Random template placement and prior information

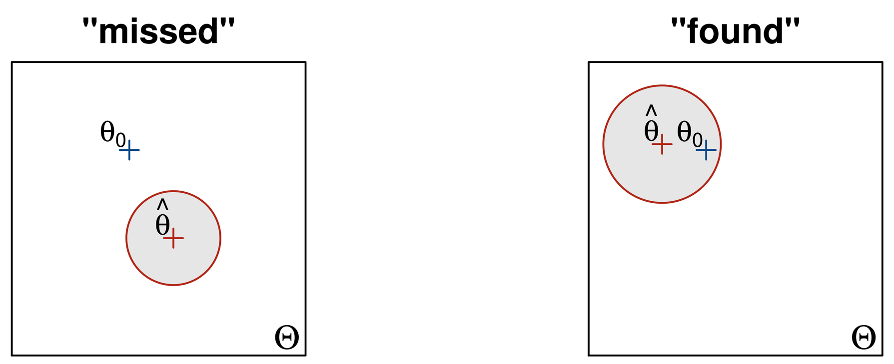
Christian Röver

Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Hannover, Germany

In signal detection and parameter estimation applications one is often faced with random template placement problems. Problems of this kind include e.g. the setup of template banks, and also Monte Carlo implementations like parallel tempering or importance resampling. Sampling strategies are then commonly set up based on figures like minimum overlap, or SNR loss. With prior information available, template placement may be formulated as a decision-theoretic problem, and optimal sampling strategies may be derived. In particular, random template banks (as in [1]) then turn out to contitute special cases of Minimax strategies.

Template placement

Suppose an unknown signal parameter $\theta_0 \in \Theta$ is to be determined. One may '**guess**' the true parameter value to be, say, $\hat{\theta}$. (This is the template placement.) Then the signal is *found* if the guess $\hat{\theta}$ is "sufficiently close" to θ_0, and *missed* otherwise:



The template's "radius" in the above sketch (defining what exactly is "sufficiently close", i.e, defining the "volume" of the above template) is defined through the **parameter space metric**, a function of θ. The metric again is usually defined through considerations of *minimum overlap* or *SNR loss*.

Random templates

Randomly placed templates are for example utilised in setting up *random template banks*. Templates $\hat{\theta}$ are (repeatedly, independently) drawn from a distribution with density $p^\star(\theta)$. The probability of success then is a function of the true value θ_0:

$$\mathrm{P}(\text{"found"}) \;\propto\; h(\theta_0)\, p^\star(\theta_0)$$

where $h(\theta)$ corresponds to the volume covered by a template placed at θ (h is the inverse template metric determinant). A common approach is to set $p^\star(\theta) \propto \big(h(\theta)\big)^{-1}$, so that the "missing probability" is constant across Θ [1].

Prior information

The above probability of a randomly placed template actually finding the true signal in general depends on the true signal location, θ_0. That location of course is unknown, but a priori information may be available in the form of a **prior probability** distribution $\mathrm{P}(\theta)$.

The prior describes where in paramater space the true signal is actually likely to be.

The problem

The placement of templates is usually associated with (computational) costs. The question is how to place templates, i.e., to decide on the template distribution $p^\star$. Due to the two **competing factors** of prior and template metric one may be tempted to choose suboptimal coverings of Θ, e.g.:

- overly dense covering of regions that are unlikely to contain the signal
- dense sampling of likely regions that are easily covered by few templates

Decision-theoretic approach

The problem is of a decision theoretic / game-theoretic nature ("*where should I invest templates in order to maximize my gain?*"). The **loss** to be minimized here is related to the number of templates. Restricting for now to independently, identically sampled templates, a **strategy** consists in a choice of $p^\star$. The **risk** associated with a strategy is the expected loss for a given true value θ_0 and a sampling strategy $p^\star$. The **Bayes risk** is the expected risk with respect to a prior distribution $\mathrm{P}(\theta)$. In order to find an optimal $p^\star$, one needs to properly define the risk and then minimize it [2].

