Best Linear Prediction: inverting a linear model plus additive noise, numerical example

Video Presentation: http://personales.upv.es/asala/YT/V/vcinv2EN.html

Objectives: understanding the relationship between formulae from "best linear prediction" (minimum variance of prediction error) based on the covariance matrix of a couple of random variables (a, b) and the linear models with additive noise $a = \theta b + \epsilon$ which we can identify from such formulae. Actually, the concrete goal of this material is obtaining the "inverse" model $b = \eta a + \varepsilon$ in an example; inverse will be understood in an "statistical" sense (minimum variance of the prediction error).

Table of Contents

Preliminaries and background concepts

Best linear prediction

Consider two random variables a and b, assuming zero mean (otherwise, we would change to increments around the mean, with no loss of generality).

If the covariance matrix is:

$\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$, with $\Sigma_{ab} = \Sigma_{ba}^T$ (in the multivariate case), then the best linear prediction of a given b is:

$\hat{a} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot b$, with a prediction error $a - \hat{a}$ having a variance given by: $\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}^T$.

En the case of a normally distributed joint distribution, as once mean and variance are known we can build the whole probability density, we can asser that the conditional probability $a|b$ is $N(\hat{a}, \Sigma_{e,a}) = N(\theta b, \Sigma_{e,a})$.

We can recreate that estimated conditional distribution with $a = \theta \cdot b + \epsilon$, being $\epsilon \sim N(0, \Sigma_{e,a})$, regardless of whether it is "physically true" or not (most likely not).

Covariance matrix associated to a linear model with additive noise

Given a model $a = \theta b + \epsilon$, with $b \sim N(0, \Sigma_{bb})$, $\epsilon \sim N(0, \Sigma_e)$, being ε statistically independent of b, then:

$$\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^T] = E[\theta b b^T \theta^T + \epsilon b^T \theta^T + \theta b \epsilon^T + \epsilon \epsilon^T]$$
$$= \theta E[bb^T]\theta^T + E[\epsilon b^T]\theta^T + \theta E[b\epsilon^T] + E[\epsilon\epsilon^T] = \theta \Sigma_{bb} \theta^T + \Sigma_e$$

$$\Sigma_{ab} = E[(\theta b + \epsilon) b^T] = \theta E[bb^T] + E[\epsilon b^T] = \theta \Sigma_{bb}$$

So the joint covariance matrix of a y b, associated to such a model, would be:

$$\Sigma := \begin{pmatrix} \theta \Sigma_{bb} \theta^T + \Sigma_e & \theta \Sigma_{bb} \\ \Sigma_{bb} \theta^T & \Sigma_{bb} \end{pmatrix}$$

Identified models from a covariance matrix

Consider $\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$.

If we denote $\theta := \Sigma_{ab} \Sigma_{bb}^{-1}$,then the best linear prediction of a given b is $\hat{a} = \theta \cdot b$, with a prediction error variance $\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1} \cdot \Sigma_{bb} \cdot \Sigma_{bb}^{-1} \Sigma_{ab}^T = \Sigma_{aa} - \theta \Sigma_{bb} \theta^T$.

if we assume a model:

$a = \theta \cdot b + \epsilon$, with $\epsilon \sim N(0, \Sigma_{e,a})$ independent of b, and b a zero-mean random variable with variance Σ_{bb}

then we would have $\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^T] = \theta \Sigma_{bb} \theta^T + \Sigma_{e,a}$ returning expressions already seen above.

Also $\Sigma_{ab} = \theta \Sigma_{bb} = \Sigma_{ab} \Sigma_{bb}^{-1} \cdot \Sigma_{bb}$ would give the correct covariance.

So, the linear model with additive noise $a = \theta b + \epsilon$, with variance of b being Σ_{bb} and variance of ϵ being $\Sigma_{e,a}$ "explains" the whole joint covariance matrix between a and b.

Example: inversion of a linear model (static)

Let us assume we have $b = coef \cdot a + \varepsilon$, with an *a priori* variance of a equal to 4, variance of ε equal to 1.75. The statistically optimal "inverse" model is NOT $a = (b - \epsilon)/coef$, understanding "inverse" as the linear prediction "$\hat{a}$" of "a" with lowest variance of the error $a - \hat{a}$.

```
vz_a=4; vzaeps=1.75;
```

Thus, covariance between a and b is

```
coef=0.8;
covab=coef*vz_a
```

```
covab = 3.2000
```

and the variance of b that the model predicts is

```
vza_b=vzaeps+coef*vz_a*coef
```

```
vza_b = 4.3100
```

```
MatrizVC_Sigma=[vz_a covab;covab vza_b]
```

```
MatrizVC_Sigma = 2×2
    4.0000    3.2000
    3.2000    4.3100
```

- Best prediction of b given a is, obviously, the model we were starting from:

```
covab/vz_a % = coef
```

```
ans = 0.8000
```

```
vzaerrb=vza_b-covab^2/vz_a % = vza eps
```

```
vzaerrb = 1.7500
```

- The best linear prediction of a given b is NOT $coef^{-1} \cdot b$:

```
eta=covab/vza_b
```

```
eta = 0.7425
```

```
1/coef
```

```
ans = 1.2500
```

```
vzaerra=vz_a-covab^2/vza_b
```

```
vzaerra = 1.6241
```

Another way to obtain the variance of the error of "a given b", arises from the fact that

$a - \eta b = [1 \quad -\eta] \cdot \begin{pmatrix} a \\ b \end{pmatrix}$, so we can evaluate

```
[1 -eta]*MatrizVC_Sigma*[1;-eta]
```

```
ans = 1.6241
```

- Indeed if I now test the variance of $a - coef^{-1}b$, I get:

```
[1 -1/coef]*MatrizVC_Sigma*[1;-1/coef]
```

```
ans = 2.7344
```

which is larger than `vzaerra`, so the "algebraic inverse" model is not the one with "least prediction error variance".

Conclusions

Inverting a model in statistics is more "involved" than solving for an unknown in an algebraic equation: adding noise makes the original data non-recoverable.

In a "time series" $x_{k+1} = \theta \cdot x_k + \epsilon_k$, this expression (or variations thereof) is named as the "forward" equation, and the "backwards" equation for optimal estimation of the past given the present is $x_k = \eta x_{k+1} + \varepsilon'_{k+1}$, where $\eta \neq \theta^{-1}$ and neither variance of ϵ_k and ε'_k are coincident with what the algebraic inversion $\theta^{-1}(x_k - \epsilon_{k-1})$ would suggest.