Toy problem

Joseph D Romano

We give a pedagogical discussion of optimal estimators and marginalization over calibration uncertainty in the context of single and multibaseline measurements for a simple toy problem. We show that the same final results are obtained working with either the full set of measurements or with the optimal estimators for the individual baselines.

I. SOME USEFUL RESULTS

A. Maximization

- Consider a set of measurements

$$\boldsymbol{x} = M \cdot \boldsymbol{\mu} + \boldsymbol{n} , \qquad \langle \boldsymbol{n} \rangle = 0 , \qquad \langle \boldsymbol{n}\boldsymbol{n}^T \rangle = C \tag{1}$$

consisting of a deterministic part defined by (known) transformation matrix M and parameters $\boldsymbol{\mu}$ (not necessarily of the same dimension as $\boldsymbol{x}$) and zero-mean, Gaussian-distributed noise $\boldsymbol{n}$ with covariance matrix C.

- Likelihood function:

$$L(\boldsymbol{x}|C, \boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi)^{N_x} \det C}} \exp\left[-\frac{1}{2} (\boldsymbol{x} - M \cdot \boldsymbol{\mu})^T \cdot C^{-1} \cdot (\boldsymbol{x} - M \cdot \boldsymbol{\mu}) \right] \tag{2}$$

- Maximizing the likelihood wrt the parameters $\boldsymbol{\mu}$ is equivalent to minimizing

$$\chi^2(\boldsymbol{\mu}) \equiv (\boldsymbol{x} - M \cdot \boldsymbol{\mu}) \cdot C^{-1} \cdot (\boldsymbol{x} - M \cdot \boldsymbol{\mu}) \tag{3}$$

The optimal estimators are:

$$\boldsymbol{\mu}_{\rm opt} = (M^T \cdot C^{-1} \cdot M)^{-1} \cdot M^T \cdot C^{-1} \cdot \boldsymbol{x} \tag{4}$$

with

$$\langle \boldsymbol{\mu}_{\rm opt} \rangle = \boldsymbol{\mu} , \qquad C_{\rm opt} \equiv \langle \boldsymbol{\mu}_{\rm opt} \boldsymbol{\mu}_{\rm opt}^T \rangle - \langle \boldsymbol{\mu}_{\rm opt} \rangle \langle \boldsymbol{\mu}_{\rm opt}^T \rangle = (M^T \cdot C^{-1} \cdot M)^{-1} \tag{5}$$

- In terms of the optimal estimators,

$$\chi^2(\boldsymbol{\mu}) = (\boldsymbol{\mu}_{\rm opt} - \boldsymbol{\mu})^T \cdot C_{\rm opt}^{-1} \cdot (\boldsymbol{\mu}_{\rm opt} - \boldsymbol{\mu}) + \chi^2(\boldsymbol{\mu}_{\rm opt}) \tag{6}$$

Since the second term is independent of $\boldsymbol{\mu}$, it is a normalization constant that can be ignored when making probabilistic statements about $\boldsymbol{\mu}$. *Hence, it is often more convenient to work with the optimal estimators* $\boldsymbol{\mu}_{\rm opt} = \boldsymbol{\mu}_{\rm opt}(\boldsymbol{x})$ *instead of* $\boldsymbol{x}$.

B. Marginalization

- Alternatively, suppose

$$\boldsymbol{x} = M \cdot \boldsymbol{\lambda} + \boldsymbol{n} , \qquad \langle \boldsymbol{n} \rangle = 0 , \qquad \langle \boldsymbol{n}\boldsymbol{n}^T \rangle = C \tag{7}$$

where the parameters of the deterministic part (now denoted by $\boldsymbol{\lambda}$) are so-called *nuisance parameters*, whose values we don't particularly care about.

