# Bayes via forward simulation: Approximate Bayesian Computation

**Jessi Cisewski**
Carnegie Mellon University

June 2014

## What is Approximate Bayesian Computing?

- "Likelihood-free" approach
- Works by simulating from the forward process

## What is Approximate Bayesian Computing?

- "Likelihood-free" approach
- Works by simulating from the forward process

Why not just use the likelihood?

## Issues with writing down a likelihood

1. Physical model too complex or unknown
2. No generally accepted theory
3. Strong dependency in data
4. Observational limitations


- Likelihood-Free Cosmological Inference with Type Ia Supernovae: Approximate Bayesian Computation for a Complete Treatment of Uncertainty (Weyant et al., 2013)
- Likelihood - free inference in cosmology: potential for the estimation of luminosity functions (Schafer and Freeman, 2012)
- Approximate Bayesian Computation for Astronomical Model Analysis: A case study in galaxy demographics and morphological transformation at high redshift (Cameron and Pettitt, 2012)

# ABC for Type Ia Supernovae

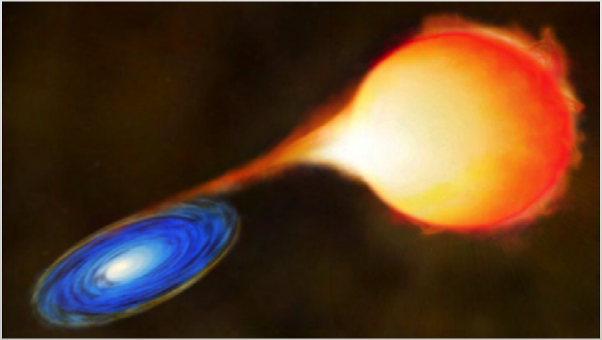


Image: NASA/CXC/M. Weiss (from http://www.universetoday.com

# ABC for Type Ia Supernovae*

- Theory predicts relationship between distance modulus ($mag_{app}$ - $mag_{abs}$), $\mu$, and redshift, $z$

http://i1-news.softpedia-static.com

$$\mu(z,\theta) = 5\log_{10}\left(\frac{c(1+z)}{H_0}\int_0^z \frac{du}{\sqrt{\Omega_m(1+u)^3+(1-\Omega_m)}}\right)+25$$

* Likelihood-Free Cosmological Inference with Type Ia Supernovae: Approximate Bayesian Computation for a Complete Treatment of Uncertainty (Weyant et al., 2013)

# ABC for Type Ia Supernovae

$$\mu(z,\theta) = 5\log_{10}\left(\frac{c(1+z)}{H_0}\int_0^z \frac{du}{\sqrt{\Omega_m(1+u)^3+(1-\Omega_m)}}\right) + 25$$

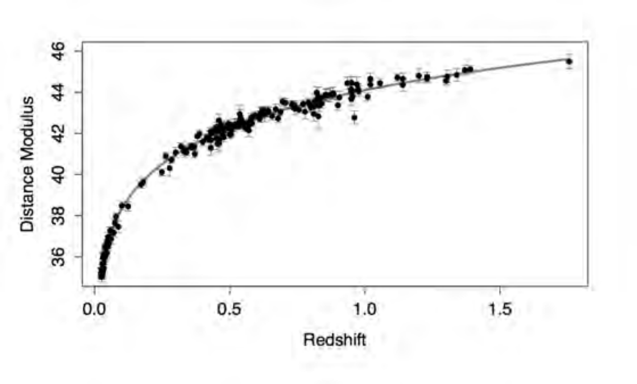


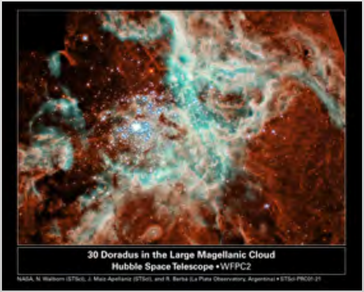

Figure: Chad Schafer/Riess et al. (2007), 182 Type Ia Supernovae

# ABC for Type Ia Supernovae: Why ABC?

## Excerpt from Weyant et al. 2013

For example, with limited spectroscopic follow-up, we must rely on light-curve classification codes and photometric redshift tools to maximize the scientific potential of SN Ia cosmology with LSST and near-future surveys. These two crucial steps alone introduce a nontrivial component to our probability models from which we construct the likelihood. Additionally, there are significant systematic uncertainties including errors from calibration, survey design and cadence, host galaxy subtraction and intrinsic dust, population evolution, gravitational lensing, and peculiar velocities. All of these uncertainties contribute to a probability model which simply cannot be accurately described by a multivariate normal distribution.

