Multivariate Normal Distribution

only iff Σ positive definite, then PDF $= \det(2\pi\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \Rightarrow N(\mu,\Sigma)$

$\frac{1}{(2\pi\sigma^2)^{N/2}}$ when Σ = identity because in every direction the same

Mutual Information

$$I(X;Y)=\sum_{Y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \Rightarrow = \int_Y\int_X p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)dx\,dy$$

Optimal Analytical parameters Ex. 4.1

Problem of fitting a parametrized curve $y=\alpha_0 f^0(x)+\dots+\alpha_M f^M(x)$ to a Dataset $D=(x_k,y_k)$, $k=1\dots N$. f^i is known, $\alpha_i\in\mathbb{R}$ are unknown params. Model: $y_k = f_k^T\alpha + e_k$, where $f_k=[f^0(x_k),f^1(x_k)\dots f^M(x_k)]^T$, α = similarly vector of $[\alpha^0,\dots,\alpha^M]$. $e_k\sim N(0,\sigma^2)$ (σ^2 is unknown), Assume Uniform Priors. (w/ large enough bounds).

a) Find MPV of (α,σ^2): Denote $s=\sigma^2$, PDF posterior is $= p(\alpha,s|D)=\frac{1}{(2\pi s)^{N/2}}\exp\left\{-\frac{1}{2s}J(\alpha;D)\right\}p(\alpha,s)$ (N because Multidimensional)

where $J(\alpha;D)=\sum_k(y_k-f_k^T\alpha)^2$ and $p(\alpha,s)$ is uniform PDF.

$\Rightarrow$ log Likelihood $L(\alpha,s|D)=\frac{N}{2}\ln(2\pi)+\frac{N}{2}\ln(s)+\frac{1}{2s}J(\alpha;D)-\ln(p(\alpha,s))$.

$\Rightarrow$ Derivative of $L(\alpha,s|D)=\frac{1}{2s}\frac{\partial}{\partial\alpha}J(\alpha;D)=-\frac{1}{s}\sum_k(y_k-f_k^T\alpha)\frac{\partial}{\partial\alpha}(f_k^T\alpha)=-\frac{1}{s}\sum_k(y_k-f_k^T\alpha)f_k=-\frac{1}{s}\left(\sum_k y_k f_k-\left(\sum_k f_k f_k^T\right)\alpha\right)\overset{!}{=}0$

jump $\Rightarrow \hat{\alpha}=B^{-1}d$, where $B=\sum_k f_k f_k^T$ and $d=\sum_k y_k f_k$. "Since $J(\alpha;D)$ is a quadratic form of α, $\hat{\alpha}$ minimizes the Log-likelihood and is thus MPV of α."

$\frac{\partial}{\partial s}L(\alpha,s|D)=\frac{N}{2s}-\frac{J(\alpha;D)}{2s^2}=\frac{1}{2s^2}(Ns-J(\alpha;D))$. $\quad\frac{\partial}{\partial s}L(\alpha,s|D)\overset{!}{=}0 \Rightarrow \hat{s}=\frac{1}{N}J(\hat{\alpha};D)$.

$\frac{\partial}{\partial s}\left(\frac{N}{2s}-\frac{J(\alpha;D)}{2s^2}\right)=-\frac{N}{2s^2}+\frac{J(\alpha;D)}{s^3}$. Substituting $s\leftarrow\hat{s}$, $\alpha\leftarrow\hat{\alpha}$ $\Rightarrow -\frac{N}{2\hat{s}^2}+\frac{J(\hat{\alpha};D)}{\hat{s}^3}=\frac{-N}{2\hat{s}^2}+\frac{N\hat{s}}{\hat{s}^3}=\frac{1}{2}\frac{N}{\hat{s}^2}>0$

2nd derivative

$\Rightarrow \hat{s}$ minimizes LogLikelihood $\Rightarrow$ is MVP of s.

b) Gaussian asymptotic Approximation of the posterior PDF (given).

Hessian: $\frac{\partial^2 f}{\partial x_1^2}\dots\frac{\partial^2 f}{\partial x_1\partial x_n}$, $\frac{\partial^2 f}{\partial x_2\partial x_1}\dots$

GAA is centered at MPV, with covariance Matrix $\Sigma=H^{-1}$, where H is the Hessian at the MPV.

Using derivatives from previous step: $\frac{\partial^2}{\partial\alpha^2}L=\frac{1}{s}B$, $\frac{\partial^2}{\partial\alpha\partial s}L=\frac{\partial^2}{\partial s\partial\alpha}=\frac{1}{2s^2}(d-B\alpha)$, $\frac{\partial^2}{\partial s^2}L=-\frac{N}{2s^2}+\frac{J(\alpha;D)}{s^3}$

Replace $\alpha\leftarrow\hat{\alpha}$, $s\leftarrow\hat{s}$ $\Rightarrow$ hessian $H=\begin{pmatrix}\frac{1}{\hat{s}}B & 0\\ 0 & \frac{1}{2}\frac{N}{\hat{s}^2}\end{pmatrix}=\frac{1}{\hat{s}}\begin{pmatrix}B & 0\\ 0 & \frac{N}{2\hat{s}}\end{pmatrix}\Rightarrow\Sigma\ \hat{s}\begin{pmatrix}B^{-1} & 0\\ 0 & \frac{2\hat{s}}{N}\end{pmatrix}$

Gaussian Distribution $X \sim \mathcal{N}(\mu, \sigma^2)$

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E[X] = \int_{-\infty}^{\infty} x f_x(x)\,dx = \int_{-\infty}^{\infty}(x-\mu) f_x(x)\,dx + \mu \int_{-\infty}^{\infty} f_x(x)\,dx$$

$y := (x-\mu)$, $\frac{dy}{dx} = 1 \Rightarrow dx = dy$

$$E[X] = \underbrace{\int_{-\infty}^{\infty}(y) f_x(y+\mu)\,dy}_{=0} + \underbrace{\mu \cdot 1}_{\mu}$$

negatively symmetric because exponent is squared and y is not

show that: $E[X] = \mu$

Nice to know: $\int e^{-ax^2}\,dx = \sqrt{\frac{\pi}{a}}$

Show $E[(X-\mu)^2] = \sigma^2$ (Variance)

Let $y = x - \mu$, $a = \frac{1}{2\sigma^2}$

$$E... = \sqrt{\frac{a}{\pi}} \int_{-\infty}^{\infty} y^2 e^{-ay^2}\,dy$$

$$= \sqrt{\frac{a}{\pi}} \frac{\partial}{\partial a} \int e^{-ay^2}\,dy$$

because $\frac{\partial}{\partial a} y^2 e^{-ay^2} = -y^4 e^{-ay^2}$ $\Rightarrow \frac{\partial}{\partial a} = -y^2$

$$= -\sqrt{\frac{a}{\pi}} \frac{\partial}{\partial a} \sqrt{\frac{\pi}{a}} = \frac{1}{2a} = \sigma^2$$

