H₀ AND ODDS ON COSMOLOGY

ANDREW JAFFE
Canadian Institute for Theoretical Astrophysics, 60 St. George Street, Toronto, Ontario, Canada M5S 1A1
Received 1995 September 18; accepted 1996 April 16

ABSTRACT

Recent observations by the Hubble Space Telescope of Cepheids in the Virgo cluster imply a Hubble constant \( H₀ = 80 \pm 17\, \text{km s}^{-1}\,\text{Mpc}^{-1} \); other recent observations find \( 70 \lesssim H₀ \lesssim 90\, \text{km s}^{-1}\,\text{Mpc}^{-1} \), with several large excursions in either direction. We attempt to clarify some issues of interpretation of these results for determining the global cosmological parameters \( \Omega \) and \( \Lambda \). In this paper, we use these results as a case study in the formalism of Bayesian model comparison, allowing a rigorous comparison of the different cosmological possibilities. We concentrate our analysis on three recent determinations of the Hubble constant, but the results are generic so long as they prefer \( H₀t₀ > 1 \), which would seem to require \( \Lambda > 0 \) within the context of Friedmann-Robertson-Walker cosmologies. With our more rigorous methods, the data do indeed suggest a universe with a nonzero cosmological constant but vanishing curvature: \( \Omega + \Lambda = 1 \).

Subject headings: cosmology: observations — cosmology: theory — distance scale — methods: statistical

1. INTRODUCTION

What have we really learned from the recent observations of Virgo Cepheids and their implications for the Hubble constant? That is, what have we learned about global cosmology from these observations, \( H₀ = 100\ h\ \text{km s}^{-1}\,\text{Mpc}^{-1} \)? Other recent observations find \( 70 \lesssim H₀ \lesssim 90\, \text{km s}^{-1}\,\text{Mpc}^{-1} \), with several large excursions in either direction. We attempt to clarify some issues of interpretation of these results for determining the global cosmological parameters \( \Omega \) and \( \Lambda \). In this paper, we use these results as a case study in the formalism of Bayesian model comparison, allowing a rigorous comparison of the different cosmological possibilities. We concentrate our analysis on three recent determinations of the Hubble constant, but the results are generic so long as they prefer \( H₀t₀ > 1 \), which would seem to require \( \Lambda > 0 \) within the context of Friedmann-Robertson-Walker cosmologies. With our more rigorous methods, the data do indeed suggest a universe with a nonzero cosmological constant but vanishing curvature: \( \Omega + \Lambda = 1 \).

Bayes’s theorem states

\[
p(\theta | D) = \frac{p(\theta) p(D | \theta)}{p(D)} ,
\]

where \( p(ab|c) \) roughly means “the probability [density] of \( a \) given \( b \) and \( c \).” Here \( \theta \) represents the parameters of the theory we are considering (or more precisely, the statement that the parameters lie in some range), \( D \) is the outcome of some experiment, and \( I \) is any background information. Then \( p(\theta | I) \) is the prior distribution for the parameters, \( p(D | \theta I) \) is the likelihood of the data, and \( p(D | I) \) is sometimes known as the evidence. Usually, this theorem is used to decide how the experiment effects our knowledge of the parameters of the theory. It can also be used in a more general context to compare theories and see which better explain the data (MacKay 1992; Garrett & Coles 1993).

This Bayesian approach to theory-testing has the advantage that it automatically incorporates aspects of Ockham’s razor, favoring the simpler theory unless the more complicated one is significantly better at explaining the data. We let \( j, k, \ldots \) represent the different models and write the background information corresponding to each as \( I = I₁ + I₂ + \ldots \), where \( + \) is “logical or.” Then the likelihood of model \( j \) is \( p(D | j I) = p(D | I₀) \), just the evidence for the model as defined above. In this formalism, the ratio of the probabilities of two models (i.e., the “odds” favoring one or the other) is given by