# ABC for Stellar Initial Mass Functions



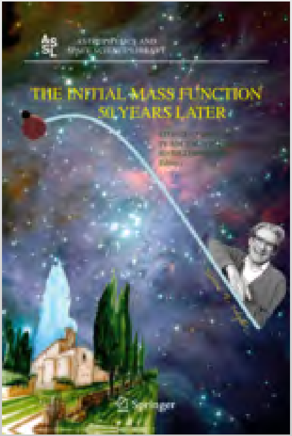




Image: `https://images.springer.com`

# Stellar cluster formation

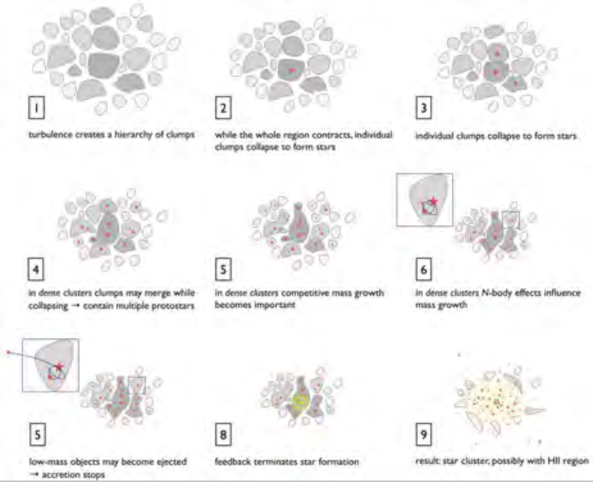



Image: `http://www.astro.ljmu.ac.uk/~ikb/research/imf-use-in-cosmology.html`

# Stellar Initial Mass Functions

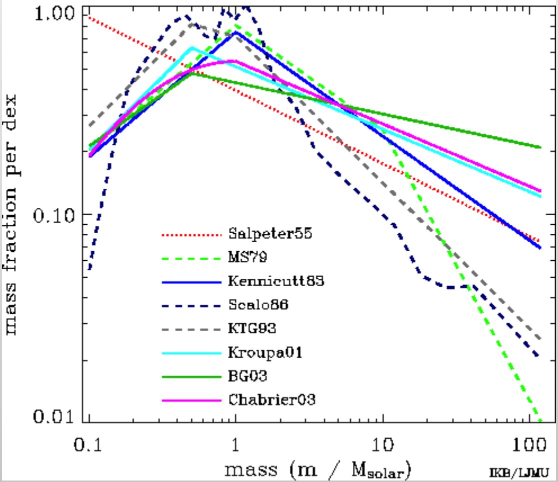



Image: http://physicsforme.files.wordpress.com

# Stellar IMF: probabilistic approach

THE ASTROPHYSICAL JOURNAL, 762:123 (21pp), 2013 January 10 doi:10.1088/0004-637X/762/2/123


THE PANCHROMATIC HUBBLE ANDROMEDA TREASURY. IV. A PROBABILISTIC APPROACH TO INFERRING THE HIGH-MASS STELLAR INITIAL MASS FUNCTION AND OTHER POWER-LAW FUNCTIONS*

DANIEL R. WEISZ[1], MORGAN FOUESNEAU[1], DAVID W. HOGG[2,3], HANS-WALTER RIX[3], ANDREW E. DOLPHIN[4], JULIANNE J. DALCANTON[1], DANIEL T. FOREMAN-MACKEY[2], DUSTIN LANG[5], L. CLIFTON JOHNSON[1], LORI C. BEERMAN[1], ERIC F. BELL[6], KARL D. GORDON[7], DIMITRIOS GOULIERMIS[3,8], JASON S. KALIRAI[7], EVAN D. SKILLMAN[9], AND BENJAMIN F. WILLIAMS[1]

Weisz et al. (2013)

- Used probabilistic model and MCMC for stellar IMF
- Difficult to build the following into a likelihood: realistic completeness function, unresolved binaries, mass segregation, cluster membership probabilities, **lack of independence among the stars within a cluster**
- ABC for stellar IMFs: Weller et al. (2014)

- Density $\rightarrow$ the function $f(y, \xi)$, where $\xi$ is fixed and $y$ is variable
- Likelihood $\rightarrow$ the function $f(y, \xi)$, where $\xi$ is variable and $y$ is fixed