Sum of $\mathcal{N}(\mu, \sigma^2)$

$Z = X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x + \sigma_y)$

$$f_{x+y}(y) = \int_{-\infty}^{\infty} f_y(y-x) f_x(x)\,dx \quad \Big\}\text{ whenever } X, Y \text{ are i.i.d.} \quad (a)$$

Ex. 02.1.a) Find posterior distribution of β given $y = \beta x + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$

$$p(\beta|D) = \frac{p(y = y_0|\beta)\, p(\beta|\text{nothing})}{p(D|\text{nothing})}$$

$p(\beta|\text{nothing}) \sim$ Uniform (prior uncertainty)

probably useless: $f_y(y) \overset{(a)}{=} \int \underbrace{f_\beta(y-x)}_{\propto 1 \text{ because Uniform } \beta} f_\epsilon(x)\,dx = \frac{1}{b-a}\int_a^b f_\epsilon(x)\,dx \sim \mathcal{N} \Rightarrow$ Ignore $P(D)$??? Maybe because Data always given?

$p(D|\beta) = p(y = y_0|\beta) = \mathcal{N}(0 + \beta x, \sigma^2) := N_c$

because $x = x_0$ is given

$$\Rightarrow p(\beta|D) \propto \frac{N_c \cdot \text{const}}{\text{Data Const}} \propto \frac{1}{\sqrt{2\pi}\sigma} e^{\left\{-\frac{1}{2\sigma^2}(y_0 - \beta x_0)^2\right\}} \propto \frac{1}{\sqrt{2\pi}\sigma} e^{\left\{-\frac{x_0^2}{2\sigma^2}\left(\beta - \frac{y_0}{x_0}\right)^2\right\}}$$

($\frac{1}{\frac{2\sigma^2}{x_0^2}}$)

$$\propto \mathcal{N}\left(\frac{y_0}{x_0}, \frac{\sigma^2}{x_0^2}\right) \Rightarrow \beta_{Max} = \frac{y_0}{x_0}$$

Neg. Log-likelihood: $-\ln(P(\beta = \beta|D)) = -\ln\left(\frac{1}{\sqrt{2\pi\frac{\sigma^2}{x_0^2}}} e^{\{\ldots\}}\right) = -\ln\left(\frac{1}{\sqrt{2\pi\frac{\sigma^2}{x_0^2}}}\right) + \frac{\left(\beta - \frac{y_0}{x_0}\right)^2}{2\frac{\sigma^2}{x_0^2}}$

Derivative $\frac{d}{d\beta}$ of that $= \frac{2(\beta x_0 - y_0)x_0}{2\sigma^2}$. Insert $\beta = \frac{y_0}{x_0} \Rightarrow$ extremum matches negative Log $\Rightarrow$ it must be a maximal β.

Alternative: $\frac{\partial^2}{\partial \beta^2} > 0$

Probability with more data: ϵ_i independent $\Rightarrow$ solution is Product

$$P(\beta|D) = \frac{P(D|\beta)P(\beta)}{P(D)} \propto \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta x_0)^2\right\} \propto \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{3}(y_i - \beta x_0)^2\right\} \sim \mathcal{N}\left(\frac{7}{6}\frac{y_0}{x_0}, \frac{\sigma^2}{3x_0^2}\right)$$

(jump)

with $D = \{y_0, 2y_0, \frac{y_0}{2}\}$ Argument: ϵ_i i.i.d $\Rightarrow y_i$ i.i.d

Log-Likelihood (Neg) Bsp.

$$P(\beta|D,I) \propto \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}}}_{(\ldots)^{-\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)^2\right\} \Rightarrow L(\beta|D,I) = \frac{N}{2}\ln(2\pi) + \frac{N}{2}\ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)^2$$

"Distributed according to f" follows from "density distribution of X is f" follows from $P(X < z) = \int \mathbb{1}_{y<z} f(y)\,dy$.

Example: $W_i = \frac{f/g}{\sum f/g}$ (sampling importance something)

$$\Rightarrow P(x<z) = \sum w_i \mathbb{1}_{y<z} = \frac{\frac{1}{N}\sum^N f/g\, \mathbb{1}_{y_i<z}}{\frac{1}{N}\sum^N f/g} \xrightarrow{N\to\infty} \frac{\int \frac{f(y)}{g(y)} \mathbb{1}_{y<z}\, g(y)\,dx}{\int \frac{f}{g} g\,dy}$$

$= P_g(y)$ = probability that we sample y according to g

$$= \frac{\int \ldots}{\underbrace{\int f(y)\,dy}_{=1}} = \int \mathbb{1}_{x<z} f(y)\,dy$$

Metropolis Hastings

(toggle max/min) (E is given function) minimum of E. Sometimes optimum = maximize instead

1. Sample $\vartheta \sim q(\cdot \mid \vartheta_{k-1})$
2. Set $h = \min(1, \exp\{-(E(\vartheta) - E(\vartheta_{k-1}))/T\})$
3. Sample $u \sim Uniform(0,1)$
4. If $u \leq h$, set $\vartheta_k = \vartheta$, else $\vartheta_k = \vartheta_{k-1}$

Needs symmetric pdf: $q(x|y) = q(y|x) = P[y|x] = P[x|y]$

With ~~detailed Balance~~ there is easily a fixpoint: ~~$p(x)t(x,y) = t(y,x)p(y)$~~

$$\int_Y t(x,y)p(y)\,dy = \int_Y t(y,x)p(x)\,dy = p(x)\underbrace{\int t(y,x)\,dy}_{=1} = p(x)$$

Bayes à trois

$$P(x,u,z|y) = P[x|u,z,y]\,P[u,z|y] = P[x|u,z,y]\,P[u|y,z]\,P[z|y]$$

Inverse CDF

large $p(x) \to$ steep $F(x)$ / $(P(X \leq x))$

$\Rightarrow u \sim U(0,1)$; $\overleftarrow{F}(u)$ gives x based on Distribution. (Because $F' = p$.)