\[
\frac{p(j | D I)}{p(k | D I)} = \frac{p(j | D)}{p(k | D)} \frac{\int d\theta_j p(\theta_j | I₀) p(D | \theta_j I₀)}{\int d\theta_k p(\theta_k | I₀) p(D | \theta_k I₀)} ,
\]

where \( \theta \) and \( I₀ \) refer to the parameters and background information required for model \( j \), and \( p(j | I) \) is the prior probability of the model. (Note that this is different than the usual definition of “odds” used in wagering.) We shall concentrate on the second factor (known as the Bayes factor), which contains the experimental information. A model is favored by this factor if the average of its likelihood with respect to the prior distribution is greater—if more of its parameter space is likely, given the data. Thus, if there are large areas of the allowed parameter space with very low likelihoods, the model as a whole may be disfavored, even if it contains a strongly favored maximum likelihood. This notion of simplicity that this formalism encourages is quite specific: it disfavors theories that require “fine tuning.” If the range of a parameter or parameters is collapsed significantly by the data (i.e., if the likelihood function is strongly peaked, while the prior probability is not), the theory is disfavored relative to theories whose prior probabilities were initially more compact (if, of course, they predicted the observed data!). Obviously, a “just so” theory with prior probabilities strongly favoring the peak of the likelihood would be supported by this method. However, these theories belong to a class that would each be given a low “model prior,” \( p(\phi | I) \) above, in the absence of prior information preferring that model.
2. BAYES'S THEOREM AND COSMOLOGICAL PARAMETERS

What are the experimental results we will be considering? Obviously, we are interested in the measured value and error of the Hubble constant, \( H \pm \sigma_H \). As a sample of recent results, we consider recent observations of the Cepheids in the Virgo cluster by \( HST \), which give \( h = 0.80 \pm 0.17 \) (Freedman et al. 1994); \( h = 0.87 \pm 0.07 \) from CFHT (Pierce et al. 1994); and \( h = 0.66 \pm 0.07 \) from SN Ia light curves (Riess et al. 1994). In addition, we care about the measured age of some "old" object in the universe, recent determinations prefer a globular cluster age of \( 14.69 \pm 2 \) Gyr (Chaboyer 1995), which we use in the following.

These are related to the theoretical parameters by the likelihood functions, which we shall take to be Gaussians:

\[
p(\tau | t) = N(\tau, \tau, \sigma_\tau) \quad \text{and} \quad p(\tilde{H} | H_0, t) = N(H_0, \tilde{H}, \sigma_H)
\]

with the normal distribution

\[
N(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).
\]

As noted by Jaynes (1995), the Gaussian distribution is appropriate when, as here, the available information is limited to the mean and variance. (We shall discuss this further below, in §3.)

We relate these measurements to the full set of theoretical parameters \((H_0, t_0, \tau, \Omega, \Lambda)\); we will almost always want to marginalize over at least a subset of these. To use Bayes's theorem, we will need to assign a prior probability to these parameters, \( p(H_0, t_0, \tau, \Omega, \Lambda) \). Obviously, there are several relationships among the parameters that we must consider. First, if \( t_0 \) is the age of the universe, then we know \( t_0 > \tau \). Second, the age and Hubble constant are related to the cosmological parameters in a matter-dominated FRW universe by

\[
H_0 t_0 = 1.02 h \frac{t_0}{10 \text{ Gyr}} = f(\Omega, \Lambda),
\]

where \( H_0 = 100 \) \( h \) km s\(^{-1}\) Mpc\(^{-1}\), \( \Omega \) is the present density of nonrelativistic matter and \( \Lambda \) is the present vacuum density in units of the critical density (see, for example, Kolb & Turner 1990, p. 528). This simplifies to the usual familiar forms in the case of \( \Lambda = 0 \). In particular, \( H_0 t_0 = \frac{\tau}{2} \) for \( \Omega = 1 \), \( \Lambda = 0 \). For \( \Lambda = 0 \), a maximum of \( f = 1 \) is attained as \( \Omega \to 0 \).

