

# On posterior probability and significance level: application to the power spectrum of HD 49 933 observed by CoRoT<sup>★</sup>

T. Appourchaux<sup>1</sup>, R. Samadi<sup>2</sup>, and M.-A. Dupret<sup>2,3</sup>

<sup>1</sup> Institut d'Astrophysique Spatiale, UMR8617, Université Paris XI, Bâtiment 121, 91405 Orsay Cedex, France  
e-mail: Thierry.Appourchaux@ias.u-psud.fr

<sup>2</sup> Observatoire de Paris, LESIA, UMR8109, 92195 Meudon Cedex, France

<sup>3</sup> Institut d'Astrophysique et de Géophysique de l'Université de Liège, Allée du 6 Août 17, 4000 Liège, Belgium

Received 19 September 2008 / Accepted 21 May 2009

## ABSTRACT

**Context.** The CoRoT mission provides asteroseismic data of very high quality allowing one to adopt new statistical approaches for mode detection in power spectra, especially with respect to testing the null hypothesis ( $H_0$ , which assumes that what is observed is pure noise).

**Aims.** We emphasize that the significance level when rejecting the null hypothesis can lead to the incorrect conclusion that the  $H_0$  hypothesis is unlikely to occur at that significance level. We demonstrate that the significance level is unrelated to the posterior probability of  $H_0$ , given the observed data set, and that this posterior probability is very much higher than implied by the significance level.

**Methods.** We use Bayes theorem to derive the posterior probability of that  $H_0$  is true assuming an alternative hypothesis  $H_1$  that a mode is present, taking some prior for the mode height, mode amplitude and linewidth.

**Results.** We compute the posterior probability of  $H_0$  for the p modes detected on HD 49 933 by CoRoT.

**Conclusions.** We conclude that the posterior probability of  $H_0$  provide a much more conservative quantification of the mode detection than the significance level. This framework can be applied to any similar stellar power spectra obtained to complete asteroseismology.

**Key words.** methods: analytical – methods: statistical – Sun: helioseismology

## 1. Introduction

In the field of helioseismology, the null hypothesis  $H_0$  was used by Appourchaux et al. (2000) to infer upper limits to the amplitude of g modes. The  $H_0$  hypothesis assumes that what is observed is pure noise. This hypothesis was applied to data of classical variable stars to detect peaks in their power spectra (Scargle 1982). In all of these papers, the authors assumed  $H_0$  and defined an upper limit corresponding to a threshold level of, e.g., 10%. Here we argue, that this threshold was arbitrarily chosen a priori. When accepting or rejecting the  $H_0$  hypothesis, there is no discussion whether borderline case should be rejected or accepted. This abrupt truncation between *good* and *bad* leads to a decision that could have been different if the threshold had been different.

In the following sections, we first lay the foundations for understanding the meaning of the  $H_0$  hypothesis, and then explain what is commonly misunderstood about the  $H_0$  hypothesis. We then derive, for specific cases encountered in helio- and asteroseismology, the posterior probability of  $H_0$ . We show how one can apply the formalism for real data, such as that gathered by the CoRoT mission for HD 49 933 and then present our conclusions.

## 2. Significance level and the $H_0$ hypothesis

Fisher (1925) devised the well known Fisher test for testing the null hypothesis ( $H_0$ ). In this test, a threshold of 5% is commonly used and the p-value quoted is the value of the test if it is less than this threshold. For instance, a result of 4.9% would result in rejecting the  $H_0$  hypothesis, while 5.1% would result in accepting the  $H_0$  hypothesis. When the  $H_0$  hypothesis is rejected, the reported p-value is used as a significance level for the validating or not accepting the  $H_0$  hypothesis. In the medical field, the so-called borderline cases led to findings related to effectiveness of medicine that were, sometimes, not proven by subsequent studies. The controversy about the use of p-values that occurred in the medical world is directly related to the abrupt and arbitrary cut-off of the threshold applied (be 5% or 10%). Although the relevance of what has been found in the medical field could seem remote to most astrophysicists, it is extremely relevant to understand that the improper use of p-values is the same as the improper use of the so-called *significance level*.

