Approximate Bayesian Computation

Jessi Cisewski
Department of Statistics
Yale University

October 26, 2016

SAMSI Astrostatistics Course

Our goals...

- Approximate Bayesian Computation (ABC)

Background: Bayesian methods

Goal: the posterior distribution of the unknown parameter(s) θ.

Posterior distribution

$$\pi(\theta \mid \overbrace{y}^{\text{Data}}) = \frac{\overbrace{f(y \mid \theta)}^{\text{Likelihood}} \cdot \overbrace{\pi(\theta)}^{\text{Prior}}}{f(y)} \propto f(y \mid \theta)\pi(\theta)$$

- The prior distribution allows you to "easily" incorporate your beliefs about the parameter(s) of interest
- Posterior is a distribution on the parameter space given the observed data

The posterior for θ given observed data x_{obs}:

$$\pi(\theta \mid x_{\text{obs}}) = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{\int f(x_{\text{obs}} \mid \theta)\pi(\theta)d\theta} = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{f(x_{\text{obs}})}$$

The posterior for θ given observed data x_{obs}:

$$\pi(\theta \mid x_{\text{obs}}) = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{\int f(x_{\text{obs}} \mid \theta)\pi(\theta)d\theta} = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{f(x_{\text{obs}})}$$

Approximate Bayesian Computation

- "Likelihood-free" approach to approximating $\pi(\theta \mid x_{\text{obs}})$
- Proceeds via simulation of the forward process

The posterior for θ given observed data x_{obs}:

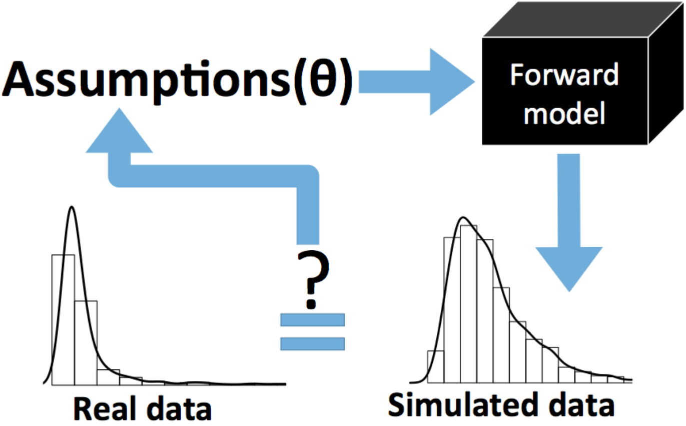
$$\pi(\theta \mid x_{\text{obs}}) = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{\int f(x_{\text{obs}} \mid \theta)\pi(\theta)d\theta} = \frac{f(x_{\text{obs}} \mid \theta)\pi(\theta)}{f(x_{\text{obs}})}$$

Approximate Bayesian Computation

- "Likelihood-free" approach to approximating $\pi(\theta \mid x_{\text{obs}})$
- Proceeds via simulation of the forward process

Why would we not know $f(x_{\text{obs}} \mid \theta)$?

1. Physical model too complex to put in analytical form
2. Strong dependency in data
3. Observational limitations


Assumptions(θ)
Forward model
?
=
Real data
Simulated data

ABC for Astronomy

- cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation (Ishida et al., 2015)
- Approximate Bayesian Computation for Forward Modeling in Cosmology (Akeret et al., 2015)
- 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)

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.

From Weyant et al. (2013)

Basic ABC algorithm

Basic ABC algorithm

For the observed data x_{obs} and prior $\pi(\theta)$:

Algorithm*

1. Sample θ_{prop} from prior $\pi(\theta)$
2. Generate x_{prop} from forward process $f(x \mid \theta_{\text{prop}})$
3. Accept θ_{prop} if $x_{\text{obs}} = x_{\text{prop}}$
4. Return to step 1

Generates a sample from an approximation of the posterior

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

Step 3: Accept θ_{prop} if $x_{\text{obs}} = x_{\text{prop}}$

- Waiting for proposals such that $x_{\text{obs}} = x_{\text{prop}}$ would be computationally prohibitive

  Instead, accept proposals with $\Delta(x_{\text{obs}}, x_{\text{prop}}) \leq \epsilon$
  for some distance Δ and some tolerance threshold ϵ

- When x is high-dimensional, will have to make ϵ too large in order to keep acceptance probability reasonable.

  Instead, reduce the dimension by comparing summaries
  $S(x_{\text{prop}})$ and $S(x_{\text{obs}})$

Simple examples

ABC illustration: binomial distribution

- Data $y_{1:n} \overset{iid}{\sim} \text{Bern}(\theta)$ where n = sample size, $\theta = P(Y = 1)$
- Forward process $F(x \mid \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, where $x = \sum_{i=1}^{n} x_i$
(In this case, we use the likelihood)
- Distance function $\rho(y, x) = |y - x|$
Hence $\rho(y, x) = 0$ if the generated dataset x has the same number of 1's as y
- Tolerance $\epsilon = 0$
- Prior $\pi(\theta) = \text{Uniform}(0,1)$

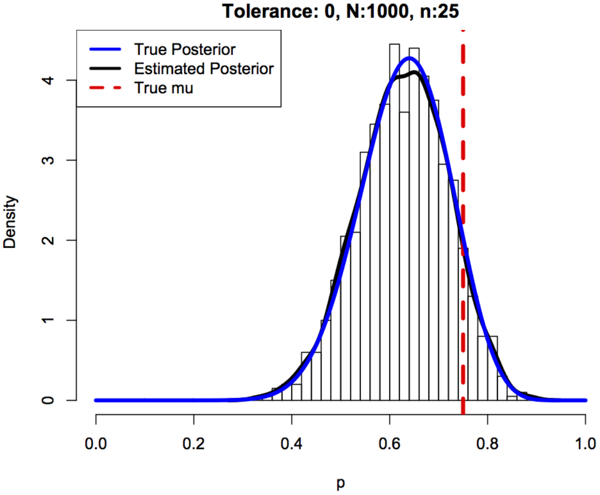
Reference: Turner and Zandt (2012)

