Exercise 8 Compulsory

Problem 1. (Time and energy)

Let the data be

$x(t)=\mu(t)+\sigma z(t)$,

with

$\mu(t)=h(t-\tau),\ t\in[t_0,t_1]\cap(\mathbb{Z}\cdot\delta)$,

and z a random sample from $N(0,1)$.

(a) Let $\bar\phi=\sum_t\frac{\phi(t)}{n}$, where n is such that $\bar 1=1$.

This means that if $\bar\phi=1$, then $\phi(t)=1$.

Then, $\bar 1=\sum_t\frac{1}{n}\Rightarrow\sum_t=n$.

So $\bar\phi=\sum_{i=1}^{n}\frac{\phi_i}{n}$.

Thus, n must be the number of timepoints t that is recorded.

$\phi(t)$ could be the length l from the center when an object is on the unit sphere at time t.

(b) Let the code h and the shift τ be known.

The expected noise energy is given by $\varepsilon_N=E\left[\overline{\sigma^2z^2}\right]$.

The UMVU estimate $\check\varepsilon_N$ of ε_N must be such that

$\check\varepsilon_N=E[\tilde\varepsilon_N\,|\,T]$,

where $\tilde\varepsilon_N$ is an unbiased estimator of ε_N, and T is a sufficient statistic for ε_N.

Then

(*) $E[\check\varepsilon_N]=\varepsilon_N$,

(**) $Var[\check\varepsilon_N]\le Var[\tilde\varepsilon_N]$.

Finding sufficient statistic:

$$\begin{aligned}\varepsilon_N&=E\left[\overline{\sigma^2z^2}\right]\\&=E\left[\sum_t\frac{\sigma^2z^2(t)}{N}\right]\\&=\frac1N E\left[\sum_t(x(t)-\mu(t))^2\right]\\&=\frac1N\sum_t\left(E[x^2(t)]-2E[\mu(t)x(t)]+E[\mu^2(t)]\right)\\&=\frac1N\sum_t\left(E[x^2(t)]-2h(t-\tau)E[x(t)]+h^2(t-\tau)\right)\end{aligned}$$

$x(t)\sim N(\mu(t),\sigma^2)=N(h(t-\tau),\sigma^2)$

$$\begin{aligned}\Rightarrow E[x^2(t)]&=Var[x(t)]+E[x]^2\\&=\sigma^2+h^2(t-\tau)\end{aligned}$$

$$\begin{aligned}\Rightarrow\varepsilon_N&=\frac1n\sum_t\left(\sigma^2+h^2(t-\tau)-2h(t-\tau)\cdot h(t-\tau)+h^2(t-\tau)\right)\\&=\frac1N\sum_t\sigma^2\\&=\sigma^2\end{aligned}$$

Sufficient statistic is s^2.

$\Rightarrow\check\varepsilon_N=s^2$.

Letting the loss l be given by

$l=[10\cdot\log(\hat\varepsilon_N/\varepsilon_0)-10\cdot\log(\varepsilon_N/\varepsilon_0)]^2$,

where $\hat\varepsilon_N$ is the scale equivariant UMRV estimate.

The risk function $R(\theta,\delta(x))$ is given by

$R(\theta,\delta)=E_\theta[L(\theta,\delta(X))]$,

where $L(\theta,\delta(X))$ is the loss function and $\delta(x)$ is an estimator of θ.

In our case,

$\theta=\varepsilon_N=\sigma^2$

$\delta(x)=\hat\varepsilon_N$.

So,

$$\begin{aligned}R(\varepsilon_N,\hat\varepsilon_N)&=E_{\varepsilon_N}\left[(10\cdot\log(\hat\varepsilon_N/\varepsilon_0)-10\cdot\log(\varepsilon_N/\varepsilon_0))^2\right]\\&=E_{\varepsilon_N}\left[(10\cdot\log(\hat\varepsilon_N)-10\cdot\log(\varepsilon_0)-10\cdot\log(\varepsilon_N)+10\cdot\log(\varepsilon_0))^2\right]\\&=E_{\varepsilon_N}\left[(10\cdot\log(\hat\varepsilon_N)-10\cdot\log(\varepsilon_N))^2\right]\\&=E_{\varepsilon_N}\left[100(\log(\hat\varepsilon_N)-\log(\varepsilon_N))^2\right]\\&=100E_{\varepsilon_N}\left[\left(\log\left(\tfrac{1}{\varepsilon_N}\right)-\log\left(\tfrac{1}{\hat\varepsilon_N}\right)\right)^2\right]\end{aligned}$$