Reporting a small number for the significance level should not be used to claim the proper rejection of the  $H_0$  hypothesis. The mistake is to ascribe a significance level to a measurement carried out only once, not repeated and spanning just a very small volume of the space of the parameters. When making an observation of a random variable  $x$ , one wants to check the probability that what is observed could be due to noise. For that purpose, a test statistic is derived called  $T(x)$ . If one observes a

<sup>★</sup> The CoRoT space mission, launched on 2006 December 27, was developed and is operated by the CNES, with participation of the Science Programs of ESA, ESA's RSSD, Austria, Belgium, Brazil, Germany and Spain.

value of  $x = X$  that would not reject the  $H_0$ , then one can compute,  $p$ , the significance level or  $p$ -value, defined to be:

$$p = P_0(T(x) \geq T(X)), \quad (1)$$

where  $P_0$  is the probability that  $T(x) \geq T(X)$  when  $H_0$  is true. The test set by Eq. (1) is to confirm that the statistical test  $T(x)$  has been compared with the value given by  $T(X)$ . When computing the statistical test  $T(x)$ , one does not span the space of  $X$ : one has  $x = X$  which is the so-called point null hypothesis. In other terms, what is the probability that one has *exactly* that value of  $x$ ? This is a completely different question from knowing how true  $H_0$  is, i.e., obtaining  $p(H_0|x)$ . Berger & Sellke (1987) provided a way of deriving  $p(H_0|x)$  with respect to  $p(x|H_0)$  and  $p(x|H_1)$ , where  $H_1$  is the alternative hypothesis (i.e., there is a signal). Using Bayes theorem, they obtained the so-called posterior probability of  $H_0$  given the observed data  $x$ :

$$p(H_0|x) = \frac{p(H_0)p(x|H_0)}{p(H_0)p(x|H_0) + p(H_1)p(x|H_1)} \quad (2)$$

where  $p(H_0)$  and  $p(H_1)$  are the probabilities related to the  $H_0$  and  $H_1$  hypothesis, respectively. We note that we would like to determine  $p(H_0|x)$ , the probability of  $H_0$  being true given the data we have, not the probability of the data  $x$  given  $H_0$ , i.e.,  $p(x|H_0)$ . To minimize the impact of the probabilities assigned to both hypotheses, we assume that these are equally probable ( $p_0 = 0.5$ ):

$$p(H_0|x) = \left(1 + \frac{p(x|H_1)}{p(x|H_0)}\right)^{-1}. \quad (3)$$

Based under these assumptions, Sellke et al. (2001) found that the probability  $p(H_0|x)$  that  $H_0$  is true given some observed data  $x$  of a random variable  $X$  has a lower bound

$$p(H_0|x) \geq \left(1 - \frac{1}{ep \ln p}\right)^{-1}. \quad (4)$$

An immediate consequence is that for a significance level of 1%, the odds against  $H_0$  are at least 10 to 1, and for 10%, the odds against  $H_0$  are at least 2.6 to 1. In both cases, the likelihood of incorrectly rejecting  $H_0$  is much higher than inferred by the  $p$ -value, by at most a factor of 10 and 4, respectively.

Sellke et al. (2001) were able to set a lower value to  $p(H_0|x)$  for almost an arbitrary alternative  $H_1$  hypothesis. We show that there is indeed a lower bound when one wants to detect peaks in power spectra. Hereafter, we give several examples of how one can derive in practice the posterior probability of  $H_0$ .

### 3. Posterior probability for peak detection

Here we provide analytical examples of how one can obtain the posterior probability of  $H_0$  for some cases encountered in astero- and helio-seismology.

#### 3.1. Long-lived modes

##### 3.1.1. Mode height known a priori

We search a peak restricted to a single frequency bin in a power spectrum which has a  $\chi^2$  with 2 d.o.f. statistics, for which a bin has reached a value  $x$ . We wish to verify whether this is caused by either a true sine wave or due to noise. We have for the  $H_0$  hypothesis,

$$p(x|H_0) = e^{-x}. \quad (5)$$

The noise is assumed to be 1. For the alternative hypothesis  $H_1$ , we assume that there is a signal of a long-lived mode, i.e., restricted to one bin, for which the mode height  $H$  is known and the mode is stochastically excited (like a stellar p mode). We then have:

$$p(x|H_1) = \frac{1}{1+H} e^{-x/(1+H)}. \quad (6)$$

Equation (3) is then rewritten for our problem as

$$p(H_0|x) = \left(1 + \frac{1}{1+H} e^{xH/(1+H)}\right)^{-1}. \quad (7)$$

Since the significance  $p = p(x|H_0)$ , we finally have that

$$p(H_0|x) = \left(1 + \frac{1}{1+H} p^{-H/(1+H)}\right)^{-1}. \quad (8)$$

It can be shown that the minimum of  $p(H_0|x)$  is reached for  $H = -\ln p - 1$  for the value:

$$p_{\min}^H(H_0|x) = \left(1 - \frac{1}{ep \ln p}\right)^{-1}. \quad (9)$$

In this case, the lower bound defined by Eq. (4) is reached.

##### 3.1.2. Mode height unknown

Most of the time one does not *know* of course the height of the mode to be detected. We can assume a prior for the mode height which can be e.g., a uniform distribution or a gaussian function. For example, if we assume that the mode height is uniformly distributed over some range  $[0, H_u]$ ,  $p(x|H_1)$  is rewritten as

$$p^{\text{uni}}(x|H_1) = \frac{1}{H_u} \int_0^{H_u} \frac{1}{1+H'} e^{-x/(1+H')} dH'. \quad (10)$$

Then Eq. (3) can be rewritten as

$$p^{\text{uni}}(H_0|x) = \left(1 + \frac{1}{H_u} \int_0^{H_u} \frac{1}{1+H'} p^{-H'/(1+H')} dH'\right)^{-1}. \quad (11)$$

It can be shown that a minimum is reached at a value of  $H_{\min}$  by solving

$$\frac{1}{H_{\min}} \int_0^{H_{\min}} \frac{1}{1+H'} p^{-H'/(1+H')} dH' = \frac{1}{1+H_{\min}} p^{-H_{\min}/(1+H_{\min})}. \quad (12)$$

The minimum is then given by

$$p_{\min}^{\text{uni}}(H_0|x) = \left(1 + \frac{1}{1+H_{\min}} p^{-H_{\min}/(1+H_{\min})}\right)^{-1}. \quad (13)$$

From Eqs. (8) and (9), we can deduce that

$$p_{\min}^{\text{uni}}(H_0|x) > \left(1 - \frac{1}{ep \ln p}\right)^{-1}. \quad (14)$$

This means that when there is more uncertainty about the possible height of the mode, one is less likely to reject the  $H_0$  hypothesis.

### 3.2. Short-lived modes

Appourchaux (2004) described how one can detect a mode having a lifetime shorter than the observation time. He suggested smoothing the power spectrum to increase the signal-to-noise ratio and provided an analytical expression for  $H_0$  related to the summation over  $n$  bins of a  $\chi^2$  statistic with constant mean, and for  $H_1$  related to the summation over  $n$  bins of a  $\chi^2$  statistic with different means due to the presence of the mode profile. Therefore, when the spectrum is summed over  $n$  bins, we can derive for the  $H_0$  hypothesis

$$p(x|H_0) = \frac{x^{n-1}e^{-x}}{\gamma(n)}, \quad (15)$$

where the mean of the power spectrum is 1, while  $n$  is the mean of the smoothed power spectrum (for simplicity, we assumed that  $S = 1$ ), and  $\gamma(n)$  is the Gamma function. The significance level  $x$  is then given by solving

$$p = \frac{1}{\gamma(n)} \int_x^{+\infty} u^{n-1} e^{-u} du. \quad (16)$$

Assuming that the mode has a known amplitude  $A$  and a known linewidth  $\Gamma$ , we can write using Eq. (8) of Appourchaux (2004) the approximation

$$p(x|H_1) = \frac{\lambda^\nu}{\gamma(\nu)} x^{\nu-1} e^{-\lambda x}, \quad (17)$$

where  $\lambda$  and  $\nu$ , defined in Appourchaux (2004) are obtained by integrating symmetrically about the central frequency of the mode, thereby ensuring that the signal is maximum;  $\lambda$  and  $\nu$  are functions of the mode height  $H(=A^2/\pi\Gamma)$  and  $\Gamma$ . If we do not know the mode amplitude and its linewidth, we can specify a prior for the amplitude and the linewidth, which, for instance, can be done using a model of mode excitation. Here we assume that the mode amplitude  $A$  and the linewidth  $\Gamma$  are independent (this would not be the case of  $H$  and  $\Gamma$ ). If we use uniform priors, we then have

