Date DA(

Data analysis (Binary systems)

For binary systems we can model shape of the signal.
If the signal weak we can search this shape in the noise → (согласованный фильтр) matched filter. The shape (signal) which we se is called template (шаблон).

Weak signal ⇒ we have two hypothesis ① data = noise ② data = noise + signal H_1

→ model testing (hypothesis testing)

- frequentist approach : signal is deterministic & noise corrupts the signal.

Neyman - Pearson criterion ; maximization of detection statistic for a chosen significance of the test.

Likelihood ratio (функционал правдоподобия) - most powerful detection statist in Gaussian noise

- we fix acceptable false alarm probability

$P(y \geq y_{threshold} \mid H_0) = \alpha$ - probability of detection stat in H_0 (data = noise) to be above threshold $y_{threshold} = \alpha$.

We choose α → determine y_{thr}. →

$$y = \frac{P(d \mid H_1, \theta_i)}{P(d \mid H_0, \theta_i)} \leftarrow \text{likelihood}$$

detection probability $P(y \geq y_{th} \mid H_1)$ assigned to observed y

$y < y_{thr}$ → not a detection ; $y \geq y_{thr}$ → we have significan of detection $P(y \geq y_{thr} \mid H_1)$ given false alarm probability α (say 1% or smaller). → maximum likelyhood estimation of parameters

- Bayesian approach : We have to assign probability to our hypothesis before looking at the data → prior probability π_0, We treat parameters of the signal as random variables & try to estimate the probability density function for parameter based on observation

Date DRAFT

$$P(H_i|d) = P(d|H_i)\,\frac{\pi(H_i)}{P(d)}$$

Both approaches are used in DA. Let us build likelyhood f-n

- Matched filter : search for signal of known form.
  - assume the we know the form of the signal exactly but we do not know the arrival time : correlation of the signal with the data

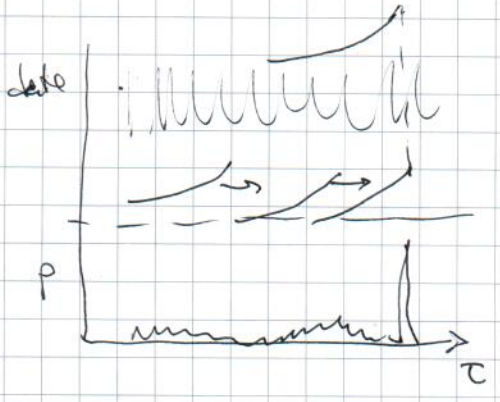
$$\rho(t) = 4\,\mathrm{Re}\int \frac{\tilde{d}(f)\,\tilde{h}^*(f)}{S_n(f)}\,e^{-2\pi i f t}\,df$$

$\tilde{d}(f)$ → data in Fourier domain

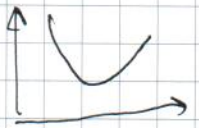
$\tilde{h}^*(f)$ → template in Fourier do (c.c.)

$$\tilde{I}(f) = \int d(t)\,e^{2\pi i f t}\,dt$$

τ → time shift



But! we detector is not equally sensitive @ all frequencies



if signa has very low or very high frequencies we will "see" them with our detector unless signal has very strong amplitude

↳ we use $S_n(f)$ → noise power spectral density as weights at different frequencies!

$$\mathrm{SNR}^2(t) \text{ (matched filter)} = 4\,\mathrm{Re}\int_0^{f_{max}} \frac{\tilde{s}(f)\,\tilde{h}^*(f)}{S_n(f)}\,e^{-2\pi i f \tau}\,df$$

f_{max} ← determined by detector

$e^{-2\pi i f \tau}$ allows search over time (shift)

"2" $S_n(f)$ → one-sided PSD (for positive freq. only)

"2" → $\int \frac{\tilde{d}(f)\tilde{h}^*(f) + \tilde{d}^*(f)\tilde{h}(f)}{S_n(f)}\,df$ (to have real $e^{-i2\pi ft}$ is absorbed in $h(t)$.

Date DA3

In practice the signal depends on parameters $\{\theta_i\}$ which we do not know à priori. Noise might also depend on parameters!