How do we include this information in the prior probability distribution in a consistent way? Because of these constraints, we cannot express it as the product of the individual priors. In this case, we can split up the prior information \( I \) into several disjoint parts: \( I = \Theta F I' \), where \( \Theta \) is the proposition that \( t_0 > \tau \), and \( F \) is the proposition that \( H_0 t_0 = f(\Omega, \Lambda) \). The information \( I' \) contains "everything else"; in particular, it contains the information that we would use to assign priors in the absence of \( F \) and \( \Theta \). We can then use Bayes's theorem in a slightly different way, considering \( F \) and \( \Theta \) as "data":

\[
p[H_0, t_0, \Omega, \Lambda | I] = p(H_0, t_0, \Omega, \Lambda | \Theta F I') \frac{p(H_0, t_0, \Omega, \Lambda | I)}{p(\Theta F | I')} \]

Putting it all together, the posterior distribution for the parameters is

\[
p[H_0, t_0, \Omega, \Lambda | \tilde{H}, \tau, \Theta] \propto \int d\Omega d\Lambda p(H_0, t_0, \Omega, \Lambda | I) \]

Heeding the usual admonitions about taking infinite limits only at the end of the calculation, we will assume that the prior information stipulates that \( 0 \leq H_0 \leq H \) and \( 0 \leq t_0 \leq 10^4 \), \( \tau \leq T \), with the upper limits much greater than the observed values. If we take a uniform prior for \( \tau \) between those limits, we can marginalize it out immediately, leaving

\[
p[H_0, t_0, \Omega, \Lambda | \tilde{H}, \tau, \Theta] \propto N(H_0, \tilde{H}, \sigma_H) \]

The new prior (given only \( I \)) is assigned without the relations among the parameters, and the factor \( p(\Theta F | I) \) is, as usual, a normalizing parameter. It remains only to calculate the likelihood, which can obviously be split into several pieces. Using the product rule and the independence of the information,

\[
p(\Theta F | H_0, t_0, \tau, \Omega, \Lambda) = p(\Theta | H_0, t_0, \tau, \Omega, \Lambda) p(F | H_0, t_0, \Omega, \Lambda) \]

The probability \( p(\Theta | t_0, \tau) \) is unity when \( t_0 > \tau \) and zero otherwise, or simply

\[
p(\Theta | t_0, \tau) = \Theta[t_0 - \tau] \]

where \( \Theta(x) \) is the unit step function. The other factor is not so obvious. It must be unity when \( H_0 t_0 = f(\Omega, \Lambda) \) and zero otherwise. Obviously, this is a set of measure zero; we are led to assign a probability proportional to a delta function. There is considerable ambiguity in the assignment, however: any function with zeroes at the same place as \( [H_0 t_0 - f(\Omega, \Lambda)] \) will suffice as the argument of the delta function. This is a common occurrence when passing to infinite or infinitesimal limits in probability theory—the precise approach to the limit must be specified (see Jaynes 1995 for an extended discussion of this point). We must search elsewhere in our prior information \( I' \) to supply the correct prescription. Based on the symmetry of the expression, and the reasonable results that follow from it below, we take

\[
p(F | H_0, t_0, \Omega, \Lambda) \propto \delta[H_0 t_0 - f(\Omega, \Lambda)]
\]

Heeding the usual admonitions about taking infinite limits only at the end of the calculation, we will assume that the prior information stipulates that \( 0 \leq H_0 \leq H \) and \( 0 \leq t_0 \leq 10^4 \), \( \tau \leq T \), with the upper limits much greater than the observed values. If we take a uniform prior for \( \tau \) between those limits, we can marginalize it out immediately, leaving

\[
p[H_0, t_0, \Omega, \Lambda | \tilde{H}, \tau, \Theta] \propto \frac{\delta[H_0 t_0 - f(\Omega, \Lambda)]}{p(\tilde{H}, \tau, \Theta)}
\]
Since we are looking for limits on $\Omega$ and $\Lambda$, we still need to marginalize out the Hubble constant and age of the universe. Unfortunately, for the obvious uniform priors, $p(H, t_0 | I) = 1/(HT)$, the required integrals cannot be done analytically and are unnormalized as $H, T \to \infty$. We can approximate the result by

