Approximate Bayesian Computation

Jessi Cisewski
Carnegie Mellon University

International Space Science Institute

May 12, 2015

Outline

1. **Approximate Bayesian Computation (ABC) Overview**
   $\longrightarrow$ Framework for inference without specifying a likelihood

2. **Simple example**: Stellar Initial Mass Function (IMF) using usual assumptions

Bayesian methods

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

Posterior distribution

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

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: $\pi(\theta)$

$\longrightarrow$ 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 ρ is a distance function.

$\pi(\theta \mid \rho(y_{1:n}, x_{1:n}) < \epsilon) \longrightarrow$ (assuming ρ preserves sufficiency)

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

Approximate Bayesian Computation

- "Likelihood-free" approach (likelihood is not explicitly specified)
- Works by simulating from the forward process

Issues with writing down a likelihood

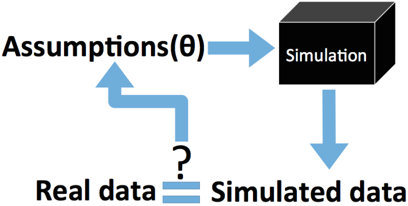
1. Physical model too complex or unknown
2. Theory is not fully understood
3. Strong dependency in data
4. Observational limitations

Approximate Bayesian Computation

- "Likelihood-free" approach (likelihood is not explicitly specified)
- Works by simulating from the forward process

Issues with writing down a likelihood

1. Physical model too complex or unknown
2. Theory is not fully understood
3. Strong dependency in data
4. Observational limitations



For observations $y_{1:n}$, distance function ρ, and (small) tolerance ϵ

Algorithm 1 Basic ABC Algorithm

for $i = 1$ to N **do**
  while $\rho(S_y, S_x) > \epsilon$ **do**
    Propose θ^* by drawing θ^* from prior $\pi(\theta)$
    Generate $x_{1:n}$ from forward process $F(x \mid \theta^*)$
    Calculate summary statistics $\{S_y, S_x\}$
  end while
  $\theta^{(i)} \leftarrow \theta^*$
end for

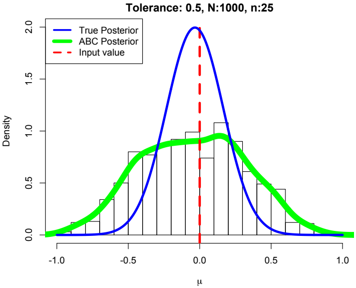
- ABC posterior based on $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}\} = \{\theta^{(i)}\}_{i=1}^{N}$
- $\{\theta^{(i)}\}_{i=1}^{N}$ are often referred to as *particles*

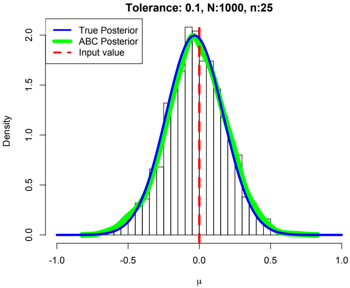
Introduced in Pritchard et al. (1999) (population genetics), Rubin (1984) (conceptually)

Gaussian illustration

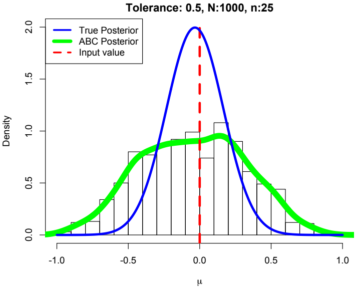
- Data $y_{1:n} \overset{iid}{\sim}$ Normal$(\mu, 1)$ where n = sample size, μ is unknown
- Forward process $F(x \mid \mu^*) \sim$ Normal$(\mu^*, 1)$
  (In this case, we use the likelihood)
- Summary statistics $\{S_y = \bar{y}, S_x = \bar{x}\}$
- Distance function $\rho(S_y, S_x) = |\bar{y} - \bar{x}|$
- Tolerance $\epsilon = 0.50$ and 0.10
- Prior $\pi(\mu) =$ Normal(0,10)

Gaussian illustration: posteriors for μ





Gaussian illustration: posteriors for μ



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)

Algorithm 2 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 $\tau_t^2 = 2 \cdot \text{var}\left(\{\theta_{t-1}^{(i)}, W_{t-1}^{(i)}\}_{i=1}^N\right)$

  for $i = 1$ to N **do**

    while $\rho\left(S(y_{1:n}), S(x_{1:n})\right) > \epsilon_t$ **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

