Evidence for extra radiation? Profile likelihood versus Bayesian posterior

Jan Hamann

Department of Physics and Astronomy
University of Aarhus, DK-8000 Aarhus C, Denmark

E-mail: hamann@phys.au.dk

Abstract. A number of recent analyses of cosmological data have reported hints for the presence of extra radiation beyond the standard model expectation. In order to test the robustness of these claims under different methods of constructing parameter constraints, we perform a Bayesian posterior-based and a likelihood profile-based analysis of current data. We confirm the presence of a slight discrepancy between posterior- and profile-based constraints, with the marginalised posterior preferring higher values of the effective number of neutrino species $N_{\text{eff}}$. This can be traced back to a volume effect occurring during the marginalisation process, and we demonstrate that the effect is related to the fact that cosmic microwave background (CMB) data constrain $N_{\text{eff}}$ only indirectly via the redshift of matter-radiation equality. Once present CMB data are combined with external information about, e.g., the Hubble parameter, the difference between the methods becomes small compared to the uncertainty of $N_{\text{eff}}$. We conclude that the preference of precision cosmological data for excess radiation is “real” and not an artifact of a specific choice of credible/confidence interval construction.
1 Introduction

In the past years, measurements of the temperature and polarisation anisotropies in the cosmic microwave background have revealed a wealth of information about the Universe. One particular quantity that can be inferred from CMB data is the relativistic energy density $\rho_r$ around decoupling, typically expressed in terms of the effective number of massless neutrino degrees of freedom $N_{\text{eff}}$:

$$\rho_r = \frac{\pi^2}{15} T_\gamma^4 (1 + \alpha N_{\text{eff}}) ,$$  

where $T_\gamma = (2.72548 \pm 0.00057) \text{ K}$ [1] is the CMB temperature and $\alpha \equiv \frac{7}{18} \left(\frac{4}{17}\right)^{1/3}$. The three standard model neutrino species are expected to contribute $N_{\text{eff}} = 3.046$ effective degrees of freedom [2]. Intriguingly however, present cosmological data show some indication for $N_{\text{eff}} > 3.046$ [3–9], hinting at the possible existence of further light particle species. These hints are based on a Bayesian statistics analysis of the data however, and as long as the evidence for $N_{\text{eff}} > 3.046$ is weak, one might also want to consider an alternative approach of constraining $N_{\text{eff}}$. A profile likelihood analysis for instance, being prior-independent and parameterisation-invariant, provides a useful cross-check of these results and is complementary to the usual Bayesian analysis based on the posterior probability density [10]. Using a profile likelihood-based analysis, it was recently claimed in [11] that the hints for $N_{\text{eff}} > 3.046$ are merely artifacts of the Bayesian construction of parameter constraints. We shall revisit this claim in the present work.

This paper is organised as follows: we will describe the details of our analysis in section 2, present our results in section 3 and conclude in section 4.

2 Analysis

2.1 Data sets

For clarity of presentation and given the considerable numerical effort required to reliably construct the profile likelihood, we will limit ourselves to two different combinations of data:
Table 1. Parameters and prior ranges for the cosmological and nuisance parameters. For each individual analysis only one out of the first three parameters is used.

| Parameter                                    | Symbol | Prior       |
|----------------------------------------------|--------|-------------|
| Hubble parameter                             | $h$    | 0.4 → 1.0   |
| Dark energy density                          | $\Omega_\Lambda$ | 0 → 1      |
| Ratio of sound horizon to angular diameter distance at decoupling | $\theta_s$ | 0.5 → 10   |
| Baryon density                               | $\omega_b$ | 0.005 → 0.1 |
| Cold dark matter density                     | $\omega_{cdm}$ | 0.01 → 0.99 |
| Amplitude of scalar spectrum @ $k=0.05\text{ Mpc}^{-1}$ | $\log[10^{10}A_s]$ | 2.7 → 4   |
| Scalar spectral index                        | $n_s$  | 0.5 → 1.5   |
| Redshift of reionisation                     | $z_{re}$ | 1 → 50     |
| Effective number of massless neutrinos       | $N_{\text{eff}}$ | 1.5 → 10  |
| Amplitude of Sunyaev-Zel’dovich contribution | $A_{SZ}$ | 0 → 3     |
| Amplitude of clustered point source contribution | $A_c$ | 0 → 20   |
| Amplitude of Poisson point source contribution | $A_P$ | 0 → 100   |

1. A CMB only set, consisting of the 7-year Wilkinson Microwave Anisotropy Probe data (WMAP7) [12] plus the 2008 Atacama Cosmology Telescope (ACT) data [13]. For this data set, the discrepancy between the Bayesian result of [13] and the profile likelihood result reported in [11] is particularly large.