$E[(X-x)^2]$ minimized when $x=E[X]$

$$\begin{aligned}E[X]&=E\left[\log\left(\tfrac{1}{\varepsilon_N}\right)\right]\\&=E\left[\log\left(\tfrac{1}{\sigma^2}\right)\right]\\&=\log\left(\tfrac{1}{\varepsilon_N}\right)\end{aligned}$$

$x_i-\mu_i=y_i=\sigma z_i$

$y_i^2=\sigma^2z_i^2$

$\overline{y^2}=\sigma^2\overline{z^2}$

$\frac{1}{\sigma^2}=\frac{\overline{z^2}}{\overline{y^2}}$

$\overline{z^2}=\frac1n\sum_{i=1}^n z_i^2$

$\sum_{i=1}^n z_i^2\sim\chi_n^2=\Gamma\left(\frac n2,2\right)$

$\Rightarrow\frac{1}{\sigma^2}\sim\Gamma\left(\frac n2\left(\frac{1}{n\overline{y^2}}\right),2\right)=\Gamma(\alpha,\beta)$

So $E\left[\log\left(\frac{1}{\sigma^2}\right)\right]=\psi(\alpha)-\log(\beta)$

$\Rightarrow\hat\varepsilon_N=e^{-(\psi(\alpha)-\log(\beta))}$

$\sqrt{Var[\check\varepsilon_N]}$ is the standard uncertainty of $\check\varepsilon_N$

$\sqrt{Var[\hat\varepsilon_N]}$ is the standard uncertainty of $\hat\varepsilon_N$

(c) Assume τ unknown, everything else known.

Let $\hat\tau=\operatorname{argmin}\overline{(x-\mu)^2}$,

$\tilde\tau=\operatorname{argmax}\overline{(x-\bar x)(\mu-\bar x)}$,

$\bar\tau=\operatorname{argmin}E[(\tau-\tau(\theta))^2\,|\,x]$,

where $\tau(\theta)$ is the Bayesian signal arrival time by assuming a uniform prior on τ

$\hat\tau$ is the τ that minimizes the mean squared error.

This is suitable, because the data is Gaussian, and by minimizing the mean squared error we get suitable estimates of the parameters.

Since the $\hat\tau$ that minimizes the MSE is the same $\hat\tau$ that maximizes the likelihood, this estimator is in harmony with the likelihood principle.

From Birnbaum's theorem we have (likelihood principle)$\Leftrightarrow$(conditionality principle + sufficiency principle), so it is in harmony with the sufficiency principle.

$\tilde\tau$ is the τ that maximizes the mean value of the dot product, and by doing that we are minimizing the angle between them.

And since $\tilde\tau=\frac1n\operatorname{argmax}\sum_{i=1}^n(x_i-\bar x)(\mu_i-\bar x)$, $\bar x$ can be taken out because it just shifts the data.

$$\begin{aligned}\Rightarrow\tilde\tau&=\frac1n\operatorname{argmax}\sum_{i=1}^n x_i\mu_i\\&=\frac1n\operatorname{argmin}\sum_{i=1}^n-2x_i\mu_i\\&=\frac1n\operatorname{argmin}\sum_{i=1}^n(x_i^2-2x_i\mu_i+\mu_i^2)\\&=\hat\tau\end{aligned}$$

So, $\tilde\tau$ is the same as $\hat\tau$, so it is also in harmony with the likelihood and sufficiency principle.

$\bar\tau$ is the τ that minimizes the square error between the τ and $\tau(\theta)$.

So $\bar\tau=E[\tau(\theta)\,|\,x]$, which is a good estimate.

$\bar\tau$ is not in harmony with the likelihood principle because no prior is uninformative.

(d) To determine the bias and standard deviations, one could use Bootstrap.

To find the standard uncertainty, one would use the standard deviations from the Bootstrap.

