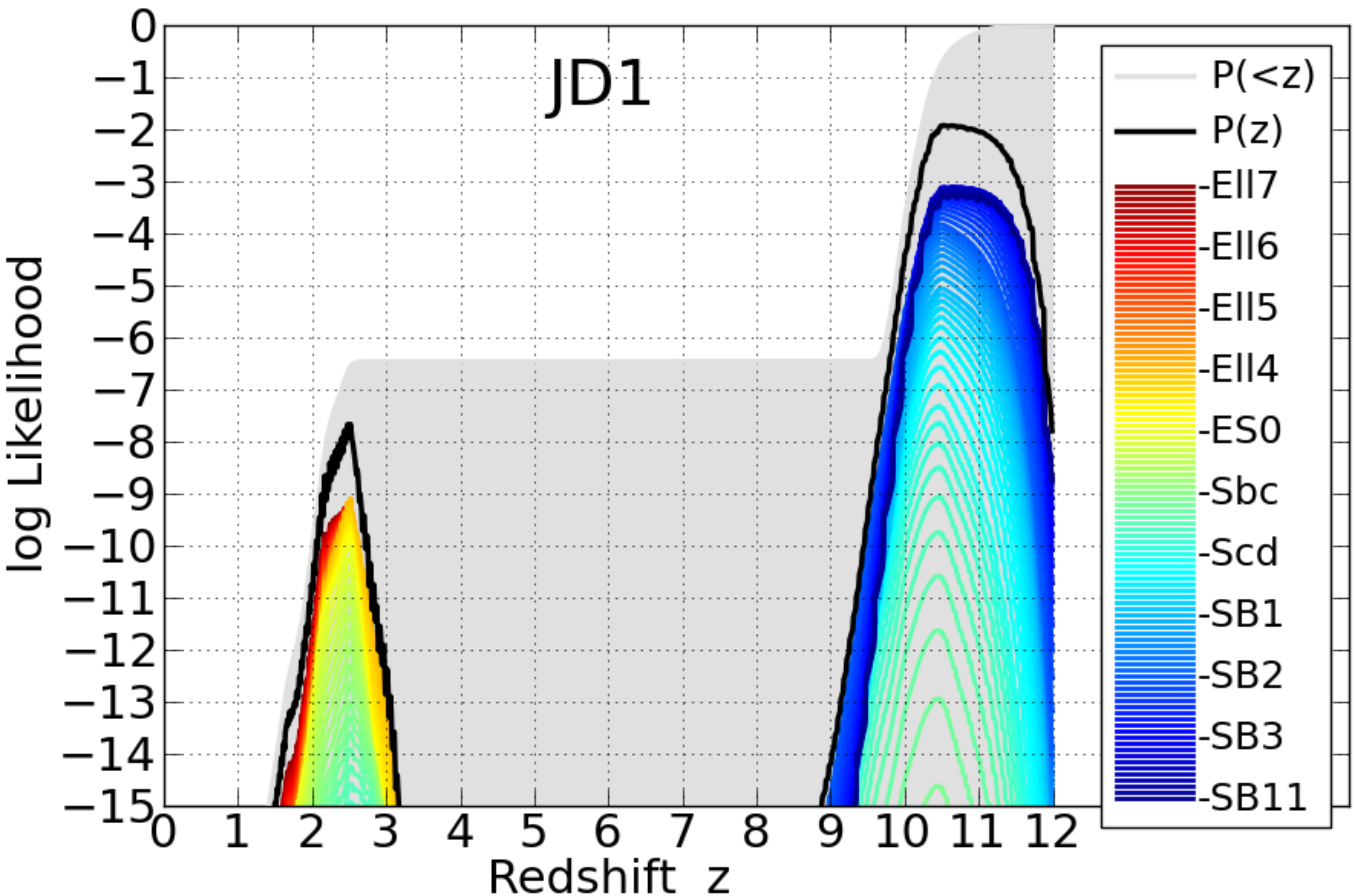
What do you report in the abstract?