Assume that we know the noise & it has Gaussian distribution & noise has zero mean! $\langle \text{noise} \rangle = \langle n \rangle = 0$

$\langle\ \rangle$ over noise realizations or assuming ergodicity $\langle\ \rangle_t$ over time

data ← noise, signal ← parameters

$$d(t) = n(t) + s(t, \vec{\theta}_t)$$

we compute $h(\vec{\theta}_1)$, $h(\vec{\theta}_2)$, $h(\vec{\theta}_3)$, ...

$d(t) - h(\vec{\theta}_i)$ → residuals if $h(\vec{\theta}_t) = s(t, \vec{\theta}_t)$ → $d(t) - h(t, \vec{\theta}_t) = n(t)$

otherwise $d(t) - h(t, \vec{\theta}_i) = n + \left[s(t, \theta_t) - h(t, \vec{\theta}_i + \vec{\theta}_t) \right] \neq n(t)$ ($\neq 0$)

↳ "deformed signal"

Likelihood of n → maximum likelihood → the residuals resemble the Gaussian noise

$$P(n) = \prod_i^N p(n(t_i) \mid \mathcal{H}_0(\text{white, Gaussian})) = \underbrace{\frac{1}{(\sigma\sqrt{2\pi})^N}}\ e^{-\frac{1}{2\sigma^2}\sum_i n(t_i)^2}$$

Gaussian distribution with zero mean & σ-variance

Non white → different variance @ different frequencies

$$P(\vec{d} = n \mid \mathcal{H}_0) \sim e^{-\frac{1}{2}(n|n)} = e^{-\frac{1}{2}(d|d)} \quad \text{generalization}$$

$$(a|b) = 2\int \frac{\tilde{a}(f)\,\tilde{b}^*(f) + \tilde{a}^*(f)\,\tilde{b}(f)}{S_n(f)} \leftarrow \text{non-white}$$

likelihood $p(d \mid \mathcal{H}_1, \theta_i) \sim e^{-\frac{1}{2}(d - h(\theta_i) \mid d - h(\theta_i))}$

if $h(\theta_i) = s$ (signal)

Λ = likelihood ratio $\dfrac{p(d \mid \mathcal{H}_1, \theta_i)}{p(d \mid \mathcal{H}_0)} \sim e^{+\frac{1}{2}(d \mid h(\theta_i)) - \frac{1}{2}(h(\theta_i) \mid h}$

(frequentist approach)

Date DAY

if $d = n$ (H_0)

$\langle \log \mathcal{L} \rangle = \langle d \mid h(\theta_i) \rangle - \frac{1}{2} \langle h(\theta_i) \mid h(\theta_i) \rangle \overset{H_0}{=} \overbrace{\langle n \mid h(\theta_i) \rangle}^{=0} - \frac{1}{2} \langle h \mid$

$= -\frac{1}{2} \langle h \mid h \rangle < 0$

if $d = n + s$

$\langle \log \mathcal{L} \rangle = \langle s \mid h(\theta_i) \rangle - \frac{1}{2} \langle h(\theta_i) \mid h(\theta_i) \rangle$

$\langle \log \mathcal{L} \rangle_{\max \theta_i} = \left[\text{when } \theta_i = \theta_{tr} \ \& \ s \approx h(\theta_{tr}) \right] = \frac{1}{2} \langle s \mid s \rangle = \frac{1}{2} \langle h(\theta_{tr}) \mid h$

> 0

$2 \langle \log \mathcal{L} \rangle_{\max \theta_i} = SNR^2$

- maximization of the likelyhood ratio (or likelihood) → $\{\hat{\theta}\}$ → ma
  likelihood estimators of parameters → unbiased : if averaged over
  noise realizations $\langle \{\hat{\theta}\} \rangle = \{\theta_{tr}\}$
- for a given noise realization $\{\hat{\theta}_i\} \neq \{\theta_{tr}\}$ - estimator →
  how close $\hat{\theta}$ to θ_{tr} → depends on the strength of the signa
  the stronger signal → less influence of the noise → $\hat{\theta}$ closer t
- If $s \neq h(\theta_{tr})$ → systematic error in modelling GW signa
  $s \approx h(\theta_{tr})$ → $\min_{\theta_i} | (s - h(\theta_i) \mid s - h(\theta_i)) \rightarrow \tilde{\theta}$

  $|\theta_{tr} - \tilde{\theta}|$ → bias in parameter estimation : we still
might detect the signal s but with bias in parameters (trying
to correct mismodelling by adjusting parameters) effectualness

faithfulness : $\dfrac{(s \mid h(\theta_{tr}))}{\sqrt{(s \mid s)\,(h(\theta_{tr}) \mid h(\theta_{tr}))}}$ - overlap $[-1, 1]$

1 - perfect match
0 - very bad
$-1 \rightarrow h = -s$

- maximisation of likelyhood - optimization problem.
  PSO, GA, grid-based methods

Date DA5

Consider only grid based: We want to cover parameter space $\{\Theta\}$ by grid points at equal distance between them. Grid should be not very coarse (loose signal) or bad parameter estimation or should be not very fine (computationally expensive).