- Given a prior distribution $\pi(\boldsymbol{\lambda})$ for $\boldsymbol{\lambda}$, we can marginalize the likelihood function over these parameters to obtain a *marginalized* likelihood, independent of $\boldsymbol{\lambda}$:

$$L(\boldsymbol{x}|C) = \int d^{N_\lambda}\boldsymbol{\lambda}\, L(\boldsymbol{x}|C, \boldsymbol{\lambda}) \pi(\boldsymbol{\lambda}) \tag{8}$$

- If we assume a multivariate-Gaussian prior:

$$\pi(\boldsymbol{\lambda}|D,\boldsymbol{\lambda}_0) = \frac{1}{\sqrt{(2\pi)^{N_\lambda}\det D}}\exp\left[-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_0)^T\cdot D^{-1}\cdot(\boldsymbol{\lambda}-\boldsymbol{\lambda}_0)\right] \tag{9}$$

the marginalization can be done analytically:

$$L(\boldsymbol{x}|C,D,\boldsymbol{\lambda}_0) \equiv \int d^{N_\lambda}\boldsymbol{\lambda}\, L(\boldsymbol{x}|C,\boldsymbol{\lambda})\pi(\boldsymbol{\lambda}|D,\boldsymbol{\lambda}_0) \tag{10}$$

$$= \frac{1}{\sqrt{(2\pi)^{N_x}\det E}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-M\cdot\boldsymbol{\lambda}_0)^T\cdot E^{-1}\cdot(\boldsymbol{x}-M\cdot\boldsymbol{\lambda}_0)\right] \tag{11}$$

where

$$E^{-1} = C^{-1} - C^{-1}\cdot M\cdot\left(M^T\cdot C^{-1}\cdot M + D^{-1}\right)^{-1}\cdot M^T\cdot C^{-1} \tag{12}$$

II. TOY MODELS

- We consider 4 different cases:
  1) Single baseline measurements with no calibration uncertainty
  2) Single baseline measurements with calibration uncertainty
  3) Two baseline measurements with no calibration uncertainty
  4) Two baseline measurements with calibration uncertainty

A. Single baseline measurements with no calibration uncertainty

- Let the vector $\boldsymbol{x}$ correspond to a single set of measurements:

$$\boldsymbol{x} \leftrightarrow x_i = \mu + n_i\,, \quad i = 1,2,\cdots,N \tag{13}$$

with common mean μ, and zero-mean, Gaussian-distributed, white noise n_i:

$$\langle n_i\rangle = 0\,, \qquad C_{ij} = \langle n_i n_j\rangle = \sigma^2\,\delta_{ij} \tag{14}$$

- In the absence of calibration uncertainty:

$$\chi^2(\mu) = \sum_i \frac{(x_i-\mu)^2}{\sigma^2} \tag{15}$$

Here the transformation matrix M is just the N-dimensional vector, $M \leftrightarrow M_i = 1$ for all i.

- Using the general results of Sec. I, the optimal estimator and corresponding uncertainity are:

$$\mu_{\rm opt} = \frac{1}{N}\sum_i x_i \equiv \bar{x}\,, \qquad C_{\rm opt} \equiv \sigma^2_{\rm opt} = \frac{\sigma^2}{N} \equiv \bar{\sigma}^2 \tag{16}$$

which are just the sample mean and its corresponding variance.

- Using Eq. (6), it follows that

$$\chi^2(\mu) = \frac{(\mu_{\rm opt}-\mu)^2}{\sigma^2_{\rm opt}} - \chi^2(\mu_{\rm opt}) = \frac{(\bar{x}-\mu)^2}{\bar{\sigma}^2} - \sum_i\frac{(x_i-\bar{x})^2}{\sigma^2} \tag{17}$$

where the last term is independent of μ.

- Thus, the μ-dependent part of the likelihood function is proportional to

$$\exp\left[-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\bar{\sigma}^2}\right] \tag{18}$$

which is a Gaussian distribution for μ centered on the sample mean $\bar{x}$ with variance $\bar{\sigma}^2 = \sigma^2/N$.