Ex 7.2.a)

Algorithm to math: (z proposed state, y old state)

$$P[\underset{\ddot{u}}{accept}|z,y] = \begin{cases} \alpha(z,y), & u=1 \\ 1-\alpha(z,y), & u=0 \end{cases}$$

$$\alpha = \min\left\{1, \frac{p(z)}{p(y)}\right\}$$

into $P[x|u,z,y] = \delta(x-z)\mathbb{1}_{u=1} + \delta(x-y)\mathbb{1}_{u=0}$

$P[z|y]$ = proposal distribution $= q(z,y)$

To get Transition Probability:

$$t(x,y) = P[x|y] \overset{\text{marginalize}}{=} \int_z \sum_{u\in\{0,1\}} P[x,z,u|y]\,dz$$

$$= \int_z \delta(x-z)\alpha(z,y)q(z,y)\,dz + \delta(x-y)\int_z (1-\alpha(z,y))\,q(z,y)\,dz$$

Goal is to minimize $E(x)$.

Define $S_{zy} := \{(z,y) : E(z) < E(y)\}$

$$t(x,y) = \alpha(x,y)\underbrace{q(x,y)}_{q(x|y)} + \delta(x-y)\left(1 - \int_{S_{zy}} q(z,y)\,dz - \int_{\overline{S}_{zy}} q(z,y)\underbrace{\frac{p(z)}{p(y)}}_{<1}\,dz\right)$$

(Because: $q(z,y) = q(y,z) \to q(y,z)p(z) = P[y,z] \to \frac{P[y,z]}{p(y)} = P[z|y] = q(z,y)$) (Magic)

$$t(x,y) = \alpha(x,y)q(x|y) + \delta(x-y)\left(1 - \int_{S_{zy}} q(z,y)\,dz - \int_{\overline{S}_{zy}} q(z,y)\,dz\right) = \alpha(x,y)\,q(x,y)$$

Detailed Balance

$p(x) = \int_Y t(x,y)p(y)\,dy$ = distribution for $k \to \infty$ in MCMC

$t(x_{k+1}, x_k) = P[x_{k+1}|x_k]$ = transition Probability

necessary for convergence: $p(x)t(x,y) = t(x,y)p(y)\,dy$

Simulated Annealing

Given data $\{x_i, y_i\}_{i=0}^N$ and model $f(x) = A\sin(wx)$

Formulate problem as Optimization Problem using an SA Approach. A: minimize difference between predicted and data = minimize Energy $E(A,w) = \frac{1}{N}\sum_{i=1}^{N}[y_i - A\sin(wx_i)]^2 \Rightarrow \min \exp\{E(A,w)\}$

To show Detailed Balance: Assume $E(x) < E(y)$. Then $t(x,y)p(y) = q(x,y)p(y)\overbrace{\min\left\{1, \frac{p(x)}{p(y)}\right\}}^{\alpha} = q(x,y)\,p(y)$.

$$t(y,x)p(x) = q(y,x)p(x)\min\left\{1, \frac{p(y)}{p(x)}\right\} = q(y,x)p(x)\frac{p(y)}{p(x)} = q(y,x)p(y)$$

$$t(x,y)p(y) = t(y,x)p(x) \Leftrightarrow q(x,y)p(y) = q(y,x)p(y) \Leftrightarrow \underbrace{q(x,y) = q(y,x)}_{\text{by definition true.}}$$

Same approach for $E(y) < E(x)$.

Rejection Sampling

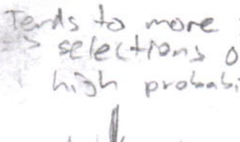
1. Sample $x^* \sim g(x)$ and $u \sim Unif(0,1)$
2. Accept $x = x^*$ if $u \leq f(x^*)/(T\cdot g(x^*))$ otherwise retry

T such that $Tg \geq f \;\forall x \to$ best T is $\sup_x \frac{f(x)}{g(x)}$

Higher probability if f close to g

Goal of Metropolis / MCMC

Tends to more x at higher probabilities $\Rightarrow$ selections over time more around high probabilities



$\Rightarrow$ throw away burn-in and consider the rest as samples because probability of selecting some point is proportional to how plausible it is

OMP Tasks

Remember: single nowait before task might be needed

task **untied** : continue on any thread after being suspended. Dont combine with threadprivate.

task final : final(n≤5) =no new tasks when n ≤5. all child tasks will also be final

taskwait : Only waits for its tasks, not those of its children. => wait in child as well

Tasks are waited for at im/ex-plicit barriers

omp_in_parallel() to check if in pragma omp parallel

Torc

x is double arr[2] / y is double arr[1]

void task(double *x (in), double *y (res))

int main(...){ torc_register_task(task); torc_init(argc, argv, MODE_MW);

```
torc_task(-1, task, 2,
          2, MPI_Double, Call_b_cop,
          1, MPI_Double, CALL_BY_RES,
          d, result[i])
```

d: double arr[2]

torc_waitall oder so

MPI Datatype

```
MPI_Datatype daT;
MPI_Type_contiguous(NumParams, MPI_DOUBLE, &daT);
Alternative: MPI_Type_Vector(count, blocklength, stride, MPI_DOUBLE, &daT
                    (count = blocks)
MPI_Type_Commit(&daT);
MPI_Bcast(Params, 1, daT, 0, MPI_COMM_WORLD)
         (buffer, count, ParamType, broadcast root)
MPI_Type_Free(&daT);
```

└ Copies from roots "Params" to all others that called this.

MPI Init

int size; int rank;

```
MPI_Init(&argc, &argv); MPI_Comm_rank(MPI_COMM_WORLD, &rank);
MPI_Comm_size(MPI_COMM_WORLD, &size);
:
MPI_Finalize();
```

MPI Groups

MPI_Comm_group(MPI_Comm, MPI_Group)

... group_union, ... group_intersection (group1, group2, newgroup)

MPI_Comm_create_group(comm, group, int tag, newcomm)

MPI Comm

Do The Init

```
int color = rank/4; // Split based on rank
MPI_Comm myComm;
MPI_Comm_split(MPI_COMM_WORLD, color, rank, &myComm);
```

Get rank within myComm like in init.

MPI Comm From Group

MPI_Comm_create(comm, group, newComm)

Nonlinear Model $x(t)=\frac{1}{\alpha}\ln(\cosh(\sqrt{g\alpha}(t-t_0)))$ Given data x at times Δt: $D=\{x_1,\dots x_N\}$ and σ, estimate uncertainty of α

Ex.02.2 Note that measurements & model predictions satisfy "Model Error Equation" $X_k = x(k\Delta t)+E_k$ ← i.i.d $N(0,\sigma^2)$ // given

1) **Find Posterior PDF**

Assume uniform prior for α.

$$p(\alpha|D,\sigma)=\frac{p(D|\alpha,\sigma)p(\alpha|\sigma)}{p(D|\sigma)}\propto p(D|\alpha,\sigma)p(\alpha|\sigma)=U_\alpha(\alpha_1,\alpha_2)\prod_{k=1}^{N}p(X_k|\alpha,\sigma)$$

$$=U_\alpha(\alpha_1,\alpha_2)\prod_{k=1}^{N}N(x(k\Delta t),\sigma^2)\propto \mathbb{1}_{[\alpha_1,\alpha_2]}\prod_{k=1}^{N}\exp\left\{-\frac{(X_k-x(k\Delta t))^2}{2\sigma^2}\right\}=\mathbb{1}_{[\alpha_1,\alpha_2]}\sum_{k=1}^{N}(X_k-x(k\Delta t))^2$$

const, const $\cdot\exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}$

Must integrate to 1 => $\frac{1}{C}=\int_{\alpha_1}^{\alpha_2}\exp(\dots)\,d\alpha$ for proportionality const. C.