Distance → we can see θ_i - as coordinates on the parameter space → introduce interval & metric

$$ds^2 = \| \hat{h}(\theta_i + d\theta_i) - \hat{h}(\theta_i) \| \approx \text{assume } d\theta_i \text{ is infinitesimal} \approx$$

$$= \left(\hat{h}(\theta_i + d\theta_i) - \hat{h}(\theta_i) \,\middle|\, \hat{h}(\theta_i + d\theta_i) - \hat{h}(\theta_i) \right) \approx \left(\frac{\partial \hat{h}}{\partial \theta_i} \,\middle|\, \frac{\partial \hat{h}}{\partial \theta_j} \right) d\theta_i$$

$$= g_{ij}\, dx^i dx^j$$

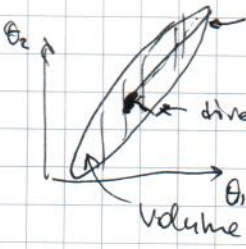
↓ metric on the parameter space !

$$\gamma_{ij} = \left(\frac{\partial \hat{h}}{\partial \theta_i} \,\middle|\, \frac{\partial \hat{h}}{\partial \theta_j} \right)$$

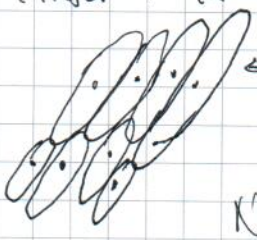
$\hat{h}$: $(\hat{h}|\hat{h}) = 1$ (normalized templates)

ds^2 → tells us how similar (correlated) two nearby templates in parameter sp

$ds^2(d\theta^i)$: $= \text{const} = 0.01$ ← sets desired level of correlation between two nearby templates



→ we want grid to be uniform in this distance



← ellipses are f-ns of central point or in geometrical language $\gamma_{ij} = \gamma_{ij}(\theta_i)$

Non trivial problem in $N_{dim} > 2$

Need coordinates (parametrization = combination of physical parameters) such that coordinates are close to Cartesian:

m_1 & m_2 → really bad, like $\frac{m_1}{m_2}$ → better → ...

Date DA6

Assume we have solved placement problem ⇒ ($\{\theta_i\}$ set of grid points in parameter space) → generate GW signal @ each point.

$(1-ds) = 0.03 \rightarrow \langle \hat{h}(\theta_i) | \hat{h}(\theta_i + \delta\theta_i) \rangle = 0.97 \rightarrow$ < 10% loss in event rate

→ template bank → compute likelihood ~ $(d|h(\theta_i)) - \frac{1}{2}\langle h(\theta_i)|h($ at each point → find maximum over $\{\theta_i\}$

→ low latency strategy for searching GW signal in LIGO/Virgo & estimation of significance of a candidate.

- Bayesian analysis: expensive computationally → often used when we know that H_1 is true. The power of this method → allows testing several models (different signal models, non-GR features

Consider several models M_i each parametrized by own set of parameters $\vec{\theta}_i$ (vector of parameters for i-th model)

$$P(M_i|d) = \frac{P(d|M_i)\,\pi(M_i)}{P(d)}$$

& for a given model the posterior pdf:

$$\text{posterior} \rightarrow P(\vec{\theta}_i|d,M_i) = \frac{\overbrace{P(d|\vec{\theta}_i,M_i)}^{\text{likelihood}}\,\pi(\theta_i)}{P(d|M_i)} \leftarrow \text{prior on parameters}$$

$P(d|M_i) \leftarrow$ evidence of model M_i

$$P(d|M_i) = \int d\vec{\theta}_i\; P(d|\vec{\theta}_i,M_i)\,\pi(\vec{\theta}_i) \quad \Big| \text{ normalization of pdf } \int P(\theta_i)\,d\vec{\theta}_i =$$

then $P(M_i|d) = \left[\int d\vec{\theta}_i\; P(d|\vec{\theta}_i,M_i)\,\pi(\vec{\theta}_i)\right] \frac{\pi(M_i)}{P(d)}$ ← probability of model M_i