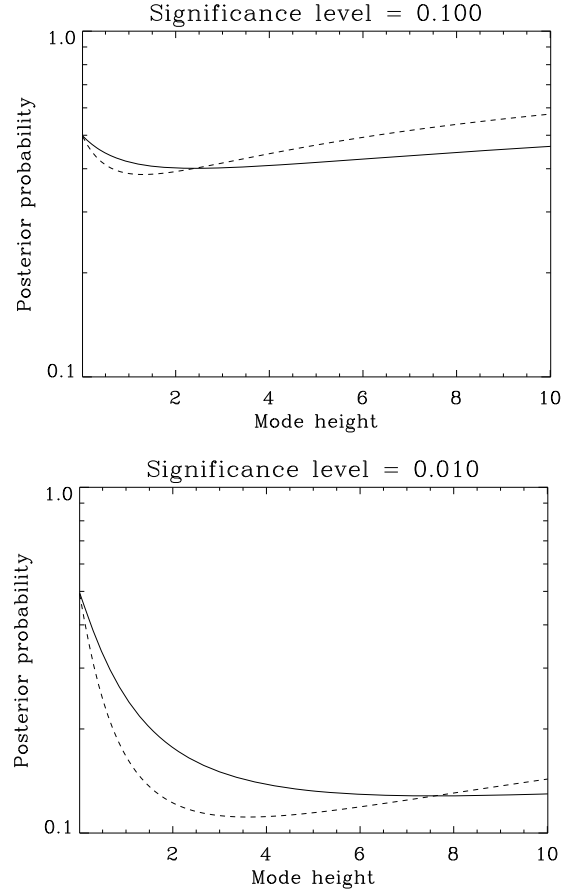
$$p(x|H_1) = \frac{1}{A_u \Gamma_u} \int_0^{A_u} \int_0^{\Gamma_u} \frac{\lambda^\nu}{\gamma(\nu)} x^{\nu-1} e^{-\lambda x} dA' d\Gamma'. \quad (18)$$

This equation would be similar if we were to have a uniform prior on the mode height, where  $A$  would then be replaced by  $H$ . Substituting Eqs. (16) and (17) into Eq. (3), we can then obtain the posterior probability  $p(H_0|x)$  when the mode amplitude and linewidth are known. If we replace Eqs. (17) by (18), we can then obtain the probability  $p(H_0|x)$  when the mode amplitude and linewidth are not known.

### 3.3. Discussion

Figure 1 shows the posterior probability for long-lived modes of a known mode height (Eq. (8)) and an unknown mode height (Eq. (11)) for two different significance levels. Figure 2 shows the posterior probability for short-lived modes of a known mode height and linewidth (Eqs. (15) and (17)) and for an unknown mode height and linewidth (Eqs. (15) and (18), with  $H$  replacing  $A$ ) for a single significance level; even a significance level of 1% does not provide a more robust rejection of the  $H_0$  hypothesis. These posterior probabilities have a lower bound which means that even a very low significance level is no guarantee for positive detection!

It is also counterintuitive that the posterior probability increases when the mode height (known or unknown) increases.



**Fig. 1.** Posterior probability as a function of the known mode height (dashed line), or as a function of the mode height range ( $H_u$ ) of the uniform prior (continuous line) for a significance level of 10%. For the known mode height, the minimum is reached for an height of 1.3 with a value of 38%. For the unknown mode height, the minimum is reached at a value of 40%. Posterior probability as a function of the known mode height (dashed line), or as a function of the maximum mode height of the uniform prior (continuous line) for a significance level of 1%. For the known mode height, the minimum is reached for an height of 3.6 with a value of 11%. For the unknown mode height, the minimum is reached at a value of 12.7%.