2) **Negative Log Likelihood**

$$L(\alpha)=-\ln(p(\alpha|D,\sigma))=-\ln(C)+\frac{1}{2\sigma^2}\sum_{k=1}^{N}(X_k-x(k\Delta t))^2$$

3) **Simple Case**

Consider One Measurement: X_1. Given g, show that α = asdf is the most probable value.

"Find max $p(\alpha)$" == "Find min $L(\alpha)$" => $\arg\min_{\alpha\in[\alpha_1,\alpha_2]}\sum_{k=1}^{N}(X_k-x(k\Delta t))^2$ ← k becomes 1

$$\frac{\partial}{\partial\alpha}(X_1-x(\Delta t))^2=0 \Leftrightarrow 2(X_1-x(\Delta t))\left(-\frac{\partial}{\partial\alpha}x(\Delta t)\right)=0$$

$$\Leftrightarrow\left(X_1-\frac{1}{\alpha}\ln\left(\cosh(\sqrt{g\alpha}(\Delta t-t_0))\right)\right)\overset{\text{cross}}{\times}\left(\frac{1}{2\alpha\sqrt{\alpha}}\sqrt{g}(\Delta t-t_0)\tanh(\sqrt{\alpha g}(\Delta t-t_0))-\ln\left(\cosh(\sqrt{\alpha g}(\Delta t-t_0))\right)\cdot\frac{1}{\alpha^2}\right)=0$$

Now substitute all const. values => asdf. To make sure it's a minimum, compute $\frac{\partial^2}{(\partial\alpha)^2}L(\alpha)$ and show >0

Ex.07.1)

Sampling Importance Resampling

Get samples from f given g. Sampling g is easy. Given Data $\{Y_i\}_{i=1}^N$ sampled from g.

Define $w_i=\frac{f(Y_i)g(Y_i)}{\sum_{i=1}^N f(Y_i)g(Y_i)}$ and $X\sim \Pr(X=Y_i|Y_1,\dots,Y_N)=w_i$. Show that the density distribution of X is f:

$$\Pr(X<z)=\sum w_i\mathbb{1}_{(Y_i<z)}=\frac{\frac{1}{N}\sum_{i=1}^{N}\frac{f(Y_i)}{g(Y_i)}\mathbb{1}_{(Y_i<z)}}{\frac{1}{N}\sum_{i=1}^{N}\frac{f(Y_i)}{g(Y_i)}}\xrightarrow{N\to\infty}\frac{\int\frac{f(Y_i)}{g(Y_i)}\mathbb{1}_{(Y_i<z)}g(Y_i)\,dx}{\int\frac{f(y)}{g(y)}g(y)\,dy}$$

(could be called dy; many samples according to $g(y)$)

$$=\frac{\int f(y)\mathbb{1}_{(y<z)}\,dy}{\int f(y)\,dy}=\int f(y)\mathbb{1}_{(y<z)}\,dy=F_f(z)=\Pr_f(x<z)$$

=> in $N\to\infty$, $X\sim$ according to f

b) Given i.i.d. Samples ~ like X $\{X_i\}_{i=1}^N$. Show that

$$E\left[\frac{1}{N}\sum_{i=1}^{N}h(X_i)\right]=E\left[\sum_{i=1}^{N}w_ih(Y_i)\right].$$

=> $=E[h(X)]$. Let p := density of X.

$$\Rightarrow p(x)=\int\cdots\int p(x|Y_1,\dots Y_n)\,p(Y_1\dots Y_n)\,dY_1\dots dY_n=\int p(x|y)g(y_1)g(y_2)\dots g(y_N)\,dy$$

← $y:=(y_1,y_2,\dots,y_n)^T$

$$p(x|y)=\sum_{i=1}^{N}w_i\delta_{\text{Dirac}}(x-y_i)=\begin{cases}w_1, & x=y_1\\ \vdots\\ w_N, & x=y_N\end{cases}$$

$$\Rightarrow E[h(x)]=\int h(x)p(x)\,dx=\iint h(x)\sum_{i=1}^{N}w_i\delta(x-y_i)g(y_1)\dots g(y_N)\,dy\,dx$$

Notice that

$$\int h(x)\sum_{i=1}^{N}w_i\delta(x-y_i)\,dx=\sum_{i=1}^{N}w_i\int h(x)\delta(x-y_i)\,dx=\sum_{i=1}^{N}w_ih(y_i)$$

we get

$$E[h(x)]=\int\sum_{i=1}^{N}w_ih(y_i)g(y_1)\dots g(y_N)\,dy=E\left[\sum_{i=1}^{N}w_ih(Y_i)\right]$$

Ableitungen

sinh' -> cosh
cosh' -> sinh
tanh' -> $\cosh^{-2}$
coth -> $-\sinh^{-2}x$

ETH zürich — High Performance Computing for Science and Engineering II
Spring semester 2018

P. Koumoutsakos, P. Hadjidoukas
ETH Zentrum, HG D1.1/G3
CH-8092 Zürich

Set 07 - Sampling: toward MCMC and TMCMC

Issued: April 16, 2018
Hand in: April 23, 2018

Question 1: Sampling Importance Resampling.

This exercise will bring you closer to the concept of Transitional Markov Chain Monte Carlo. The goal here is to get samples from the distribution f having already in hand samples from a distribution g. Usually sampling g is easier than sampling f.

Assume we have samples $\{Y_i\}_{i=1}^N$ from g. We define the weights

$$\omega_i = \frac{f(Y_i)/g(Y_i)}{\sum_{i=1}^N f(Y_i)/g(Y_i)}, \tag{1}$$

and a random variable X with $\Pr(X = y_i \mid y_1, \ldots, y_N) = \omega_i$.

a) Show that in the limit $N \to \infty$, the random variable X is distributed according to f.