Binomial illustration: R code

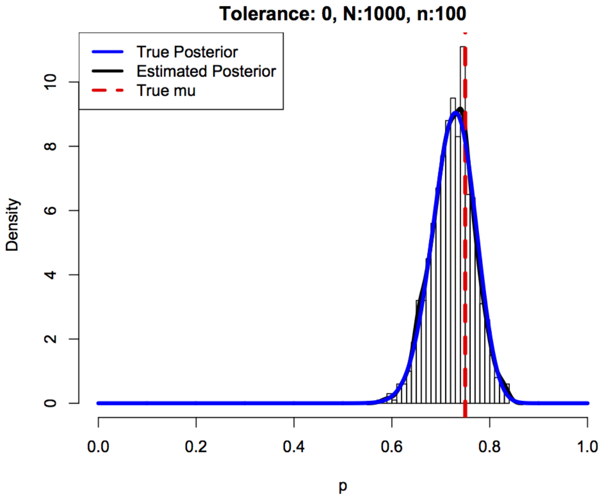
```
n <- 25    #number of observations
N <- 1000  #generated sample size
true.theta <- .75
data <- rbinom(n,1,true.theta)
epsilon <- 0   #Tolerance
theta <- numeric(N)
rho <- function(y,x) abs(sum(y)-sum(x))/n #Distance function


for(i in 1:N){
    d <- epsilon+1
    while(d>epsilon) {
            proposed.theta <- rbeta(1,1,1)   #Prior
            x <- rbinom(n,1,proposed.theta)  #Forward process
            d <- rho(data,x)
    }
    theta[i] <- proposed.theta
}
```

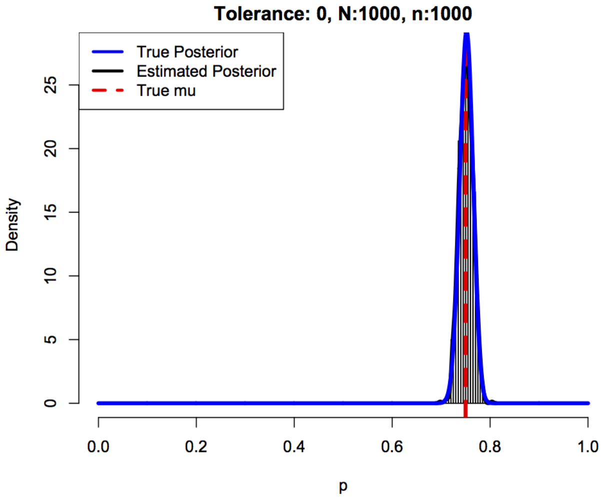
Binomial illustration: ABC posterior



Binomial illustration: ABC posterior



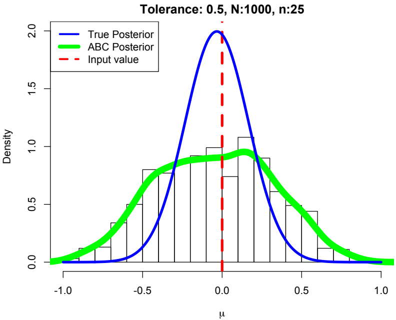
Binomial illustration: ABC posterior

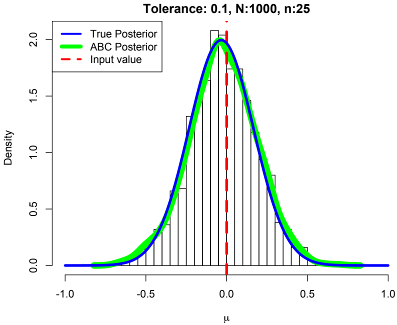


Gaussian illustration

- Data x_{obs} consists of 25 iid draws from Normal$(\mu, 1)$
- Summary statistics $S(x) = \bar{x}$
- Distance function $\Delta(S(x_{prop}), S(x_{obs})) = |\bar{x}_{prop} - \bar{x}_{obs}|$
- Tolerance $\epsilon = 0.50$ and 0.10
- Prior $\pi(\mu) =$ Normal(0,10)

Gaussian illustration: posteriors for μ





Some intuition about ABC

If one wants to generate a draw from the posterior:

1. Draw θ_{prop} from the prior $\pi(\theta)$.
2. Draw x_{prop} from the density $f(x \mid \theta_{\text{prop}})$.
3. Accept θ_{prop} if $x_{\text{prop}} = x_{\text{obs}}$.

Why?

If one wants to generate a draw from the posterior:

1. Draw θ_{prop} from the prior $\pi(\theta)$.
2. Draw x_{prop} from the density $f(x \mid \theta_{\text{prop}})$.
3. Accept θ_{prop} if $x_{\text{prop}} = x_{\text{obs}}$.

Why? Let θ_{acc} denote an accepted θ_{prop}. Then, for any θ,

$$\begin{aligned}
P(\theta_{\text{acc}} = \theta) &= P(\theta_{\text{prop}} = \theta \mid x_{\text{prop}} = x_{\text{obs}}) \\
&= P(x_{\text{prop}} = x_{\text{obs}} \mid \theta_{\text{prop}} = \theta) P(\theta_{\text{prop}} = \theta) \Big/ P(x_{\text{prop}} = x_{\text{obs}}) \\
&= f(x_{\text{obs}} \mid \theta)\pi(\theta) \Big/ f(x_{\text{obs}}) = \pi(\theta \mid x_{\text{obs}})
\end{aligned}$$

in the case where **θ is discrete.**

Illustration from Chad Schafer (CMU)

If one wants to generate a draw from the posterior:

1. Draw θ_{prop} from the prior $\pi(\theta)$.
2. Draw x_{prop} from the density $f(x \mid \theta_{\text{prop}})$.
3. Accept θ_{prop} if $x_{\text{prop}} = x_{\text{obs}}$.

