

# The Mass Distribution of Stellar-Mass Black Holes

Will M. Farr

and

Niharika Sravan

*Northwestern University Center for Interdisciplinary Exploration and Research in Astrophysics  
2145 Sheridan Rd., Evanston, IL 60208*

`w-farr@northwestern.edu, niharika.sravan@gmail.com`

Andrew Cantrell

and

Laura Kreidberg

and

Charles Bailyn

*Yale University Department of Astrophysics  
P.O. Box 208101, New Haven, CT 06520*

`andrew.cantrell@yale.edu, laura.kreidberg@yale.edu, charles.bailyn@yale.edu`

Vicky Kalogera

*Northwestern University Center for Interdisciplinary Exploration and Research in Astrophysics  
2145 Sheridan Rd., Evanston, IL 60208*

`vicky@northwestern.edu`

## ABSTRACT

We perform a Bayesian analysis of the mass distribution of stellar-mass black holes based on the observed masses of the compact object in 17 low mass X-ray binaries. Modeling the mass distribution both parametrically—as a power law, exponential, gaussian, or combination of two gaussians—and non-parametrically—as histograms with varying numbers of bins—we explore the distribution in the context of Markov Chain Monte Carlo calculations of model parameters. We give confidence bounds on the shape of the mass distribution in the context of each model and compare the models with each other by calculating their Bayesian evidence. We address the existence of a “gap” between the most massive neutron stars and least massive black holes, finding that the best model (the power law), the second-best model (a one-bin histogram, or flat distribution), and

the third-best model (a two-bin histogram) have minimum black hole masses above 5.00, 3.23, and 3.82 solar masses (95% confidence), respectively. We therefore conclude that our sample of black hole masses provides strong evidence of a gap between the maximum neutron star mass and minimum black hole mass.

## 1. Introduction

## 2. Systems

The 17 low mass X-ray binary systems on which this study is based are listed in Table 1. In each of these systems, spectroscopic measurements of the secondary star provide an orbital period for the system and a half-amplitude for the secondary’s velocity curve. These measurements can be combined into the mass function,

$$f(M) = \frac{PK^3}{2\pi G} = \frac{M \sin^3 i}{(1 + q)^2}, \quad (1)$$

where  $P$  is the orbital period,  $K$  is the secondary’s velocity semi-amplitude,  $M$  is the black hole mass,  $i$  is the inclination of the system, and  $q \equiv M_2/M$  is the mass ratio of the system.

The mass function defines a lower limit on the mass:  $f(M) < M$ . To accurately determine the mass of the black hole, the inclination  $i$  and mass ratio  $q$  must be measured. Ideally, this can be accomplished by fitting ellipsoidal light curves from these systems but even in the ideal case (see, e.g., Cantrell et al. (2010) on A0620) this procedure can be complicated by possible contributions from an accretion disk and hot spots in the disk. For some systems (e.g. GS 1354 (Casares et al. 2009)) strong variability prevents determination of the inclination from the lightcurve at all; in these cases an upper limit on the inclination often comes from the observed lack of eclipses in the lightcurve. In general, accurately determining  $q$  and  $i$  requires a careful system-by-system analysis.

For the purposes of this paper, we adopt the following simplified approach to the estimation of the black hole mass from the observed data. When an observable is well-constrained, we assume that the true value is normally distributed about the measured value with a standard deviation equal to the quoted observational error. This is the case for the mass function in all the systems we use, and for many systems’ mass ratios and inclinations. When a large range is quoted in the literature for an observable, we take the true value to be distributed uniformly (for the mass ratio) or isotropically (for the inclination) within the quoted range. Table 1 gives the assumed distribution for the observables in the 17 systems we use. We do not attempt to deal with the systematic biases in the observational determination of  $f$ ,  $q$ , and  $i$  in any realistic way, instead leaving this for future work.

From these assumptions, we can generate probability distributions for the true mass of the black hole given the observations and errors via the Monte Carlo method: drawing samples of  $f$ ,  $q$ , and  $i$  from the assumed distributions and computing the mass implied by Equation (1) gives samples of

Source	$f$ ( $M_{\odot}$ )	$q$	$i$ (degrees)	References
GRS 1915	$N(9.5, 3.0)$	$N(0.0857, 0.0284)$	$N(70, 2)$	Greiner et al. (2001)
XTE J1118	$N(6.44, 0.08)$	$N(0.0264, 0.004)$	$N(68, 2)$	Gelino et al. (2008)
				Harlaftis & Filippenko (2005)
Cyg X1	$N(0.251, 0.007)$	$N(2.778, 0.386)$	$I(23, 38)$	Gies et al. (2003)
XTE J1650	$N(2.73, 0.56)$	$U(0, 0.5)$	$I(50, 80)$	Orosz et al. (2004)
GRS 1009	$N(3.17, 0.12)$	$N(0.137, 0.015)$	$I(37, 80)$	Filippenko et al. (1999)
M33 X7	$N(0.46, 0.08)$	$N(4.47, 0.61)$	$N(74.6, 1)$	Orosz et al. (2007)
A0620	$N(2.76, 0.036)$	$N(0.06, 0.004)$	$N(50.98, 0.87)$	Cantrell et al. (2010)
				Neilsen et al. (2008)
GRO J0422	$N(1.13, 0.09)$	$U(0.076, 0.31)$	$N(45, 2)$	Gelino & Harrison (2003)
Nova Mus 1991	$N(3.01, 0.15)$	$N(0.128, 0.04)$	$N(54, 1.5)$	Gelino et al. (2001)
GRO J1655	$N(2.73, 0.09)$	$N(0.3663, 0.04025)$	$N(70.2, 1.9)$	Greene et al. (2001)
4U 1543	$N(0.25, 0.01)$	$U(0.25, 0.31)$	$N(20.7, 1.5)$	Orosz (2003)
XTE J1550	$N(7.73, 0.4)$	$U(0, 0.04)$	$N(74.7, 3.8)$	Orosz et al. (2010)
V4641 Sgr	$N(3.13, 0.13)$	$U(0.42, 0.45)$	$N(75, 2)$	Orosz (2003)
GS 2023	$N(6.08, 0.06)$	$U(0.056, 0.063)$	$N(55, 4)$	Charles & Coe (2006)
GS 1354	$N(5.73, 0.29)$	$N(0.12, 0.04)$	$I(50, 80)$	Casares et al. (2009)
Nova Oph 77	$N(4.86, 0.13)$	$U(0, 0.053)$	$I(60, 80)$	Charles & Coe (2006)
GS 2000	$N(5.01, 0.12)$	$U(0.035, 0.053)$	$I(43, 74)$	Charles & Coe (2006)

Table 1: The source parameters for the 17 low mass X-ray binaries used in this work.  $f$  is the mass function for the compact object,  $q$  is the mass ratio  $M_2/M$ , and  $i$  is the inclination of the system to the line of sight. We indicate the distribution used for the true parameters when computing the probability distributions for the masses of these systems:  $N(\mu, \sigma)$  implies a Gaussian with mean  $\mu$  and standard deviation  $\sigma$ ,  $U(a, b)$  is a uniform distribution between  $a$  and  $b$ , and  $I(\alpha, \beta)$  is an isotropic distribution between the angles  $\alpha$  and  $\beta$ . [FIXME: Orosz for XTE J1550 needs a real reference (Orosz et al. 2010).]

$M$  from the distribution induced by the relationship in Equation (1). Mass distributions generated in this way for the 17 systems used in this work are shown in Figure 1. These mass distributions constitute the “observational data” we will use in Section 3 to explore various underlying mass distributions.

In Figure 2 we show the combination of these individual mass distributions into one single distribution. (Note that this combination is not a good estimator of the true mass distribution of black hole masses—see, e.g., Mandel (2010). We include it here only to give a sense of the range of masses of black holes in our sample; we will come to estimators for the mass distribution.)