Bayes problem is to evaluate $P(d)$ & rare cases when we have limited number of models that are mutually exclusive $P(d) = \sum_i P(d|M_i)\,\pi(M_i)$ → hardly ever possib

⇒ usually consider posterior odd ratios →

$$O_{ij}(d) = \frac{P(M_i|d)}{P(M_j|d)} = \underbrace{\left[\frac{\int d\vec{\theta}_i\, P(d|\vec{\theta}_i, M_i)\,\pi(\vec{\theta}_i)}{\int d\vec{\theta}_j\, P(d|\vec{\theta}_j, M_j)\,\pi(\vec{\theta}_j)}\right]}_{\text{Bayes factor}} \underbrace{\frac{\pi(M_i)}{\pi(M_j)}}_{\text{prior odds}}$$

If we do not have preference of one model over other → we assign equal π - priors to models → Bays factor tells us if the data prefers one or another model.

The problem is to ~~def~~ set the threshold on th Bayes factor besides common sence : $>10-20$ → likely >100 strong evidence >1000 definite, few - wat conclusi

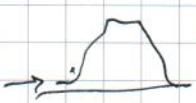
In reality (weak signal, small ~~defn~~ deviations from GR) = we do not expect very high Bayes factor : often used frequentist approach to set the threshold.

→ Problem to evaluate $p(d|M_i) = \int p(d|\vec{\theta}, M)\,\pi(\vec{\theta}|M)\,d\vec{\theta}$

→ multidimensional integral

& ~~partial~~ another problem to get the posterior :

$$p(\vec{\theta}|d, M_i) = \frac{p(d|\vec{\theta}, M)\,\pi(\vec{\theta}|M)}{p(d|M)}$$



⇒ pdf for all parameters → can marginalize fo each parameter & say about mean, median, confide interval, can compare with prior ⇒ how informative were observations (did they update our prior knowle

How to ~~solve~~ get $p(\vec{\theta}|d, M)$ & $p(d|M)$

Date DA8

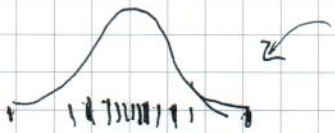
P(dicll) - evaluation of multidimensional integrals

- Vegas algorith → monte-carlo integration

$$I = \int_V f(\vec{x})\, d\vec{x} \approx \frac{V}{N} \sum_{i=1}^{N} f(\vec{x}_i)$$

$$\sigma_I^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(f(x_i) - \bar{f}\right)^2$$

divide parameter space into equal hypercubes, ~~then~~ choose ran points in each hypercube (box), evaluate integral, resize cubes according to density $g = \frac{|f|}{I(|f|)}$: make boxes smaller (denser) in the parts of parameter space where we have largest contrib to the integral. & less densed (larger) ~~other~~ in other parts



- Markov chain Monte Carlo

We construct Markov chain : stochastic process where the ne ~~point~~ point $\vec{\theta}_{k+1}$ in the chain depends only on previous point $\vec{\theta}_k$. (i) We want chain to move towards the region of parameter space with high likelihood

(ii) We ~~want~~ need to introduce transitional probability $P(\vec{\theta}_{k+1} | \vec{\theta}_k)$ → way of moving chain ~~from~~ $\vec{\theta}_k$ to $\vec{\theta}_{k+1}$. (iii) If transitional probability satisfies the ballance equation:

$$\Lambda(\vec{\theta}_k)\, P(\vec{\theta}_{k+1} | \vec{\theta}_k) = \Lambda(\vec{\theta}_{k+1})\, P(\vec{\theta}_k | \vec{\theta}_{k+1})$$ then after some "burn-in" length chain samples from the distribution $\Lambda(\vec{\theta}_k)$

Date DA9

Consider a particular implementation (Metropolis & Hastings) → particular way of building the transitional probability which satisfies balance equation. (i) introduce a proposal distribution $q(\vec{\theta}_{k+1}|\vec{\theta}_k)$ (could be arbitrary) & (ii) introduce the acceptance probability:

M-H ratio

$$\alpha(\vec{\theta}_{k+1}|\vec{\theta}_k) = \min\left\{1, \frac{p(d|\vec{\theta}_{k+1})}{p(d|\vec{\theta}_k)} \frac{q(\vec{\theta}_k|\vec{\theta}_{k+1})}{q(\vec{\theta}_{k+1}|\vec{\theta}_k)} \frac{\pi(\vec{\theta}_{k+1})}{\pi(\vec{\theta}_k)}\right\}$$

likelihood ratio; ratio of priors

- Easier to understand if we assume uniform priors & use symmetric proposal distribution ($q(\theta_k|\theta_{k+1}) = q(\theta_{k+1}|\theta_k)$ – Metropolis ratio) →

$$\tilde{\alpha}(\vec{\theta}_{k+1}|\vec{\theta}_k) = \min\left\{1, \frac{p(d|\vec{\theta}_{k+1})}{p(d|\vec{\theta}_k)}\right\}$$

α → probability of acceptance $\vec{\theta}_{k+1}$: if $p(d|\vec{\theta}_{k+1}) > p(d|\vec{\theta}_k$

point is always accepted, if $p(d|\vec{\theta}_{k+1}) < p(d|\vec{\theta}_k$ point is accepted with probability α (likelihood ratio) → (draw random number $x \in U(0,1)$ accept in $\alpha \geq x$) ⇒ chain moves towards the maximum likelihood, but (!) it can also go down the slope.

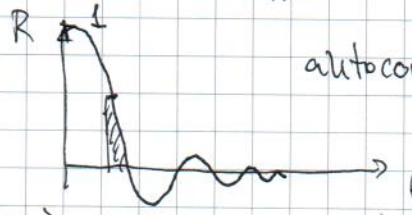
However The theorem tells us that the chain will sample from posterior pdf (after some burn-in length) independ of the proposal distribution $q(\vec{\theta}_{k+1}|\vec{\theta}_k)$ however

① the efficiency of the sampling very strongly depends on the proposal: ideally proposal is similar / close to the posterior, if we know multimodality → take it into proposal. proposal can be as complex as you want

Date DA 10

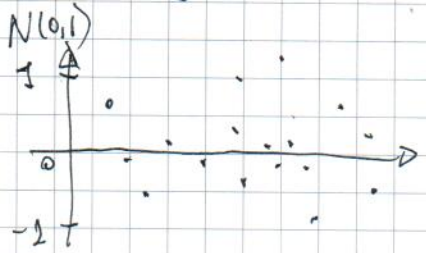
# of independent samples (uncorrelated) : defined by auto correlation length of the chain

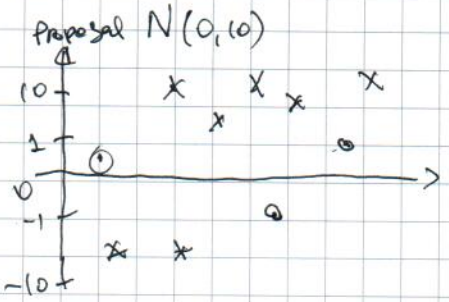
$$R_{xx}(l) = \sum_n x(n)\, x^*(n-l) \qquad \int_{-\infty}^{\infty} f(u+\tau)\, f^*(u)\, du = \int f(u)\, f^*(u-\tau)$$



autocorrelation length → not unique definition

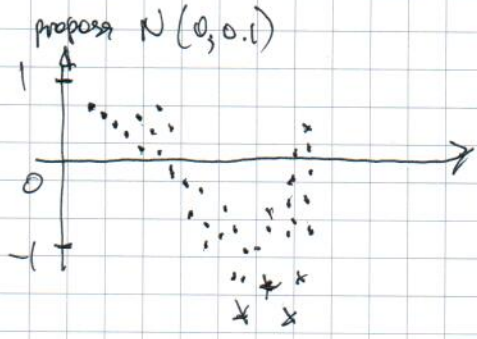
$R = \frac{1}{2}$; $R = 0$; $R = \frac{1}{e}$





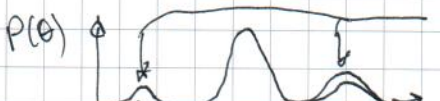
* rejected
⊙ accepted

low acceptance rate



high acceptance but highly correlated chain (large autocorrelation length)
low computational efficiency

② Multimodal posterior distribution



local maxima ← chain can get stuck there, it will move & find the global/all maxima if we wait >100 years