$$p(\Omega \Lambda | \hat{H} \hat{t} \Omega \Lambda I) \approx \frac{p(\Omega \Lambda | I)}{p(\hat{H} \hat{t} | \Omega \Lambda I)} \Phi(f(\Omega, \Lambda), \hat{H} \hat{t}, \sigma_j), \quad (12)$$

where the width is $\sigma_j \approx \hat{H} \sigma_H + \hat{t} \sigma_T$ (although the equivalent quadrature sum might seem more appropriate). That is, we now have the likelihood for the observations given $f(\Omega, \Lambda)$:

$$p(\hat{H} \hat{t} | f) \approx \Phi(f, \hat{H} \hat{t}, \sigma_j). \quad (13)$$

This is what we naively expect: the experiment gives an approximate lower limit to $f(\Omega, \Lambda)$. Moreover, this expression is exact (up to constants of proportionality) in the limit $\sigma_H \to 0$. (In the figures and tables below, we use a numerical integration instead of the approximation of eq. [12]).

This likelihood, coupled with priors for $\Omega$ and $\Lambda$, is just what we need for the model-comparison formalism. In this case, we shall examine the following “theories,” or classes of models corresponding to different prior distributions for $\Omega$ and $\Lambda$:

$$p(\Omega \Lambda | I) = \begin{cases} 
\delta(\Omega - 1)\delta(\Lambda) & \text{model 1: } \Omega = 1, \Lambda = 0; \\
\delta(\Omega + \Lambda - 1) & \text{model 2: } \Omega + \Lambda = 1; \\
\delta(\Lambda) & \text{model 3: } 0 \leq \Omega \leq 1, \Lambda = 0; \\
1 & \text{model 4: } 0 \leq \Omega \leq 1, 0 \leq \Lambda \leq 1. 
\end{cases} \quad (14)$$

Model 1 has no parameters (and so is the “simplest”), models 2 and 3 both have one parameter, and model 4 has two parameters. If we chose one of the models with a parameter, we may of course use the usual Bayesian techniques to find confidence intervals for $\Omega$ and $\Lambda$. For now, we are not allowing for the possibility that $\Omega \geq 1$.) We have also not allowed any “cosmic variance,” that is, the possibility that we live in a large underdense region and that we are only measuring the local expansion rate (Shi, Woodrow, & Dursi 1995).

For a given model, we can now plot the posterior distribution or we can integrate over the posteriors to compare the models. In Figures 1, 2, and 3, we show the unnormalized posterior (or evidence) for cases 2–4,

![Figure 1](image1.png)  
**Fig. 1.**—Posterior distribution for $\Omega$ for model 3: $\Omega \leq 1$, $\Lambda = 0$, given each of the three data sets, as labeled.

![Figure 2](image2.png)  
**Fig. 2.**—Posterior distribution for $\Lambda$ for model 2: $\Omega + \Lambda = 1$, given each of the three data sets, as labeled.

![Figure 3](image3.png)  
**Fig. 3.**—Posterior distribution for $\Omega$ and $\Lambda$ for model 4: $\Omega, \Lambda \leq 1$, given the HST data, $H_0 = 80 \pm 17$ km s$^{-1}$ Mpc$^{-1}$; contour spacing is 0.2 and the most likely value is the upper left corner ($\Lambda = 1, \Omega = 0$).
likelihood in the region with $\Omega + \Lambda \gtrsim 1$; if we instead use a prior $p(\Omega|I) = 2\delta(\Omega + \Lambda - 1)$, defining model 4, the evidence for this model drops considerably. Even the truly open model 2 ($\Omega < 1, \Lambda = 0$) fits the data quite poorly; the maximum value in that model is $f = 1$ at $\Omega = \Lambda = 0$, which still requires $H_0 \lesssim 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, marginally in conflict with most of the data. If the data suggested even more strongly that $H_0 t_0 > 1$, theories without a cosmological constant would be completely disfavored (since the likelihood for $f$ would be negligible over the domain of the theory); obviously this is a conclusive test for $\Lambda$ (within the standard hot big bang, FRW universe paradigm).