### 3. Models

In this section we apply a Bayesian analysis to various models for the underlying mass distribution from which the systems in Table 1 were drawn. The end result will be the probability distribution for the parameters of each model implied by the data from Section 2 in combination with our prior assumptions about the probability distribution for the parameters. Bayes’ rule relates these quantities. For a model with parameters  $\theta$  in the presence of data  $d$ , Bayes’ rule states

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}. \quad (2)$$

Here,  $p(\theta|d)$ , called the posterior probability distribution function, is the probability distribution for the parameters  $\theta$  implied by the data  $d$ ;  $p(d|\theta)$ , called the likelihood, is the probability of observing data  $d$  given that the model parameters are  $\theta$ ;  $p(\theta)$ , called the prior, reflects our estimate of the probability of the various model parameters in the absence of any data; and  $p(d)$ , called the evidence, is an overall normalizing constant ensuring that

$$\int d\theta p(\theta|d) = 1, \quad (3)$$

whence

$$p(d) = \int d\theta p(d|\theta)p(\theta). \quad (4)$$

In our context, the data are the mass distributions given in Section 2:  $d = \{p_i(m)|i = 1, 2, \dots, 17\}$ . We assume that the measurements in Section 2 are independent, so the complete likelihood is given by a product of the likelihoods for the individual measurements. For a model with parameters  $\theta$  that predicts a mass distribution  $p(m|\theta)$  for black holes, we have

$$p(d|\theta) = \prod_i \int dm p_i(m)p(m|\theta). \quad (5)$$

That is, the likelihood of an observation is the average over the individual mass distribution implied by the observation,  $p_i(m)$ , of the probability for a black hole of that mass to exist according to the

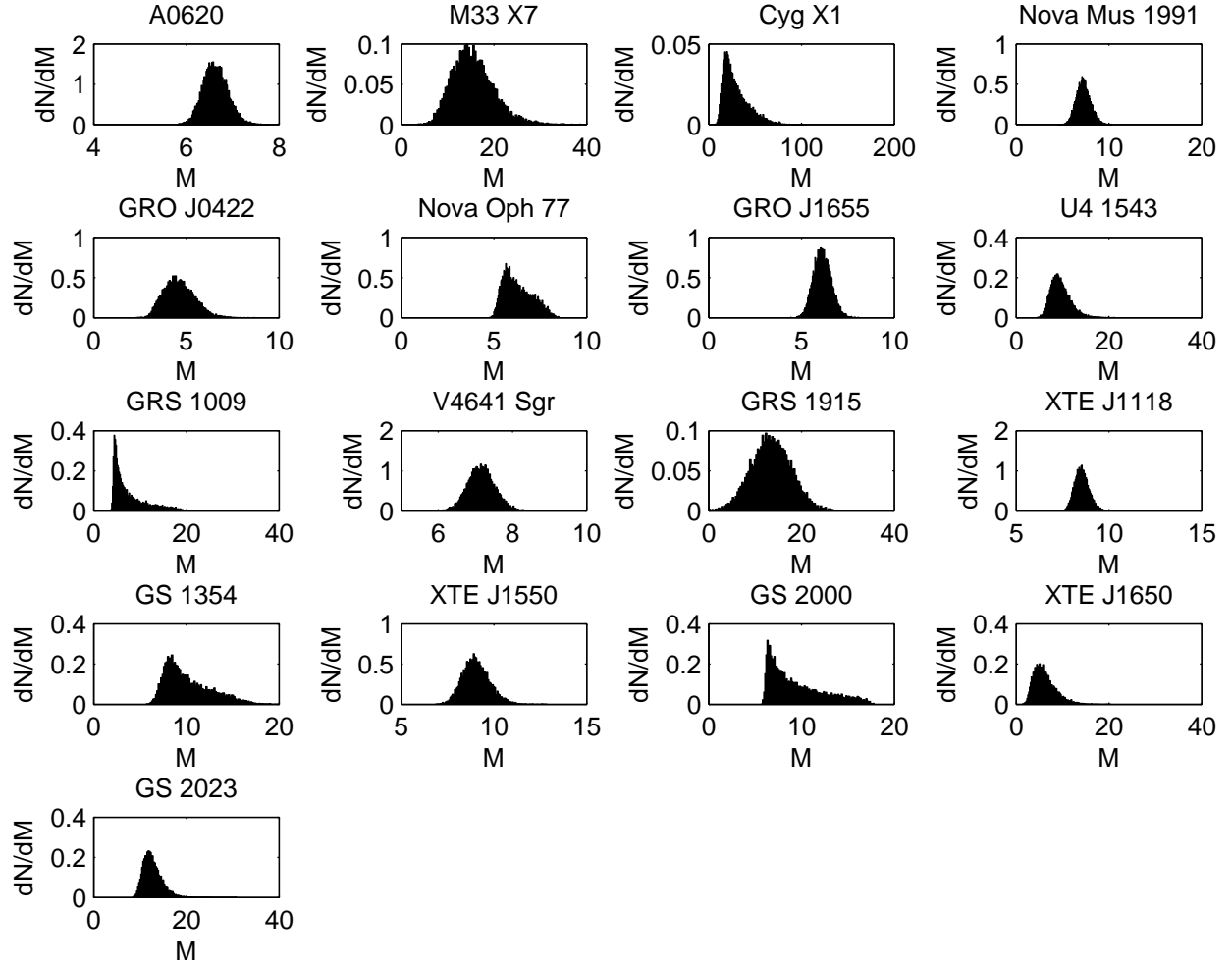


Fig. 1.— The mass distributions implied by Equation (1) and the assumed distributions on observational parameters given in Table 1. The significant asymmetry and long tails in many of these distributions are the result of the non-linear relationship (Equation (1)) between  $M$ ,  $f$ ,  $q$ , and  $i$ .

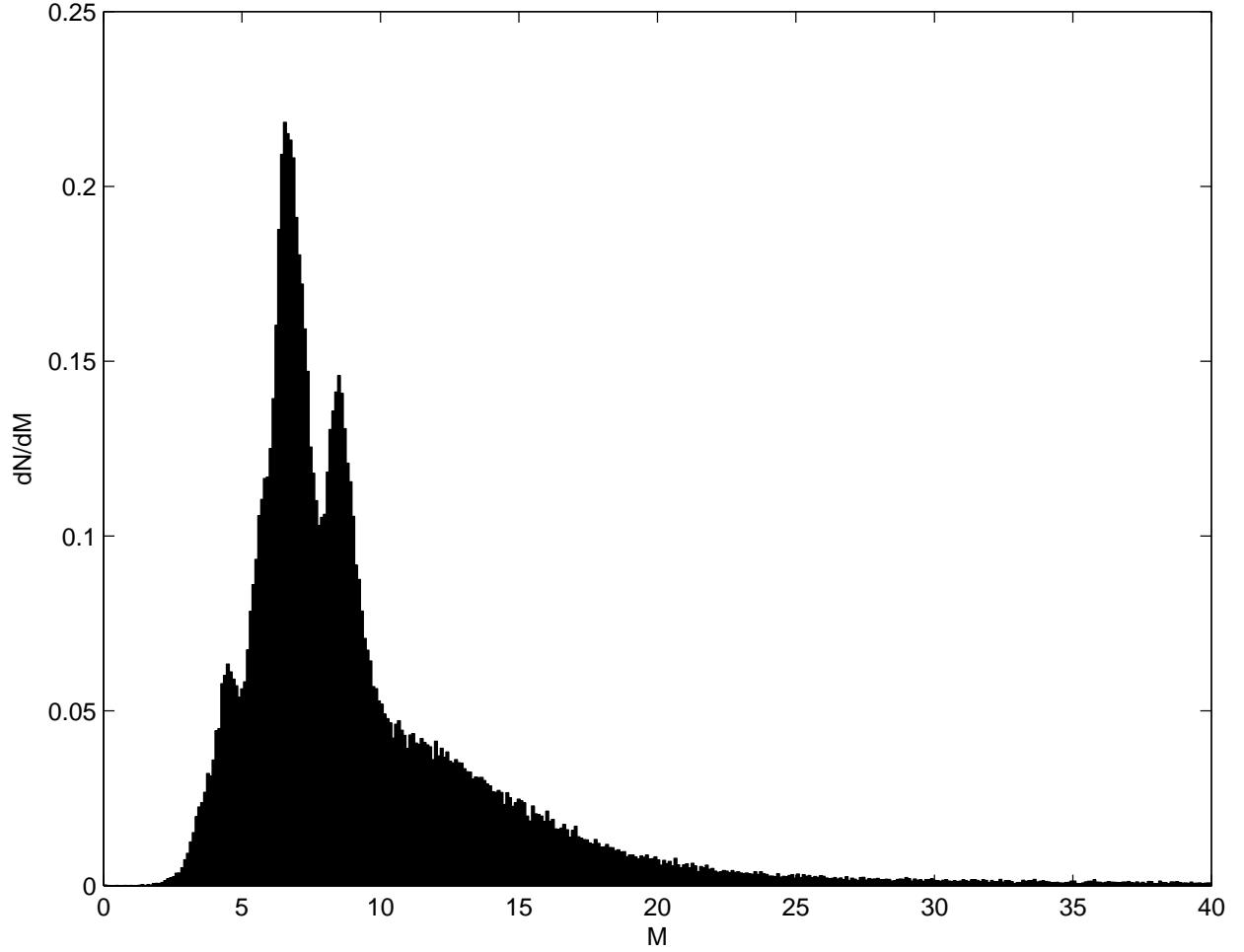


Fig. 2.— The mass distributions in Figure 1 combined into a single distribution. Note that this combination is not a good estimator of the true mass distribution of black hole masses in the galaxy—see, e.g., Mandel (2010). We include it here only to give a sense of the range and general shape of the mass distribution of black holes in our sample.

model of the mass distribution,  $p(m|\theta)$ . We approximate the integrals as averages of  $p(m|\theta)$  over the Monte Carlo mass samples drawn from the distributions in Table 1 (also see Figure 1):

$$p(d|\theta) \approx \prod_i \frac{1}{N_i} \sum_{j=1}^{N_i} p(m_{ij}|\theta), \quad (6)$$

where  $m_{ij}$  is the  $j$ th sample (out of a total  $N_i$ ) from the  $i$ th individual mass distribution.