Why? Let θ_{acc} denote an accepted θ_{prop}. Then, for any $T \subseteq \mathbb{R}$,

$$\begin{aligned}
P(\theta_{\text{acc}} \in T) &= P(\theta_{\text{prop}} \in T \mid x_{\text{prop}} = x_{\text{obs}}) \\
&= \int_T P(x_{\text{prop}} = x_{\text{obs}} \mid \theta_{\text{prop}} = \theta)\pi(\theta)\, d\theta \Big/ P(x_{\text{prop}} = x_{\text{obs}}) \\
&= \int_T f(x_{\text{obs}} \mid \theta)\pi(\theta)\, d\theta \Big/ f(x_{\text{obs}}) = \int_T \pi(\theta \mid x_{\text{obs}})\, d\theta
\end{aligned}$$

in the case where **θ is continuous.**

If one wants to generate a draw from the posterior:

1. Draw θ_{prop} from the prior $\pi(\theta)$.
2. Draw x_{prop} from the density $f(x \mid \theta_{\text{prop}})$.
3. Accept θ_{prop} if $x_{\text{prop}} = x_{\text{obs}}$.

Why? Let θ_{acc} denote an accepted θ_{prop}. Then, for any $T \subseteq \mathbb{R}$,

$$\begin{aligned} P(\theta_{\text{acc}} \in T) &= P(\theta_{\text{prop}} \in T \mid x_{\text{prop}} = x_{\text{obs}}) \\ &= K \int_T P(x_{\text{prop}} = x_{\text{obs}} \mid \theta_{\text{prop}} = \theta)\pi(\theta)\, d\theta \\ &= K \int_T f(x_{\text{obs}} \mid \theta)\pi(\theta)\, d\theta = \int_T \pi(\theta \mid x_{\text{obs}})\, d\theta \end{aligned}$$

in the case where **θ is continuous.**

If one wants to generate a draw from the posterior:

1. Draw θ_{prop} from the prior $\pi(\theta)$.
2. Draw x_{prop} from the density $f(x \mid \theta_{\text{prop}})$.
3. Accept θ_{prop} if **???**

Why? Let θ_{acc} denote an accepted θ_{prop}. Then, for any $T \subseteq \mathbb{R}$,

$$\begin{aligned} P(\theta_{\text{acc}} \in T) &= P(\theta_{\text{prop}} \in T \mid \text{Accept } \theta_{\text{prop}}) \\ &= K \int_T P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta)\pi(\theta)\, d\theta \\ &\stackrel{?}{=} K' \int_T f(x_{\text{obs}} \mid \theta)\pi(\theta)\, d\theta = \int_T \pi(\theta \mid x_{\text{obs}})\, d\theta \end{aligned}$$

in the case where θ **is continuous.**

The Point:

θ_{acc} is a draw from the posterior if

$$P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta) \propto f(x_{\text{obs}}|\theta) \ \ \text{(the likelihood)}$$

This creates a basis for assessing the quality of the approximation, irrespective of the prior.

To achieve this, we could accept θ_{prop} if $x_{\text{prop}} = x_{\text{obs}}$.
Of course, this is not practical.

Clear alternative is to accept θ_{prop} if x_{prop} is "close to" x_{obs} using some chosen distance metric Δ.

Clear alternative is to accept θ_{prop} if x_{prop} is “close to” x_{obs} using some chosen distance metric Δ.

What is the price of this approximation?

Clear alternative is to accept θ_{prop} if x_{prop} is "close to" x_{obs} using some chosen distance metric Δ.

What is the price of this approximation?

Define:

$$\phi_\epsilon(x_{\text{prop}}, x_{\text{obs}}) = \begin{cases} 1, & \text{if } \Delta(x_{\text{prop}}, x_{\text{obs}}) < \epsilon \\ 0, & \text{if } \Delta(x_{\text{prop}}, x_{\text{obs}}) \geq \epsilon \end{cases}$$

In other words, $\phi_\epsilon(x_{\text{prop}}, x_{\text{obs}})$ is an indicator as to whether or not x_{prop} is close to x_{obs}.

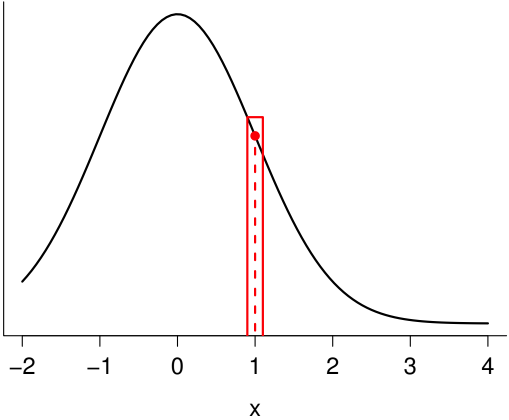
Hence,

$$
\begin{aligned}
P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta) &= P(\Delta(x_{\text{prop}}, x_{\text{obs}}) < \epsilon \mid \theta_{\text{prop}} = \theta) \\
&= \int \phi_\epsilon(x, x_{\text{obs}}) f(x \mid \theta)\, dx \\
&\longrightarrow Kf(x_{\text{obs}} \mid \theta) \ \text{ as } \epsilon \to 0
\end{aligned}
$$

Hence, for ϵ small,

$$
P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta) \approx Kf(x_{\text{obs}} \mid \theta)
$$

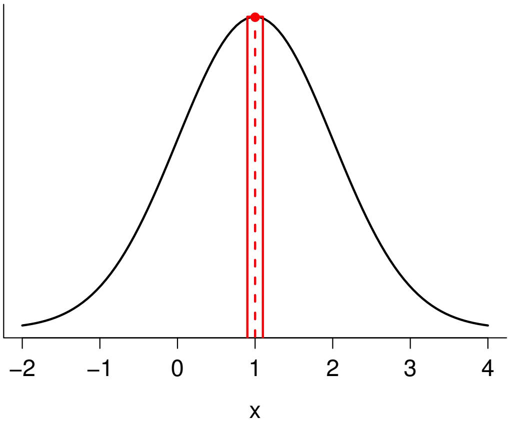
Toy Example: Assume that x_{obs} is Gaussian with mean θ and variance one.