2. The same data combined with a constraint on the Hubble parameter (HST) derived by Riess et al. [14].

2.2 Model and priors

We consider a one-parameter extension of the 6-parameter $\Lambda$CDM vanilla model, varying also the effective number of massless degrees of freedom $N_{\text{eff}}$ on top of the standard parameters. Additionally, three parameters describing the foreground contribution to the small-scale CMB temperature spectrum are required. The parameterisation of the vanilla model is not unique, and there are a number of different choices commonly used in the literature. Since the parameterisation implicitly determines the prior probability distribution, these choices can affect the inference of parameters, even though the physical models are equivalent. In this work we explicitly compare three parameterisation choices: a flat prior on the Hubble parameter $H_0$, a flat prior on the dark energy density $\Omega_\Lambda$ and a flat prior on the ratio of sound horizon to angular diameter distance at decoupling $\theta_s$. We list all free parameters and their associated prior ranges in table 1. The primordial Helium fraction is fixed to $Y_p = 0.24$ in order to facilitate comparison with other authors’ results.

2.3 Marginalised posterior and profile likelihood

Given a model with $n$ free parameters, the full information of the data is contained in the $n$-dimensional likelihood function $\mathcal{L}$. If one wants to construct constraints on a single parameter $\varphi$, the dimensionality obviously needs to be reduced. Most commonly, this is done in a Bayesian framework, by first promoting $\mathcal{L}$ to a probability density function (through multiplication with a prior probability density), and then integrating (“marginalising”) over the nuisance directions (see [10] for a more detailed discussion), resulting in the marginalised posterior. The marginalised posterior can easily be extracted from Markov chains and has
a straightforward interpretation as the probability density of the true value of \( \varphi \), given the model, data and priors.

Since the choice of priors (or equivalently, the choice of parameter basis [15]) may be somewhat subjective, one might also want to consider a prior-independent construction, such as the profile likelihood \( \mathcal{L}^p \). Here, instead of integrating over the nuisance directions, one takes the maximum value of \( \mathcal{L} \) for a fixed value of \( \varphi \). Though the profile likelihood does not have a formal probabilistic interpretation, it is often used to construct approximate frequentist confidence intervals based on the likelihood ratio, by identifying the region for which \( \Delta \chi^2_{\text{eff}} = -2 \ln(\mathcal{L}^p(\varphi_{\text{max}}) - \mathcal{L}^p(\varphi)) < 1 \) with the 68% confidence interval. We note that this interval may not have the desired frequentist coverage properties if the profile likelihood is not Gaussian [16].

2.4 Construction of the profile likelihood

We construct the marginalised posteriors from Markov chains generated with a modified version of the public Markov chain Monte Carlo sampler CosmoMC [17], using a conservative Gelman-Rubin convergence criterion [18] of \( R - 1 < 0.01 \), and making sure that the numerical precision settings are sufficient for the data sets considered.

Naïvely, one might think that one could use the same chains to construct the profile likelihood, by binning the data in \( N_{\text{eff}} \) and identifying the best-fitting point in each bin. Unfortunately, this method does not turn out to be suitable for the case at hand. The reason is that the standard Metropolis-Hastings algorithm samples the region near the maximum of the posterior very poorly. In Appendix A we present a rough analytical estimate of the probability of finding at least one sample of a Markov chain within a given \( \Delta \chi^2_{\text{eff}} \) of the best-fit. As shown in the bottom panel of figure 4, for our 10-parameter model one would need of order \( 10^5 \) independent samples to even have a 50% chance of the best-fitting sample to lie within 0.5 of the true best-fit \( \chi^2_{\text{eff}} \). This should be compared to the typically few times \( 10^4 \) correlated samples one usually has in Markov chains used for parameter estimation. The problem is exacerbated by the binning: in particular the estimate of \( \mathcal{L}^p \) for the bins in the tails of the marginalised posterior would be extremely inaccurate.

We therefore employ a different, numerically somewhat more demanding, construction that avoids under-sampling of the tails and is immune to biases introduced by a binning procedure. On a grid of fixed values of \( N_{\text{eff}} \), we estimate the respective maxima of the likelihood by generating Markov chains at temperatures \( T \ll 1 \), with the temperature and length of chains chosen such that \( \ln \mathcal{L}^p \) is estimated with an accuracy of at least 0.1. In addition, we determine the global best-fit by letting \( N_{\text{eff}} \) vary as well.