We are going to show that the density of the distribution of X is f,

$$\begin{aligned} \Pr(X < z) &= \sum \omega_i \mathbb{1}_{Y_i<z} \\ &= \frac{\frac{1}{N}\sum \frac{f(Y_i)}{g(Y_i)} \mathbb{1}_{Y_i<z}}{\frac{1}{N}\sum \frac{f(Y_i)}{g(Y_i)}} \xrightarrow{N\to\infty} \frac{\int \frac{f(y)}{g(y)} \mathbb{1}_{y<z} g(y)\, dy}{\int \frac{f(y)}{g(y)} g(y)\, dy} \\ &= \frac{\int f(y) \mathbb{1}_{y<z}\, dy}{\int f(y)\, dy} \\ &= \int \mathbb{1}_{y<z} f(y)\, dy, \end{aligned} \tag{2}$$

where the limit is true due to the Central Limit Theorem. Thus, in the limit $N \to \infty$ the random variable X is distributed according to f.

b) Let $\{X_i\}_{i=1}^N$ be i.i.d. samples following the same distribution as X. Show that

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N h(X_i)\right] = \mathbb{E}\left[\sum_{i=1}^N \omega_i h(Y_i)\right]. \tag{3}$$

Let p be the density of the random variable X. Since X_i are drawn independently,

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N h(X_i)\right] = \mathbb{E}\left[h(X)\right]. \tag{4}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_N)$ and $p(\boldsymbol{y}) = g(y_1)\ldots g(y_N)$ since Y_i are i.i.d. By the law of total probability,

$$\begin{aligned} p(x) &= \int \cdots \int p(x \mid y_1, \ldots, y_n) p(y_1, \ldots, y_N)\, dy_1 \ldots dy_N \\ &= \int p(x \mid \boldsymbol{y})\, g(y_1) \ldots g(y_N)\, d\boldsymbol{y}. \end{aligned} \tag{5}$$

The conditional probability of X on $\boldsymbol{y}$ is given by,

$$p(x \mid \boldsymbol{y}) = \sum_{i=1}^N \omega_i\, \delta(x - y_i) = \begin{cases} \omega_1, & x = y_1 \\ \vdots & \\ \omega_N, & x = y_N \end{cases}, \tag{6}$$

where δ is the slightly violated Dirac function[1]. Substituting 6 to 5 and then to 4

$$\begin{aligned} \mathbb{E}\left[h(X)\right] &= \int h(x) p(x)\, dx \\ &= \int \int h(x) \sum_{i=1}^N \omega_i\, \delta(x - y_i)\, g(y_1) \ldots g(y_N)\, d\boldsymbol{y}\, dx. \end{aligned} \tag{7}$$

By noticing that

$$\int h(x) \sum_{i=1}^N \omega_i\, \delta(x - y_i)\, dx = \sum_{i=1}^N \omega_i \int h(x) \delta(x - y_i)\, dx = \sum_{i=1}^N \omega_i\, h(y_i), \tag{8}$$

we get

$$\begin{aligned} \mathbb{E}\left[h(X)\right] &= \int \sum_{i=1}^N \omega_i\, h(y_i)\, g(y_1) \ldots g(y_N)\, d\boldsymbol{y} \\ &= \mathbb{E}\left[\sum_{i=1}^N \omega_i\, h(Y_i)\right], \end{aligned} \tag{9}$$

that concludes the proof.

Grading scheme:
Total: 10 pts

- 5 points: question a
- 5 points: question b

[1] The Dirac is not a function but a distribution and is defined only through the action on test functions, i.e., $\delta_{x_0}[\varphi] = \int \varphi(x)\, \delta_{x_0}(dx) = \varphi(x_0)$

Question 2: MCMC for optimization

Markov Chain Monte Carlo (MCMC) is a sampling technique used in Bayesian inference to obtain samples from the posterior. It can also be used in discrete and continuous optimization problems.

An optimization problem has the following form:

$$\min_{\vartheta \in \Theta} E(\vartheta), \quad E : \mathcal{E} \to \mathcal{V}, \Theta \subset \mathcal{E}. \tag{10}$$

The set $\mathcal{V}$ should be totally ordered. Usually it is the set of real or natural numbers. If $\Theta \quad \mathcal{E}$, the minimization is *unconstrained*, and if $\Theta \subsetneq \mathcal{E}$, the problem is called *constrained*.

If $\mathcal{V}$ is numeric, the problem can be reformulated as follows:

$$\min_{\vartheta \in \Theta} E(\vartheta) \Leftrightarrow \max_{\vartheta \in \Theta} \exp\{-E(\vartheta)/T\}, \quad T > 0. \tag{11}$$

$$\Leftrightarrow \max_{\vartheta \in \Theta} \frac{\exp\{-E(\vartheta)/T\}}{\int_\vartheta \exp\{-E(\vartheta)/T\} \mathrm{d}\vartheta} := \max_{\vartheta \in \Theta} P(\vartheta; T). \tag{12}$$

where we silently ignore any integrability issues and assume E behaves nicely enough. The quantity P in (12) is a valid probability density function. If $P(\vartheta; T)$ had its mass highly concentrated around the minima of E, we could just sample ϑ from this distribution. The sampling is done by the evolution of a Markov chain using appropriate transition probabilities as explained below. The high concentration of mass around the minima is achieved by a technique known as *simulated annealing* which reduces the value of T over time as new samples are being generated.

Overview of MCMC

As we know from Exercise 5, a Markov chain is characterized by its *transition probability* $\mathbb{P}\left[x_{k+1} \mid x_k\right] = t(x_{k+1}, x_k)$. The *marginal distribution* of state $k+1$ can be written as:

$$\mathbb{P}\left[x_{k+1}\right] = \int_{x_k} \mathbb{P}\left[x_{k+1} \mid x_k\right] \mathbb{P}\left[x_k\right] \mathrm{d}x_k. \tag{13}$$

If we can let $k \to \infty$, then we obtain a distribution $p(x)$ which satisfies the following relation:

$$p(x) = \int_y t(x, y) p(y) \mathrm{d}y. \tag{14}$$

As an aside, you can check that if the Markov Chain is discrete, the finding of the stationary distribution corresponds to solving an eigenvector problem.

There are in general no guarantees that such a distribution exists. If it does, the marginal distributions will eventually converge to it. When *detailed balance* holds:

$$p(x) t(x, y) = t(y, x) p(y), \tag{15}$$

then the fixed point exists almost trivially:

$$\int_y t(x, y) p(y) \mathrm{d}y = \int_y t(y, x) p(x) \mathrm{d}y = p(x) \underbrace{\int_y t(y, x) \mathrm{d}y}_{=1} = p(x). \tag{16}$$

$P(\vartheta; T)$ is what we want our stationary distribution to be. Various MCMC algorithms are thus concerned with picking the appropriate transition probability $t(\cdot, \cdot)$.