Our calculation of the likelihood of each observation does not include any attempt to account for selection effects in the observations. We simply assume (almost certainly incorrectly) that any black hole drawn from the underlying mass distribution is equally likely to be observed, leaving the analysis of selection effects to future work.

For a mass distribution with several parameters,  $p(\theta|d)$  lives in a multi-dimensional space. Exploring the entirety of this space systematically rapidly becomes prohibitive as the number of parameters increases. A more efficient way to explore the distribution  $p(\theta|d)$  is to use a Markov Chain Monte Carlo (MCMC) method. MCMC methods produce a chain (sequence) of parameter samples,  $\{\theta_i, i = 1, \dots\}$ , such that a particular parameter,  $\theta$ , appears in the sequence with a frequency equal to its probability according to  $p(\theta|d)$ . In this way, regions of parameter space where  $p(\theta|d)$  is large are sampled densely while regions where  $p(\theta|d)$  is small are effectively ignored. A Markov chain has the property that the transition probability from one element to the next,  $p(\theta_i \rightarrow \theta_{i+1})$ , depends only on the value of  $\theta_i$ , not on any previous values in the chain.

One way to produce a sequence of MCMC samples is via the following algorithm, first proposed by [FIXME: Metropolis citation needed]:

1. Begin with the current sample,  $\theta_i$ .
2. Propose a new sample,  $\theta_p$ , by drawing randomly from an arbitrary “jump proposal distribution” with probability  $Q(\theta_i \rightarrow \theta_p)$ .
3. Compute the “acceptance” probability,

$$p_{\text{accept}} \equiv \frac{p(\theta_p|d) Q(\theta_p \rightarrow \theta_i)}{p(\theta_i|d) Q(\theta_i \rightarrow \theta_p)} \quad (7)$$

4. With probability  $\min(1, p_{\text{accept}})$  set  $\theta_{i+1} = \theta_p$ ; otherwise set  $\theta_{i+1} = \theta_i$ .

Note that this algorithm is more likely to accept a proposed jump when it increases the posterior (the first term in Equation (7)) and when it is to a location in parameter space from which it is easy to return (the second term in Equation (7)). As  $i \rightarrow \infty$  the samples  $\theta_i$  are distributed according to  $p(\theta|d)$ . In practice the number of samples required before the chain appropriately samples  $p(\theta|d)$  depends strongly on the jump proposal distribution; proposal distributions that often propose jumps toward or within regions of large  $p(\theta|d)$  can be very efficient, while poor proposal distributions can require prohibitively large numbers of samples before convergence.

Once we have a chain of samples from  $p(\theta|d)$ —a distribution of distributions—the distribution for any quantity of interest can be computed. For example, Figure 3 shows the median values, and the 10%, and 90% quantiles for value of the BH mass probability distribution at various masses in each of the models discussed in the following subsections. Such a figure can be generated from the MCMC output by computing the value of the parameterized BH mass distribution at various masses for each parameter sample output by the MCMC, and then plotting the quantiles at each mass.

It is also common to look at the one-dimensional distribution for a single parameter obtained by integrating over all other dimensions in parameter space; this is approximated by a histogram of the MCMC sample values for that parameter. Such a distribution is called the “marginalized” distribution.

We defer discussion of the appropriateness of the different models given the data (i.e. the “goodness of fit”) until Section 4.

### 3.1. Parametric Models for the Black Hole Mass Distribution

In this subsection, we discuss the various parametric models of the underlying black hole mass distribution that we have analyzed using the data from Section 2.

#### 3.1.1. Power-Law Models

Many astrophysical distributions are power-laws. Let us assume that the BH mass distribution is given by

$$p(m|\theta) = p(m|\{M_{\min}, M_{\max}, \alpha\}) = \begin{cases} Am^\alpha & M_{\min} \leq m \leq M_{\max} \\ 0 & \text{otherwise} \end{cases}. \quad (8)$$

The normalizing constant  $A$  is given by

$$A = \frac{1 + \alpha}{M_{\max}^{1+\alpha} - M_{\min}^{1+\alpha}}. \quad (9)$$

To perform the Bayesian analysis, we need to impose a prior on the parameters  $M_{\min}$ ,  $M_{\max}$ ,  $\alpha$ . We choose to impose a broad prior that represents our significant uncertainty in the appropriate values of these parameters. We use

$$p(\theta) = p(\{M_{\min}, M_{\max}, \alpha\}) = \begin{cases} \frac{1}{2} \frac{1}{40^2} \frac{1}{20} & 0 \leq M_{\min} \leq M_{\max} \leq 40, \quad -12 \leq \alpha \leq 8 \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

where the masses are measured in solar masses. This prior is uniform in the masses between 0 and  $40M_{\odot}$ , and uniform in power-law slope between  $-12$  and  $8$ .



Our MCMC analysis output is a list of  $\{M_{\min}, M_{\max}, \alpha\}$  values distributed according to the posterior

$$p(\theta|d) = p(\{M_{\min}, M_{\max}, \alpha\}|d) \propto p(d|\{M_{\min}, M_{\max}, \alpha\})p(\{M_{\min}, M_{\max}, \alpha\}), \quad (11)$$

with the likelihood  $p(d|\{M_{\min}, M_{\max}, \alpha\})$  defined in Equation (5). Figure 3 presents the resulting distribution of mass distributions inferred from the MCMC distribution of parameters.

In Figure 4, we display a histogram of the resulting samples in  $\alpha$ ; this represents the one-dimensional “marginalized” distribution

$$p(\alpha|d) = \int dM_{\min} dM_{\max} p(\{M_{\min}, M_{\max}, \alpha\}|d). \quad (12)$$

The resulting distribution is quite broad with

$$-7.66 < \alpha < -1.36 \quad (13)$$

enclosing 90% of the probability. The median value is  $\alpha = -4.28$ , and we have  $\alpha < 0$  (a falling underlying mass distribution) at 98.7% confidence. The distribution of the minimum mass is also quite interesting, but we defer this discussion until Section 5, where we discuss the distribution for the minimum black hole masses in the context of all our models.

### 3.1.2. Decaying Exponential

Theoretical arguments in Fryer & Kalogera (2001) suggest that the black-hole mass distribution may be well-represented by a decaying exponential with a minimum mass:

$$p(m|\theta) = p(m|\{M_{\min}, M_0\}) = \begin{cases} \frac{e^{-\frac{m}{M_0}}}{M_0} \exp\left[-\frac{m}{M_0}\right] & M_0 \leq m \\ 0 & \text{otherwise} \end{cases}. \quad (14)$$

We choose again a broad, flat prior

$$p(\{M_{\min}, M_0\}) = \begin{cases} \frac{1}{40^2} & 0 \leq M_{\min}, M_0 \leq 40 \\ 0 & \text{otherwise} \end{cases}, \quad (15)$$

with  $M_{\min}$  and  $M_0$  measured in solar masses.

Figure 5 displays the marginalized posterior distribution for the scale mass of the exponential,  $M_0$ . The median scale mass is  $M_0 = 3.00$ , and  $1.57 \leq M_0 \leq 5.71$  with 90% confidence. We defer discussion of the distribution of  $M_{\min}$  until Section 5. Refer to Figure 3 for the resulting shape of the distribution of distributions.

### 3.1.3. Gaussian and Two-Gaussian Models

The prototypical single-peaked probability distribution is a Gaussian:

$$p(m|\theta) = p(m|\{\mu, \sigma\}) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ - \left( \frac{m - \mu}{\sqrt{2}\sigma} \right)^2 \right]. \quad (16)$$

We choose a broad, flat prior

$$p(\{\mu, \sigma\}) = \begin{cases} \frac{1}{40^2} & 0 \leq \mu, \sigma \leq 40 \\ 0 & \text{otherwise} \end{cases}, \quad (17)$$

where both  $\mu$  and  $\sigma$  are measured in solar masses. Figure 6 shows the resulting marginalized distributions for the parameters  $\mu$  and  $\sigma$ . We constrain  $\mu$  in  $7.02 \leq \mu \leq 9.23$  with 90% confidence. (Refer to Figure 3 for the shape of the resulting distribution of distributions.)