Finding the Cramer-Rao lower bound

$Var[\hat\tau]\ge\frac{1}{I(\tau)}$

$$\begin{aligned}I(\tau)&=-\sum_{i=1}^nE\left[\frac{\partial^2}{\partial\tau^2}l(x;\tau)\right]\\&=-\sum_{i=1}^nE\left[\frac{\partial^2}{\partial\tau^2}\left(-\frac{1}{2\sigma^2}(x_i-h_i)^2\right)\right]\\&=-\sum_{i=1}^nE\left[\frac{\partial}{\partial\tau}\left(\frac{1}{\sigma^2}(x_i-h_i)h_i'\right)\right]\\&=-\sum_{i=1}^nE\left[\frac{1}{\sigma^2}\left((-h_i')(h_i')+(x_i-h_i)h_i''\right)\right]\\&=\frac{1}{\sigma^2}\sum_{i=1}^n(h_i')^2\end{aligned}$$

$\Rightarrow Var[\hat\tau]\ge\sigma^2\left(\sum_{i=1}^n(h_i')^2\right)^{-1}$

Then $\sqrt{Var[\hat\tau]}$ can be used as the optimistic standard uncertainty.

When $h(t)=(1-|\frac t\alpha|)(|\frac t\alpha|<1)h_0$:

$$\begin{aligned}h(t-\tau)&=\left(1-\left|\tfrac{t-\tau}{\alpha}\right|\right)\left(\left|\tfrac{t-\tau}{\alpha}\right|<1\right)h_0\\&=\left(1-\left|\tfrac t\alpha-\tfrac\tau\alpha\right|\right)\left(\left|\tfrac t\alpha-\tfrac\tau\alpha\right|<1\right)h_0\end{aligned}$$

When h_0 increases $Var[\hat\tau]$ decreases

When α increases $Var[\hat\tau]$ increases

(e) Letting $H_0:\tau=\tau_0$ with level $\alpha=5\%$.

The likelihood of H_0 is

$\lambda(x)=\frac{\sup_{H_0}L}{\sup L}$,

where L is the likelihood function of τ.

Often $\sup_{H_0}L=L(\tau_0)$ and $\sup L=L(\hat\tau)$

H_0 is rejected if $\lambda\le\lambda_\alpha$.

By plugging our data x into L, we get the likelihood using $\hat\tau$ and τ_0.

Then, if the likelihood using τ_0 smaller than the likelihood using $\hat\tau$, we reject H_0.

In other words, the likelihood of getting the data that we have by using τ_0 is smaller than when using $\hat\tau$, we reject H_0.

$P(\hat\tau_1(X)\le\tau\le\hat\tau_2(X))=P(\tau\le\hat\tau_2(X))-P(\tau\le\hat\tau_1(X))$

$\alpha=\sup P(\lambda(X)\le\lambda_\alpha)$

$$\begin{aligned}P(\hat\tau_1(X)\le\tau\le\hat\tau_2(X))&=1-P(\hat\tau_2(X)\le\tau)-1+P(\hat\tau_1(X)\le\tau)\\&=P(\hat\tau_1(X)\le\tau)-P(\hat\tau_2(X)\le\tau)\\&=1-\alpha\end{aligned}$$

When $\alpha\uparrow1$, $P(\hat\tau_1(X)\le\tau\le\hat\tau_2(X))\ge0$, which means that H_0 is never rejected.

(f) Assuming the situation is safe if $H_0:\tau\ge\tau_0$.

The test $B=(T\le t_\alpha)$ is 1 if T is lower than some value, i.e. we reject H_0. So it is reasonable.

Problem 2. (Space)