In Table 1, we also show the odds ratio for the combined data set using all three measurements, which combine to give $H_0 \approx 77 \pm 5 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Because of this small error, a value of $H_0 \approx 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ would be a "5σ" fluctuation, and so the odds ratio for the $\Omega = 1, \Lambda = 0$ model that requires it is $10^5 : 1$! How seriously we should take this result depends crucially on whether we trust the experiments involved (so the likelihood is correct) and whether we have legitimate reasons for believing in particular values or ranges of the cosmological parameters (so the prior distribution is correct). If the assumptions we have made are incorrect (that is, if they do not correctly account for the information we have gathered in other observations of the universe) the inferences we draw from them will be incorrect. Here this shows up in the preference for a "wide open" universe (\(\Omega = 0\)); we have other information, not accounted for in this analysis, that implies, for example, \(\Omega \gtrsim 0.2\). And of course, if our error estimates or the shapes of the distributions in which they occur are incorrect, the inferences cannot be trusted.

In Table 2, we present confidence limits on the parameters $\Lambda$ and $\Omega$, assuming $\Omega + \Lambda = 1$ and $\Omega < 1, \Lambda = 0$, (models 2 and 3), respectively. We define the confidence limit $C$ as usual to be the point such that $\int_0^C d\Omega p(\Omega | DI) = C$ or $\int_0^C d\Lambda p(\Lambda | DI) = C$. Note that the naive combinations of the data sets seem to require quite unusual parameters with the universe dominated by either the cosmological constant or by curvature. That is, the limits on the parameters are quite strict, disfavoring values that are preferred by unrelated analyses [for example, the "best-fit" value of $\Lambda \approx 0.7$, which some have quoted (Krauss & Turner 1995)]. We also note that changing the globular cluster age $\tau$ to the more conservative (from a cosmological standpoint, not according to stellar evolution) value of 12 Gyr has no qualitative effect on the results.

3. COMBINING INFORMATION FROM DIFFERENT EXPERIMENTS—SYSTEMATICS

So far, we have considered each experimental result for the Hubble constant separately and in combinations. However, we must be more careful than before about the possible systematic errors unaccounted for in each result; are the experiments each really measuring the same quantity? Our formalism allows us to approach an answer to this question when we phrase it a bit differently: are the experimental results consistent with one another and with our prior notions of what those results should be? Not surprisingly, the tool we need is again the quantity known as the evidence: the probability of the data given a particular model.

In Figures 1 and 2 we showed the unnormalized posterior probability, the evidence, $p(\Omega | I)p(H|I)\Omega|\Lambda I)$. This allows us to compare different experiments: for example, the CFHT results are always "less likely" than those of the $HST$ given these models, since a smaller part of the $(H_0, t_0)$ parameter space is allowed. Simply put, these models have difficulty producing a larger value of $H_0$—the well-known concern with these results implying a large Hubble con-
stant. The Bayesian analysis has not given us a qualitatively different understanding of the implications of the data; rather, it allows us to put that knowledge on a more quantitative footing, spelling out all of the assumptions (and problems).

Before, we used the evidence to compare models to one another given particular data or combinations thereof; now, we compare the likelihood of having observed particular data given a specific model. If we wish to combine models, we simply use the appropriate joint likelihood function, which is the product of the likelihoods for each experiment. (Because we have assumed Gaussian likelihoods for the experimental results, we can equivalently just use the appropriate quadrature combinations of the individual $H_i$, as long as $H_i^2/\sigma_i^2 \geq 1$.) We must also decide on what to use as our prior distribution for $\Omega$ and $\Lambda$—in the light of which model or combination of models will we consider the data? For example, if we only consider $\Omega = 1$, $\Lambda = 0$, the more strongly the data disfavor $f = \hat{\Omega}$, the smaller the evidence.