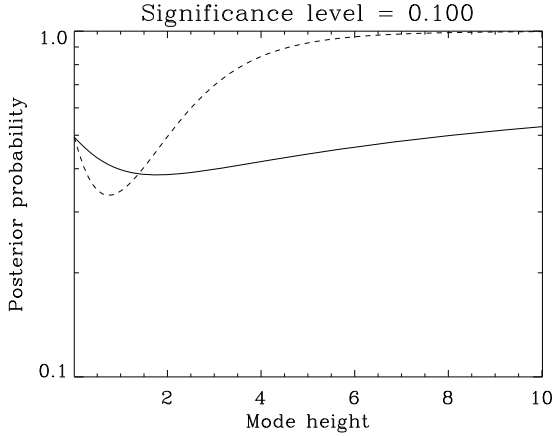
We recall that the significance level  $p$  corresponds to the level  $x$  at which the peak has been observed (i.e., it is  $p = e^{-x}$  for long-lived modes). If we assume a priori that the mode has a high mode height, then the observation at a low significance level indicates that our assumption about the high mode height is incorrect, and that the data dismisses the a priori assumptions made about the mode height. In other words, it is more probable that the  $H_0$  hypothesis is true.

In the absence of an alternative hypothesis, it is advisable to define a low significance value that will probably reject the  $H_0$  hypothesis. Figure 3 shows the lower bound<sup>1</sup> set by Eq. (4) compared to the minimum found using Eq. (11) for the uniform prior on the mode height.

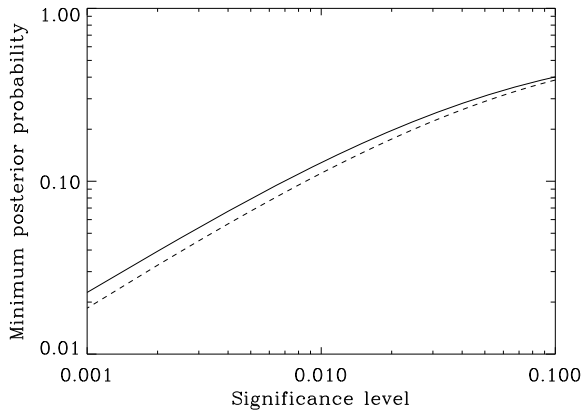
## 4. Application to the CoRoT data: HD 49 933

The computation of the posterior probability was applied to CoRoT data for illustrative purposes. The data used are those of the first initial run performed on HD 49 933 (Appourchaux et al. 2008). The objective was to provide an objective way of

<sup>1</sup> The same lower bound as Eq. (9) for a known mode height.



**Fig. 2.** Posterior probability for a short-lived mode after smoothing the power spectra over 10 bins corresponding to a window twice as large as the mode linewidth. The posterior probability is given as a function of the known mode height (dashed line), or as a function of the mode height range ( $H_u$ ) of the uniform prior (continuous line) and a uniform prior for the linewidth, for a significance level of 10%.

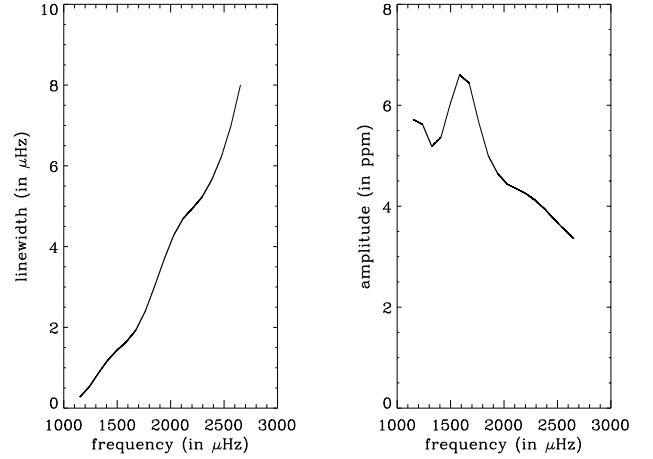


**Fig. 3.** Lower bound to the posterior probability as function of the significance level for the known mode height (dashed line) and for the uniform prior (solid line).

detecting oscillation modes in HD 49 933 that could be applied to any other star.

The methodology used for deriving the posterior probability was as follows:

- we computed the power spectrum from the detrended time series, as in Appourchaux et al. (2008);
- we smoothed the spectrum over  $n$  bins using a boxcar;
- we selected 30 50- $\mu$ Hz wide windows starting at 1200  $\mu$ Hz (the 50- $\mu$ Hz window corresponds roughly to half the large frequency separation);
- for each window, we computed the median in the window of the smoothed spectrum which provides an estimate of the mean noise level if the modes are not present;
- the smoothed spectrum was normalized in each window by dividing by the median and multiplying by the number of smoothing bins  $n$  which provides values commensurate with these of Eq. (15);
- we applied the  $H_0$  hypothesis for a detection probability that a signal caused by noise of 10% over all the 30 windows, taking into account the fact that in each window the number



**Fig. 4.** (Left) Theoretical mode linewidth of HD 49 933 as a function of frequency. (Right) Theoretical mode amplitude of HD 49 933 as a function of frequency.