**Definition: the likelihood function**

Given observations $y_{1:n} = (y_1, y_2, \ldots, y_n) = D$ of random variables $Y_{1:n} = (Y_1, Y_2, \ldots, Y_n)$ with joint pdf $f(y_{1:n} \mid \theta)$ for $\theta \in \mathbb{R}^p$, the likelihood function is defined as

$$L(\theta \mid y_{1:n}) = f(y_{1:n} \mid \theta).$$

Assuming observations are independent, the likelihood is written

$$L(\theta \mid y_{1:n}) = \prod_{i=1}^{n} f(y_i \mid \theta).$$

Typically, when we collect data there some data-generating equation connecting the quantity(ies) of interest (e.g. the unknown parameters) with the observations.

Goal: the posterior distribution of the unknown parameter(s) $\theta$.

## Posterior distribution

$$\pi(\theta \mid y_{1:n}) = \frac{L(\theta \mid y_{1:n})\pi(\theta)}{\int L(\theta' \mid y_{1:n})\pi(\theta')d\theta'}$$

Prior: $\pi(\theta)$

# Example: Mean of a Gaussian with known $\sigma^2$

Suppose $\mu$ is unknown and $\sigma^2$ is known and data are sampled from a $N(\mu, \sigma^2)$. Then the **likelihood** is

$$L(\mu, \sigma^2 \mid y_{1:n}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2}$$

If the prior on $\mu$ is $\pi(\mu) \sim N(\mu_0, \sigma_0^2)$, then the **posterior** is

$$\pi(\mu \mid y_{1:n}) \sim N(\mu_1, \sigma_1^2)$$

where $\mu_1 = \dfrac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \qquad \sigma_1^2 = \dfrac{1}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}$

### Posterior distribution

$$\pi(\theta \mid y_{1:n}) = \frac{L(\theta \mid y_{1:n})\pi(\theta)}{\int L(\theta' \mid y_{1:n})\pi(\theta')d\theta'}$$

Prior: $\pi(\theta)$

$\longrightarrow$ In the standard Bayesian set-up, the **likelihood** is required

## Posterior distribution

$$\pi(\theta \mid y_{1:n}) = \frac{L(\theta \mid y_{1:n})\pi(\theta)}{\int L(\theta' \mid y_{1:n})\pi(\theta')d\theta'}$$

Prior: $\pi(\theta)$

With ABC, generate $x_{1:n}$ from the forward process that produced $y_{1:n}$, then approximate the posterior using

$$\pi(\theta \mid \rho(y_{1:n}, x_{1:n}) < \epsilon)$$

where $\rho$ is a distance function.
$\pi(\theta \mid \rho(y_{1:n}, x_{1:n}) < \epsilon) \longrightarrow$ [1]

- $\pi(\theta \mid y_{1:n})$ (the posterior) as $\epsilon \longrightarrow 0$
- $\pi(\theta)$ (the prior) as $\epsilon \longrightarrow \infty$

[1] assuming $\rho$ preserves sufficiency

Basic idea: A statistic is said to be **sufficient** for some parameter $\theta$ if it contains all the information in the data for estimation of $\theta$.

**Definition: Sufficiency**

Let $Y_{1:n}$ be an independent and identically distributed sample of size $n$ from a distribution parameterized by $\theta$. A statistic, $T = T(Y_{1:n})$ is *sufficient* for $\theta$ if the distribution of $Y_{1:n} \mid T$ does not depend on $\theta$.

Examples:

1. $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ is a sufficient statistic for $\mu$ where $Y_i$ are iid $N(\mu, 1)$, $i = 1, \ldots, n$
2. $Y_{(n)} = \max\{Y_1, \ldots, Y_n\}$ is a sufficient statistic for $\theta$ where $Y_i$ are iid $U(0, \theta)$, $i = 1, \ldots, n$

# Sufficient statistics: Bernoulli example

Given a sample $Y_1, Y_2, Y_3, Y_4 \sim Bern(p)$:

$$Y_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Then $Y = \sum_{i=1}^{4} Y_i \sim Binomial(4, p)$ and

$$P(Y = y) = \binom{4}{y} p^y (1 - p)^{1-y}$$

where $y = 0, 1$.

*For example, p could the the proportion of stars in the Milky Way that have at least one planet orbiting it.*

We might wonder, is $\hat{p} = \frac{1}{4} \sum_{i=1}^{4} y_i$ a sufficient statistic for $p$?

