Best Linear Prediction: identified linear model with additive noise (II), numerical example

Video:

http://personales.upv.es/asala/YT/V/vcinv1EN.html, http://personales.upv.es/asala/YT/V/vcinv1bEN.html

Objectives: Understand the relationship between the "best (minimum-variance) linear predictor", $\widehat{a} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot b$, based in variance-covariance formulae of, say, two random variables (a, b), versus linear models with additive noise $a = \theta b + \epsilon$ which can be identified from the linear predictor formulae.

*This code ran without errors in Matlab `R2022a`

Table of Contents

Theory

From "best linear prediction" to "linear model with additive noise"

Given two random variables a and b, we assume zero mean to simplify developments (there is no loss of generality, as we can make a change of variable to "increments around the mean").

If their symmetric variance-covariance matrix is:

$\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$, with $\Sigma_{ab} = \Sigma_{ba}^T$ (so we are also dealing with a multivariate case, if needed),

then, the best linear predictor of a given b is:

$$\widehat{a} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot b,$$

with a prediction error $a - \widehat{a}$ having zero conditional mean (given b) , with variance given by:

$\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}^T.$

In the case of a normal distribution, as knowledge of mean and variance is enough to build the whole density function, then the conditional probability distribution of $a|b$ has the form $N(\widehat{a}, \Sigma_{e,a})$ which, well, amounts to $N(\theta \cdot b, \Sigma_{e,a})$ being $\theta := \Sigma_{ab}\Sigma_{bb}^{-1}$.

For instance, a way to generate the above conditional distribution would be with a "linear model with additive noise" expressed as $a = \theta \cdot b + \epsilon$, being $\epsilon \sim N(0, \Sigma_{e,a})$... Clearly, this may or not be the "true" underlying model that generated Σ, of course.

The reverse: covariance matrix associated to a linear model with additive noise

Consider now an arbitrary θ, and a linear model $a = \theta b + \epsilon$, where $b \sim N(0, \Sigma_{bb})$, and additive noise is $\epsilon \sim N(0, \Sigma_e)$, 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 covarianve matrix of the vector random variable $\begin{pmatrix} a \\ b \end{pmatrix}$ would be:

$$\Sigma := E\left[\begin{pmatrix} a \\ b \end{pmatrix}\begin{pmatrix} a^T & b^T \end{pmatrix}\right] = \begin{pmatrix} \theta \Sigma_{bb} \theta^T + \Sigma_e & \theta \Sigma_{bb} \\ \Sigma_{bb} \theta^T & \Sigma_{bb} \end{pmatrix}$$

There is, evidently a relationship between the first idea (from VC matriz to linear model) and the second one (from linear model to VC matrix), which will be explored in this material.

Example: identifying a model from a VC matrix

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

Predicting "a" given "b", with linear plus additive noise model

If we call $\theta := \Sigma_{ab}\Sigma_{bb}^{-1}$, the best linear prediction 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 now assume a model

$a = \theta \cdot b + \epsilon$, being $\epsilon \sim N(0, \Sigma_{e,a})$ an additive noise, statistically independent of a and b, and b is assumed to be a random variable with zero mean and variance Σ_{bb}, then we would have $\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^{T}] = \theta\Sigma_{bb}\theta^{T} + \Sigma_{e,a}$ which indeed holds..

Furthermore, $\Sigma_{ab} = \theta\Sigma_{bb} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot \Sigma_{bb}$ also renders the covariance that appeared in matrix Σ.

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

```
Sigma=[4 3;3 8]; %cov. matrix of (a,b)
standarised=1;
if(standarised) %scale to  stddev = 1
    dsvt=sqrt(diag(Sigma)); %standard dev. of each variable
    EscM=inv(diag(dsvt)); %scaling matrix
    Sigma=EscM*Sigma*EscM'
end
```

```
Sigma = 2×2
    1.0000    0.5303
    0.5303    1.0000
```

```
eig(Sigma) %must be positive (semi)definite to have a statistical meaning.
```

```
ans = 2×1
    0.4697
    1.5303
```

```
theta=Sigma(1,2)*inv(Sigma(2,2))
```

```
theta = 0.5303
```

```
vce_ea=Sigma(1,1)-Sigma(1,2)*inv(Sigma(2,2))*Sigma(2,1)
```

```
vce_ea = 0.7188
```

```
std_desv_ea=sqrt(vce_ea)
```

```
std_desv_ea = 0.8478
```

So, covariance between a and b is explained $a = \theta \cdot b + \epsilon$... this does not mean that this model is the "underlying physics" generating the data coming from "real world" stuff... we need validation sets, testing actual whiteness and uncorrelation of the residuals, etc. (out of the scope of this material)... and, well, anything will be ultimately invalidated (nature is very complex and nonlinear).