Depicts the convolution

$$\int \phi_\epsilon(x, x_{\text{obs}}) f(x \mid \theta)\, dx = P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta)$$

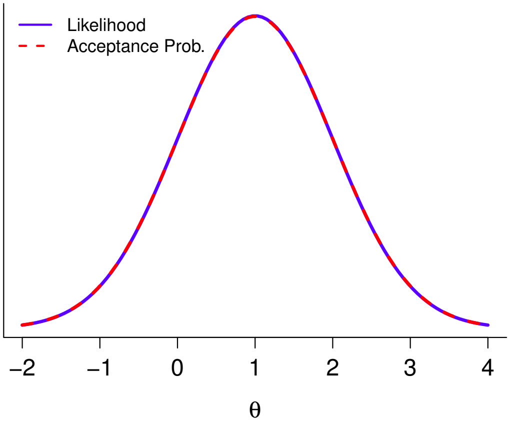
for case where $x_{\text{obs}} = 1$, $\theta = 0$, $\epsilon = 0.1$.



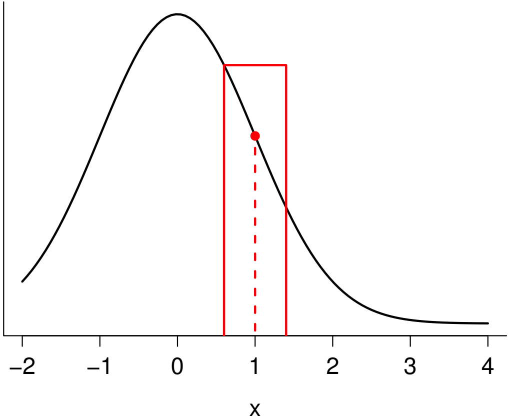
Depicts the convolution

$$\int \phi_\epsilon(x, x_{\text{obs}}) f(x \mid \theta)\, dx = P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta)$$

for case where $x_{\text{obs}} = 1$, $\theta = 1$, $\epsilon = 0.1$.



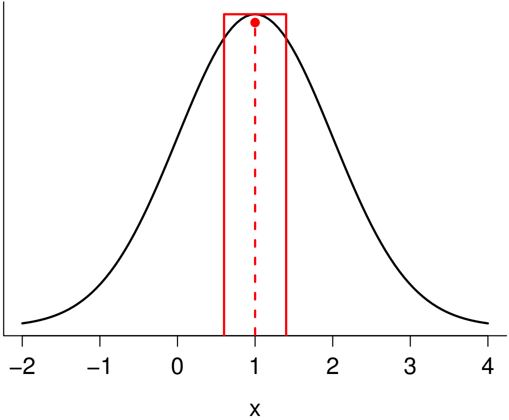
Compare these quantities for all θ. Case where $x_{\text{obs}} = 1$, $\epsilon = 0.1$.



Depicts the convolution

$$\int \phi_\epsilon(x, x_{\text{obs}}) f(x \mid \theta)\, dx = P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta)$$

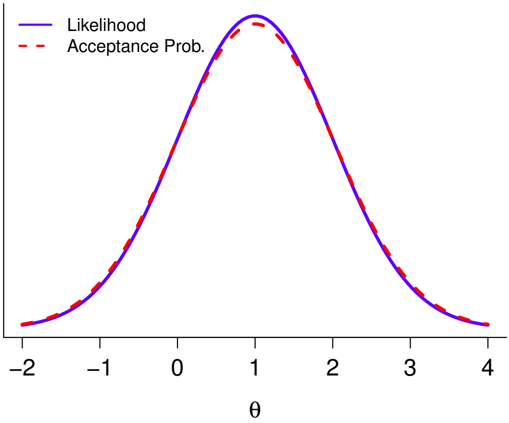
for case where $x_{\text{obs}} = 1$, $\theta = 0$, $\epsilon = 0.4$.



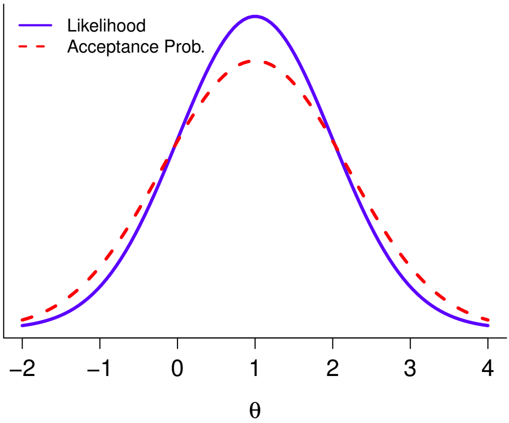
Depicts the convolution

$$\int \phi_\epsilon(x, x_{\text{obs}}) f(x \mid \theta)\, dx = P(\text{Accept } \theta_{\text{prop}} \mid \theta_{\text{prop}} = \theta)$$

for case where $x_{\text{obs}} = 1$, $\theta = 1$, $\epsilon = 0.4$.



Compare these quantities for all θ. Case where $x_{\text{obs}} = 1$, $\epsilon = 0.4$.



Case where $x_{\text{obs}} = 1$, $\epsilon = 1$.

Data summaries

Ideally, S_x is sufficient, i.e.

$$f(x \mid S_x, \theta) = f(x \mid S_x)$$

as this implies that

$$f(x \mid \theta) = f(x \mid \theta, S_x) f(S_x \mid \theta) \propto f(S_x \mid \theta)$$

Examples:

1. $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ is a sufficient statistic for μ 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 θ where Y_i are iid $U(0, \theta)$, $i = 1, \ldots, n$

In a nutshell

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

Marin et al. (2012)

In a nutshell

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

Marin et al. (2012)

How to pick a tolerance, ϵ?

Sequential ABC

Sequential ABC

Main idea

Instead of starting the ABC algorithm over with a smaller tolerance (ϵ), use the already sampled particle system as a proposal distribution *rather* than drawing from the prior distribution.