# Sufficient statistics: Bernoulli example, continued.

$Y \sim Binomial(4, p)$ and

$$P(Y = y) = \binom{4}{y} p^y (1-p)^{1-y}$$

Suppose we observe $Y_1 = 1,\ Y_2 = 1,\ Y_3 = 0,\ Y_4 = 0$, then

$$\hat{p} = \frac{1}{4}\sum_{i=1}^{4} y_i = 0.5.$$

$$P((1,1,0,0) \mid \hat{p} = 0.5) = \frac{P\left((1,1,0,0) \cap \hat{p} = 0.5\right)}{P(\hat{p} = 0.5)} = \frac{P(1,1,0,0)}{P(\hat{p} = 0.5)}$$
$$= \frac{p^2(1-p)^2}{\binom{4}{2} p^2 (1-p)^2} = \frac{1}{\binom{4}{2}}$$

Because $P((1,1,0,0) \mid \hat{p} = 0.5)$ does not depend on the unknown parameter $p$, $\hat{p}$ is a sufficient statistic for $p$.

# Sufficient statistics: Gaussian example with known $\sigma^2$

Consider an independent and identically distributed sample of size $n$ from a $N(\mu, \sigma^2)$ where $\mu$ is unknown and $\sigma^2$ is known.

$$
\begin{aligned}
L(\mu \mid y_{1:n}, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} y_i^2 - 2\mu\sum_{i=1}^{n} y_i + n\mu^2)} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2} \cdot e^{-\frac{1}{2\sigma^2}(-2\mu\sum_{i=1}^{n} y_i + n\mu^2)}
\end{aligned}
$$

By the Factorization Criterion, $\sum_{i=1}^{n} y_i$ is a sufficient statistic for $\mu$.

# Basic ABC algorithm

For the observed data $y_{1:n}$, prior $\pi(\theta)$ and distance function $\rho$:

**Algorithm***

1. Sample $\theta^*$ from prior $\pi(\theta)$
2. Generate $x_{1:n}$ from forward process $f(y \mid \theta^*)$
3. Accept $\theta^*$ if $\rho(y_{1:n}, x_{1:n}) < \epsilon$
4. Return to step 1

Generates a sample from an approximation of the posterior:

$$f(x_{1:n} \mid \rho(y_{1:n}, x_{1:n}, \theta) < \epsilon) \cdot \pi(\theta) \approx f(y_{1:n} \mid \theta)\pi(\theta) \propto \pi(\theta \mid y_{1:n})$$

*Introduced in Pritchard et al. (1999) (population genetics)

# Binomial illustration

- Example from Turner and Zandt (2012)
- Data are a sample of 1's and 0's coming from $Y_i \sim \text{Bern}(p)$ where $n =$ sample size, $p = P(Y = 1)$.
- Likelihood is $L(p \mid y) = \binom{n}{y} p^y (1-p)^{n-y}$, where $y = \sum_{i=1}^{n} y_i$ (but we will pretend we do not know this).

Need to determine a distance function, $\rho$. Use the following:

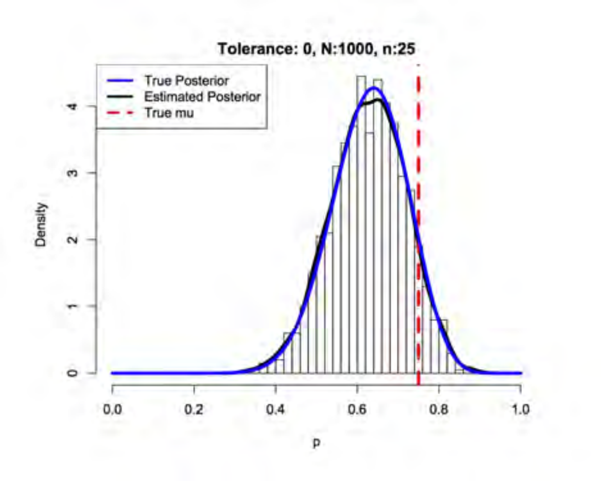
$$\rho(y, x) = \frac{1}{n}|y - x|$$

Hence $\rho(y, x) = 0$ if the generated dataset $x$ has the same number of 1's as $y$.