Let the data be

$x_i=\mu_i+\sigma z_i$

where

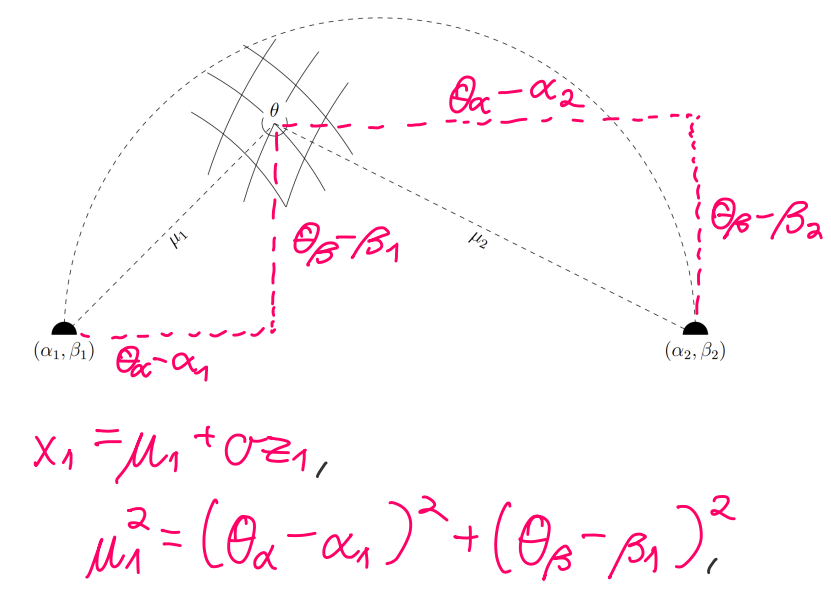
$z_1,\dots,z_n\sim N(0,1)$,

$\mu_i^2=(\theta_\alpha-\alpha_i)^2+(\theta_\beta-\beta_i)^2$.

Assume that the position $\theta=(\theta_\alpha,\theta_\beta)$ is the unknown model parameter.

(a) In the case where $n=2$, we have two sattelites at positions (α_1,β_2) and (α_2,β_2).

By Pytagoras we have



$x_1=\mu_1+\sigma z_1$,

$\mu_1^2=(\theta_\alpha-\alpha_1)^2+(\theta_\beta-\beta_1)^2$,

$x_2=\mu_2+\sigma z_2$,

$\mu_2^2=(\theta_\alpha-\alpha_2)^2+(\theta_\beta-\beta_2)^2$.

(b) The MLE can be found using the fact that $x_i\sim N(\mu_i,\sigma^2)$, and with many sattelites,

$$\begin{aligned}L&=\prod_{i=1}^n f(x_i\,|\,\dots)\\&=\prod_{i=1}^n(2\pi\sigma^2)^{-\frac12}e^{-\frac{1}{2\sigma^2}(x_i-\mu_i)^2}\\&=(2\pi\sigma^2)^{-\frac n2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu_i)^2}\end{aligned}$$

where $\mu_i=\sqrt{(\theta_\alpha-\alpha_i)^2+(\theta_\beta-\beta_i)^2}$.

So,

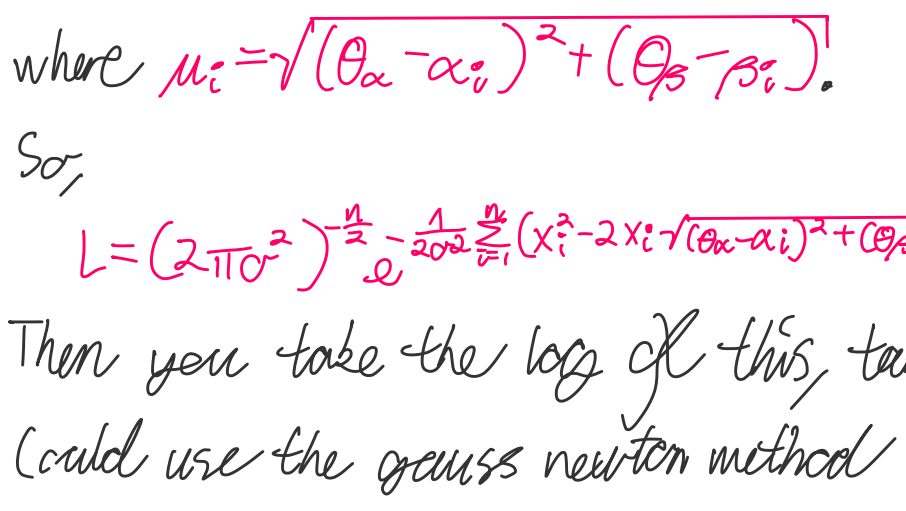
$L=(2\pi\sigma^2)^{-\frac n2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i^2-2x_i\sqrt{(\theta_\alpha-\alpha_i)^2+(\theta_\beta-\beta_i)^2}+(\theta_\alpha-\alpha_i)^2+(\theta_\beta-\beta_i)^2)}$

Then you take the log of this, take the derivative of θ_α and θ_β, then put equal to zero.

Could use the gauss newton method to minimize $\sum_{i=1}^n(x_i-\mu_i)^2$.