Optimality criteria

The *loss* incurred in any instance may e.g. be defined as:

- the waiting time T until the true value is found
- a constant loss in case value is not found within N trials
- …

Minimizing the expected waiting time

For given θ_0 and repeated independent sampling from $p^\star$, the probability of success remains constant, and the number of templates (or 'waiting time') T until the true θ_0 is found follows a geometric distribution with

$$\mathrm{P}(T=t|\theta_0) \;=\; \big(1 - c\,h(\theta_0)\,p^\star(\theta_0)\big)^{t-1}\big(c\,h(\theta_0)\,p^\star(\theta_0)\big)$$

for some $c \in \mathbb{R}^+$. The **risk** in this case is:

$$\mathrm{E}[T|\theta_0] \;=\; \frac{1}{c\,h(\theta_0)\,p^\star(\theta_0)}$$

For a given prior with density $p(\theta)$, the prior expected risk (**Bayes risk**) is:

$$\mathrm{E}[T] \;=\; \frac{1}{c}\int_\Theta \frac{p(\theta_0)}{h(\theta_0)\,p^\star(\theta_0)}\,\mathrm{d}\theta_0$$

which is minimized by choosing the **Bayes rule**

$$p^\star(\theta) \;\propto\; \sqrt{\frac{p(\theta)}{h(\theta)}}.$$

The Minimax rule

Instead of minimizing the *expected* risk, one might instead wish to minimize the **worst-case risk**. This leads to a sampling density

$$p^\star(\theta) \;\propto\; \big(h(\theta)\big)^{-1}$$

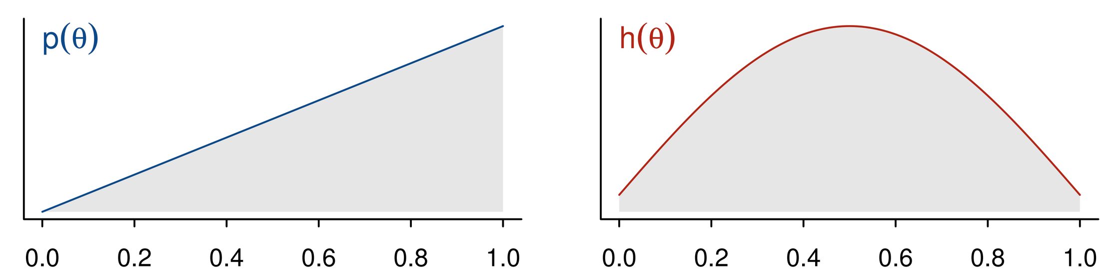
which makes the probability of "finding" the true signal a constant, independent of the true value θ_0. This **Minimax rule** again also constitutes the Bayes rule for a particular prior ($p(\theta) \propto \big(h(\theta)\big)^{-1}$).

Toy example 1

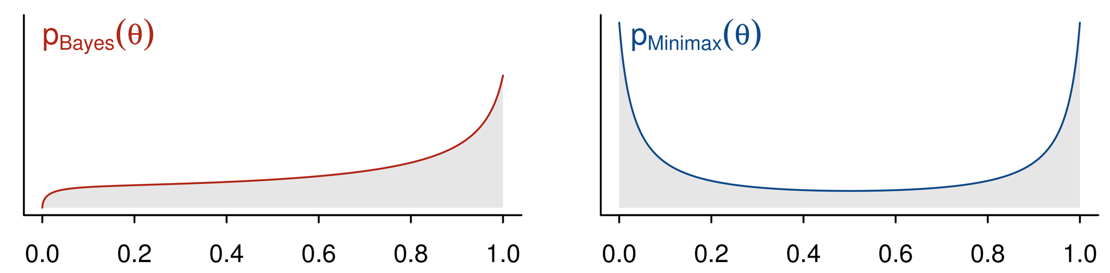
Consider the simple case of a '*flat*' template metric ($h(\theta) \propto 1$), and a Gaussian prior distribution with mean μ and variance σ^2 ($p(\theta) \propto \exp\big(-\frac{(\theta-\mu)^2}{2\sigma^2}\big)$). In this case, the Bayes rule would again be a Gaussian $p^\star$ with mean μ and an inflated variance of $2\sigma^2$. The Minimax rule on the other hand does not exist.