Particle system:

(1) retained sampled values, (2) importance weights

Some references:

Beaumont et al. (2009); Moral et al. (2011); Bonassi and West (2004)

Monte Carlo integration $\longrightarrow$ Importance sampling

MC Integration

General idea

Monte Carlo methods are a form of stochastic integration used to approximate expectations by invoking the law of large numbers.

$$I = \int_a^b h(y)dy = \int_a^b w(y)f(y)dy = E_f(w(Y))$$

where $f(y) = \frac{1}{b-a}$ and $w(y) = h(y) \cdot (b-a)$

- $f(y) = \frac{1}{b-a}$ is the pdf of a U(a,b) random variable
- By the LLN, if we take an iid sample of size N from $U(a,b)$, we can estimate I as

$$\hat{I} = N^{-1} \sum_{i=1}^{N} w(Y_i) \longrightarrow E(w(Y)) = I$$

MC Integration: Gaussian CDF example*

- Goal: estimate $F_Y(y) = P(Y \leq y) = E\left[I_{(-\infty,y)}(Y)\right]$ where $Y \sim N(0,1)$:

$$F(Y \leq y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \int_{-\infty}^{\infty} h(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

where $h(t) = 1$ if $t < y$ and $h(t) = 0$ if $t \geq y$

MC Integration: Gaussian CDF example*

- Goal: estimate $F_Y(y) = P(Y \leq y) = E\left[I_{(-\infty,y)}(Y)\right]$ where $Y \sim N(0,1)$:

$$F(Y \leq y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \int_{-\infty}^{\infty} h(t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

where $h(t) = 1$ if $t < y$ and $h(t) = 0$ if $t \geq y$

- Draw an iid sample $Y_1, \ldots, Y_N$ from a $N(0,1)$, then the estimator is

$$\hat{I} = N^{-1} \sum_{i=1}^{N} h(Y_i) = \frac{\#\text{ draws } < x}{N}$$

★ Example 24.2 of Wasserman (2004)

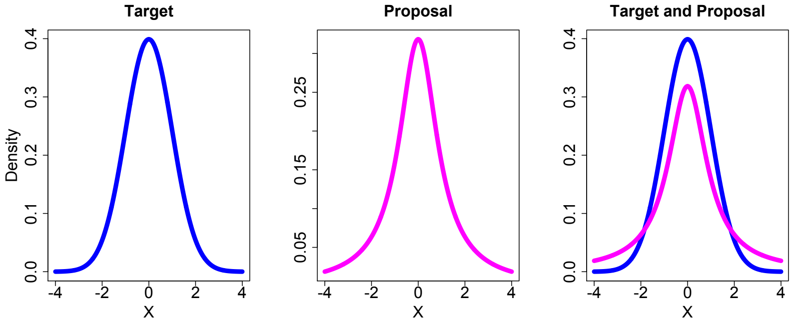
Importance Sampling: motivation

- Standard Monte Carlo integration is great if you can sample from the *target* distribution (i.e. the desired distribution)
  $\longrightarrow$ But what if you can't sample from the target?

Importance Sampling: motivation

- Standard Monte Carlo integration is great if you can sample from the *target* distribution (i.e. the desired distribution)
  $\longrightarrow$ But what if you can't sample from the target?

- Idea of importance sampling: draw the sample from a *proposal* distribution and re-weight the integral using *importance weights* so that the correct distribution is targeted



MC Integration $\longrightarrow$ Importance Sampling

$$I = \int h(y)f(y)dy$$

- h is some function and f is the probability density function of Y
- When the density f is difficult to sample from, importance sampling can be used

MC Integration $\longrightarrow$ Importance Sampling

$$I = \int h(y)f(y)dy$$

- h is some function and f is the probability density function of Y
- When the density f is difficult to sample from, importance sampling can be used
- Rather than sampling from f, you specify a different probability density function, g, as the proposal distribution.

$$I = \int h(y)f(y)dy = \int h(y)\frac{f(y)}{g(y)}g(y)dy = \int \frac{h(y)f(y)}{g(y)}g(y)dy$$

Importance Sampling

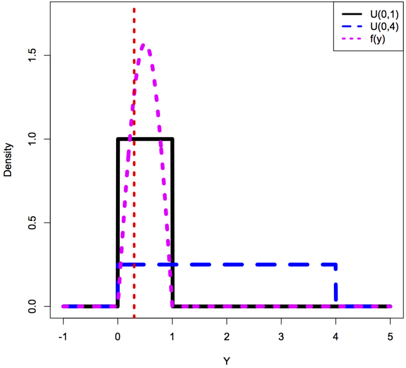
$$I = E_f\left[h(Y)\right] = \int \frac{h(y)f(y)}{g(y)} g(y)dy = E_g\left[\frac{h(Y)f(Y)}{g(Y)}\right]$$

Hence, given an iid sample $Y_1, \ldots, Y_N$ from g, our estimator of I becomes

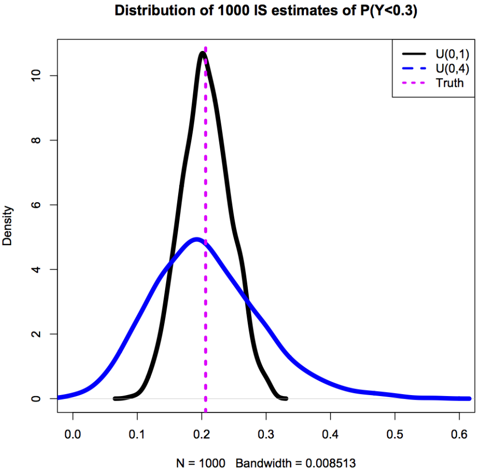
$$\hat{I} = N^{-1}\sum_{i=1}^{N} \frac{h(Y_i)f(Y_i)}{g(Y_i)} \longrightarrow E_g\left[\frac{h(Y)f(Y)}{g(Y)}\right] = I$$

Importance sampling: Illustration

- Goal: estimate $P(Y < 0.3)$ where $Y \sim f$
- Try two proposal distributions: U(0,1) and U(0,4)



If we take 1000 samples of size 100, and find the IS estimates, we get the following *estimated* expected values and variances.



	Expected Value	Variance
Truth	0.206	0
g_1: U(0,1)	0.206	0.0014
g_2: U(0,4)	0.211	0.0075

Extensions of Importance Sampling

- Sequential Importance Sampling
- Sequential Monte Carlo (Particle Filtering)
  $\longrightarrow$ See Doucet et al. (2001)
- Approximate Bayesian Computation $\longrightarrow$ See Turner and Zandt (2012) for a tutorial, and Cameron and Pettitt (2012); Weyant et al. (2013) for applications to astronomy

Importance sampling can be used in ABC to improve the computational efficiency.