Summary Statistics



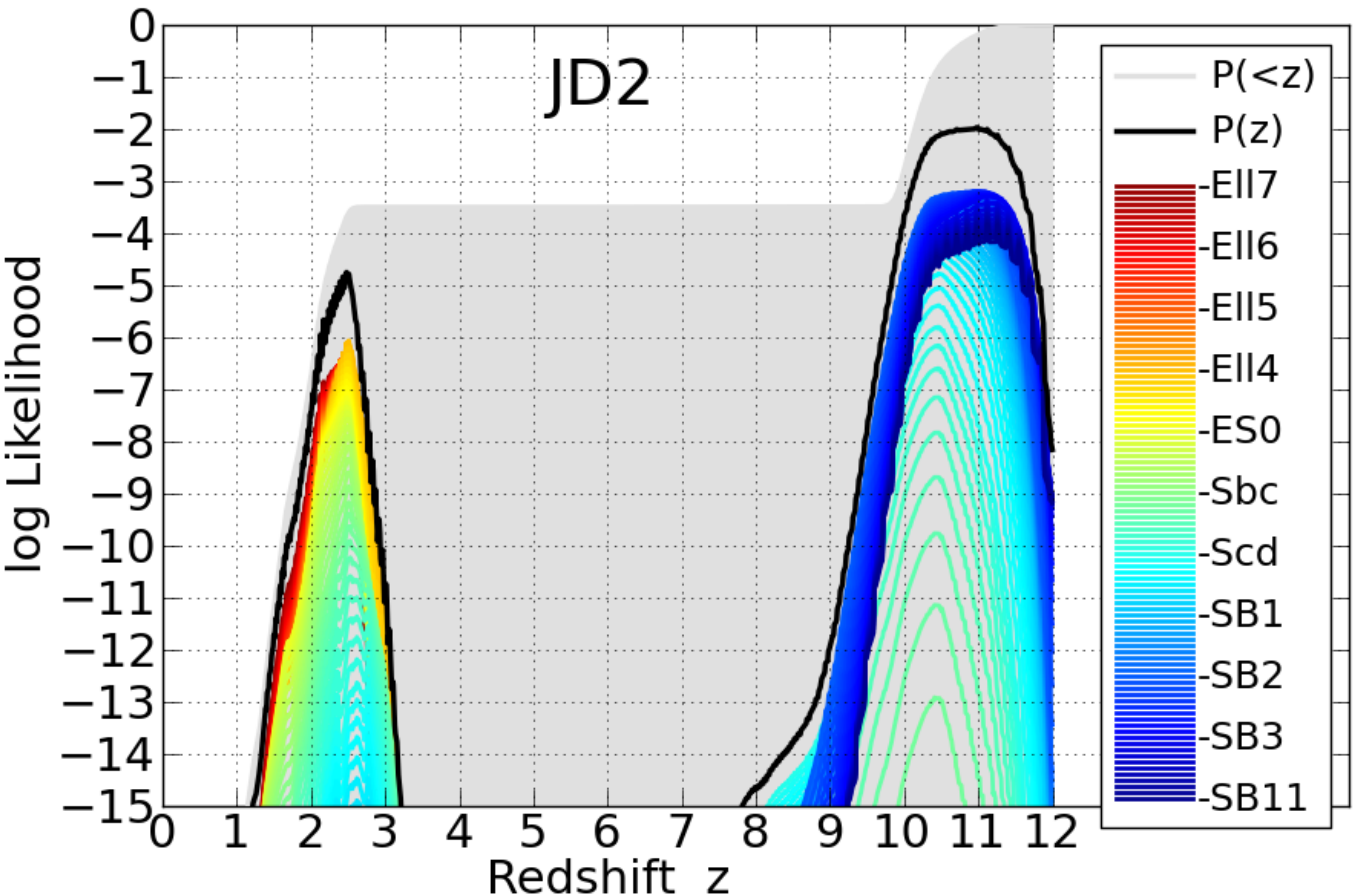
Weisz+ 2015

[CLASH: Three Strongly Lensed Images of a Candidate z ~ 11 Galaxy]



