(2019)

(SUPPLEMENT)

Metropolis Hastings in Statistical Data analysis

from "Baysian Stat Analysis

See also: EMCMCE Python package

The Problem: Baysian Inference

D: given Data set

I: prior information (Jeffrey Priors)

H_i: Model to describe data / more frequent Model params θ of model M.

Baysian inference:

$$P(H_i|D,I) = \frac{\overset{\text{PRIOR}}{p(H_i|I)}\, p(D|H_i,I)}{P(D|I)}$$

Hypothesis in absence of D: prior prob. $P(H_i|I)$

Hypothesis with Data (D): posterior or Likelihood

$p(D|I)$: "prior predictive prob" GLOBAL LIKELIHOOD.

Baysian solution to parameter estimation:

Posterior PDF: $p(\theta|D,M)$ full distribution.

this leads to "best fit" & "error bars":

$$\boxed{\begin{aligned} \langle\theta\rangle &= \int d\theta\, \theta\, p(\theta|D,M) \\ C &= \int d\theta\, p(\theta|D,M) = C \end{aligned}}$$ CREDIBLE REGION.

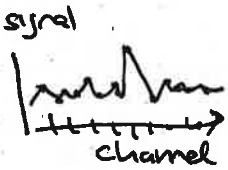
Nuissance parameters: another params Φ: data

$$p(\theta|D,M) = \int d\Phi\, p(\theta,\Phi|D,M)$$

→ Margenilization MARGINAL POSTERIOR PDF θ.

→ Note: this also leads to OCCAM'S RAZOR

Example Problem: Spectral line estimation



Theory: existing lineshape:

$$\underset{\text{Data}}{\underline{d}} = T\cdot f_i + \underset{\text{error}}{\underbrace{e}}$$

Theory predicts: $T\exp\left(\frac{v_i - v_0}{2\sigma_L^2}\right) \hat{=} T f_i$ Model

Note that $$P(D,M) = \int d\theta\, P(\theta|M)\, p(D|\theta,M) = \mathcal{L}(M)$$

is the GLOBAL LIKELIHOOD FUNCTION where all fitting parameters θ are integrated over

For Model verification we consider multiple (M_i).

For Global likelihood:

integration over all possible values.

$$p(D|M_1,I) = \int dT \underbrace{p(T|M_1,I)}_{\text{Prior knowledge}} p(D|M_1,T,I)$$

Choice of Prior: $p(T|M,I)$

In this case it refers to possible values of T. Key part in BAYSIAN STAT. ANALYSIS.

uniform prior: T_{max}, T_{min} $\quad \Delta T = T_{max} - T_{min}$

$$p(T|M_1,I) = \frac{1}{\Delta T} \quad \wedge \int d\,p(T|M_1,I)dT = 1$$

Jeffrey Priors: Equal probability.

$$\int_{0.1}^{1} p(T|M,I)dT = \int_{1.0}^{10} p(T|M_1,I)\,dT$$

$$\text{for } p(T|M_1,I) = \frac{k}{T} \quad \wedge\ k: \text{norm const.}$$

Probability, global likelihood function.

For the case that all errors are independent (e_i) then:

$$p(D_i|M_1,T,I) \triangleq p(E_i|M_1,T,I) \leftarrow \text{Prob density function of error}$$

$$\Rightarrow p(D|M_1,T,I) = p(D_1 \ldots D_N|M_1,T,I)$$

$$= \prod_{i=1}^{N} p(D_i|M,T,I) = \prod_{i=1}^{N} p(E_i|M,T,I)$$

Thus:

subtract Model

$$\boxed{\begin{aligned} p(D|M_1,T,I) &= \prod_{i=1}^{N} \frac{1}{\sqrt{\sigma^2 2\pi}} \exp\left(\frac{(d_i - Tf_i)^2}{2\sigma^2}\right) \\ &= (2\pi)^{-N/2} \sigma^{-N} \exp\left\{\sum \frac{(d_i - Tf_i)^2}{2\sigma^2}\right\} \end{aligned}}$$

here we assume the errors are distributed:

$$p(E_i|M,T,I) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{e_i^2}{2\sigma^2}\right) \quad \text{(Gaussian error)}$$

Note: in principe a very high dimensional integral, evaluated for all possible model parameters

We need $P(D|M_1 I)$ the global likelihood function.

$p(T|D, M_1, I)$: posterior PDF of signal strength, given Model (M_1), Data (D) and prior information (I).

$$\boxed{\underbrace{p(T|D, M_1, I)}_{\text{POSTERIOR PDF}} = \frac{p(T|M_1, I) \cdot p(D|M_1, T, I)}{p(D|M_1, I)}}$$

BAYES THEOREM

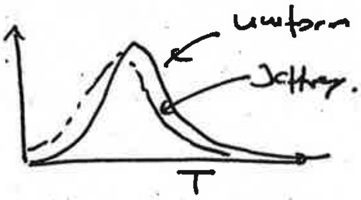
$P(T|M_1 I)$: prior

$p(D|M_1, T, I)$: global likelyhood ~~global likelihood~~ (probability of entire data set)

$p(D|M_1 I)$: GLOBAL LIKELIHOOD FUNCTION

$= \int dT \; p(D|M_1, I, T)\, dT$

Note: $p(T/D, M_1 I)$



MARKOV CHAIN MONTE CARLO : MCMC

(1953. Metropolis, Teller et al).