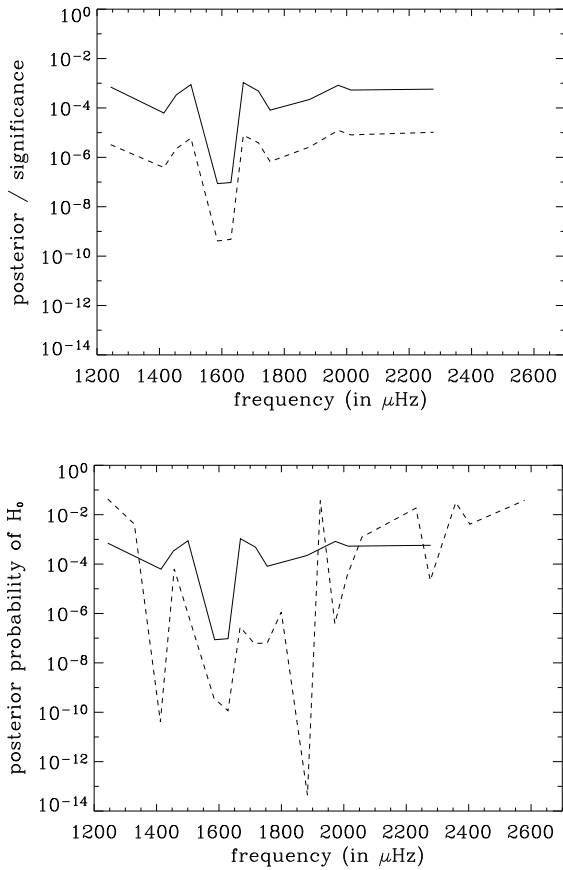
of independent bins is  $50 \delta\nu^{-1} n^{-1}$  ( $\delta\nu$  is the frequency resolution of the original power spectrum). The detection probability is then  $0.1 (N_w)^{-1} (50 \delta\nu^{-1} n^{-1})^{-1}$  per independent bin;

- we then solved Eq. (16) for  $x_{\text{det}}$  given the detection level given above;
- in each window, we then selected the bins that are greater than  $x_{\text{det}}$ , i.e., we accepted or rejected the  $H_0$  hypothesis;
- after the selection, we retained the greatest value  $x_{\text{max}}$  found in the window corresponding to the central frequency of the mode (See Eq. (17));
- we then computed the posterior probability of  $H_0$  given by Eq. (3) using Eqs. (15) and (18) assuming some prior on the mode height and linewidth as described below;
- for comparison, we also computed the significance level as given by Eq. (16) from the value of  $x_{\text{max}}$ .

The theoretical amplitudes for HD 49 933 are derived from Samadi et al. (2009a) using an adiabatic treatment of radiative transfer, and the excitation rate as calculated in Samadi et al. (2009b). The theoretical linewidths were computed with the non-adiabatic pulsation code MAD. This code includes a time-dependent convection treatment described in Grigahcène et al. (2005): it takes into account the role played by the variations in the convective flux, the turbulent pressure, and the dissipation rate of turbulent kinetic energy. This treatment is non-local, with three free parameters  $a$ ,  $b$ , and  $c$  corresponding to the non-locality of the convective flux, the turbulent pressure, and the entropy gradient. We assume the values  $a = 10$ ,  $b = 3$ , and  $c = 3.5$  obtained by fitting the convective flux and turbulent pressure of 3D hydrodynamic simulations in the upper overshooting region of the Sun (Dupret et al. 2006). According to Grigahcène et al. (2005), we introduced a free complex parameter  $\beta$  in the perturbation of the energy closure equation. We used here the value  $\beta = -3i$ , which provides a good agreement between the theoretical and observed linewidths and phase lags in the range of solar pressure modes. A 1D stellar model obtained with the code CESAM was used for our non-adiabatic computations. It has a solar metallicity and reproduces the effective temperature and gravity of HD 49 933 (Samadi et al. 2009a). The amplitudes and linewidths are shown on Fig. 4. We linearly interpolated both curves to provide a continuous prior with frequency.