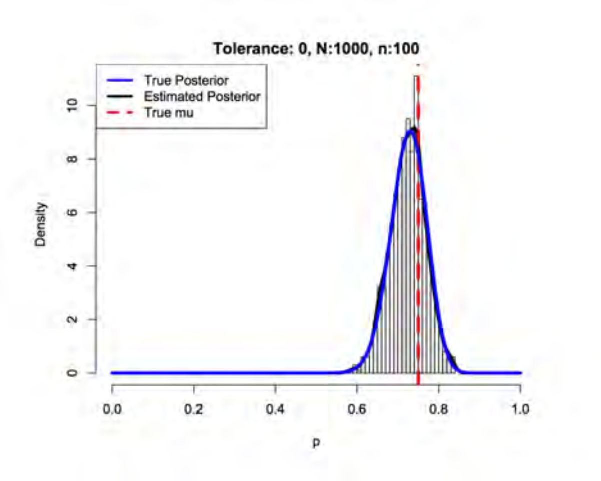
# Binomial illustration: R code (Turner and Zandt 2012)

```
n=1000  #number of observations
N=1000  #generated sample size
true.p=.75
data=rbinom(n,1,true.p)
epsilon=0
alpha.hyper = 1
beta.hyper = 1
p=numeric(N)
rho=function(y,x) abs(sum(y)-sum(x))/n
for(i in 1:N){
    d=epsilon+1
    while(d>epsilon) {
        proposed.p=rbeta(1,alpha.hyper,beta.hyper)
        x=rbinom(n,1,proposed.p)
        d=rho(data,x)}
    p[i]= proposed.p}
```

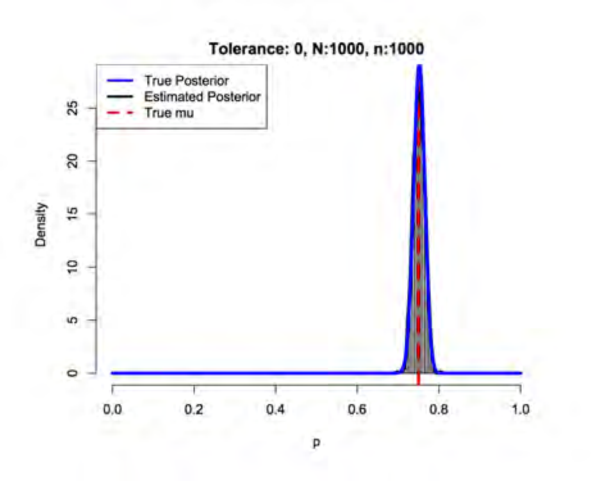
# Binomial illustration: posterior



# Binomial illustration: posterior



# Binomial illustration: posterior



# Mean of a Gaussian with known variance

Consider the following model:

$$\mu \sim N(\mu_0, \sigma_0^2)$$
$$Y_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

**Recall**: the **posterior** is

$$\pi(\mu \mid y_{1:n}) \sim N(\mu_1, \sigma_1^2)$$

where

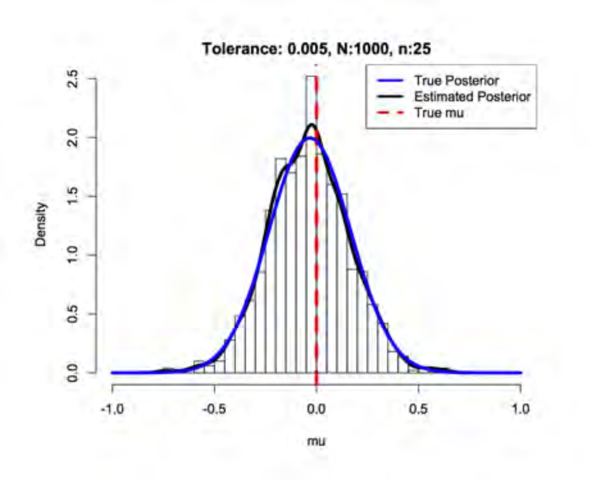
$$\mu_1 = \frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \qquad \sigma_1^2 = \frac{1}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}$$