To compute the Bayesian estimates, we can use MCMC on $\pi(\theta|x)\propto\pi(x|\theta)\pi(\theta)$ where $\pi(x|\theta)$ is the likelihood and $\pi(\theta)$ is the prior.

(c) In the case $n=2$, we have



$\Rightarrow$ (1) $\theta_\alpha^2+\theta_\beta^2=\mu_1^2$

(2) $(\alpha_2-\alpha_1-\theta_\alpha)^2+\theta_\beta^2=\mu_2^2$

$\Rightarrow(\theta_\alpha^2+\theta_\beta^2)-((\alpha_2-\alpha_1-\theta_\alpha)^2+\theta_\beta^2)=\mu_1^2-\mu_2^2$

$\Rightarrow2(\alpha_2-\alpha_1)\theta_\alpha-(\alpha_2-\alpha_1)^2=\mu_1^2-\mu_2^2$

$\Rightarrow\theta_\alpha=\frac{1}{2(\alpha_2-\alpha_1)}(\mu_1^2-\mu_2^2+(\alpha_2-\alpha_1)^2)$

$\Rightarrow\frac{1}{2(\alpha_2-\alpha_1)}(\mu_1^2-\mu_2^2+(\alpha_2-\alpha_1)^2)+\theta_\beta^2=\mu_1^2$

$$\begin{aligned}\Rightarrow\theta_\beta&=\sqrt{\mu_1^2-\theta_\alpha}\\&=\sqrt{\mu_1^2-\frac{1}{2(\alpha_2-\alpha_1)}(\mu_1^2-\mu_2^2+(\alpha_2-\alpha_1)^2)}\end{aligned}$$

Then, using that $E[x_i]=\mu_i$

$\Rightarrow\hat\theta_\alpha=\frac{1}{2(\alpha_2-\alpha_1)}(x_1^2-x_2^2+(\alpha_2-\alpha_1)^2)$

$\hat\theta_\beta=\sqrt{x_1^2-\hat\theta_\alpha}$

$$\begin{aligned}E[\hat\theta_\alpha]&=\frac{1}{2(\alpha_2-\alpha_1)}(E[x_1^2]-E[x_2^2]+(\alpha_2-\alpha_1)^2)\\&=\frac{1}{2(\alpha_2-\alpha_1)}(\sigma^2+\mu_1^2-\sigma^2-\mu_2^2+(\alpha_2-\alpha_1)^2)\\&=\theta_\alpha\end{aligned}$$

$\hat\theta_\alpha$ is unbiased.

$$\begin{aligned}L(x|\theta)&=\prod_{i=1}^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x_i-\mu_i)^2}\\&=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^2(x_i-\mu_i)^2}\end{aligned}$$

$$\begin{aligned}l(x|\theta)&=\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)-\frac{1}{2\sigma^2}\sum_{i=1}^2(x_i-\mu_i)^2\\&\propto-\frac{1}{2\sigma^2}\sum_{i=1}^2(x_i-\mu_i)^2\\&=-\frac{1}{2\sigma^2}\sum_{i=1}^2\left(x_i-\sqrt{(\theta_\alpha-\alpha_i)^2+(\theta_\beta-\beta_i)^2}\right)^2\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\theta_\alpha}l(x|\theta)&=-\frac{1}{2\sigma^2}\sum_{i=1}^2 2(x_i-\mu_i)\frac{\partial}{\partial\theta_\alpha}\mu_i\\&=-\frac{1}{2\sigma^2}\sum_{i=1}^2 2(x_i-\mu_i)\frac12\mu_i^{-1}2(\theta_\alpha-\alpha_i)\\&=\frac{-1}{\sigma^2}\sum_{i=1}^2\mu_i^{-1}(x_i-\mu_i)(\theta_\alpha-\alpha_i)\\&=0\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\theta_\beta}l(x|\theta)&=\frac{-1}{\sigma^2}\sum_{i=1}^2\mu_i^{-1}(x_i-\mu_i)(\theta_\beta-\beta_i)\\&=0\end{aligned}$$