Toy example 2

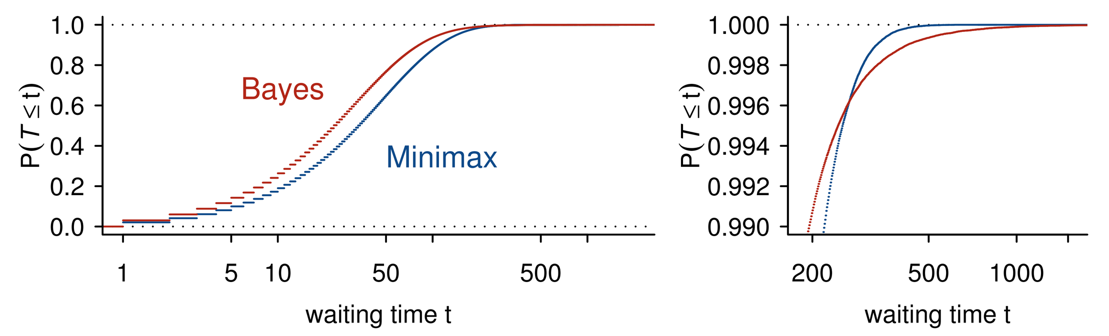
Consider the case of $\Theta = [0,1]$, where prior and template metric are given by:



so the prior suggests $\theta \approx 1$ to be most likely, while the template metric demands a denser sampling at both $\theta \approx 0$ and $\theta \approx 1$. Then the densities corresponding to **Bayes** and **Minimax** strategies look like:



The performances of the different strategies may be compared by a Monte Carlo simulation—by repeated drawing of 'true' parameters from the prior p and then trying to recover it using either the *Bayes* or *Minimax* strategies. The following plots illustrate the cumulative distributions of the time it takes until the true signal is found:



The signal is usually found faster by the Bayes rule (left panel), while—not surprisingly—the Minimax rule shows a better worst-case behaviour (right panel). The mean numbers of trials used are 36 for the Bayes rule, and 48 for the Minimax rule.

In practice, the discrepancies may be arbitrarily large, up to the case where the Minimax rule does not exist, as in the previous example.

Parallel tempering example

Parallel tempering is a Markov chain Monte Carlo (MCMC) technique used for parameter estimation and closely related to *simulated annealing* methods. A '*temperature*' parameter $T \geq 1$ is introduced to smoothen the likelihood surface and avoid getting trapped in local optima [3]. Instead of considering the "plain" posterior density function $p(\theta|y)$, the *tempered* version may be set up as

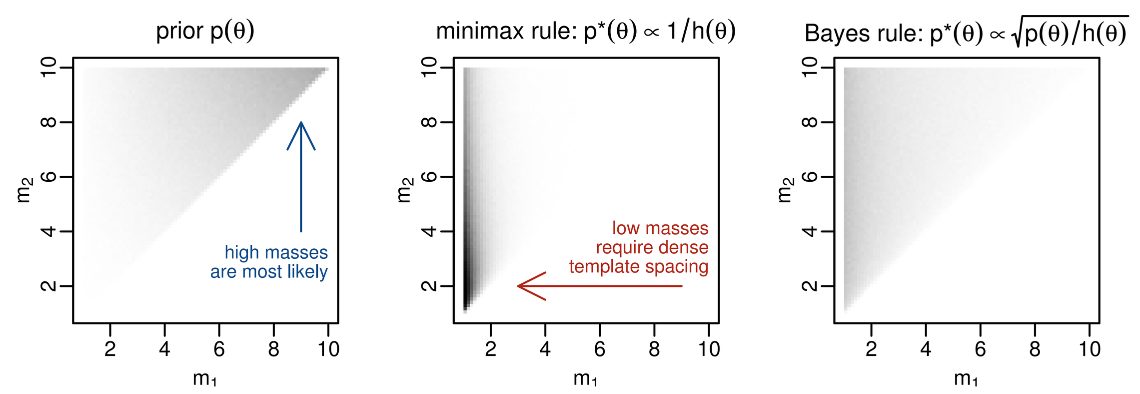
$$f_{(T)}(\theta) \;=\; p^\star(\theta)^{1-\frac{1}{T}}\,p(\theta|y)^{\frac{1}{T}}$$

which allows to adjust between the two density functions $p(\theta|y)$ and $p^\star(\theta)$ by increasing the temperature T. An increase in T is supposed to enhance the stochastic search capabilities of the MCMC sampler, and so the limiting distribution $p^\star$ should **optimize the chances of convergence**, i.e., of finding the true parameter value θ_0.