3 Results

In figure 1 we show the results for WMAP7+ACT data. Firstly, we note that the posteriors differ very little for the different priors, indicating a remarkable robustness of the results to the choice of prior. Secondly, the profile \( \Delta \chi^2_{\text{eff}} \) clearly deviates from the parabolic shape one would expect for a Gaussian profile likelihood, showing an obvious skew towards the large-\( N_{\text{eff}} \) side, so we refrain from mapping it to frequentist confidence limits. Thirdly, \( \mathcal{L}^p \) is markedly shifted (by up to about two thirds of a standard deviation) towards lower values of \( N_{\text{eff}} \) compared to the marginalised posteriors. A similar tendency was also observed in [10], and, more recently, in [11] — however, their results for the same data set (both mode and likelihood ratio-based bounds) differ considerably from ours, possibly due to them attempting
to construct the profile likelihood from Markov chains that were originally generated for the purpose of Bayesian parameter inference. For instance, the individual best-fit estimates of the eight $T = 1$ WMAP+ACT Markov chains (each containing about $3 \times 10^4$ samples) we generated for constructing the marginalised posterior display a considerable spread, with a standard deviation of 0.57 – indicating the unreliability of this method.

Is there an explanation for why larger values of $N_{\text{eff}}$ have a high posterior probability despite apparently not fitting the data too well (and vice versa for smaller $N_{\text{eff}}$)? In [11], it was claimed that the effect, and, by association, also any possible hints for a deviation of $N_{\text{eff}}$ from the standard model expectation, is “driven by prior effects”. This is a very generic statement however; it should be clear that any Bayesian credible intervals are always to some extent prior-dependent. We would like to propose a slightly more specific explanation here, namely that the shift of the marginalised posterior towards larger $N_{\text{eff}}$-enhancement is caused by a volume effect in the marginalisation process.

Let us, for a moment, imagine the full posterior were Gaussian. In that case, marginalisation and profiling would lead to the same result. Also, for a Gaussian posterior, the variance of the other parameters’ marginalised posteriors on slices of constant $N_{\text{eff}}$ would not depend on $N_{\text{eff}}$. If, however, these variances did depend on $N_{\text{eff}}$, and happened to be positively correlated with $N_{\text{eff}}$, then at larger (smaller) $N_{\text{eff}}$ there would be more (less) volume in the nuisance directions, and the marginalised posterior would be enhanced (suppressed) compared to the profile. We shall see that this is indeed the case here, and there is in fact a simple physical argument for why it should be so.

As discussed in [7, 19], $N_{\text{eff}}$ impacts the CMB power spectra in several ways; most

\[ \Delta \chi^2_{\text{eff}} \]

Figure 1. Constraints on $N_{\text{eff}}$ from WMAP7+ACT data. Thin black lines denote the posterior probability density marginalised over the other parameter directions for three different choices of prior (solid: $H_0$, dashed: $\theta_s$, dotted: $\Omega_\Lambda$). The profile likelihood is plotted in thick red lines, both in terms of $L^p/L^p_{\text{max}}$ (solid) and $\Delta \chi^2_{\text{eff}}$ (dotted).
importantly through the redshift of matter-radiation equality

\[ 1 + z_{eq} \equiv \frac{\rho_m}{\rho_r} = \frac{\omega_m}{\omega_\gamma} \frac{1}{1 + \alpha N_{eff}}, \]  

which determines the magnitude of the early integrated Sachs-Wolfe effect. It is actually \( z_{eq} \) (not \( N_{eff} \) or the matter density \( \omega_m \)) that is directly constrained by the CMB [20], and hence essentially uncorrelated with \( \omega_m \) and \( N_{eff} \). Ignoring the tiny uncertainty in the photon energy density \( \omega_\gamma \), the variance of \( \omega_m \) for fixed \( N_{eff} \) is given by

\[ \text{Var}(\omega_m)|_{N_{eff}} \simeq \text{Var}(z_{eq})(\omega_\gamma(1 + \alpha N_{eff}))^2, \]  

and thus the posterior becomes wider in the \( \omega_m \)-direction for larger \( N_{eff} \). Since \( \omega_m \) has degeneracies with other parameters, such as \( H_0 \), the widening is propagated to those directions as well, amplifying the total volume effect. In figure 2 we show the \( N_{eff} \)-dependence of the posterior’s width: using our original Markov chains, we evaluate the variance of the marginalised posterior of \( \omega_m \) on slices of width \( \delta N_{eff} = 1 \). This is compared to the expectation from the measurement of \( z_{eq} = 3180 \pm 129 \), which can easily be calculated from the same chains. The variance on these slices is composed of two components, the intrinsic one of equation (3.2), and a constant piece due to the bin width, given by

\[ \text{Var}_b(\omega_m) = \frac{1}{12} \omega_\gamma^2 z_{eq}^2 \alpha^2 \delta N_{eff}^2. \]  