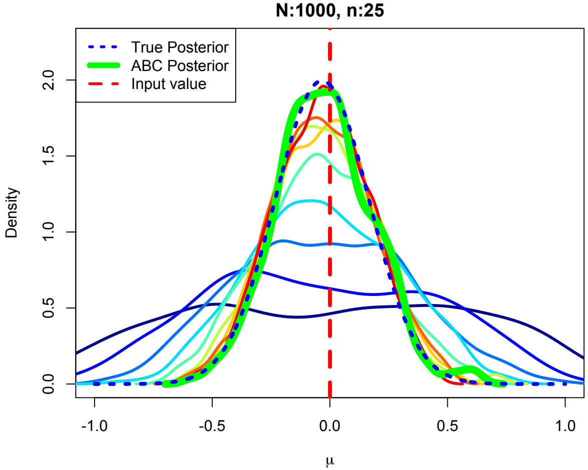
Algorithm 1 ABC - Population Monte Carlo algorithm*

At iteration $t = 1$
Algorithm 1: Basic ABC sampler to obtain $\{\theta_1^{(i)}\}_{i=1}^N$
Set importance weights $W_1^{(i)} = 1/N$ for $i = 1, \ldots, N$
for $t = 2$ to T **do**
  Set $\tau_t^2 = 2 \cdot \mathrm{var}\left(\{\theta_{t-1}^{(i)}, W_{t-1}^{(i)}\}_{i=1}^N\right)$
  for $i = 1$ to N **do**
    while $\rho(S(y_{1:n}), S(x_{1:n})) > \epsilon_t$ **do**
      Draw θ_0 from $\{\theta_{t-1}^{(i)}\}_{i=1}^N$ with probabilities $\{W_{t-1}^{(i)}\}_{i=1}^N$
      Propose $\theta^* \sim N(\theta_0, \tau_t^2)$
      Generate $x_{1:n}$ from $F(x \mid \theta^*)$
      Calculate summary statistics $\{S_y, S_x\}$
    end while
    $\theta_t^{(i)} \leftarrow \theta^*$
    $\widetilde{W}_t^{(i)} \leftarrow \dfrac{\pi\left(\theta_t^{(i)}\right)}{\sum_{j=1}^N W_{t-1}^{(j)} \phi\left[\tau_t^{-1}(\theta_t^{(i)} - \theta_{t-1}^{(j)})\right]}$
  end for
  $\{W_t^{(i)}\}_{i=1}^N \leftarrow \{\widetilde{W}_t^{(i)}\}_{i=1}^N / \sum_{i=1}^N \widetilde{W}_t^{(i)}$
end for

Decreasing tolerances $\epsilon_1 \geq \cdots \geq \epsilon_T$, $\phi(\cdot)$ is the density function of a $N(0, 1)$

*From Beaumont et al. (2009)

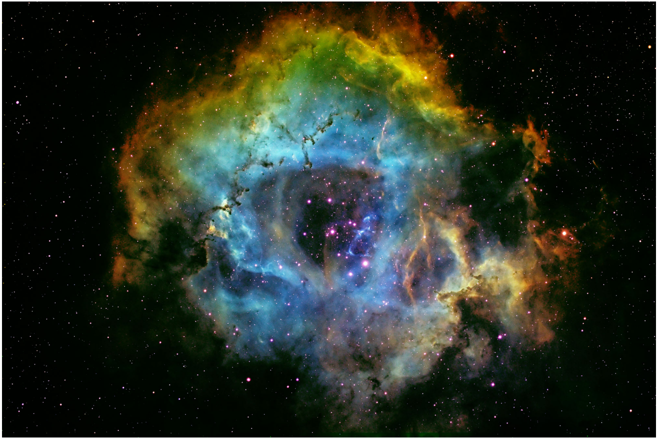
Gaussian illustration: sequential posteriors



Tolerance sequence, $\epsilon_{1:10}$:

1.00 0.75 0.53 0.38 0.27 0.19 0.15 0.11 0.08 0.06

Stellar Initial Mass Function



Stellar Initial Mass Function: the distribution of star masses after a star formation event within a specified volume of space

Molecular cloud → Protostars → Stars

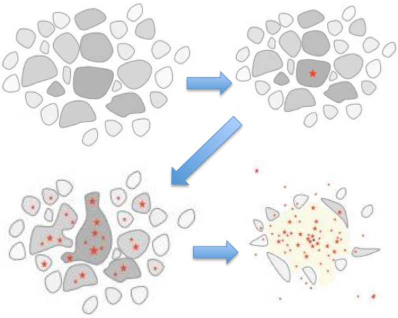
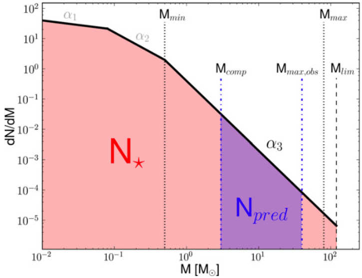


Image: adapted from `http://www.astro.ljmu.ac.uk`

Broken power-law

(Kroupa, 2001)

$$\Phi(M) \propto M^{-\alpha_i},$$
$$M_{1i} \leq M \leq M_{2i}$$

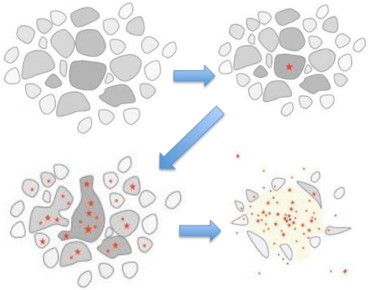


$\alpha_1 = 0.3$ for $0.01 \leq M/M^*_{Sun} \leq 0.08$ [Sub-stellar]
$\alpha_2 = 1.3$ for $0.08 \leq M/M_{Sun} \leq 0.50$
$\alpha_3 = 2.3$ for $0.50 \leq M/M_{Sun} \leq M_{\mathrm{max}}$

Many other models, e.g. Salpeter (1955); Chabrier (2003)

*1 M_{Sun} = 1 Solar Mass (the mass of our Sun)