One of the popular MCMC algorithms is the Metropolis-Hastings algorithm. For it we need a symmetric proposal distribution $\mathbb{P}\left[x \mid y\right] = q(x, y) = q(y, x) = \mathbb{P}\left[y \mid x\right]$ (note: this is not the same as $t(x, y)$):

1. Sample $\vartheta \sim q(\cdot, \vartheta_{k-1})$.
2. Set $h = \min\left(1, \exp\{-(E(\vartheta) - E(\vartheta_{k-1}))/T\}\right)$
3. Sample $u \sim$ Unif$(0, 1)$
4. If $u \leq h$, set $\vartheta_k = \vartheta$, else set $\vartheta_k = \vartheta_{k-1}$.

Questions

a) Show that the Metropolis Hasting algorithm induces a transition probability $t(x, y)$ that satisfies the detailed balance condition.

First, we need to derive the transition probability $\mathbb{P}\left[x \mid y\right] = t(x, y)$. Starting from the algorithm, we may start from the joint distribution $\mathbb{P}\left[x, u, z \mid y\right]$, where x represents the next state, u the Bernoulli variable sampled according to the Metropolis choice, z the proposed state, and y the old state.

$$\mathbb{P}\left[x, u, z \mid y\right] = \mathbb{P}\left[x \mid u, z, y\right] \mathbb{P}\left[u \mid z, y\right] \mathbb{P}\left[z \mid y\right]. \tag{17}$$

The conditional probability $\mathbb{P}\left[x \mid u, z, y\right]$ can be written as:

$$\mathbb{P}\left[x \mid u, z, y\right] = \delta(x - z) \mathbb{1}_{u=1} + \delta(x - y) \mathbb{1}_{u=0}, \tag{18}$$

which we interpret as follows: if $u = 1$ (proposal is accepted), then x will be the same as z, hence the delta function $\delta(x - z)$. When $u = 0$ (proposal is not accepted), then x will be the same as y, hence the delta $\delta(x - y)$.

Further, u is a Bernoulli distributed variable, so $\mathbb{P}\left[u \mid z, y\right]$ reads as:

$$\mathbb{P}\left[u \mid z, y\right] = \begin{cases} \alpha(z, y), & u = 1 \\ 1 - \alpha(z, y), & u = 0. \end{cases}, \tag{19}$$

$$\alpha(z, y) = \min\left\{1, \frac{p(z)}{p(y)}\right\}. \tag{20}$$

Finally, the proposal distribution is simply $\mathbb{P}\left[z \mid y\right] = q(z, y)$.

To get the transition probability $t(x, y) = \mathbb{P}\left[x \mid y\right]$ we marginalize:

$$t(x, y) = \int_z \mathrm{d}z \sum_{u \in \{0,1\}} \mathbb{P}\left[x, z, u \mid y\right] \tag{21}$$

$$= \int_z \mathrm{d}z\ \delta(x - z) \alpha(z, y) q(z, y) + \delta(x - y) \int_z \mathrm{d}z\ (1 - \alpha(z, y)) q(z, y). \tag{22}$$

Define $\mathcal{S}(z, y) := \{(z, y) : E(z) < E(y)\}$, then:

$$t(x, y) = \alpha(x, y)q(x, y) + \delta(x - y)\left(1 - \int_z \mathrm{d}z \; \alpha(z, y)q(z, y)\right) \tag{23}$$

$$= \alpha(x, y)q(x \mid y) + \delta(x - y)\left(1 - \int_{\mathcal{S}(z,y)} \mathrm{d}z \; q(z, y) - \int_{\mathcal{S}'(z,y)} \mathrm{d}z \; q(z, y)\frac{p(z)}{p(y)}\right). \tag{24}$$

Using $q(z, y) = q(y, z) \rightarrow q(y, z)p(z) = \mathbb{P}[y, z] \rightarrow \frac{\mathbb{P}[y,z]}{p(y)} = \mathbb{P}[z \mid y] = q(z, y)$.

$$t(x, y) = \alpha(x, y)q(x \mid y) + \delta(x - y)\left(1 - \int_{\mathcal{S}(z,y)} \mathrm{d}z \; q(z, y) - \int_{\mathcal{S}'(z,y)} \mathrm{d}z \; q(z, y)\right) \tag{25}$$

$$= \alpha(x, y)q(x, y) \tag{26}$$

To show detailed balance, we start by assuming $E(x) < E(y)$. Then:

$$t(x, y)p(y) = q(x, y)p(y) \min\left\{1, \frac{p(x)}{p(y)}\right\} = q(x, y)p(y). \tag{27}$$

$$t(y, x)p(x) = q(y, x)p(x) \min\left\{1, \frac{p(y)}{p(x)}\right\} = q(y, x)p(x)\frac{p(y)}{p(x)} = q(y, x)p(y). \tag{28}$$

$t(x, y)p(y) = t(y, x)p(x) \Leftrightarrow q(x, y)p(y) = q(y, x)p(y) \Leftrightarrow q(x, y) = q(y, x)$. The last equality holds by definition. The same approach may be used for the case $E(x) > E(y)$.

In the next questions you will implement the Metropolis Hastings algorithm maximum to find the maximum likelihood solution the linear regression problem

$$\min_{\beta} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2. \tag{29}$$

b) Implement a routine `metropolis_step`. It takes the current sample $\boldsymbol{\beta}_k$ (and other needed values, such as X, y, T) and generates the $\boldsymbol{\beta}_{k+1}$ according to the Metropolis Hastings algorithm. It also returns the new value of the loss function $E(\boldsymbol{\beta}_{k+1})$.
   Pick an appropriate symmetric proposal distribution, such as $q(\boldsymbol{\beta}_{k+1} \mid \boldsymbol{\beta}_k) = \mathcal{N}(\boldsymbol{\beta}_{k+1}; \boldsymbol{\beta}_k, \Lambda)$.

c) Implement the `mcmc_routine` which runs the `metropolis_step` K times and keeps track of:

   1. the $\boldsymbol{\beta}_{k^*}$ and k^* for which $E(\cdot)$ is the smallest among all the generated samples,
   2. all the values $k, E(\boldsymbol{\beta}_k)$.

   The routine should take as parameter the function $T(k)$ used to control the parameter T in the Metropolis Hastings step.

d) Find experimentally how many iterations $\bar{k}$ are needed *on average* to reach the minimum within a given tolerance: $\|\boldsymbol{\beta}_{k^*} - \hat{\boldsymbol{\beta}}\| < \varepsilon$. Repeat the experiment for multiple values of ε. The $\hat{\boldsymbol{\beta}}$ can be obtained via least squares solutions. Produce a plot of $\bar{k}(\varepsilon)$ vs. ε. Do this for different decreasing functions $T(k)$, such as $T_0/\log(k), T_0/k, T_0\exp(-cT)$.