Their sum is found to be in excellent agreement with the variances from the chains.

The constraints on \( N_{eff} \) can be improved by adding non-CMB data to break some of the parameter degeneracies, and most of the recent hints for \( N_{eff} > 3.046 \) are based on such
combinations of data. As an example, we add the HST-constraint on $H_0$ here, which breaks the $N_{\text{eff}}-H_0$ degeneracy. Our results for WMAP7+ACT+HST data are shown in figure 3. The profile likelihood is closer to Gaussian now, and the magnitude of the volume effect has become much smaller – $L_p$ is shifted by roughly 0.2 with respect to the marginalised posteriors. If we compare this to the posterior standard deviation of $\sim 0.7$, we see that the volume effect by itself cannot account for the observed deviation from the standard model expectation. We summarise our results in table 2.

![Figure 3](image-url)  

**Figure 3.** Same as figure 1, for WMAP7+ACT+HST.

**Table 2.** Summary of constraints on $N_{\text{eff}}$ from different analysis methods. For the marginalised posterior we list the mean $\langle N_{\text{eff}} \rangle$, mode $P_{\text{max}}$, standard deviation $\sigma_{N_{\text{eff}}}$, and the minimal 68%- and 95%-credible intervals [10]. For the profile likelihood, we list the mode $L_{p_{\text{max}}}$ and the intervals in which $\Delta \chi^2_{\text{eff}} \leq 1$ and 4, respectively.

| Analysis   | WMAP7+ACT       | WMAP7+ACT+HST     |
|------------|-----------------|-------------------|
|            | $\langle N_{\text{eff}} \rangle$ | $P_{\text{max}}$ | $\sigma_{N_{\text{eff}}}$ | 68% MCI | 95% MCI | $\langle N_{\text{eff}} \rangle$ | $P_{\text{max}}$ | $\sigma_{N_{\text{eff}}}$ | 68% MCI | 95% MCI |
| $H_0$-prior| 5.78            | 5.68              | 1.45               | 4.18–7.12 | 3.03–8.76 | 4.37      | 4.30     | 0.72            | 3.61–5.03 | 2.96–5.80 |
| $\theta_s$-prior | 5.69      | 5.20              | 1.44               | 4.02–6.92 | 3.01–8.59 | 4.37      | 4.28     | 0.75            | 3.57–5.05 | 2.89–5.86 |
| $\Omega_\Lambda$-prior | 5.67       | 5.20              | 1.46               | 4.05–6.98 | 2.90–8.65 | 4.39      | 4.28     | 0.74            | 3.60–5.08 | 2.98–5.89 |
| Profile    | $L^p_{\text{max}}$ | $\Delta \chi^2_{\text{eff}} \leq 1$ | $\Delta \chi^2_{\text{eff}} \leq 4$ | $L^p_{\text{max}}$ | $\Delta \chi^2_{\text{eff}} \leq 1$ | $\Delta \chi^2_{\text{eff}} \leq 4$ |
|            | 4.73            | 3.29–6.14        | 2.12–8.09         | 4.07      | 3.43–4.76 | 2.79–5.50 |
4 Discussion

We have demonstrated that constraints on the effective number of neutrino species, inferred from CMB data, can be subject to a slight discrepancy between the Bayesian marginalised posterior and the profile likelihood. This can be attributed to a volume effect primarily in the $\omega_m$ direction, caused by the fact that the CMB data are directly sensitive mostly to the redshift of equality, not $N_{\text{eff}}$ itself.