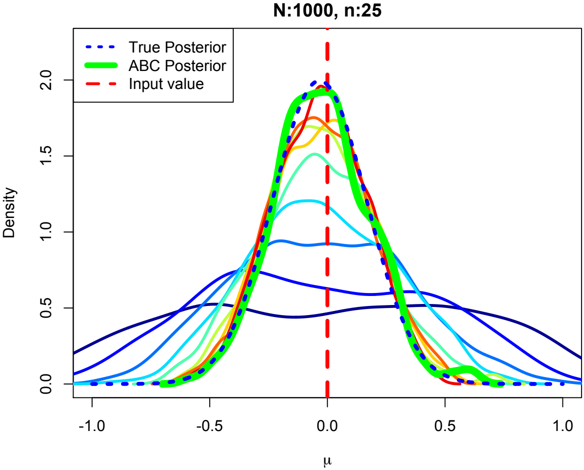
  $\{W_t^{(i)}\}_{i=1}^N \leftarrow \{\widetilde{W}_t^{(i)}\}_{i=1}^N / \sum_{i=1}^N \widetilde{W}_t^{(i)}$

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

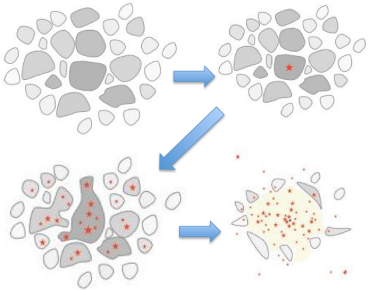


Tolerance sequence, $\epsilon_{1:10}$:

1.00 0.75 0.53 0.38 0.27 0.19 0.15 0.11 0.08 0.06

Star Cluster Formation

IMF: The distribution of star masses after a star formation event within a specified volume of space



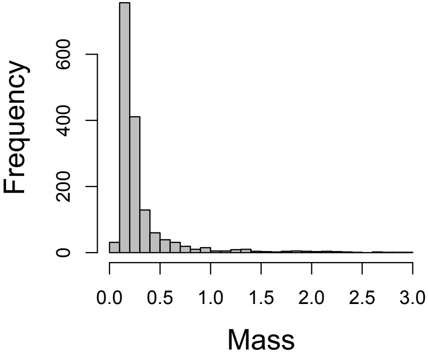


Image (left): Adapted from `http://www.astro.ljmu.ac.uk`

Examples of IMF models

- **Power-law: Salpeter (1955)**
  - Used a power law with $\alpha = 2.35$
- **Broken power-law: Kroupa (2001)**

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

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

- **Log-Normal model: Chabrier (2003a)**

$$\xi(\log m) = \frac{dn}{d \log m} = 0.158 \times \exp\left(-\frac{(\log m - \log 0.08)^2}{2(0.69)^2}\right)$$

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

ABC for the stellar initial mass function

1. We propose an ABC algorithm using the canonical IMF model as the forward process
   - Assume stellar masses are independent draws from the IMF
   - Useful for selecting appropriate summary statistics and distance functions
   - Can account for various observational limitations and uncertainties

2. We propose a new data-generating model
   - Account for the dependency in the stellar masses due to the cluster formation process
   - New model can be used in other settings

Cisewski, Weller, Schafer, and Hogg (Submitted)

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 $< T_{age} \approx \tau^{-2/5} \times 10^{8/5}$

  Then the likelihood is

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

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

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

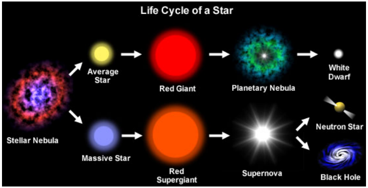


Image: `http://scioly.org`

Observational limitations: completeness

- Completeness function:

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

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

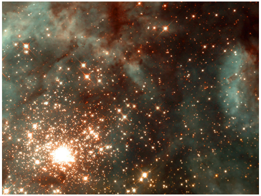
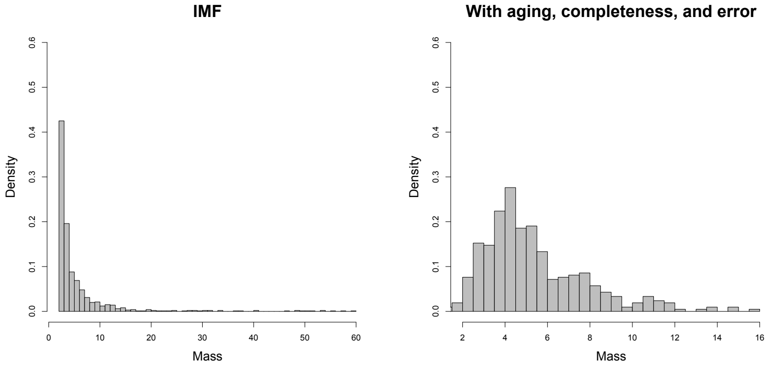


Image: NASA, J. Trauger (JPL), J. Westphal (Caltech)

Observational limitations: measurement error

Incorporating log-normal measurement error gives our final likelihood:

$$
\begin{aligned}
&L(\alpha \mid m_{1:n_{obs}}, n_{tot}) = \\
&\quad \left(P(M > T_{age}) + \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}} \right) \int_{C_{\min}}^{C_{\max}} M^{-\alpha} \times \left(1 - \frac{M - C_{\min}}{C_{\max} - C_{\min}} \right) dM \right)^{n_{tot} - n_{obs}} \\
&\quad \times \prod_{i=1}^{n_{obs}} \left\{ \int_2^{T_{age}} (2\pi\sigma^2)^{-\frac{1}{2}} m_i^{-1} e^{-\frac{1}{2\sigma^2}(\log(m_i) - \log(M))^2} \left(\frac{1-\alpha}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}} \right) M^{-\alpha} \right. \\
&\quad \left. \times \left(I\{M > C_{\max}\} + \left(\frac{M - C_{\min}}{C_{\max} - C_{\min}} \right) I\{C_{\min} \leq M \leq C_{\max}\} \right) dM \right\}
\end{aligned}
$$



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

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

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

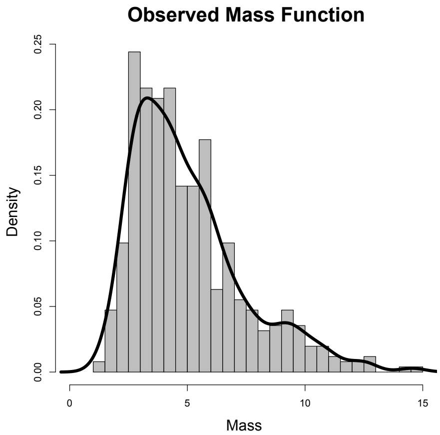
2. Number of stars observed

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

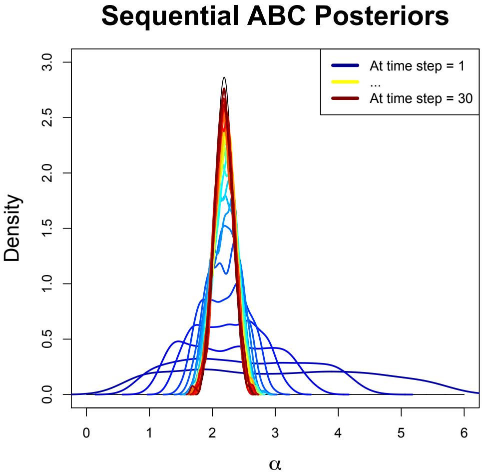
m_{sim} = masses of the stars simulated from the forward model
m_{obs} = masses of observed stars
n_{sim} = number of stars simulated from the forward model
n_{obs} = number of observed stars

Simulation Study

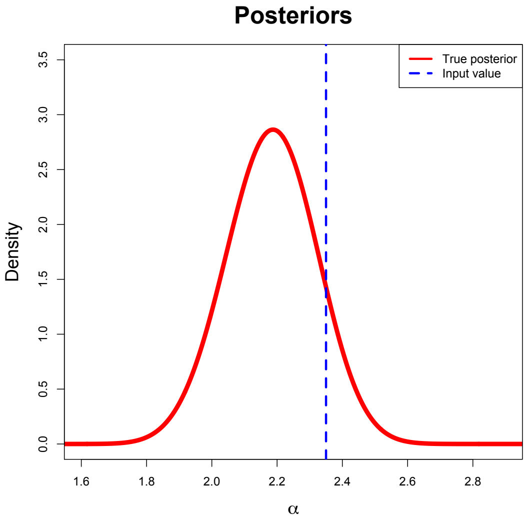
1. Draw $n = 10^3$ stars
2. IMF slope $\alpha = 2.35$ with $M_{min} = 2$ and $M_{max} = 60$
3. $N = 10^3$ particles
4. $T = 30$ sequential time steps