In a parallel tempering algorithm one then basically runs several MCMC chains at increasing temperatures in parallel, where only the first one ($T=1$) is used for Monte Carlo integration, while the additional chains ($T>1$) are supposed to improve the algorithm's performance (mixing and convergence).

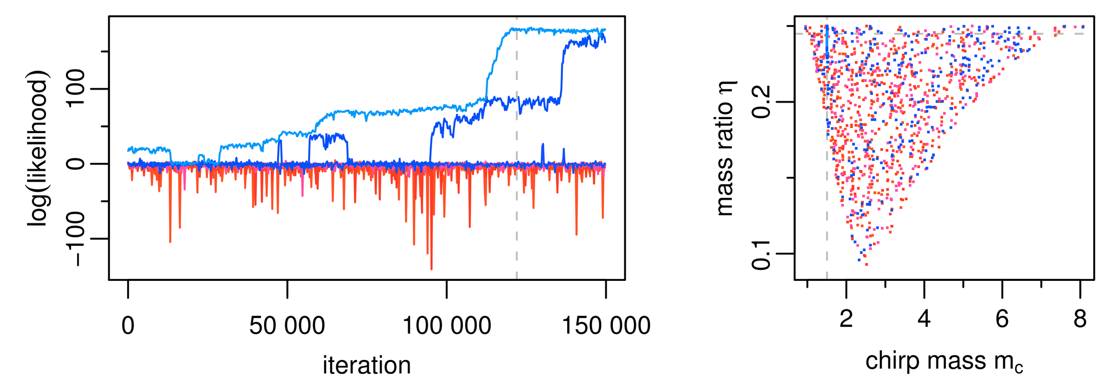
Binary inspiral context

In the context of binary inspiral signals, prior and template metric are particularly **contradictory**: templates need to be spaced densely at low masses (these are the long-lived signals of many cycles), while *a priori* one rather expects high masses (which are the high-amplitude signals that are detectible out to large distances) [3].



The optimized strategy (right panel) naturally leads to a compromise between the two extremes.

The following plots illustrate actual MCMC runs for parameter estimation on simulated data using the above sampling scheme. The low-temperature chains converge to the true values yielding large likelihoods (left panel, blue lines). At the same time, high-temperature chains sample from the complete prior range (reddish dots, right panel).



Summary

In the parallel tempering application, the decision theoretic approach leads to an optimal solution to the problem of choosing a limiting distribution $p^\star$.

An analogous procedure should be helpful in related problems, e.g. in setting up (random or deterministic) template banks for searches when prior information is available. An obvious consequence is that if one were to place a *single* template, one would choose the location θ^* in parameter space where the template covers the greatest prior probability, i.e., where $h(\theta^*) \times p(\theta^*)$ is largest.

A closely related question then is where to place templates given only limited (computational) resources, which should allow to optimally draw bounds on the parameter space (instead of imposing arbitrary constraints). When resources are not an issue, one might at least want to order templates by their chances of success. Either way, the concept of a minimax strategy (independent of a prior distribution) may also be of interest, in particular the question of its existence, or its performance in comparison to a corresponding Bayes rule.

References

[1] C. Messenger, R. Prix, and M. A. Papa. Random template banks and relaxed lattice coverings. *Physical Review D*, 79(10):104017, May 2009.

[2] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, 2nd edition, 1985.

[3] C. Röver. *Bayesian inference on astrophysical binary inspirals based on gravitational-wave measurements*. PhD thesis, The University of Auckland, 2007. URL http://hdl.handle.net/2292/2356.

LIGO-G0900562-v3