To look for a second peak in the black-hole mass distribution, we also examined a two-Gaussian model:

$$p(m|\theta) = p(m|\{\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha\}) = \frac{\alpha}{\sigma_1\sqrt{2\pi}} \exp \left[ - \left( \frac{m - \mu_1}{\sqrt{2}\sigma_1} \right)^2 \right] + \frac{1 - \alpha}{\sigma_2\sqrt{2\pi}} \exp \left[ - \left( \frac{m - \mu_2}{\sqrt{2}\sigma_2} \right)^2 \right], \quad (18)$$

where we restrict  $\mu_1 < \mu_2$  with our broad, flat prior:

$$p(\{\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha\}) = \begin{cases} \frac{1}{2} \frac{1}{40^4} & 0 \leq \mu_1 \leq \mu_2 \leq 40, \quad 0 \leq \sigma_1, \sigma_2 \leq 40, \quad 0 \leq \alpha \leq 1 \\ 0 & \text{otherwise} \end{cases}. \quad (19)$$

The one-Gaussian model is a subset of the Gaussian model: when  $\alpha$  is small, then the presence of the first Gaussian is irrelevant, and the second gives the same mass distribution as a single Gaussian; when  $\alpha$  is large, the reverse situation again reproduces the single-Gaussian model; and when  $\alpha$  is in the middle of its range, the two means and standard deviations can be nearly equal, reproducing again the single-Gaussian model. This is precisely what we observe from our MCMC samples in the two-Gaussian model. In Section 4, we will quantify more precisely how much redundancy we have in this model relative to the one-Gaussian model, but we should emphasize here that there is no evidence of a significant contribution from a second peak in the two-Gaussian model.

## 3.2. Non-Parametric Models for the Black Hole Mass Distribution

The previous subsection discussed models for the underlying black hole mass distribution that assumed particular parameterized shapes for the distribution. In this subsection, we will discuss models that do not assume a priori a shape for the black hole mass distribution. The fundamental

non-parameteric distribution in this section is a histogram with some number of bins,  $N_{\text{bin}}$ . Such a distribution is piecewise-constant in  $m$ .

One choice for representing such a histogram would be to fix the bin locations, and allow the heights to vary. With this approach, one should be careful not to “split” features of the mass distribution across more than one bin in order to avoid diluting the sensitivity to such features; similarly, one should avoid including more than “one” feature in each bin. The locations of the bins, then, are crucial. An alternative representation of histogram mass distributions avoids this difficulty.

We choose to represent a histogram mass distribution with  $N_{\text{bin}}$  bins by allocating a fixed total probability,  $1/N_{\text{bin}}$ , to each bin. The lower and upper bounds for each bin are allowed to vary; when these are close to each other (i.e. the bin is narrow), the distribution will have a large value, and conversely when the bounds are far from each other. We assume that the non-zero region of the distribution is contiguous, so we can represent the boundaries of the bins as a non-decreasing array of masses,  $w_0 \leq w_1 \leq \dots \leq w_{N_{\text{bin}}}$ , with  $w_0$  the minimum and  $w_{N_{\text{bin}}}$  the maximum mass for which the distribution has support. This gives the distribution

$$p(m|\theta) = p(m|\{w_0, \dots, w_{N_{\text{bin}}}\}) = \begin{cases} 0 & m < w_0 \text{ or } w_{N_{\text{bin}}} \leq m \\ \frac{1}{N_{\text{bin}}} \frac{1}{w_{i+1} - w_i} & w_i \leq m < w_{i+1} \end{cases}. \quad (20)$$

For priors on the histogram model with  $N_{\text{bin}}$  bins, we assume that the bin boundaries are uniformly distributed between 0 and  $40M_{\odot}$  subject only to the constraint that the boundaries are non-decreasing from  $w_0$  to  $w_{N_{\text{bin}}}$ :

$$p(\{w_0, \dots, w_{N_{\text{bin}}}\}) = \begin{cases} \frac{(N_{\text{bin}}+1)!}{40^{N_{\text{bin}}+1}} & 0 \leq w_0 \leq w_1 \leq \dots \leq w_{N_{\text{bin}}} \leq 40 \\ 0 & \text{otherwise} \end{cases}. \quad (21)$$

The median values of the histogram mass distributions that result from the MCMC samples of the posterior distribution for the  $w_i$  parameters for one-, two-, three-, four-, and five-bin histogram models are shown in Figure 3. As the number of bins increases, the models are better able to capture features of the mass distribution. (But note that in Section 4, we will find that the one-bin histogram—which is a flat distribution between a maximum and minimum mass—is the best model from this group.) We defer discussion of the distribution of the minimum mass from the histogram distributions until Section 5.

## 4. Model Selection

In Section 3, we discussed a series of models for the underlying black hole mass distribution. For each model, we have assumed that the underlying mass distribution corresponds to the model, and we ask what distributions are implied by the data for the parameters in the model. We have

not yet asked which models are more likely to correspond to the actual distribution. That is the topic of this section.

To compare models in the context of a Bayesian analysis, consider a set of models,  $\{M_i|i = 1, \dots\}$ , each with corresponding parameters  $\theta_i$ . Re-writing Equation (2) to be explicit about the assumption of a particular model, we have

$$p(\theta_i|d, M_i) = \frac{p(d|\theta_i, M_i)p(\theta_i|M_i)}{p(d|M_i)}. \quad (22)$$

This gives the posterior probability of the parameters  $\theta_i$  in the context of model  $M_i$ . But, the particular model itself can be regarded as a parameter in a larger “super-model” that encompasses all the  $M_i$ . If we let the model become a parameter, Bayes’ rule gives us

$$p(\theta_i, M_i|d) = \frac{p(d|\theta_i, M_i)p(\theta_i|M_i)p(M_i)}{p(d)}, \quad (23)$$

where we have introduced the model prior  $p(M_i)$ , which represents our estimate on the probability that model  $M_i$  is correct in the absence of the data  $d$ . The normalizing evidence is now

$$p(d) = \sum_i \int d\theta_i p(\theta_i, M_i|d). \quad (24)$$

To compare the various models  $M_i$ , we are interested in the marginalized posterior probability of  $M_i$  (also called the evidence for model  $M_i$ ):

$$p(M_i|d) \equiv \int d\theta_i p(\theta_i, M_i|d). \quad (25)$$

This quantity can be re-written in terms of the single-model evidence (see Equations (22) and (4)):

$$p(M_i|d) = \frac{p(M_i)}{p(d)} \int d\theta_i p(d|\theta_i, M_i)p(\theta_i|M_i) = \frac{p(d|M_i)p(M_i)}{p(d)}. \quad (26)$$

There are two broad ways to compare models in the context of the MCMC analysis of Section 3. The first way, which we discuss in Section 4.1, is to use the MCMC samples from the various models to estimate the value of the integral in Equation (26). The second way is to treat the model literally as a parameter in a second MCMC analysis. One then proposes jumps (see the discussion in Section 3) not only for the  $\theta_i$  within a particular model, but from parameters  $\theta_i$  in model  $M_i$  to parameters  $\theta_j$  in  $M_j$ . This technique is called a “reversible-jump MCMC” (Green 1995); we discuss it in Section 4.2.

#### 4.1. Direct Evidence Calculations

In this section we discuss techniques for using the MCMC samples from Section 3 to estimate the integral in Equation (26). The first technique we use comes from re-writing Equation (22):

$$\frac{p(\theta_i|d, M_i)}{p(d|\theta_i, M_i)} = \frac{p(\theta_i|M_i)}{p(d|M_i)}. \quad (27)$$

Integrating both sides over all  $\theta_i$ , recalling that the prior is normalized, we have

$$\int d\theta_i \frac{p(\theta_i|d, M_i)}{p(d|\theta_i, M_i)} = \frac{1}{p(d|M_i)}. \quad (28)$$

We can estimate the integral using our MCMC samples (which are distributed with local density  $p(\theta_i|d, M_i)d\theta_i$ ) via

$$\int d\theta_i \frac{p(\theta_i|d, M_i)}{p(d|\theta_i, M_i)} \approx \frac{1}{N_i} \sum_j \frac{1}{p(d|\theta_{ij}, M_i)}, \quad (29)$$

where the  $j$ th MCMC parameter sample (from a total  $N_i$ ) is  $\theta_{ij}$ . Thus we have

$$\frac{1}{p(d|M_i)} \approx \frac{1}{N_i} \sum_j \frac{1}{p(d|\theta_{ij}, M_i)}; \quad (30)$$

the evidence for model  $M_i$  can be estimated from the harmonic mean of the likelihood over the MCMC samples.