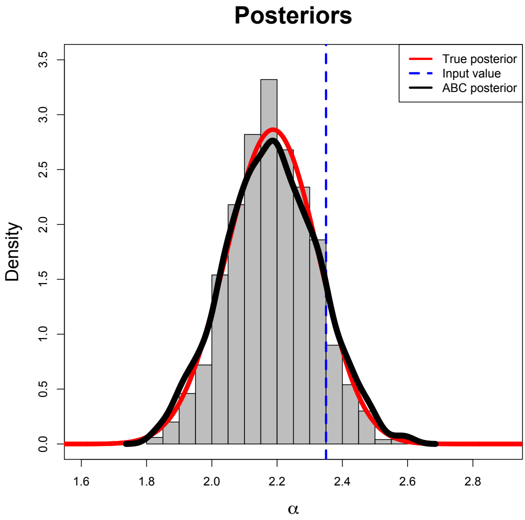
Results


Sequential ABC Posteriors
At time step = 1
...
At time step = 30
Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
α

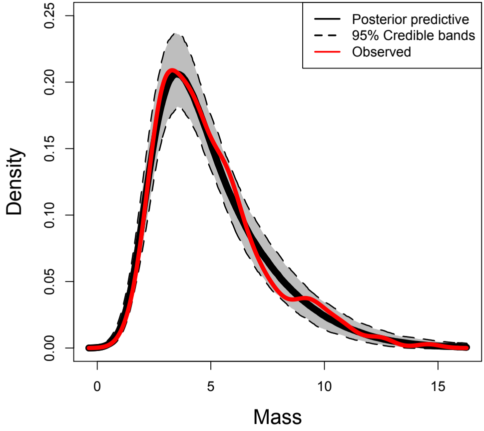
Results


Posteriors
True posterior
Input value
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1.6
1.8
2.0
2.2
2.4
2.6
2.8
α

Results


Posteriors
True posterior
Input value
ABC posterior
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1.6
1.8
2.0
2.2
2.4
2.6
2.8
α

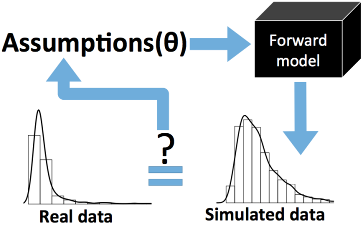
Results



Credible bands based on 1000 draws from the ABC posterior

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



Bibliography I

Bastian, N., Covey, K. R., and Meyer, M. R. (2010), "A Universal Stellar Initial Mass Function? A Critical Look at Variations," *Annual Review of Astronomy and Astrophysics*, 48, 339–89.

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

— (2003b), "Galactic Stellar and Substellar Initial Mass Function1," *Publications of the Astronomical Society of the Pacific*, 115, 763–795.

— (2005), "The initial mass function: from Salpeter 1955 to 2005," in *The Initial Mass Function 50 years later*, Springer, pp. 41–50.

Da Rio, N., Robberto, M., Hillenbrand, L. A., Henning, T., and Stassun, K. G. (2012), "The Initial Mass Function of the Orion Nebula Cluster across the H-burning Limit," *The Astrophysical Journal*, 748, 14.

D'Souza, R. M., Borgs, C., Chayes, J. T., Berger, N., and Kleinberg, R. D. (2007), "Emergence of tempered preferential attachment from optimization," *Proceedings of the National Academy of Sciences*, 104, 6112–6117.

Elmegreen, B. G. (2000), "Modeling a high-mass turn-down in the stellar initial mass function," *The Astrophysical Journal*, 539, 342.

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

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.

Bibliography II

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.

Rubin, D. B. (1984), "Bayesianly justifiable and relevant frequency calculations for the applied statistician," *The Annals of Statistics*, 12, 1151–1172.

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

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

Simon, H. A. (1955), "On a class of skew distribution functions," *Biometrika*, 425–440.

Weidner, C. and Kroupa, P. (2004), "Evidence for a fundamental stellar upper mass limit from clustered star formation," *Monthly Notices of the Royal Astronomical Society*, 348, 187–191.

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.

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.

Yule, G. U. (1925), "A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS," *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 21–87.