In Table 1, we show the evidence for each of our models individually, as well as the combination of models $1 + 2 + 3 + 4$, each weighted equally, all for each possible combination of data sets. Note that combining of models inevitably decreases the evidence, because the quadrature combination of data sets decreases the effective error in the Hubble constant measurement, $\sigma_p$, thereby decreasing the width of the step in the likelihood for $f$.

To really decide if a given data set is “correct” (i.e., that it is measuring the quantity claimed by the observers), we would again need more information about what they should have seen, given a particular theory, or other— independent—observations. We have so far taken the observers at their word; we have assumed that the quoted results for the Hubble constant really represent “1 $\sigma$” confidence intervals, so the $\sigma_i^2$ gives the variance. In this case, the appropriate distribution to use for the likelihood, which takes into account the information given and nothing more, is indeed a Gaussian (Jaynes 1995). However, there are perhaps reasons to question whether this is “all of the information available.” Freedman et al. (1994) provide an extensive error budget in their paper; some items, like the location of the galaxy in question with respect to the cluster center, are best modeled by distributions other than a Gaussian (e.g., an isothermal sphere, which has considerably larger tails). For the data of Pierce et al. (1994), on the other hand, fewer systematic effects have been included, so we should perhaps be wary of the small size of their error bars.

4. “THEORIST’S PRIORS”

Even if we accept all of the Hubble constant data as correct, there may be other information available to us that we are not considering, resulting in the unphysically extreme limits that seem to require a universe dominated by a cosmological constant or curvature. So far, we have used uniform priors for $\Omega$ and $\Lambda$. With these distributions, we are guaranteed to never prefer the simplest model with $\Omega = 1$, $\Lambda = 0$. This is because the function $f(\Omega, \Lambda)$ has a minimum, with a value of $\frac{7}{6}$ at that point, and the likelihood for $f$ is approximately uniform (actually, it is monotonically increasing) for values of greater than about $f = \hat{H}t$, so $\Omega = 1$, $\Lambda = 0$ is always the least likely point in our parameter space! Practically speaking, once $f$ is below $\frac{7}{6}$, $f(\Omega, \Lambda)$ is always located in the constant probability part of the likelihood, and the data cannot distinguish between the theories, as we would expect. Moreover, the data also seem to require a universe dominated by curvature or a cosmological constant, as we saw in Table 2.

This sort of data, giving only a lower limit on the quantity $f(\Omega, \Lambda)$, will never prefer a value $\Omega = 0$ or $\Lambda = 1$—it will try to make the universe as old as possible. As usual, what we really need is independent information on $\Omega$ or $\Lambda$.

Even without any specific measurements of the density of the universe (although such estimates abound), we would not choose uniformly between the different theories. Rather, we may instead prefer probability distributions that are more strongly peaked toward the theoretically favored cases of a flat universe ($\Omega + \Lambda = 1$), or a universe without a cosmological constant ($\Lambda = 0$), typical biases held by theorists. To this end, we can use

$$p(\Omega | I) = 1/(\Omega | 1 - \Omega)) \ ; \ 0 < \Omega < 1 - e^{-a} ;$$

(15)

$$p(\Lambda | I) = 1/(a\Lambda) \ ; \ e^{-a} < \Lambda < 1 .$$

(16)

These distributions are uniform in the logarithm of $|1 - \Omega|$ and $\Lambda$, emphasizing the “scaling” nature of these quantities. We have cut the distributions off with the small parameter $e^{-a}$ to normalize them; inflation, for example, only specifies that $\Omega + \Lambda - 1$ is exponentially suppressed, so this is an appropriate normalization.

Unfortunately, when we use these prior distributions in our formalism, we find that they overwhelm the data! That is, the posterior distributions strongly peak at the prior favored values, $\Omega = 1$ or $\Lambda = 0$; the likelihood peaks are too gentle. Actually, there are still local maxima in the posterior from the likelihoods, at $\Omega = 0$ and $\Lambda = 1$, but they are strongly suppressed relative to the prior maxima.