Alternately, we can attempt to perform the integral over  $\theta_i$  for the model evidence directly using our MCMC samples. If we could assign non-overlapping volumes,  $V_j$ , to each of the  $N_i$  MCMC samples,  $\theta_{ij}$ , then we could approximate the integral in Equation (26) via

$$\int d\theta_i p(d|\theta_i, M_i)p(\theta_i|M_i) \approx \sum_j p(d|\theta_{ij}, M_i)p(\theta_{ij}|M_i)V_j. \quad (31)$$

To associate a volume to each MCMC sample, we use a data structure known as a kD-tree. (Our algorithm is similar to Algorithm P of Weinberg (2010).) A kD-tree decomposes a volume in  $\mathbb{R}^n$  into non-overlapping boxes, each of which contains some number of points from a given sample. The decomposition proceeds recursively: given a region to subdivide and a sample of points, a dimension is chosen (we choose the dimension of largest extent of the sample, but other choices are possible) and the region is split along this dimension such that the numbers of points to each side of the partition are equal or differ by one. Each of the resulting sub-regions and the sub-samples contained in them are further split in the same manner, until every region contains fewer than some chosen number of points. The sequence of splittings forms a binary tree, with the initial region and point sample at the root, and the final boxes and subsamples at the leaves. From experimentation on artificial point distributions, we find that terminating the splitting when each box contains fewer than  $\sim 64$  to  $\sim 256$  samples is reasonably effective in estimating evidence integrals in  $\sim 5$  dimensions. (See also Weinberg (2010).) To estimate the integral as in Equation (31), we walk the leaves of the kD-tree; for each leaf, we compute the volume spanned by the samples in that leaf, and assign an equal fraction of this volume to each sample (note that the volume spanned by the samples is *smaller* than the leaf region itself). The integral is computed using the assigned volumes as in Equation (31).

The harmonic mean estimate of the evidence is known for having large statistical uncertainty due to the tendency for a few low-likelihood points to dominate the sum (Weinberg 2010); the

statistics of the kD-tree estimate are less well-known. In any case, it is useful to have multiple independent estimates of the evidence for a given model. In Figure 7 we use both methods to estimate the evidence for the various models discussed in Section 3. We assume that no model is a priori more likely than any other, so the model priors  $p(M_i)$  are equal. We also ignore the model-independent constant  $p(d)$  in Equation (26), so only ratios of evidence are relevant. Due to the large statistical uncertainty expected in the harmonic mean estimate of the evidence, we perform bootstrap estimates of the range of the harmonic mean, reporting the 10% and 90% values in the error bars of Figure 7.

The results of the direct evidence calculation appear to point to the power law and one-bin histogram as the two most favored models, but the error ranges and inconsistency between the two techniques for calculating evidence cast some doubt on this conclusion. In the next section, we will confirm these results with a more accurate method for estimating the evidence.

## 4.2. Reversible-Jump MCMC

The technique of reversible-jump MCMC (Green 1995) follows the “super-model” concept, in which the choice of model,  $M_i$ , becomes a parameter, to its logical conclusion. One performs an MCMC that proposes jumps between the parameter spaces of the various models that are under consideration (i.e. from parameters  $\theta_i$  in model  $M_i$  to parameters  $\theta_j$  in model  $M_j$ ). For this MCMC to be efficient, proposed jumps into a model should favor regions with large posterior; when the posterior is highly-peaked in a small region of parameter space, proposed jumps outside this region are unlikely to be accepted, and the reversible-jump MCMC samples will require a very long chain to properly sample the “super-model” posterior. The evidence for a particular model is proportional to the number of reversible-jump MCMC samples that live in that model’s parameter space.

We use the kD-trees from the last subsection to interpolate the posterior between MCMC samples from single-model MCMCs when proposing jumps in our reversible-jump MCMC. Thus, proposed jumps track the posteriors computed from the single-model MCMCs. (For an alternative approach to generating jump proposals for reversible-jump MCMCs, see Littenberg & Cornish (2009).) The interpolation algorithm works as follows. We construct a kD-tree as described in the last subsection, continuing the subdivision until each region contains exactly one sample point. To propose a jump, we first choose a model uniformly from the set of models (i.e. according to the model priors). Then we choose a point uniformly from the sample chain (i.e. a point is chosen according to the posterior). Then we find the leaf region in the kD-tree that contains this point, and propose a jump uniformly within this region. This procedure for proposing a jump admits a straightforward calculation of the jump proposal distribution. The advantage of using a kD-tree instead of the method in Littenberg & Cornish (2009) is that the tree is locally adaptive to the structure of the MCMC samples, and therefore more efficiently tracks the posterior from each of the single-model MCMCs.

We have performed a reversible-jump MCMC that jumps between all the models (both parameteric and non-parameteric) in Section 3. The counts for the number of samples in each model appear in Figure 8; relative probabilities of the models are given in Table 2. The most favored model is the power law from Section 3.1.1, followed by the one- and two-bin histograms from Section 3.2. This confirms the results from the direct integration and harmonic mean evidence estimators in the last subsection. Interestingly, the theoretical curve from Fryer & Kalogera (2001) places fourth in the ranking of evidence. However, while the exponential distribution is a factor of 3.9 times less likely than the best-fit power law model, it is only 1.2 times less likely than the two best histogram models. A choice of model prior that only slightly favored the exponential model over the non-parameteric histograms would put it in second place.

Assuming that the reversible-jump MCMC has converged (and we have tested ours for convergence), the standard deviation on the number of counts in each model is given by

$$\sigma_{N_i} = \sqrt{N_i \left(1 - \frac{N_i}{N}\right)}, \quad (32)$$

where  $N_i$  is the number of counts in model  $M_i$ , and  $N$  is the total number of reversible-jump counts. This quantity is extremely small relative to the statistical error in the harmonic mean estimator or the likely truncation error on the direct integration estimate, so we consider the reversible-jump evidence results to be definitive.

## 5. The Minimum Mass of the Black Hole Mass Distribution

Using the MCMC samples for the parameters of the distributions in Section 3, we can form probability distributions for the minimum black hole mass implied by the model posteriors. (For the Gaussian models without a sharp cutoff we use the mass corresponding to the 1% quantile of the distribution as the “minimum black hole mass.”) Figure 9 displays the probability distributions for the minimum mass in the various models.

For the three most-favored models (the power law, one-, and two-bin histograms), the minimum black hole mass is above 5.00, 3.23, and 3.82 solar masses, respectively, with 95% confidence. While these limits are lower than those of a previous study of fewer systems (Bailyn et al. 1998), they are still significantly above the maximum neutron star mass. The measured masses of the systems in Section 2 show strong evidence of a “gap” between the maximum neutron star and minimum black hole mass.

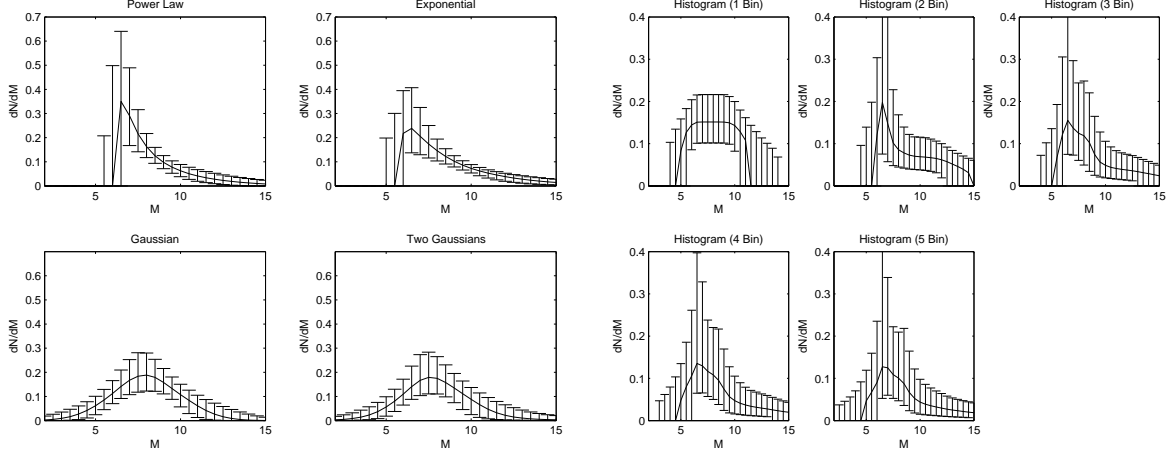


Fig. 3.— The median values of the black hole mass distribution,  $p(m|\theta)$ , at various masses implied by the posterior  $p(\theta|d)$  for the models discussed in Sections 3.1 and 3.2. Error bars span the 10% to 90% range. Note that these “distributions of distributions” are not necessarily normalized, and need not be “shaped” like the underlying model distributions.

Model	Relative Probability
Power Law (Section 3.1.1)	0.441660
Histogram (1 Bin, Section 3.2)	0.143435
Histogram (2 Bin, Section 3.2)	0.135854
Exponential (Section 3.1.2)	0.112514
Histogram (3 Bin, Section 3.2)	0.075592
Histogram (4 Bin, Section 3.2)	0.038839
Gaussian (Section 3.1.3)	0.022999
Histogram (5 Bin, Section 3.2)	0.021011
Two Gaussian (Section 3.1.3)	0.008096

Table 2: Relative probabilities of the various models from Section 3 computed from reversible-jump MCMC samples.