Predicting "b" given "a", with linear plus additive noise model

If we interchange "a" and "b" everywhere we would get a model $b = \eta a + \varepsilon$, with $\eta = \Sigma_{ba}\Sigma_{aa}^{-1}$ (best linear prediction), and additive noise $\varepsilon \sim N(0, \Sigma_{e,b})$, independent from a and b, with variance $\Sigma_{e,b} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ba}^T = \Sigma_{bb} - \eta\Sigma_{aa}\eta^T$

Hence, analogously to the above, the model $b = \eta a + \varepsilon$ with a having variance Σ_{aa} and ε with variance $\sigma_{e,b}$ also explains the observed variances and covariance between a and b.

```
eta=Sigma(2,1)*inv(Sigma(1,1))
```

```
eta = 0.5303
```

```
vza_eb=Sigma(2,2)-Sigma(2,1)*inv(Sigma(1,1))*Sigma(1,2)
```

```
vza_eb = 0.7187
```

```
desvtip_eb=sqrt(vza_eb)
```

```
desvtip_eb = 0.8478
```

Relationship between both identified models

Which one is the "true" model? $a = \theta \cdot b + \epsilon$ or $b = \eta a + \varepsilon$?

Well, in a "statistical" sense "both" are true... Physically maybe "none" is true...

If we were in a "deterministic" case, assuming scalar a and b if we believe the model $a = \theta b + \epsilon$, this would imply that we believe $b = \theta^{-1}a - \theta^{-1}\epsilon$, so $\eta = \theta^{-1}$, but this is NOT the case with the models we obtained:

```
inv(theta) %no es "eta"
```

```
ans = 1.8856
```

```
inv(theta)*std_desv_ea %no es desvtip_eb
```

```
ans = 1.5986
```

Let us draw the output of both models, with some additional stuff to get a glimpse of all this:

- Some samples of the bidimensional normal random variable (a, b).

```
R = mvnrnd([0 0],Sigma,5000);
```

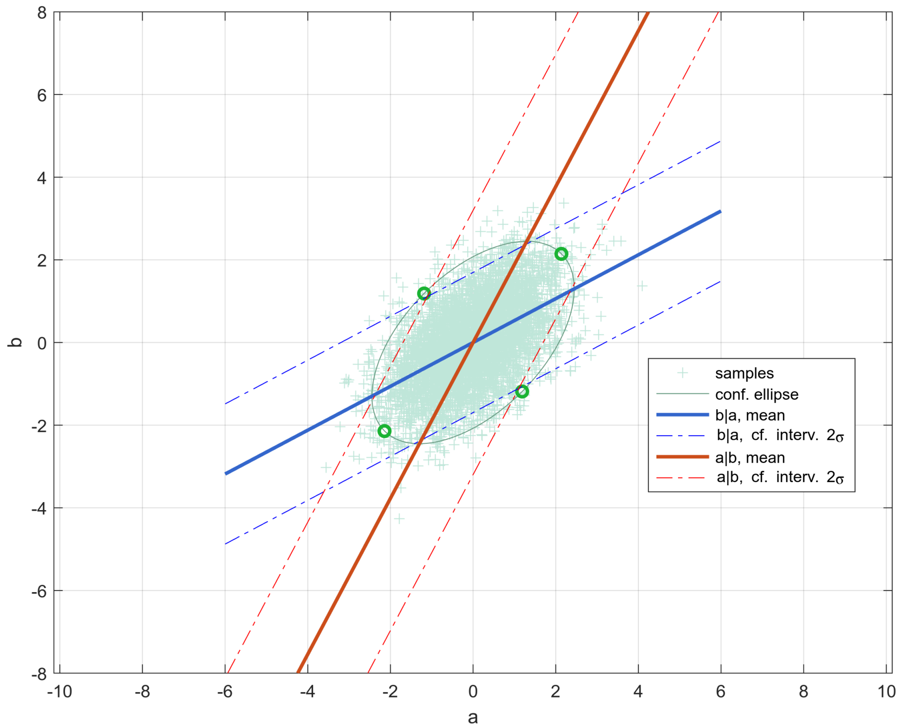
```
plot(R(:,1),R(:,2),'+',Color=[0.75 0.9 0.85]) %muestras
hold on
```

- 95% confidence ellipsoid (if normally-distributed data assumed)

```
[V,D]=eig(Sigma);
alpha=(0:360)*2*pi/360;
circle_o=[sin(alpha); cos(alpha)];
ell=V*sqrt(D)*sqrt(chi2inv(0.95,2))*circle_o;
plot(ell(1,:),ell(2,:),Color=[.4 .6 .5])
%axis of the elipse
L=round((length(alpha)-1)/4);
plot(ell(1,1:L:end),ell(2,1:L:end),'o',Color=[.1 .7 .2],LineWidth=2)
```

- Models of "a given b", and "b given a" (mean plus confidence intervals, $\pm 2\sigma$)

```
range_a=[-6 6];
range_b=[-8 8];
plot(range_a,eta*range_a,LineWidth=2,Color=[0.2 0.4 0.8]), %hold on
plot(range_a,eta*range_a+2*desvtip_eb,'-.b')
plot(range_a,eta*range_a-2*desvtip_eb,'-.b')
plot(theta*range_b,range_b,LineWidth=2,Color=[.8 .3 .1]),
plot(theta*range_b-2*std_desv_ea,range_b,'-.r')
plot(theta*range_b+2*std_desv_ea,range_b,'-.r')
hold off, grid on, axis equal
legend("samples","conf. ellipse","","b|a, mean","b|a, cf. interv. 2\sigma","","a|b, mea
xlabel("a"),ylabel("b")
```



*Note the "b|a" line cutting the ellipsopid at the "vertical" tangents, and the "a|b" one cutting it at the "horizontal" tangent. The model given by the major axis of the ellipse is called the "total least squares" fit, or "first principal component"; details are out of the scope of this material, for brevity.