We should not be surprised by this result. Mathematically, Figures 1–3 show a very gently varying likelihood distribution, easily dominated by the strong prior peaks. Moreover, if we accept Bayesian probability theory as a model for our actual reasoning processes, this explains our reluctance to give up on the viability of the $\Omega = 1$ universe. Only when and if such data arrives as to overwhelm our prior convictions will we be convinced.

5. CONCLUSION

So what, in the end, does all this mean? We have quantified the conventional wisdom (Krauss & Turner 1995): the “best-fit model” is one with $\Omega + \Lambda = 1$, favoring a cosmological constant $\Lambda \gtrsim 0.6$ (considering only the HST measurement). Alternately, a low $\Omega$ is possible, although less strongly favored, requiring $\Omega \lesssim 0.3$, at least with current data. Recently, Leonard & Lake (1995) have performed a similar analysis, without the explicit emphasis on probabilistic methods, coming to the similar conclusion that $\Omega < 1$ and $\Lambda > 0$ is the most likely interpretation of the current data. As our knowledge of the age of the universe and the Hubble constant increase (if the central values we have used here are indeed correct), cosmological parameters further from the simplest flat universe with no cosmological constant will be required. One novel feature of this analysis is its “automatic” use of Ockham’s razor; for example, the model with $\Omega + \Lambda \leq 1$ is disfavored with respect to the one-parameter models, even though it includes them as a subset, because on average, it predicts smaller values for $H_0 t_0$. 
Finally, we must stress that we have taken pains to concentrate on the information garnered from the data alone (and a very limited amount of data at that). We have not analyzed any of the myriad other determinations of the Hubble constant that find values different from those used here. It is clear that the field is still dominated by systematics; it is not clear (although it is hoped) that these HST results in particular are free from these problems. In a separate work, we will conduct a historical survey of age and Hubble constant data using these techniques.

This paper has been intended both as a tutorial in the application of Bayesian model-comparison techniques to astrophysical and cosmological problems, and as a cautionary tale: inference from invalid or inconsistent assumptions can result in incorrect results. At the very least, the assumptions that go into the solution of a problem, embodied in the forms of the prior and likelihood distributions, must be explicitly stated. Bayesian theory has the advantage that it requires such an explicit formulation—but it in no way forces reasonable and consistent use of such information without careful accounting. Here some of the results we have gathered by, for example, combining the different Hubble constant data in a naive fashion, seem to give far too stringent limits on the amount of mass in the universe (as in Table 2 above); is this because our prior information about \( \Omega \) and \( \Lambda \) was incorrectly utilized? Or is it because we have not correctly understood the results of one or more experiments?

The author would like to thank Brian Chaboyer and Nick Kaiser for helpful discussion, and Peter Coles and Ned Wright for comments on an earlier draft. Bill Press, the referee of a previous version, provided many suggestions.

REFERENCES

Chaboyer, B. 1995, ApJ, 444, L9
Freedman, W. L., et al. 1994, Nature, 371, 757
Garrett, P., & Coles, P. 1993, Comm. Astrophys., 17, 23
Jaynes, E. T. 1995, preprint
Kolb, E. W., & Turner, M. S. 1989, The Early Universe (Redwood City: Addison-Wesley)
Krauss, L., & Turner, M. S. 1995, Gen. Relativ. Grav., 27, 1137
Leonard, S., & Lake, K. 1995, Queens’ preprint
MacKay, D. J. C. 1992, in Maximum Entropy and Bayesian Methods, ed. C. R. Smith et al. (Dordrecht: Kluwer), 39
Pierce, M. J., et al. 1994, Nature, 371, 385
Riess, A. G., Press, W. H., & Kirshner, R. P. 1994, ApJ, 438, L17
Shi, X., Widrow, L. M., & Dursi, L. J. 1995, MNRAS, 281, 565