# Mean of a Gaussian with known variance: R code

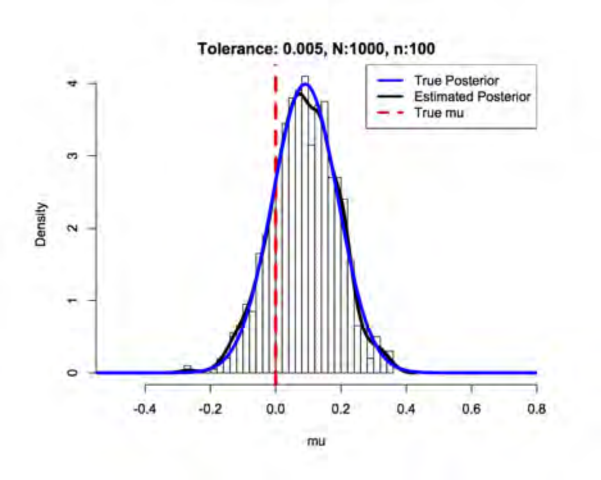
```
n=25        #number of observations
N=1000      #particle sample size
true.mu = 0; sigma = 1
mu.hyper = 0; sigma.hyper = 10
data=rnorm(n,true.mu,sigma)
epsilon=0.005
mu=numeric(N)
rho=function(y,x) abs(sum(y)-sum(x))/n

for(i in 1:N){
d= epsilon +1
   while(d>epsilon) {
       proposed.mu=rnorm(1,0,sigma.hyper) #<--prior draw
       x=rnorm(n, proposed.mu, sigma)
       d=rho(data,x)}
mu[i]= proposed.mu}}
```

# Mean of a Gaussian with known variance: posterior



# Mean of a Gaussian with known variance: posterior



# Some key questions in ABC:

1. What **summary** of the data will preserve useful information for estimating $\theta$?
   $\longrightarrow$ **selecting a summary statistic**

2. What **distance metric** should be used to compare the simulated and observed summary statistics?
   $\longrightarrow$ **selecting** $\rho$

3. What **threshold** should be used to accept/reject candidates?
   $\longrightarrow$ **selecting** $\epsilon$

# ABC for Type Ia Supernovae: algorithm

Type Ia supervovae - theory predicts relationship between redshift and distance modulus

$$\mu(z,\theta) = 5\log_{10}\left(\frac{c(1+z)}{H_0}\int_0^z \frac{du}{\sqrt{\Omega_m(1+u)^3+(1-\Omega_m)}}\right)+25$$

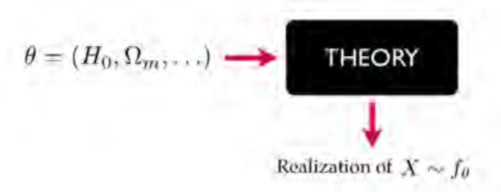



Figure: Chad Schafer

Observations $(z_i, \mu_i)$ are drawn from $\mu_i = \mu(z_i,\theta) + \sigma_i\epsilon_i$

⋆ Weyant et al. (2013)

# ABC for Type Ia Supernovae: algorithm

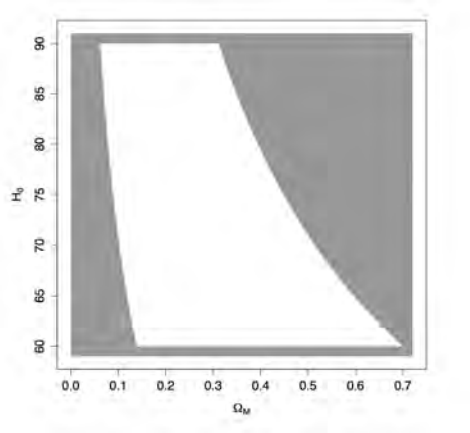



Figure: Chad Schafer

★ Weyant et al. (2013)

# ABC for Type Ia Supernovae: algorithm

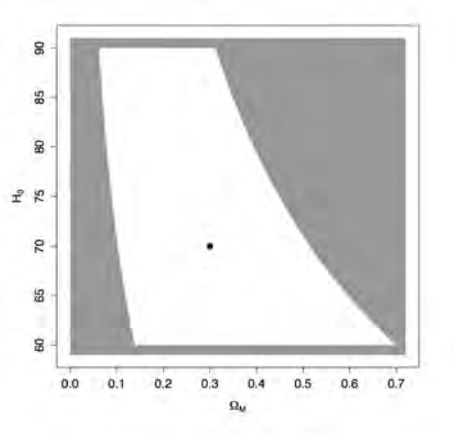



Figure: Chad Schafer

★ Weyant et al. (2013)