Before we come to an interpretation, let us illuminate the statistical aspect of this result. Regarded from a sampling theory perspective, the mode of the full multi-dimensional posterior can be regarded as an unbiased estimator of the true parameter values (since in the present problem it coincides by construction with the maximum of the likelihood). In the process of marginalisation, this property is lost – the most probable value of $N_{\text{eff}}$ does not provide the best possible fit to the data, or, in other words, the mode of $P(N_{\text{eff}})$ becomes a biased estimator of $N_{\text{eff}}$ (see also [21] for a discussion, or [22] for another applied example). The profile likelihood on the other hand retains the unbiasedness of the mode estimator, but, unlike the marginalised posterior, it is not sensitive to volume effects, and thus does not have a formal statistical interpretation.

In general it should not come as a surprise that, whenever the full posterior/likelihood’s dimensionality is reduced, loss of information will be incurred. Marginalisation and profiling simply preserve different properties of their related multi-dimensional objects, and can thus be a good diagnostic of unusual features. A discrepancy between the two would point to a deviation from Gaussianity, and, from a Bayesian perspective, could for instance indicate that a certain amount of fine-tuning relative to the prior expectation is required in order to optimise the fit to the data.\footnote{We remark that one could, in principle, choose the priors such that the discrepancy would vanish (e.g., here, a flat prior on $z_{\text{eq}}$ instead of $N_{\text{eff}}$ might be a good guess). But with $N_{\text{eff}}$ arguably being a more fundamental quantity than $z_{\text{eq}}$, it is doubtful whether such a choice could be reasonably justified from a theoretical point of view.}

Finally, to evaluate the relevance of this effect, the magnitude of the bias should be set in relation to the intrinsic width of the marginal distribution. For WMAP+ACT data, the difference between profile and posterior is of order two thirds of a standard deviation, thus not contributing the dominant – but certainly a non-negligible – part to the indication for a non-standard $N_{\text{eff}}$. With the addition of HST data, however, the bias is reduced to less than one third of a standard deviation, and a similar trend is to be expected if one added, for instance, large scale structure data, or improved measurements of the CMB damping tail – be it existing ones from the South Pole Telescope [5], or upcoming ones from Planck.

We conclude that the recent indication for a deviation of $N_{\text{eff}}$ from its standard model expectation cannot be accounted for by this statistical effect alone (though the presence of an additional statistical bias introduced, e.g., by the modelling of foregrounds, remains a possibility).

Acknowledgements

The author thanks Steen Hannestad and Yvonne Wong for helpful comments on the manuscript and gratefully acknowledges support from a Feodor Lynen-fellowship of the Alexander von Humboldt Foundation and the use of computing resources from the Danish Center for Scientific Computing (DCSC).
References

[1] D. J. Fixsen, Astrophys. J. 707 (2009) 916-920, [arXiv:0911.1955].
[2] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti, P. D. Serpico, Nucl. Phys. B729 (2005) 221-234, [hep-ph/0506164].
[3] J. Hamann, S. Hannestad, G. G. Raffelt, I. Tamborria and Y. Y. Y. Wong, Phys. Rev. Lett. 105 (2010) 181301, [arXiv:1006.5276].
[4] E. Giusarma, M. Corsi, M. Archidiacono, R. de Putter, A. Melchiorri, O. Mena and S. Pandolfi, Phys. Rev. D83 (2011) 115023, [arXiv:1102.4774].
[5] R. Keisler et al., Astrophys. J. 743 (2011) 28, [arXiv:1105.3182].
[6] T. L. Smith, S. Das and O. Zahn, Phys. Rev. D 85 (2012) 023001, [arXiv:1105.3246].
[7] Z. Hou, R. Keisler, L. Knox, M. Millea, C. Reichardt, [arXiv:1104.2333].
[8] J. Hamann, S. Hannestad, G. G. Raffelt, Y. Y. Y. Wong, JCAP 1109 (2011) 034, [arXiv:1108.4136].
[9] M. Archidiacono, E. Calabrese and A. Melchiorri, Phys. Rev. D 84 (2011) 123008, [arXiv:1109.2767].
[10] J. Hamann, S. Hannestad, G. G. Raffelt, Y. Y. Y. Wong, JCAP 0708 (2007) 021, [arXiv:0705.0440].
[11] A. X. Gonzalez-Morales, R. Poltis, B. D. Sherwin and L. Verde, [arXiv:1106.5052].
[12] E. Komatsu et al. (WMAP Collaboration), Astrophys. J. Suppl. 192 (2011) 18, [arXiv:1001.4538].
[13] J. Dunkley et al., Astrophys. J. 739 (2011) 52, [arXiv:1009.0866].
[14] A. G. Riess et al., Astrophys. J. 730 (2011) 119, [arXiv:1103.2976].
[15] W. Valkenburg, J. Hamann, L. M. Krauss, Phys. Rev. D78 (2008) 063521, [arXiv:0804.3390].
[16] F. Porter, Nucl. Instrum. Meth. A368 (1996) 793-803.
[17] A. Lewis and S. Bridle, Phys. Rev. D 66 (2002) 103511, [astro-ph/0205436].
[18] A. Gelman, D. B. Rubin, Statist. Sci. 7 (1992) 457-472.
[19] S. Bashinsky, U. Seljak, Phys. Rev. D69 (2004) 083002, [astro-ph/0310198].
[20] E. Komatsu et al. (WMAP Collaboration), Astrophys. J. Suppl. 180 (2009) 330-376, [arXiv:0803.0547].
[21] J. Lesgourgues and S. Pastor, Phys. Rept. 429 (2006) 307, [astro-ph/0603494].
[22] R. Stompor, S. M. Leach, F. Stivoli and C. Baccigalupi, Mon. Not. Roy. Astron. Soc. 392 (2009) 216, [arXiv:0804.2645].
A Profiling with Markov chains