P. Koumoutsakos, P. Hadjidoukas
ETH Zentrum, HG D1 1/G3
CH-8092 Zürich

Spring semester 2018

Set 02 - Bayesian inference

Issued: March 05, 2018
Hand in: March 12, 2018

Question 1: Linear Model

You are given the linear regression model that describes the relation between variables x, y.

$$y = \alpha + \beta x + \epsilon,$$

where α and β are the regression parameters, y is the output quantity of interest (QoI) of the system, x is the input variable and ϵ is a term accounting for model and measurement errors. For all following subquestions consider that the prior uncertainty for the parameter β is quantified by a uniform distribution with large enough bounds. The regression parameter α is not considered uncertain. The model error is quantified by a Gaussian distribution $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

a) First consider the model, where $\alpha = 0$, given one measurement data point, $D = \{x_0, y_0\}$.

$$y = \beta x + \epsilon$$

Find the posterior uncertainty in the model parameter β (i.e. determine the posterior distribution, the negative log-likelihood function and the most probable value (MPV) of β).
The posterior distribution of β given the dataset D is given by Bayes' therorem

$$p(\beta|D, I) = \frac{p(y = y_0|\beta, I)p(\beta|I)}{p(y = y_0|I)}$$

Since $\epsilon \sim \mathcal{N}(0, \sigma^2)$, $p(y = y_0|\beta, I) = \mathcal{N}(\beta x, \sigma^2)$ and since β follows a uniform distribution with large bounds, $p(\beta|I) \propto 1$. Therefore

$$p(\beta|D, I) \propto \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2\sigma^2}(y_0 - \beta x_0)^2\right\}.$$

With re-ordering of terms, we can prove that

$$p(\beta|D, I) \propto \frac{1}{\sqrt{2\pi}\sigma} exp\left\{\frac{x_0^2}{2\sigma^2}(\beta - \frac{y_0}{x_0})^2\right\}$$

Therefore, the posterior PDF is a normal distribution $\sim \mathcal{N}(\frac{y}{x}, \frac{\sigma}{x_0^2})$.

The negative log-likelihood function then reads

$$\begin{aligned} L(\beta|D, I) &= -ln(p(\beta|D, I)) \\ &= \frac{1}{2}ln(2\pi) + ln(\sigma) + \frac{1}{2\sigma^2}(y_0 - \beta x_0)^2 \end{aligned}$$

The most probable value is calculated as the maximum of the posterior PDF or, equivalently, as the minimum of the negative log-likelihood function. We compute the derivative of the negative log-likelihood function with respect to β

$$\frac{dL}{d\beta} = \frac{1}{\sigma^2}(y_0 - \beta x_0)(-x_0)$$

By setting the 1st derivative equal to zero we compute

$$\beta = \frac{y_0}{x_0}$$

Lastly,

$$\begin{aligned} \frac{d^2L}{d\beta^2} &= \frac{1}{\sigma^2}(x_0^2) \\ &> 0 \end{aligned}$$

therefore, β is indeed a minimum of $L(\beta|D, I)$, and thus, the MPV.

b) Second, using the same model as above, now consider a dataset of three output points, with the same input value x_0, $D = \{y_0, y_0/2, 2y_0\}$. Each observation is therefore fitted by the model

$$y_i = \beta x_i + \epsilon_i,$$

where $i = 1...N$ and ϵ_i are independent and identically distributed error terms: $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.
Find the posterior uncertainty in the model parameter β using the new dataset.
As the error terms ϵ_i are statistically independent, the output QoI's y_i are also statistically independent. Therefore, we write the joint distribution of the output QoI

$$\begin{aligned} p(\beta|D, I) &= \frac{p(D|\beta, I)p(\beta|I)}{p(D|I)} \\ &\propto \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta x_0)^2\right\} \\ &\propto \frac{1}{(2\pi\sigma^2)^{3/2}} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{3}(y_i - \beta x_0)^2\right\} \end{aligned}$$

By reordering terms and simplifying, we can see that the the posterior PDF of β is again a normal distribution: $\sim \mathcal{N}(\frac{7}{6}\frac{y_0}{x_0}, \frac{\sigma^2}{3x_0^2})$.

The negative log-likelihood function reads:

$$\begin{aligned} L(\beta|D, I) &= -ln(p(\beta|D, I)) \\ &= \frac{3}{2}ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{3}(y_i - \beta x_0)^2 \end{aligned}$$

We compute the most probable value of β

$$\begin{aligned} \frac{dL}{d\beta} &= \frac{1}{\sigma^2}\sum_{i=1}^{3}(y_i - \beta x_0)(-x_0) = 0 \\ \Rightarrow \beta &= \frac{\sum_{i=1}^{3} y_i}{3x_0} \\ &= \frac{y_0 + y_0/2 + 2y_0}{3x_0} \\ &= \frac{7}{6}\frac{y_0}{x_0} \end{aligned}$$

and since $d^2L/d\beta^2 > 0$, β is the minimum of $L(\beta|D, I)$, and thus, the MPV.

c) Here consider the complete regression model, where only β is an uncertain model parameter. You are given a dataset of N measurement points $D = \{X, Y\}$, where $X = \{x_1, ..., x_N\}$ and $Y = \{y_1, ..., y_N\}$. Each observation is, therefore, fitted by the model

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

where $i = 1...N$ and ϵ_i are independent and identically distributed error terms, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.
Find the posterior distribution, negative log-likelihood function and the most probable value (MPV) of the model parameter β, given the dataset D.
Because of independent error terms, the joint likelihood function is

$$\begin{aligned} p(Y|\beta, I) &= \prod_{i=1}^{N} p(y_i|\beta, I) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right\} \\ &= \frac{1}{(2\pi)^{N/2}\sigma^N}\prod_{i=1}^{N} exp\left\{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right\} \\ &= \frac{1}{(2\pi)^{N/2}\sigma^N} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)^2\right\} \end{aligned}$$

The posterior PDF for the uncertain parameter β reads

$$\begin{aligned} p(\beta|D, I) &\propto \frac{1}{(2\pi\sigma^2)^{N/2}} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)^2\right\} \cdot p(\beta|I) \\ &\propto \frac{1}{(2\pi\sigma^2)^{N/2}} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \alpha - \beta x_i)^2\right\}, \end{aligned}$$

since $p(\beta|I) \propto 1$.