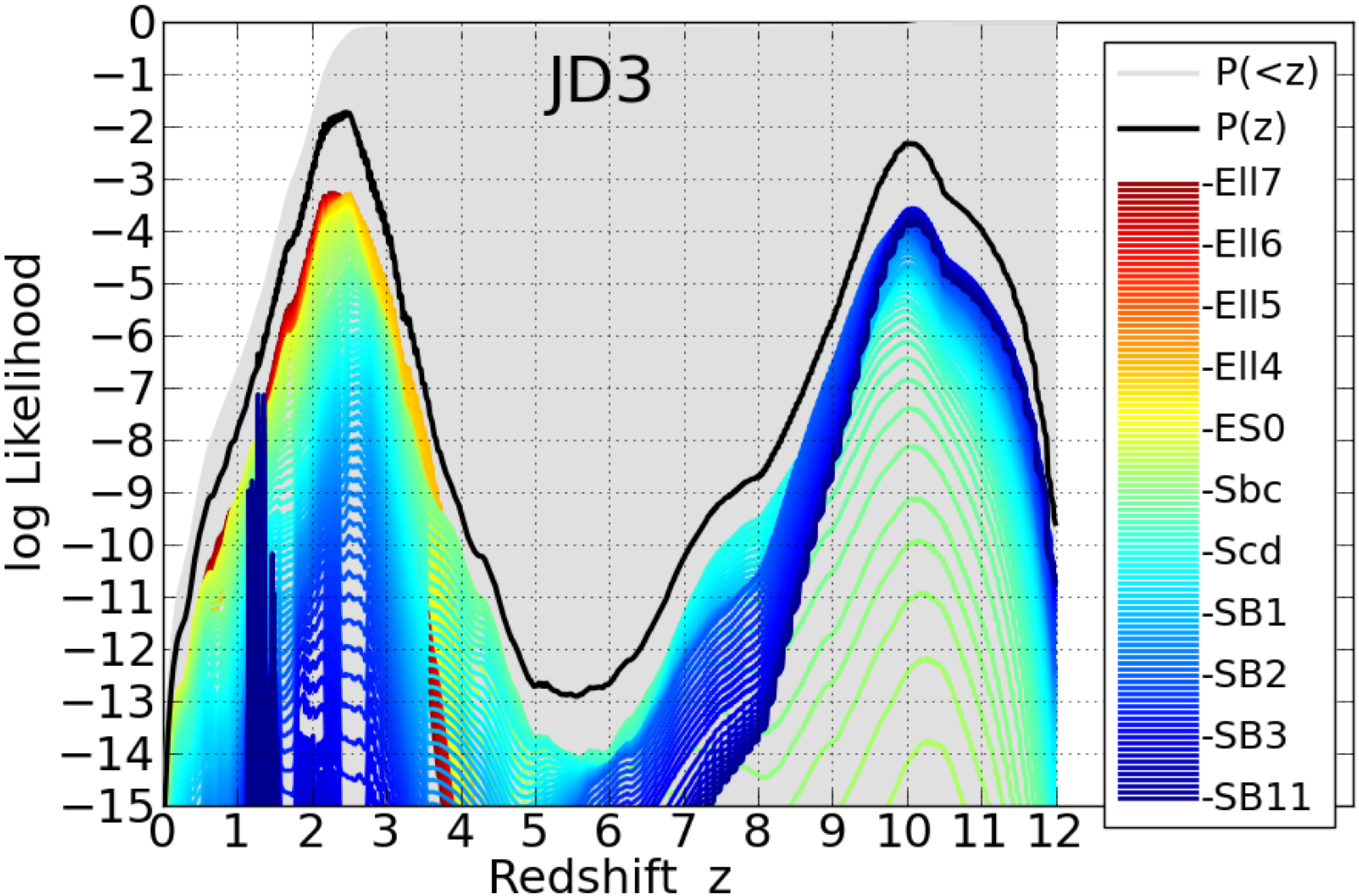
Coe+ 2013

[CLASH: Three Strongly Lensed Images of a Candidate z ~ 11 Galaxy]