The uniform priors for amplitude and linewidth are derived from the theory by taking into account an *uncertainty factor* in





**Fig. 5.** (Top) Comparison of the posterior probability (solid line) with the significance level (dashed line) as a function of frequency for a spectrum smoothed over 5 bins ( $\approx 1\mu\text{Hz}$ ). (Bottom) Posterior probability as a function of frequency for a spectrum smoothed over 5 bins (solid line) and over 50 bins (dashed line), of sizes  $\approx 1\mu\text{Hz}$  and  $\approx 10\mu\text{Hz}$ , respectively.

the theoretical model. For amplitude, we assumed that the maximum is  $\sqrt{2}$  larger than given in Fig. 4 (twice in energy); for linewidth, we assumed that the maximum is twice as great as that given in Fig. 4. We note that a larger prior increases the posterior probability, as shown on Figs. 1 and 2. We assumed that the noise floor in HD 49 933 is given by the photon noise, which is about  $0.15\text{ ppm}^2/\mu\text{Hz}$  (Appourchaux et al. 2008).

Figure 5 shows the results of the procedure described above. It is clear that the posterior probability is higher than the significance level, i.e., the posterior probability provides a more

conservative number ( $H_0$  more likely). The smoothing procedure also shows two effects that were predicted by Appourchaux (2004): first, short lived modes are easier to detect when the spectrum is smoothed, second, long lived modes are more difficult to detect when the spectrum is smoothed. The first effect manifests itself in the larger number of detected modes at higher frequency and by the decrease of the prior probability (i.e., the signal is more likely). The second effect is seen at low frequency where a couple of modes have their prior probability increased to non-negligible value after smoothing (i.e., the signal is less likely). When we compare with modes reported by Appourchaux et al. (2008), we find that more than 85% of the  $l = 0-2$  mode pairs and  $l = 1$  modes are recovered. An additional mode at  $2579\mu\text{Hz}$  is detected that could be an  $l = 1$  mode according to the identification of Appourchaux et al. (2008).

## 5. Conclusion

The significance level refers to the significance of the data *given* the hypothesis, while we are interested in the posterior probability of the null hypothesis *given* the data. Here we have shown that for a significance level of 10%, the posterior probability of the null hypothesis is at least 38% when there is no alternative hypothesis. We have illustrated how one can in practice calculate and compute the posterior probability for the null hypothesis. This has been applied to several theoretical examples and to the CoRoT data. For the first time, we have shown how one can assess the detectability of short-lived p modes in a power spectrum. The methodology can be applied to any stellar power spectrum for which theoretical expectations are available.

*Acknowledgements.* We are grateful to John Leibacher for checking the English and for correcting several typos. We are thankful to the referee for making the paper clearer.

## References

- Appourchaux, T. 2004, A&A, 428, 1039
- Appourchaux, T., Fröhlich, C., Andersen, B., et al. 2000, ApJ, 538, 401
- Appourchaux, T., Michel, E., Auvergne, M., et al. 2008, A&A, 488, 705
- Berger, J., & Sellke, T. 1987, J. Am. Stat. Assoc., 82(397), 112
- Dupret, M.-A., Samadi, R., Grigahcène, A., Goupil, M.-J., & Gabriel, M. 2006, Commun. Asteroseismol., 147, 85
- Fisher, R. A. 1925, Statistical Methods for Research Workers (Edinburgh, Scotland: Oliver and Boyd), 299
- Grigahcène, A., Dupret, M.-A., Gabriel, M., Garrido, R., & Scuflaire, R. 2005, A&A, 434, 1055
- Samadi, R., Ludwig, H.-G., Belkacem, K., et al. 2009a, A&A, submitted
- Samadi, R., Ludwig, H.-G., Belkacem, K., Goupil, M., & Dupret, M.-A. 2009b, A&A, submitted
- Scargle, J. D. 1982, ApJ, 263, 835
- Sellke, T., Bayarri, M. J., & Berger, J. 2001, The American Statistician, 55, 62