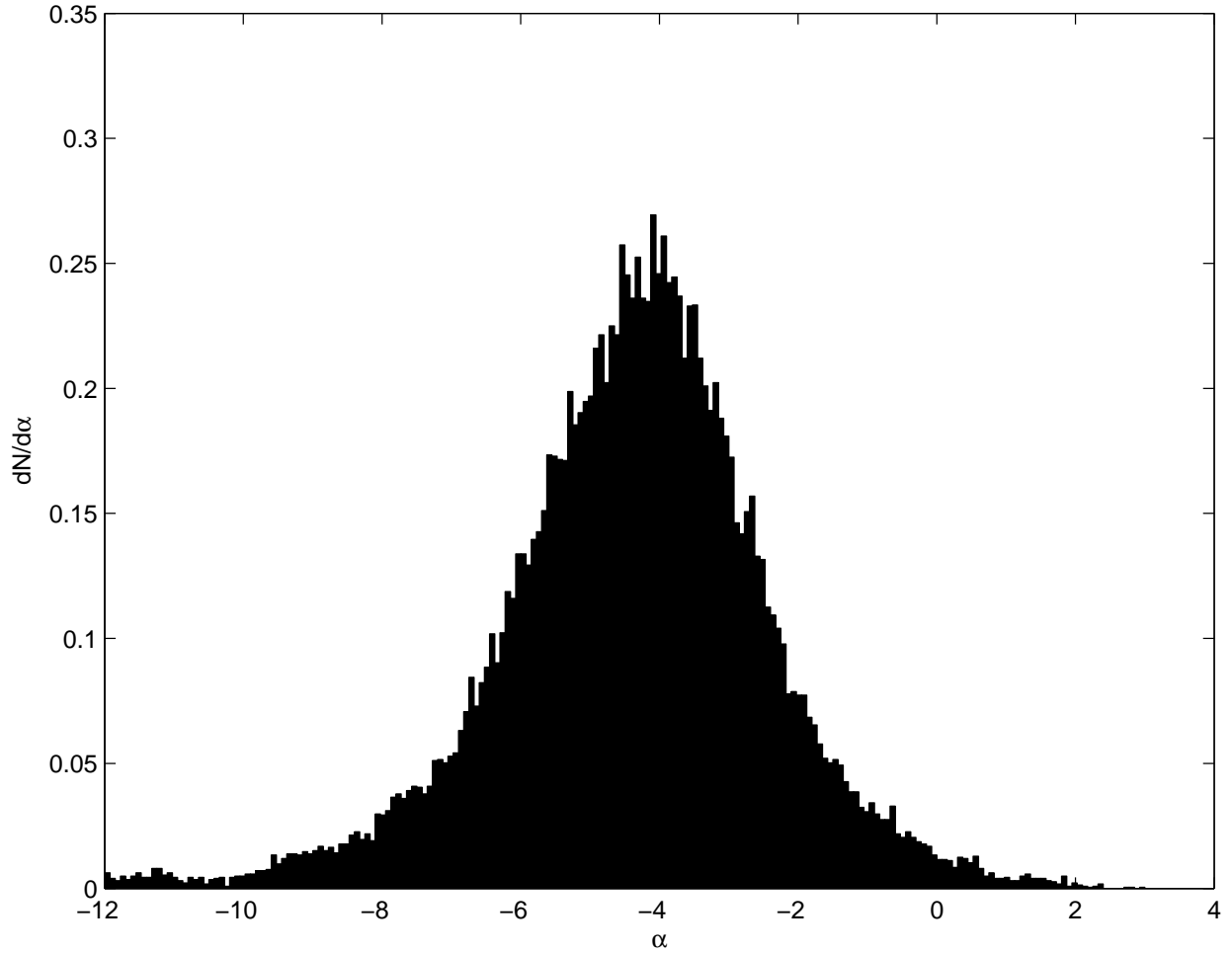


Fig. 4.— A histogram of the MCMC samples for  $\alpha$  from the power-law distribution defined in Equation (8). The histogram is an approximation to the one-dimensional marginalized PDF in Equation (12). The favored value is near  $\alpha = -4$ . The fraction of the distribution with  $\alpha < 0$  is 98.7%. The 5% quantile is  $-7.66$ , the median is  $-4.28$ , and the 95% quantile is  $-1.36$ .

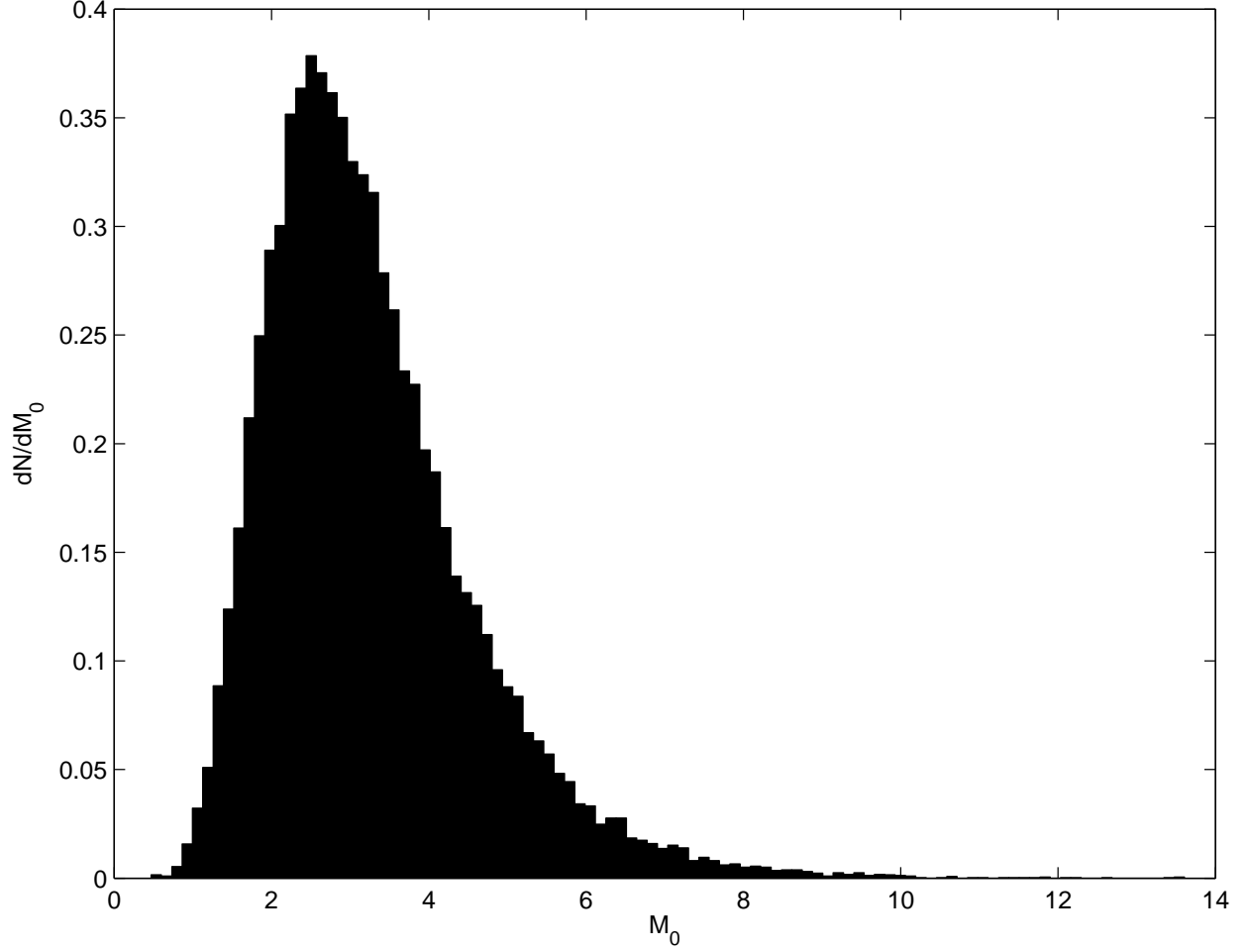


Fig. 5.— The distribution of scale masses,  $M_0$ , measured in units of a solar mass for the exponential underlying mass distribution defined in Equation (14). The median scale mass is  $M_0 = 3.00$ ; with 90% confidence  $1.57 \leq M_0 \leq 5.71$ .