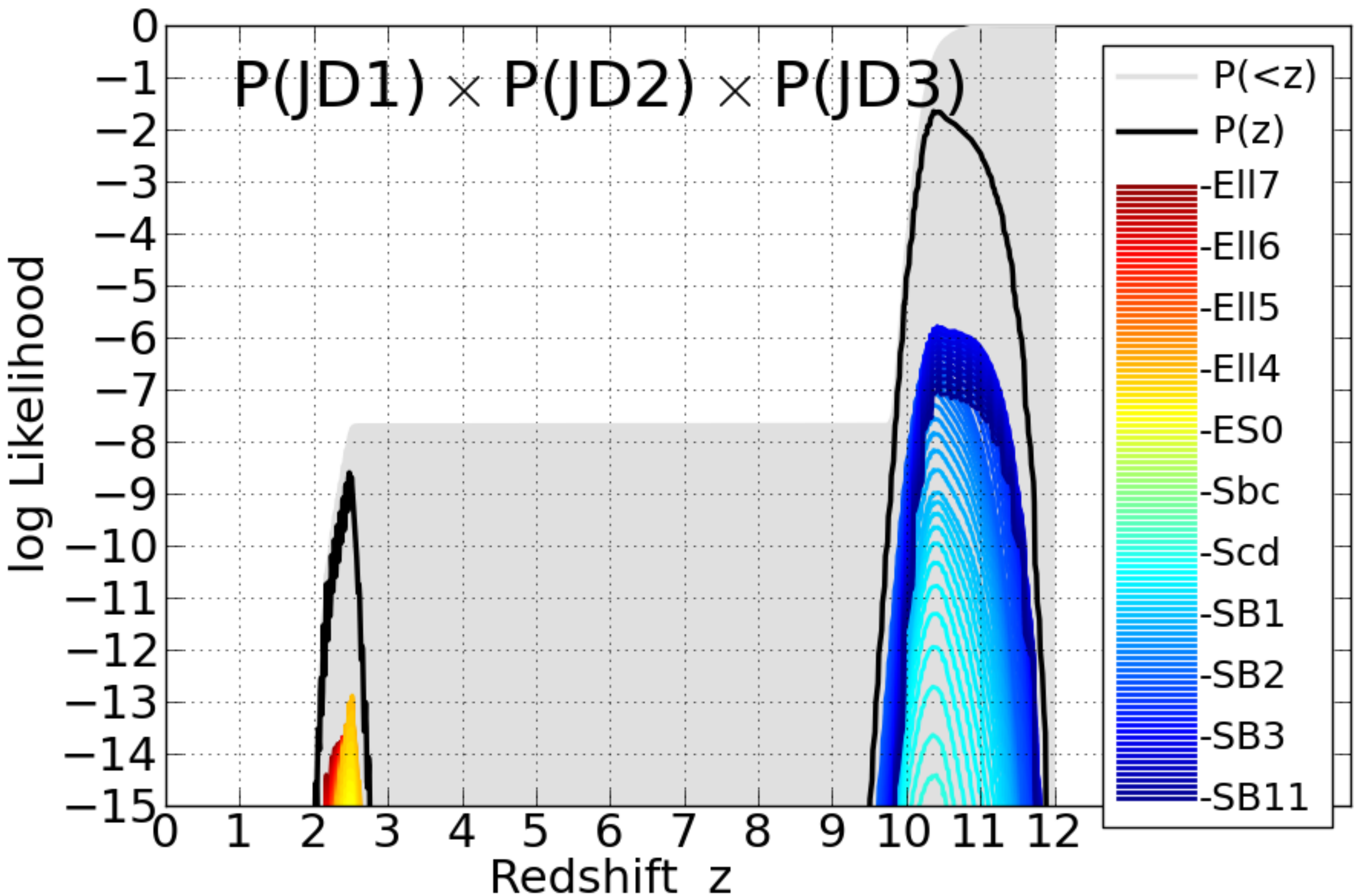
Coe+ 2013

[CLASH: Three Strongly Lensed Images of a Candidate z ~ 11 Galaxy]



Coe+ 2013

[CLASH: Three Strongly Lensed Images of a Candidate z ~ 11 Galaxy]



Coe+ 2013

Hierarchical Modeling

$$P(\theta|D) \propto \; P(D|\theta)P(\theta)$$

Hierarchical Modeling

$$P(\theta) = \frac{P(\theta|\alpha)}{\int P(\theta|\alpha)d\theta} = P(\theta|\alpha)$$

Hierarchical Modeling

$$P(\theta|D) \propto \; P(D|\theta)P(\theta|\alpha)$$

Hierarchical Modeling

$$P(\alpha, \theta | D) \propto \; P(D|\theta) P(\theta|\alpha) P(\alpha)$$

Hierarchical Modeling

$$P(\alpha, \theta | D) \propto \; P(D|\theta)P(\theta|\alpha)P(\alpha)$$

D = set of fluxes for a star
Theta = Model for mass of star
alpha = Stellar IMF

Hierarchical Modeling

Let's say I don't care about the mass of the star(s), but want to learn about the IMF

$$P(\alpha|D) \propto \ [\int P(D|\alpha,\theta)P(\theta|\alpha)d\theta]P(\alpha) = P(D|\alpha)P(\alpha)$$

The mass of the star is now a "nuisance parameter".

Hierarchically estimate the IMF from SED of a star!