By reordering terms and simplifying, we can see that the the posterior PDF of β is again a normal distribution: $\sim \mathcal{N}(\frac{\sum_{i=1}^{N} x_i(y_i - \alpha)}{\sum_{i=1}^{N} x_i^2}, \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2})$.

The negative log-likelihood function then reads:

$$\begin{aligned} L(\beta|D, I) &= -ln(p(\beta|D, I)) \\ &= \frac{N}{2}ln(2\pi) + \frac{N}{2}ln(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i}(y_i - \alpha - \beta x_i)^2 \end{aligned}$$

The most probable value is calculated as the maximum of the posterior PDF or the minimum of the negative log-likelihood function:

$$\begin{aligned} \frac{dL(\beta|D, I)}{d\beta} &= 0 \\ \Rightarrow \frac{1}{\sigma^2}\sum_{i}(y_i - \alpha - \beta x_i)(-x_i) &= 0 \\ \Rightarrow -\frac{1}{\sigma^2}\sum_{i}(y_i x_i - \alpha x_i - \beta x_i^2) &= 0 \\ \Rightarrow \beta &= \frac{\sum_i y_i x_i - \alpha\sum_i x_i}{\sum_i x_i^2} \end{aligned}$$

Along with the fact that

$$\frac{d^2L(\beta|D, I)}{d\beta^2} = \frac{\sum_i x_i^2}{\sigma^2} > 0$$

β is the minimum of $L(\beta|D, I)$ and, thus, the MPV.

Grading scheme:
Total: 30 pts

- 10 points: subquestion a
- 10 points: subquestion b
- 10 points: subquestion c

Question 2: Non-Linear Model

Consider the mathematical model of a falling object with mass m, acceleration of gravity g and air resistance force $F_{air} = m\alpha v^2$ where α is the air resistance coefficient. Using Newton's law, the equation of motion of the falling object is:

$$m\frac{dv}{dt} = mg - m\alpha v^2 \quad (1)$$

The solution for the velocity obtained from the nonlinear differential equation (1) is

$$v(t) = v_\infty \tanh\left(\frac{g(t - t_0)}{v_\infty}\right), \quad (2)$$

where $v_\infty = \sqrt{g/\alpha}$ and t_0 is the initial time. Integrating the velocity $v(t)$ with respect to time, the solution for the vertical displacement x of the falling object is finally obtained as

$$x(t) = \frac{1}{\alpha}\ln\cosh(\sqrt{g\alpha}(t - t_0)) \quad (3)$$

Measurements for the position of the falling object are obtained by a digital camera making snapshots with an interval of Δt sec. Given the observation data $D = \{\hat{X}_1, ..., \hat{X}_N\}$ of the location of the falling object at time instances $t = \{\Delta t, ..., N\Delta t\}$, respectively, we are interested in estimating the uncertainty of the parameter α of the system given the value of the variance σ^2. Note that the measurements and the model predictions satisfy the model error equation

$$\hat{X}_k = x(k\Delta t) + E_k \quad (4)$$

where the measurement error terms E_k are independent identically distributed (i.i.d.) and follow the zero-mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

Assume a uniform prior for α and derive the expressions for the

1) Posterior PDF (probablity density function) $p(\alpha|D, \sigma)$.

$$\begin{aligned} p(\alpha|D, \sigma) &\propto p(\alpha|\sigma)p(D|\alpha, \sigma) \\ &= \mathcal{U}(\alpha; \alpha_1, \alpha_2)\prod_{k=1}^{N} p(\hat{X}_k|\alpha, \sigma) \\ &= \mathcal{U}(\alpha; \alpha_1, \alpha_2)\prod_{k=1}^{N} \mathcal{N}(\hat{X}_k; x(k\Delta t), \sigma^2) \\ &\propto \mathbb{1}_{[\alpha_1, \alpha_2]}(\alpha)\prod_{k=1}^{N} \exp\left(-\frac{(\hat{X}_k - x(k\Delta t))^2}{2\sigma^2}\right) \\ &= \mathbb{1}_{[\alpha_1, \alpha_2]}(\alpha)\exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(\hat{X}_k - x(k\Delta t))^2\right) \end{aligned}$$

The proportionality costant C can be found from the normalization property of a PDF

$$C = \int_{\alpha_1}^{\alpha_2} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(\hat{X}_k - x(k\Delta t))^2\right) d\alpha$$

2) The negative log-likelihood function $L(\alpha) = -\ln p(\alpha|D, \sigma)$.
The negative log-likelihood function is defined for $\alpha \in [\alpha_1, \alpha_2]$ and equals to:

$$L(\alpha) = -\ln p(\alpha|D, \sigma) = -\ln C - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(\hat{X}_k - x(k\Delta t))^2$$

3) Consider a simple case of only one measurement $\hat{X}_1 = 1[m]$ taken 1 s after the beginning of the fall. Assuming $g = 9.8\,[m/s^2]$, show that $\alpha = 8.366\,[1/m]$ is the most probable value of the air resistance coefficient (the value which maximizes the posterior PDF).
Hint: You do not need to solve the resulting equation. Only show that the given value is indeed a minimum.
Finding a value which maximizes the posterior is equivalent to finding a value which minimizes the negative log-likelihood.

$$\arg\min_{\alpha\in[\alpha_1, \alpha_2]} L(\alpha) = \arg\min_{\alpha\in[\alpha_1, \alpha_2]} \sum_{k=1}^{N}(\hat{X}_k - x(k\Delta t))^2$$

The minimality condition for the expression above and for one observation data point is:

$$\begin{aligned} &\frac{\partial}{\partial\alpha}(\hat{X}_1 - x(\Delta t))^2 = 0 \Leftrightarrow \\ &2(\hat{X}_1 - x(\Delta t))(-\frac{\partial}{\partial\alpha}x(\Delta t)) = 0 \Leftrightarrow \\ &\left(\hat{X}_1 - \frac{1}{\alpha}\ln\cosh(\sqrt{g\alpha}(\Delta t - t_0))\right) \times \\ &\left(\frac{\sqrt{g}(\Delta t - t_0)\tanh(\sqrt{\alpha g}(\Delta t - t_0))}{2\alpha\sqrt{\alpha}} - \frac{\ln\cosh(\sqrt{\alpha g}(\Delta t - t_0))}{\alpha^2}\right) = 0 \end{aligned}$$

After substituting all constant values, one obtains $\frac{\partial L(\alpha)}{\partial\alpha} = 8.363 \cdot 10^{-7}$ which is approximately 0. To check that $\alpha = 8.366\,[1/m]$ is a minimum, one can take the second derivative and show that it's greater than 0.

Grading scheme:
Total: 20 pts

- 8 points: subquestion 1
- 5 points: subquestion 2
- 7 points: subquestion 3