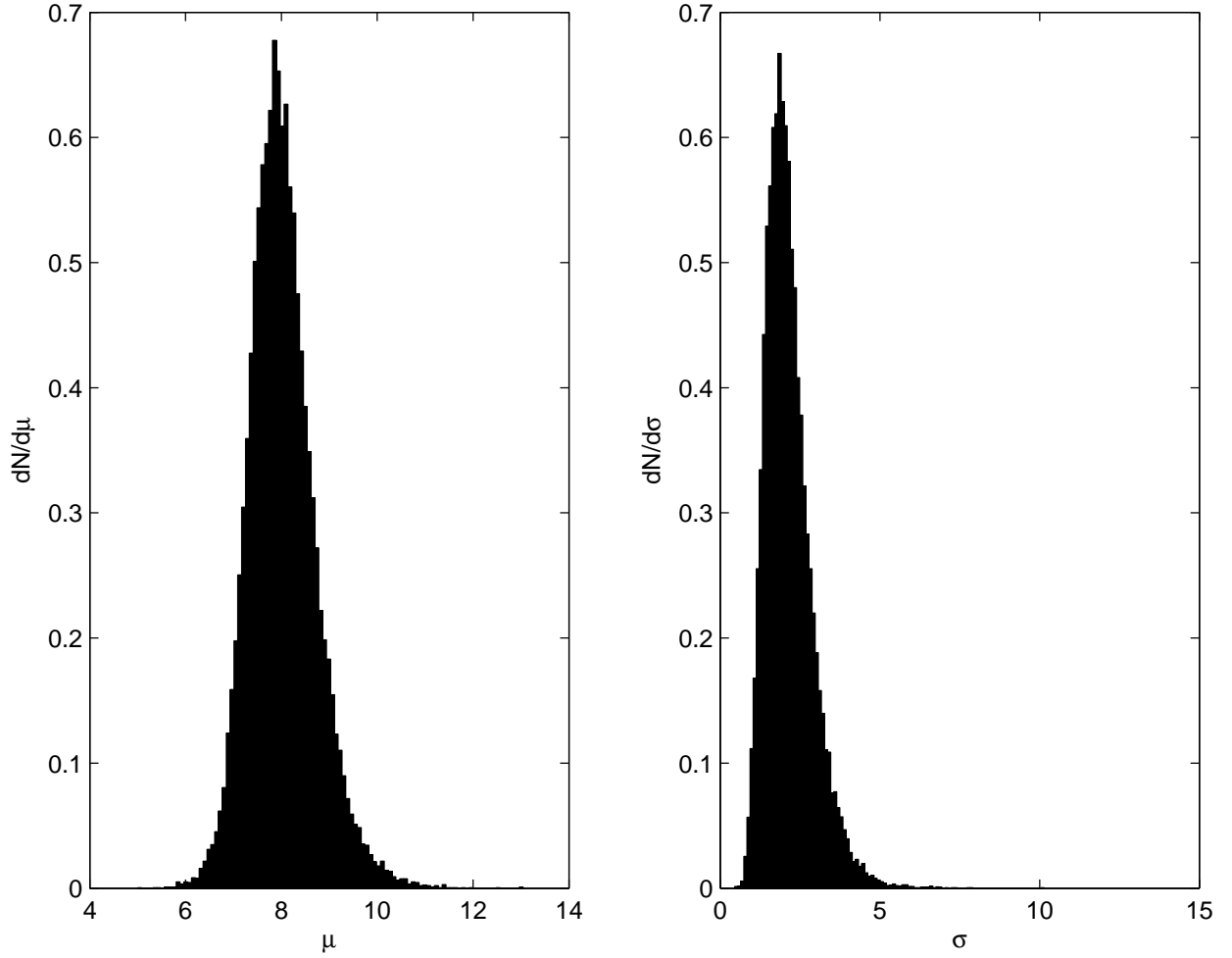


Fig. 6.— Marginalized posterior distributions for the mean  $\mu$  and standard deviation  $\sigma$  (both in solar masses) for the Gaussian underlying mass distribution defined in Equation (17). The peak of the Gaussian,  $\mu$ , is constrained in  $7.02 \leq \mu \leq 9.23$  with 90% confidence.

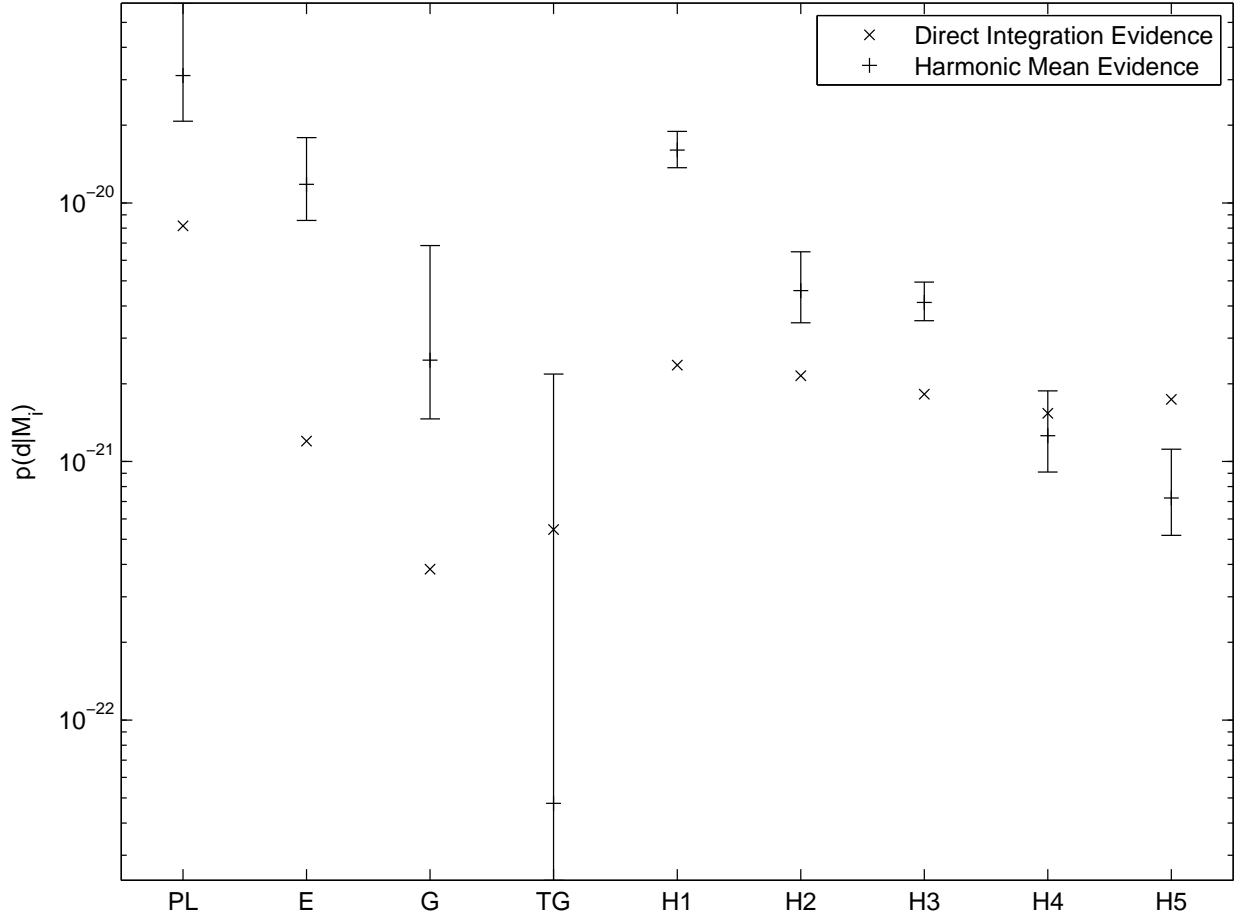


Fig. 7.— Direct integration and harmonic mean evidence for the various models from Section 3 (see Section 4.1). In increasing order along the  $x$ -axis, the models are the power-law of Section 3.1.1 (PL), the decaying exponential of Section 3.1.2 (E), the single Gaussian of Section 3.1.3 (G), the double Gaussian of Section 3.1.3 (TG), and the one-, two-, three-, four-, and five-bin histogram models of Section 3.2 (H1, H2, H3, H4, H5, respectively). The cross symbols give the value of the direct integration; the plus symbols give the harmonic mean estimator with bootstrap estimates of the 10% and 90% bounds on the harmonic mean. It appears that the power law and one-bin histogram are the two best models, but the error ranges and inconsistency between the two techniques for calculating evidence cast some doubt on this conclusion.