In this section we present an estimate of how well the maximum of a probability distribution can be determined by using a Markov chain of length $N$.

Let $\mathcal{P}$ be a probability distribution on an $n$-dimensional parameter space $\mathcal{P}$, and $\varphi \in \mathcal{P}$ be a point in this parameter space. We shall make two simplifying assumptions at this point: first, $\mathcal{P}(\varphi)$ can be approximated by an $n$-variate Gaussian distribution, and second, all the samples in the chain are independent. Without loss of generality one can then take $\mathcal{P}(\varphi)$ to have unit variance and be centered around $\varphi_{\text{max}} = 0$. Define

$$\Delta \chi^2_{\text{eff}}(\varphi) \equiv -2(\ln \mathcal{P}(\varphi_{\text{max}}) - \ln \mathcal{P}(\varphi)), \quad (A.1)$$

and the volume fraction $f_x$ of $\mathcal{P}$ for which $\Delta \chi^2_{\text{eff}}(\varphi) < x$,

$$f_x = \int_{V_x} d\varphi \mathcal{P}(\varphi), \quad (A.2)$$

with the volume $V_x$ implicitly given by the condition $\varphi \in V_x \iff \chi^2_{\text{eff}}(\varphi) < x$. If one expresses $\varphi$ in spherical coordinates, it can easily be shown that

$$f_x(n) = \int_0^{\sqrt{x}} dr \, \frac{1}{(\sqrt{2\pi})^{n/2}} \exp \left[ -\frac{r^2}{2} \right] \frac{r^{n-1}(2\pi)^{n/2}}{\Gamma(\frac{n}{2})}, \quad (A.3)$$

where $\Gamma$ is the Gamma function. If, instead of sampling from $\mathcal{P}$, one generates the Markov chain with a temperature parameter $T$ by sampling from $\mathcal{P}^{1/T}$, equation (A.3) can be generalised to

$$f_x(n, T) = \int_0^{\sqrt{T}} dr \, \frac{T^{-n/2}}{(\sqrt{2\pi})^{n/2}} \left( \exp \left[ -\frac{r^2}{2} \right] \right)^{1/T} \frac{r^{n-1}(2\pi)^{n/2}}{\Gamma(\frac{n}{2})}. \quad (A.4)$$

If all $N$ samples of the chain are independent, then the probability $\bar{p}$ that none of the points lie within $f_x$ is given by

$$\bar{p}(x, n, T, N) = (1 - f_x(n, T))^N. \quad (A.5)$$

It follows trivially that the probability of at least one point of the chain being within $\Delta \chi^2_{\text{eff}} = x$ of $\chi^2_{\text{eff}}(\varphi_{\text{max}})$ is $p(x, n, T, N) \equiv 1 - \bar{p}(x, n, T, N)$. For a few selected slices in $(n, T, N)$-space, $p(x, n, T, N)$ is plotted in figure 4.
Figure 4. Probability of finding at least one sample within $\Delta \chi^2_{\text{eff}} = x$ of the true maximum of the $n$-dimensional Gaussian posterior $P$, if the Markov chain was generated at a temperature $T$ and contains $N$ independent samples. Top left: dependence on $N$ and $n$ for $T = 1$ and $x = 0.2$. Top right: dependence on $N$ and $T$ for $n = 10$ and $x = 0.2$. Bottom: dependence on $N$ and $x$ for $T = 1$ and $n = 10.$