Metropolis – Hastings:

MCMC algorithm generate a random walk in the parameter (estimation) space such that the probability for being in the region is proportional to the POSTERIOR PDF

Problem: large number of parameters $p(\underset{\sim}{X}|D, M, I)$
parameters (e.g fit param.)

Step 1 Initialize X_0 (e.g $\{T_0, V_0\}$ for model parameters)

Step 2 obtain sample $Y = \{T_0', V_0'\}$

e.g $\{T_0', V_0'\} \in \{N(T_0, \sigma_T); N(V_0, \sigma_V)\}$.
Gaussian distributions.

Step 3 METROPOLIS RATIO

$$\boxed{r = \frac{P(Y|D I)}{P(X_t|D I)}}$$

Acceptance propability.

$$\underbrace{W_{x_t,Y}}_{\text{Transition rate}} = \min(1, r)$$

<u>Step 3</u> choose random number (if $r > 1$) called u and accept Y if : $u \leq r$ otherwise $X_{t+1} = X_t$

<u>Step 4</u> → return to step 2.

Relation to MCMC to computational physics: define $\Lambda(X) = -\ln(p(X|I)\, p(D|X,I))$ then posterior is

$$p(X|D,I) = e^{-\Lambda(x)}/Z \quad \text{and} \quad Z = \int dX\, e^{-\Lambda(x)}$$

→ evaluating of Boltzman factors and partition fct.

(cf Loredo'99)

<u>Why Metropolis algorithm works.</u> (Chapter 12 Bayesian Stat. Analysis)

proof: $\underset{\text{joint prob. distribution}}{P(X_t, X_{t+1})} = p(X_t|D,I)\, \underbrace{p(X_{t+1}|X_t)}_{\text{transition prob.}}$

$$= p(X_t|D,I)\, W_{X_t,X_{t+1}} = p(X_t|D,I) \cdot \min\left(1, \frac{P(X_{t+1}|D,I)}{P(X_t|D,I)}\right)$$

$$= \min(p(X_t|D,I), p(X_{t+1}|D,I))$$

$$= p(X_{t+1}|D,I)\, W_{X_{t+1},X_t} \quad \text{← ie reversed.} \quad = p(X_{t+1}|D,I)\, p(X_t|X_{t+1})$$

$$\Leftrightarrow \boxed{\frac{p(X_t|D,I)}{p(X_{t+1}|D,I)} = \frac{W_{X_t,X_{t+1}}}{W_{X_{t+1},X_t}}} \quad \text{DETAILED BALANCE}$$

<u>Next</u>: marginalize over X_t: to show we obtain $PDF(X_t)$ ↙ use detailed balance *

$$\int dX_t\, \underbrace{p(X_t|D,I)\, p(X_{t+1}|X_t)}_{\text{joint prob. distr}} \overset{*}{=} \int p(X_{t+1}|D,I)\, p(X_t|X_{t+1})\, dX_t$$

$$= p(X_{t+1}|D,I) \cdot \int \underbrace{p(X_t|X_{t+1})\, dX_t}_{=1}$$

$$= p(X_{t+1}|D,I) = \pi(X_t|D,I) \text{ posterior}$$

Thus joint prob. distribution converges to actual one.

Occam's Razor

Odd's ratio of two model M_1, M_2: (e.g. with different parameters)

$$O_{ij} = \underbrace{\frac{p(M_i|I)}{p(M_j|I)}}_{\text{Prior odds}} \underbrace{B_{ij}}_{\text{BAYES FACTOR}}$$

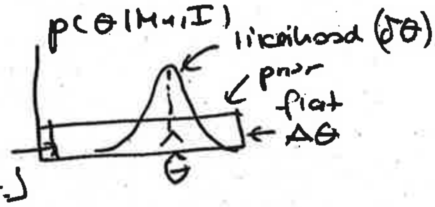
First Note that $\sum_{i=1}^{N} p(M_i | D, I) = 1$ (N models)

Example: two models. M_1 with single parameter θ
M_0: parameter θ fixed at θ_0 (no free param).

We need $P(D|M_1, I)$ the global likelihood of M_1.

e.g. assume that $P(\theta|M_1, I)$ is Gaussian centered at $\hat{\theta}$ the max. likelihood.

$\Rightarrow p(\theta | M_1, I) = \frac{1}{\Delta\theta}$: PRIOR [note: NO DATA). → model likelihood at θ]



global likelihood:

$$p(D|M_1,I) = \int d\theta \, p(\theta|M_1,I) \, p(D|\theta, M_1, I)$$

$$\simeq \frac{1}{\Delta\theta} \int d\theta \, p(D|\theta, M_1, I) \quad \text{peaked around } \hat{\theta}$$

$$\cong p(D|\hat{\theta}, M_1, I) \left(\frac{\delta\theta}{\Delta\theta}\right)$$

2nd Model No free parameters

$$p(\theta | M_0, I) = p(D | \underset{\uparrow \text{ fixed}}{\theta_0}, M_1, I)$$

important! special case of M_1!

Bayes factor

$$B_{10} = \underbrace{\frac{p(D|\hat{\theta}, I, M_1)}{p(D|\theta_0, I, M_1)}}_{\text{can never favor the simpler Model as } M_0 \text{ is contained as a special case in } M_1} \cdot \underbrace{\left(\frac{\delta\theta}{\Delta\theta}\right)}_{\text{PENALTY!}}$$

→ OCCAM'S RAZOR