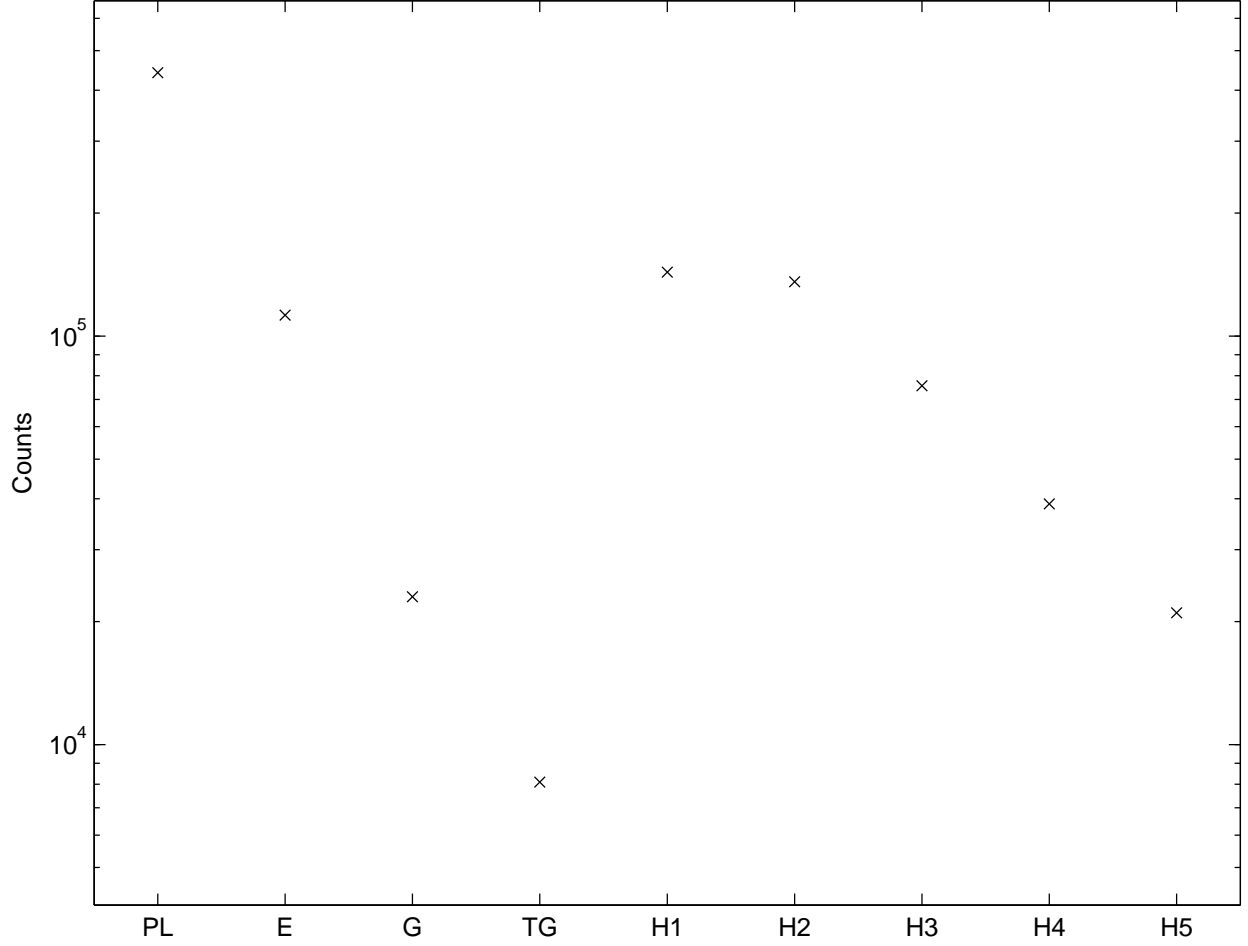


Fig. 8.— Counts of samples in the parameter space of each model from Section 3 in the reversible-jump MCMC of Section 4.2. The models are the power law of Section 3.1.1 (PL), the exponential of Section 3.1.2 (E), the Gaussian (G) and two Gaussian (TG) models of Section 3.1.3, and the one-, two-, three-, four-, and five-bin histograms of Section 3.2 (H1, H2, H3, H4, H5, respectively). The number of counts in each model is proportional to the evidence for that model; therefore, the most favored model is the power-law, followed by the one- and two-bin histogram.

## 6. Conclusion

This work was supported by grant XXX. WMF would like to thank Ilya Mandel for illuminating discussions and encouragement.

## REFERENCES

- Bailyn, C. D., Jain, R. K., Coppi, P., & Orosz, J. A. 1998, *ApJ*, 499, 367
- Cantrell, A. G. et al. 2010, *ApJ*, 710, 1127, arXiv:1001.0261
- Casares, J. et al. 2009, *ApJS*, 181, 238
- Charles, P. A., & Coe, M. J. 2006, *Compact stellar X-ray sources* (Cambridge University Press), 215–265
- Filippenko, A. V., Leonard, D. C., Matheson, T., Li, W., Moran, E. C., & Riess, A. G. 1999, *PASP*, 111, 969, arXiv:astro-ph/9904271
- Fryer, C. L., & Kalogera, V. 2001, *ApJ*, 554, 548, arXiv:astro-ph/9911312
- Gelino, D. M., Balman, S., Kiziloglu, U., Yilmaz, A., Kalemci, E., & Tomsick, J. A. 2008, in *AIP Conference Proceedings*, Vol. 1010, *A Population Explosion: The Nature and Evolution of X-Ray Binaries in Diverse Environments*, 82–84
- Gelino, D. M., & Harrison, T. E. 2003, *ApJ*, 599, 1254, arXiv:astro-ph/0308490
- Gelino, D. M., Harrison, T. E., & McNamara, B. J. 2001, *AJ*, 122, 971, arXiv:astro-ph/0105064
- Gies, D. R. et al. 2003, *ApJ*, 583, 424, arXiv:astro-ph/0206253
- Green, P. J. 1995, *Biometrika*, 82, 711
- Greene, J., Bailyn, C. D., & Orosz, J. A. 2001, *ApJ*, 554, 1290, arXiv:astro-ph/0101337
- Greiner, J., Cuby, J. G., & McCaughrean, M. J. 2001, *Nature*, 414, 522, arXiv:astro-ph/0111538
- Harlaftis, E., & Filippenko, A. 2005, in *Astronomical Society of the Pacific Conference Series*, Vol. 332, *The Fate of the Most Massive Stars*, ed. R. Humphreys and K. Stanek, 416
- Littenberg, T. B., & Cornish, N. J. 2009, *Phys. Rev. D*, 80, 063007, arXiv:0902.0368
- Mandel, I. 2010, *Phys. Rev. D*, 81, 084029, arXiv:0912.5531
- Neilsen, J., Steeghs, D., & Vrtilik, S. D. 2008, *MNRAS*, 384, 849, arXiv:0710.3202

- Orosz, J. A. 2003, in IAU Symposium, Vol. 212, A Massive Star Odyssey, from Main Sequence to Supernova, ed. K. A. van der Hucht, A. Herrero, & C. Esteban, 365, arXiv:astro-ph/0209041
- Orosz, J. A. et al. 2007, Nature, 449, 872, arXiv:0710.3165
- Orosz, J. A., McClintock, J. E., Remillard, R. A., & Corbel, S. 2004, ApJ, 616, 376, arXiv:astro-ph/0404343
- Orosz, J. A., Steiner, J. F., McClintock, J. E., Torres, M. A. P., Remillard, R. A., & Bailyn, C. D. 2010, ApJ, Submitted
- Weinberg, M. D. 2010, submitted to Bayesian Analysis, arXiv:0911.1777

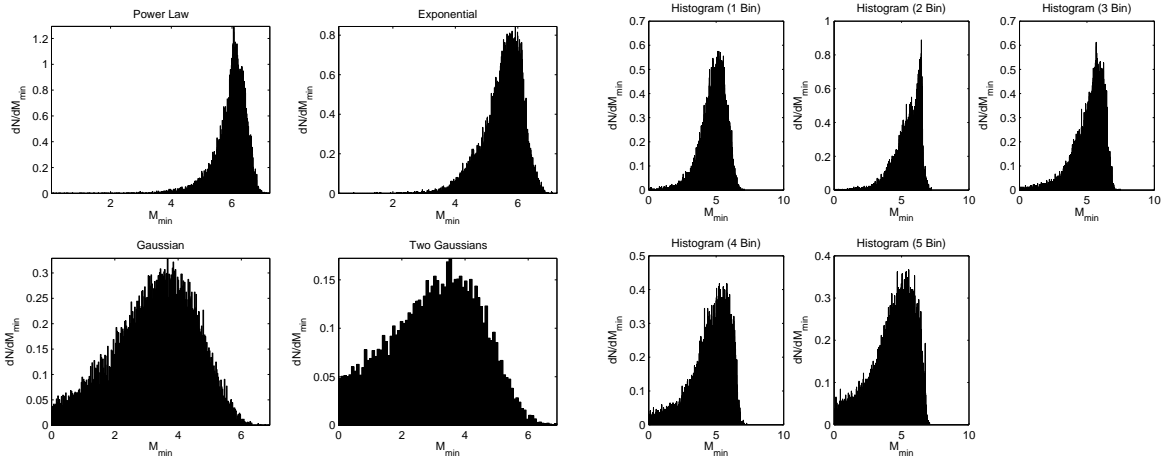


Fig. 9.— The distributions for the minimum black hole mass calculated from the MCMC samples for the models in Section 3. For the Gaussian distributions without a sharp cutoff, we use the mass corresponding to the 1% mass quantile for the “minimum black hole mass.” For the three most-favored models (the power law, one-, and two-bin histogram), the minimum black hole mass is above 5.00, 3.23, and 3.82 solar masses, respectively (at 95% confidence).