# ABC for Type Ia Supernovae: algorithm

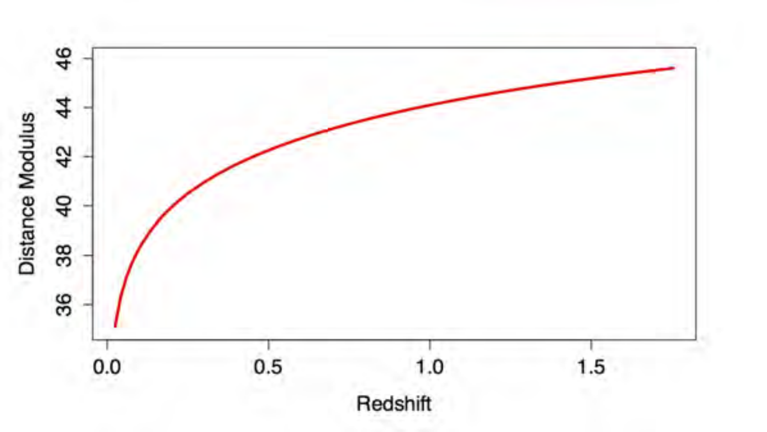



Figure: Chad Schafer

★ Weyant et al. (2013)

# ABC for Type Ia Supernovae: algorithm

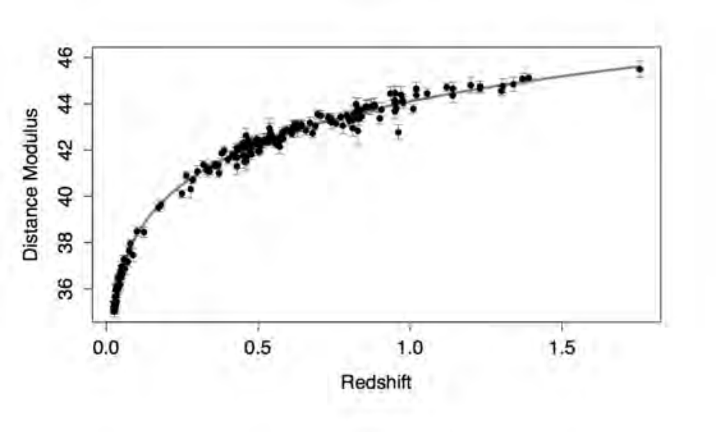



Figure: Chad Schafer

★ Weyant et al. (2013)

# ABC for Type Ia Supernovae: algorithm

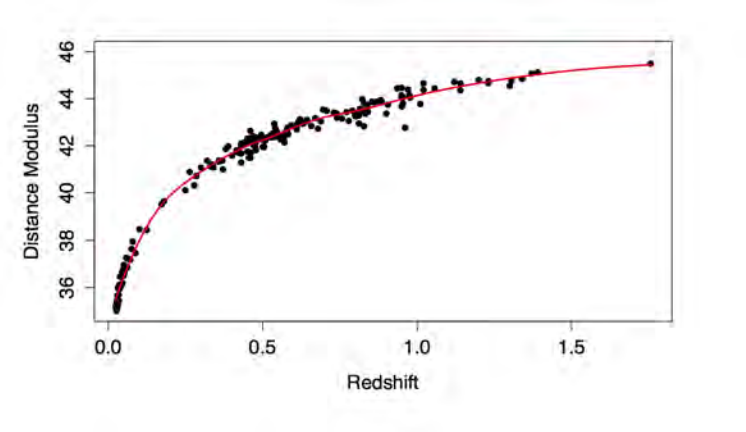



Figure: Chad Schafer

★ Weyant et al. (2013)

# ABC for Type Ia Supernovae: algorithm

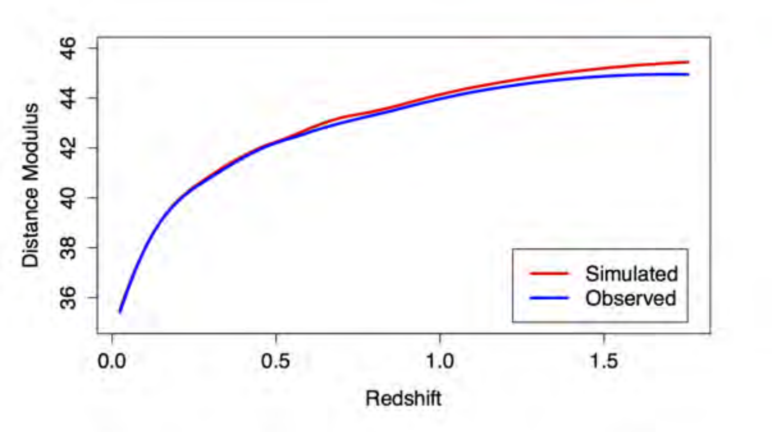



Figure: Chad Schafer

⋆ Weyant et al. (2013)