ABC for the Stellar Initial Mass Function



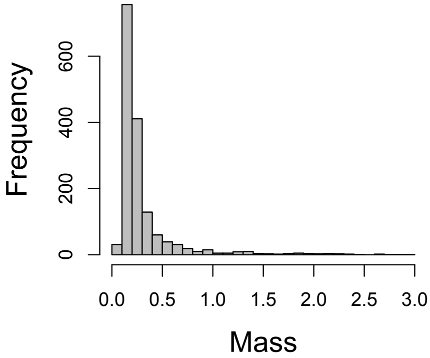


Image (left): Adapted from `http://www.astro.ljmu.ac.uk`

IMF Likelihood

- **Start with a power-law distribution**: each star's mass is independently drawn from a power law distribution with density

$$f(m) = \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}}\right) m^{-\alpha}, \quad m \in (M_{\min}, M_{\max})$$

- Then the likelihood is

$$\boxed{L(\alpha \mid m_{1:n_{tot}}) = \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}}\right)^{n_{tot}} \times \prod_{i=1}^{n_{tot}} m_i^{-\alpha}}$$

n_{tot} = total number of stars in cluster

Observational limitations: aging

- Lifecycle of star depends on mass $\rightarrow$ more massive stars die faster
- Cluster age of τ Myr $\rightarrow$ only observe stars with masses $< T_{age} \approx \tau^{-2/5} \times 10^{8/5}$

If age $= 30$ Myr so the aging cutoff is $T_{age} \approx 10\ M_{Sun}$

Then the likelihood is

$$\boxed{L(\alpha \mid m_{1:n_{obs}}, n_{tot}) = \left(\frac{1-\alpha}{T_{age}^{1-\alpha} - M_{\min}^{1-\alpha}}\right)^{n_{obs}} \left(\prod_{i=1}^{n_{obs}} m_i^{-\alpha}\right) \times P(M > T_{age})^{n_{tot}-n_{obs}}}$$

$n_{tot} = \#$ of stars in cluster

$n_{obs} = \#$ stars observed in cluster

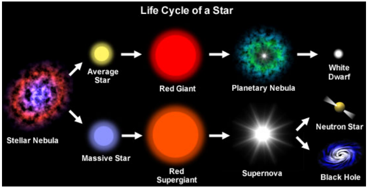


Image: `http://scioly.org`

Observational limitations: completeness

- Completeness function:

$$P(\text{observing star} \mid m) = \begin{cases} 0, & m < C_{\text{min}} \\ \frac{m - C_{\text{min}}}{C_{\text{max}} - C_{\text{min}}}, & m \in [C_{\text{min}}, C_{\text{max}}] \\ 1, & m > C_{\text{max}} \end{cases}$$

- Probability of observing a particular star given its mass
- Depends on the flux limit, stellar crowding, etc.

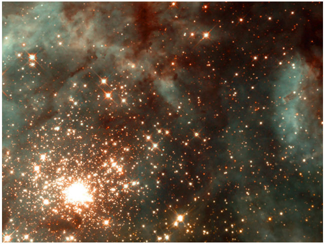


Image: NASA, J. Trauger (JPL), J. Westphal (Caltech)

Observational limitations: measurement error

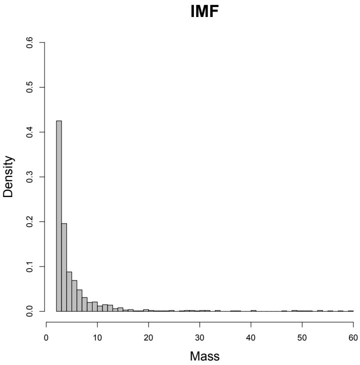
Incorporating log-normal measurement error gives our final likelihood:

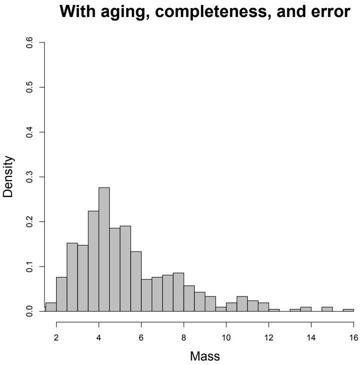
$$L(\alpha \mid m_{1:n_{obs}}, n_{tot}) =$$

$$\left(P(M > T_{age}) + \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}} \right) \int_{C_{\min}}^{C_{\max}} M^{-\alpha} \times \left(1 - \frac{M - C_{\min}}{C_{\max} - C_{\min}} \right) dM \right)^{n_{tot} - n_{obs}}$$

$$\times \prod_{i=1}^{n_{obs}} \left\{ \int_{2}^{T_{age}} (2\pi\sigma^2)^{-\frac{1}{2}} m_i^{-1} e^{-\frac{1}{2\sigma^2}(\log(m_i) - \log(M))^2} \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}} \right) M^{-\alpha} \right.$$

$$\left. \times \left(I\{M > C_{\max}\} + \left(\frac{M - C_{\min}}{C_{\max} - C_{\min}} \right) I\{C_{\min} \le M \le C_{\max}\} \right) dM \right\}$$





Sample size = 1000 stars, $[C_{\mathrm{min}}, C_{\mathrm{max}}] = [2, 4]$, $\sigma = 0.25$

Simulation Study: forward model

- Draw from

$$f(m) = \left(\frac{1-\alpha}{60^{1-\alpha} - 2^{1-\alpha}}\right) m^{-\alpha}, \quad m \in (2, 60)$$

- Aged 30 Myrs
- Observational completeness:

$$P(obs \mid m) = \begin{cases} 0, & m < 4 \\ \frac{m-2}{2}, & m \in [2, 4] \\ 1, & m > 4. \end{cases}$$

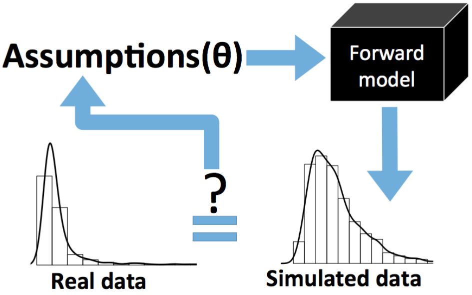
- Uncertainty: $\log M = \log m + 0.25\eta$ (with $\eta \sim N(0, 1)$)
- Prior: $\alpha \sim U[0, 6]^*$