$\Rightarrow\sum_{i=1}^2\mu_i^{-1}(x_i-\mu_i)(\theta_\alpha-\alpha_i)=\sum_{i=1}^2\mu_i^{-1}(x_i-\mu_i)(\theta_\beta-\beta_i)$

other idea,

(d) $Var[T]=E[(T-\theta)(T-\theta)^T]$

$Var[T]\ge\frac{1}{I(\theta)}$

$I(\theta)_{ij}=E\left[\left(\frac{\partial}{\partial\theta_i}\log(f(x|\theta))\right)\left(\frac{\partial}{\partial\theta_j}\log(f(x|\theta))\right)\Big|\,\theta\right]$

$\frac{\partial}{\partial\theta_i}\log(f(x|\theta))=-\frac{1}{2\sigma^2}\sum_{j=1}^n\frac{\partial}{\partial\theta_i}(x_k-\mu_j)^2$

$$\begin{aligned}\frac{\partial}{\partial\theta_i}(x_j-\mu_j)^2&=2(x_j-\mu_j)\frac{\partial}{\partial\theta_i}\sqrt{(\theta_\alpha-\alpha_j)^2+(\theta_\beta-\beta_j)^2}\\&=2(x_j-\mu_j)\frac12\left((\theta_\alpha-\alpha_j)^2+(\theta_\beta-\beta_j)^2\right)^{-\frac12}2(\theta_\gamma-\overset{\substack{\text{Either }\alpha\text{ or }\beta\\\downarrow}}{\gamma_j})\end{aligned}$$

$$\begin{aligned}\Rightarrow\frac{\partial}{\partial\theta_i}\log(f(x|\theta))&=-\frac{1}{\sigma^2}\sum_{j=1}^n\frac{(x_j-\mu_j)(\theta_i-\gamma_j)}{\sqrt{(\theta_\alpha-\alpha_j)^2+(\theta_\beta-\beta_j)^2}}\\&=-\frac{1}{\sigma^2}\sum_{j=1}^n\mu_j^{-1}(x_j-\mu_j)(\theta_\gamma-\gamma_j)\end{aligned}$$

$$\begin{aligned}I(\theta)_{ij}&=\frac{1}{\sigma^4}E\left[\left(\sum_{k=1}^n\mu_k^{-1}(x_k-\mu_k)(\theta_i-i_k)\right)\left(\sum_{l=1}^n\mu_l^{-1}(x_l-\mu_l)(\theta_j-j_l)\right)\right]\\&=\frac{1}{\sigma^4}\left(\sum_{k=1}^n\mu_k^{-1}(\theta_i-i_k)\right)\left(\sum_{l=1}^n\mu_l^{-1}(\theta_i-i_l)\right)\underbrace{E[(x_k-\mu_k)(x_l-\mu_l)]}_{k=l:\ \sigma^2,\ \text{zero otherwise}}\\&=\frac{1}{\sigma^2}\left(\sum_{k=1}^n\mu_k^{-1}(\theta_i-i_k)\right)\left(\sum_{l=1}^n\mu_l^{-1}(\theta_i-i_l)\right)\end{aligned}$$

$$I(\theta)=\frac{1}{\sigma^2}\begin{pmatrix}\sum_{k=1}^n\mu_k^{-1}(\theta_\alpha-\alpha_k)^2&\sum_{k=1}^n\mu_k^{-1}(\theta_\alpha-\alpha_k)(\theta_\beta-\beta_k)\\\sum_{k=1}^n\mu_k^{-1}(\theta_\alpha-\alpha_k)(\theta_\beta-\beta_k)&\sum_{k=1}^n\mu_k^{-1}(\theta_\beta-\beta_k)^2\end{pmatrix}$$

$\Rightarrow Var[T]=I^{-1}(\theta)$

(e) $\sqrt{tr(Var[T])}$ is a reasonable measure of the accuracy because we measure the standard deviation of the sum of variances.

$$\begin{aligned}\sqrt{tr(Var[T])}&=\sqrt{tr(I^{-1}(\theta))}\\&=\sqrt{\sigma^2tr((M^TM)^{-1})}\\&=\sigma k_D\end{aligned}$$

k_D is high if the sattelites are on the same line

(f) The likelihood of H_0 is

$\lambda(x)=\frac{\sup_{H_0}L}{\sup L}$

where L is the likelihood.

We reject H_0 if λ is smaller than some threshold λ_α.

In this case, $H_0:\theta\in A$

An alternative test is $B=(T\in R)$.

The region R is an area around A, and it is similar to A if we want to test if θ is in A

The best test is the one with the largest power.

A real life situation could be someone with an ankle monitor.