– Simulated annealing

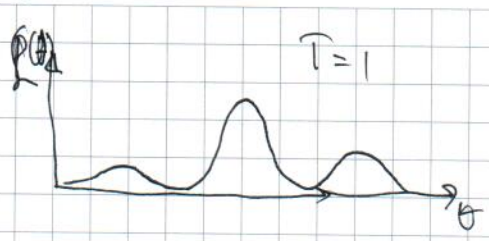
likelihood → $\int \sim e^{\langle d|h(\theta)\rangle - \frac{1}{2}\langle h|h\rangle}$ ← similar to Maxwell distribution

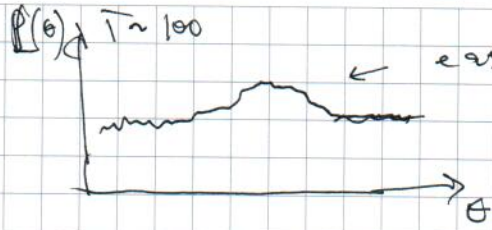
$f(E) \sim \exp\left[-\frac{E}{kT}\right]$ $\quad f(v) \sim e^{-\frac{mv^2}{2kT}}$

introduce "temperature" $e^{\frac{\langle d|h(\theta)\rangle - \frac{1}{2}\langle h|h\rangle}{T}}$ ← "E"

$T = 1$ → likelihood we introduce $T > 1$ → "heating the likelihood" → makes likelihood surface smoother by increasing the noise level

Date DA11





← easy to find real maximum

simulated annealing: we start with high temperature → can explore large part of parameter space (random walk) until we find the largest maximum in the likelihood → temperature is gradually decreases to "1".

- parallel tempering

we run multiple chains (always good) ~~with~~ in parallel with different temperature T_i & make chains "talk" to each other. We can explore parameter space simultaneously @ different scales (high temper. → large scale exploration, mid temp → mid scale, low temper → vicinity of maxima).

We perform cross-talk between chains by exchanging the points $\vec{\theta}_k^{T_i} \leftrightarrow \vec{\theta}_k^{T_j}$ (or equivalently swaps the temperature) with probability

$$p = \min\left\{1, e^{(E_i - E_j)\left(\frac{1}{T_i} - \frac{1}{T_j}\right)}\right\}$$

$$= \frac{\exp\left(-\frac{E_j}{T_i} - \frac{E_i}{T_j}\right)}{\exp\left(-\frac{E_i}{T_i} - \frac{E_j}{T_j}\right)}$$

& we define the frequency of cross-talk (how often we ~~swap~~ attemp to swap)

In addition parallel tempering allows us to evaluate the evi → introduce $\beta = \frac{1}{T}$, The evidence corresponding to each chain with the temperature T is

$$Z(\beta) = \int d\vec{\theta}\, L(\vec{\theta}, T)\, \pi(\vec{\theta})$$ (T assumed to be continuous variable

We are interested in chain with $T=1$ (true likelihood!)

Date DAY2

→ $Z(1)$ ← evidence we are interested in. Take derivative!

$$\frac{d \log Z(\beta)}{d\beta} = \frac{1}{Z(\beta)} \int \frac{\partial}{\partial \beta} \mathcal{L}(\theta, \beta)\, \pi(\theta) d\theta$$

$$\rightarrow \log \mathcal{L}(\theta, T=1) \cdot \mathcal{L}(\theta\beta)$$

$$\langle f(x) \rangle = \frac{\int f(x)\, p(x) dx}{\int p(x) dx} \leftarrow \text{average over pdf } p(x)$$

$\int p(x) dx$ — normalization

$$\frac{d \log Z(\beta)}{d\beta} = \left\langle \log \mathcal{L}(\theta, T=1) \right\rangle_\beta$$ ← averaging over posterior at temperature $T = \frac{1}{\beta}$

Integrating $\int_0^1 \frac{d \log Z(\beta)}{d\beta} d\beta = \log Z(\beta) \Big|_0^1$; $Z(0) = \int \pi(\vec{\theta})\, d\vec{\theta} = 1$

$$\rightarrow \log Z(1) = \int_0^1 \left\langle \log \mathcal{L}(\theta, T=1) \right\rangle_\beta d\beta$$ ← can be evaluated numerically if we have properly choosen Temperature ladder.

(thermodynamic integration)

— Nested models $\vec{\theta}_0$ is subset of $\vec{\theta}$ (deviations from GR, non-spining BHs

$$B \approx \frac{P(\vec{\theta} = \vec{\theta}_0 | d)}{\pi(\vec{\theta}_0)}$$