Simulation Study: summary statistics

We want to account for the following with our summary statistics and distance functions:

1. Shape of the observed Mass Function

$$\rho_1(m_{sim}, m_{obs}) = \left[\int \left\{ \hat{f}_{\log m_{sim}}(x) - \hat{f}_{\log m_{obs}}(x) \right\}^2 dx \right]^{1/2}$$



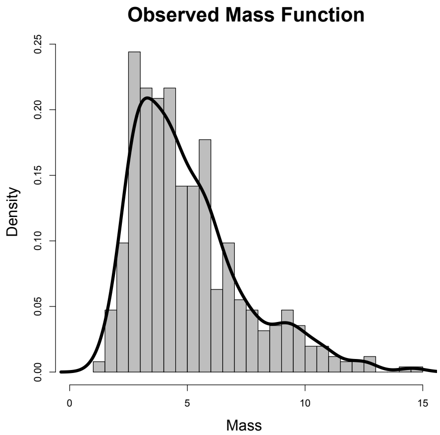
2. Number of stars observed

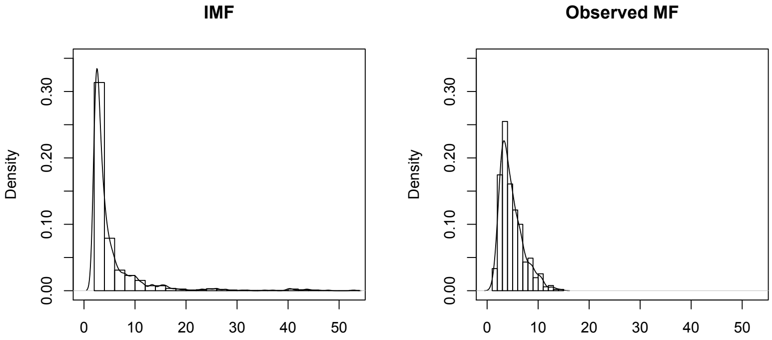
$$\rho_2(m_{sim}, m_{obs}) = |1 - n_{sim}/n_{obs}|$$

m_{sim} = masses of the stars simulated from the forward model
m_{obs} = masses of observed stars
n_{sim} = number of stars simulated from the forward model
n_{obs} = number of observed stars

Simulation Study

1. Draw $n = 10^3$ stars
2. IMF slope $\alpha = 2.35$ with $M_{min} = 2$ and $M_{max} = 60$
3. $N = 10^3$ particles
4. $T = 30$ sequential time steps

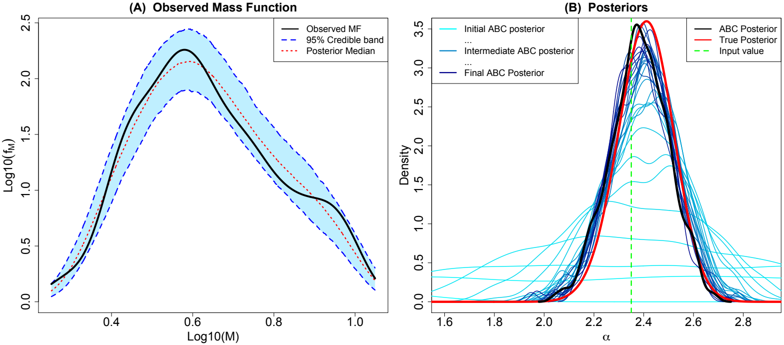




Sample size $= 1000$ stars, $[C_{\text{min}}, C_{\text{max}}] = [2, 4]$, $\sigma = 0.25$

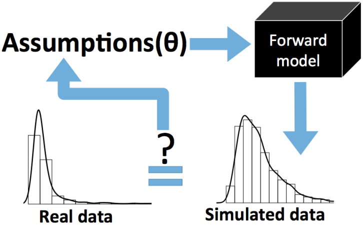
Simulation Study results

1. Draw $n = 10^3$ stars
2. IMF slope $\alpha = 2.35$ with $M_{min} = 2$ and $M_{max} = 60$
3. $N = 10^3$ particles
4. $T = 30$ sequential time steps



Summary

- ABC can be a useful tool when data are too complex to define a reasonable likelihood
- Selection of good summary statistics is crucial for ABC posterior to be meaningful



THANK YOU!!!

Bibliography I

Akeret, J., Refregier, A., Amara, A., Seehars, S., and Hasner, C. (2015), "Approximate Bayesian computation for forward modeling in cosmology," *Journal of Cosmology and Astroparticle Physics*, 2015, 043.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009), "Adaptive approximate Bayesian computation," *Biometrika*, 96, 983 – 990.

Bonassi, F. V. and West, M. (2004), "Sequential Monte Carlo with Adaptive Weights for Approximate Bayesian Computation," *Bayesian Analysis*, 1, 1–19.

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.

Chabrier, G. (2003), "The galactic disk mass function: Reconciliation of the hubble space telescope and nearby determinations," *The Astrophysical Journal Letters*, 586, L133.

Doucet, A., De Freitas, N., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, New York: Springer-Verlag.

Ishida, E. E. O., Vitenti, S. D. P., Penna-Lima, M., Cisewski, J., de Souza, R. S., Trindade, A. M. M., Cameron, E., and Busti, V. C. (2015), "cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation," *Astronomy and Computing*, 13, 1 – 11.

Kroupa, P. (2001), "On the variation of the initial mass function," *Monthly Notices of the Royal Astronomical Society*, 322, 231–246.

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

Moral, P. D., Doucet, A., and Jasra, A. (2011), "An adaptive sequential Monte Carlo method for approximate Bayesian computation," *Statistics and Computing*, 22, 1009–1020.

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.

Salpeter, E. E. (1955), "The luminosity function and stellar evolution." *The Astrophysical Journal*, 121, 161.

Bibliography II

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.

Wasserman, L. (2004), *All of statistics: a concise course in statistical inference*, Springer.

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.