# ABC for Type Ia Supernovae

Goal: posteriors for $\Omega_M$ and $\omega$

## Initial ABC Algorithm steps

1. Simulate from priors: $\Omega_M^* \sim U(0,1)$ and $\omega^* \sim U(-3,0)$
   $z^* \sim (1+z)^\beta$, $\beta = 1.5 \pm 0.6$
2. Obtain sample of $\mu^*$ via Ia light curve generating forward model $f(\Omega_M^*, \omega^*, z^*, \eta)$
3. Nonparametric smoothing of generated sample $(z^*, \mu^*)$: $(\tilde{z}, \tilde{\mu})$
4. If $\rho((\tilde{z}, \tilde{\mu}), (z, \mu)) \leq \epsilon \longrightarrow$ keep $\Omega_M^*$ and $\omega^*$

- $\eta =$ nuisance parameters, $(z, \mu)$ are the *smoothed* real observations
- Forward model uses SNANA/MLCS2k2 to get the Ia SN light curves

★ Weyant et al. (2013)

# In a nutshell

"The basic idea behind ABC is that using a representative (enough) summary statistic $\eta$ coupled with a small (enough) tolerance $\epsilon$ should produce a good (enough) approximation to the posterior..."

Marin et al. (2012)

# Additional useful resources

- http://approximatebayesiancomputational.wordpress.com/
- Csilléry et al. (2010): Approximate Bayesian Computation (ABC) in practice
- Csillery et al. (2012): abc: an R package for approximate Bayesian computation (ABC)
- Jabot et al. (2013): EasyABC: performing efficient approximate Bayesian computation sampling schemes (R package)

# Concluding remarks

$^{\star}$Approximate Bayesian Computation could be a useful tool in astronomy, **but** it should be used with care$^{\star}$

# Bibliography

Cameron, E. and Pettitt, A. N. (2012), "Approximate Bayesian Computation for Astronomical Model Analysis: A Case Study in Galaxy Demographics and Morphological Transformation at High Redshift," *Monthly Notices of the Royal Astronomical Society*, 425, 44–65.

Csilléry, K., Blum, M. G., Gaggiotti, O. E., and François, O. (2010), "Approximate Bayesian Computation (ABC) in practice," *Trends in ecology & evolution*, 25, 410 – 418.

Csillery, K., Francois, O., and Blum, M. G. B. (2012), "abc: an R package for approximate Bayesian computation (ABC)," *Methods in Ecology and Evolution*.

Jabot, F., Faure, T., and Dumoullin, N. (2013), *EasyABC: performing efficient approximate Bayesian computation*.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012), "Approximate Bayesian computational methods," *Statistics and Computing*, 22, 1167 – 1180.

Pritchard, J. K., Seielstad, M. T., and Perez-Lezaun, A. (1999), "Population Growth of Human Y Chromosomes: A study of Y Chromosome Microsatellites," *Molecular Biology and Evolution*, 16, 1791 – 1798.

Riess, A. G., Strolger, L.-G., Casertano, S., Ferguson, H. C., Mobasher, B., Gold, B., Challis, P. J., Filippenko, A. V., Jha, S., Li, W., et al. (2007), "New Hubble space telescope discoveries of type Ia supernovae at z$\geq$ 1: narrowing constraints on the early behavior of dark energy," *The Astrophysical Journal*, 659, 98.

Schafer, C. M. and Freeman, P. E. (2012), *Statistical Challenges in Modern Astronomy V*, Springer, chap. 1, Lecture Notes in Statistics, pp. 3 – 19.

Turner, B. M. and Zandt, T. V. (2012), "A tutorial on approximate Bayesian computation," *Journal of Mathematical Psychology*, 56, 69 – 85.

Weisz, D. R., Fouesneau, M., Hogg, D. W., Rix, H.-W., Dolphin, A. E., Dalcanton, J. J., Foreman-Mackey, D. T., Lang, D., Johnson, L. C., Beerman, L. C., et al. (2013), "The Panchromatic Hubble Andromeda Treasury. IV. A Probabilistic Approach to Inferring the High-mass Stellar Initial Mass Function and Other Power-law Functions," *The Astrophysical Journal*, 762, 123.

Weller, G. B., Cisewski, J., Schafer, C. M., and Hogg, D. W. (2014), "Approximate Bayesian Computation for the High Mass End of the Stellar Initial Mass Function," In preparation.

Weyant, A., Schafer, C., and Wood-Vasey, W. M. (2013), "Likeihood-free cosmological inference with type Ia supernovae: approximate Bayesian computation for a complete treatment of uncertainty," *The Astrophysical Journal*, 764, 116.