PROJECT 'ESTIMATION OF RESISTANCE'

INTRODUCTION

The resistance of carbon film resistors with a cylindrical form are specified by means of colour bands that are painted on the surface. The colours encode the so-called nominal value R_{nom} of the resistor. There is also a colour band that indicates the tolerance of the nominal value. If the tolerance is 5%, then the true resistance R is somewhere between $0.95R_{nom}$ and $1.05R_{nom}$.

Suppose that a more accurate assessment of the true resistance is needed. A multimeter of a given type can measure the resistance with an overall uncertainty (standard deviation) of 2% of the true resistance. 'Overall' means: averaged over all possible resistors, and over all conditions (temperature, etc.). Figure 1 shows a scatter diagram coming from a batch of 223 resistors. For each resistor the relative measurement is plotted against the relative error of the nominal value.

We have two sources of knowledge:

- The nominal value indicated by the colour bands.
- The measured value.

Both sources provide uncertain information. The goal of this project is to design and evaluate an estimator that combines both sources of information. For that purpose, the dataset shown in figure 1 must be used.

Literature:

- Chapter 3
- Chapter 5
- Section 9.2

Requirements:

- Matlab
- Datafile: CPESE_PROJ_PARM1.ZIP (contains the dataset)

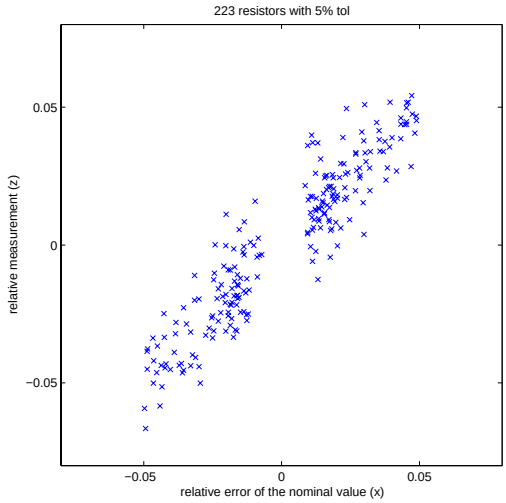

Figure 1. Scatter diagram of relative errors.

THE DESIGN

Problem definition

In order to have a solution that does not depend on the actual value of the resistance R, we rephrase the problem in terms of relative errors and relative measurements. For that purpose we define the following variables:

- The relative error of the nominal value: $x \equiv \dfrac{R_{nom} - R}{R}$ (1)

- The (relative) measurement: $z \equiv \dfrac{R_{nom} - R_{meas}}{R_{nom}}$ (2)

In these expressions, z is regarded as the measurement that will be used to estimate the unknown parameter x. z is an indirect measurement derived from the result of the multimeter. The latter is modelled by:

$$R_{meas} = R + w \tag{3}$$

where w is the error of the multimeter. Its uncertainty is $\sigma_w = 0.02R$.

With an estimate $\hat{x}$ of x we can reconstruct R using the inverse formula of (1):

$$R = \frac{R_{nom}}{1+x} = R_{nom}\left(1 - x + x^2 + \cdots\right) \approx R_{nom}\left(1 - x\right) \tag{4}$$

Here, a truncated Taylor series expansion has been applied. Since $|\hat{x}| < 0.05$, the approximation will be close, and we may estimate R by:

$$\hat{R} = R_{nom}(1 - \hat{x}) \tag{5}$$

The influence of not using the higher order terms falls far below the influence of the estimation error of $\hat{x}$, as we may verify later.

Because $\hat{R}$ depends linearly on $\hat{x}$ any cost function applied to $\hat{R}$ can simply be transformed into a cost function for $\hat{x}$. Thus, for instance, if $\hat{x}$ is the MMSE-estimate of x, then $\hat{R} = R_{nom}(1 - \hat{x})$ is the MMSE-estimate of R.

The dependence of the measurement, z, on the true value R follows from (2), (3) and (4):

$$z = \frac{x - v}{1+x} = x - x^2 + \cdots - v\left(1 - x + x^2 + \cdots\right) \approx x - v \tag{6}$$

v is the relative error of the multimeter, $v = w/R$. Again, a truncated Taylor series expansion of $1/(1+x)$ has been applied to get the approximation.

The task in this project is to find a variety of estimators $\hat{x}(z)$, i.e.:

- the MMSE estimator
- the MAP estimator
- the MMSA estimator
- the linear MMSE estimator
- the linear, unbiased MMSE estimator,

and to evaluate these designs with the appropriate evaluation criterions. For that purpose, a batch of 223 resistors with a tolerance of 5% has been selected. With an accurate ohmmeter (uncertainty less than 0.2%) the relative errors, x_n, of the nominal values haves been determined. In addition, for each resistor the (relative) resistance z_n has been determined with an inaccurate multimeter (standard deviation 2%)[1]. The set is available in the datafile `CPESE_PROJ_PARM1.mat`.

Preferably, the solutions given are analytic, but if this appears too difficult numerical solutions

[1] In fact, the set z_n is the result of a simulation. The measurement error of a single multimeter is a random constant. The proper experiment would involve the usage of 223 multimeters, one for each sample. However, this is impracticable.

also suffice. In any case, the found estimators must be presented graphically as a curve showing $\hat{x}(z)$ versus z.

Design strategy

In order to find the best estimators under the different cost functions both the prior probability density $p(x)$ and the conditional density $p(z \mid x)$, or equivalently $p(v)$, must be determined. For the first one the dataset x_n is available. For the second one, we can use the set $v_n = z_n - x_n$. Once we have assessed both probability densities, Sections 3.1 applies, and the development of the estimator is straightforward (perhaps not analytically tractable, but then a further numerical development must be set up).

Models for the probability density functions

For the assessment of the probability functions two different types of models exist (see Chapter 5):

- Parametric models
- Nonparametric models

For parametric models we assume that that the density has a given functional form that depends on one or more parameters. Suggestions for densities are:

1. Gaussian distributions
2. Uniform distributions
3. Bimodal distributions consisting of two Gaussians
4. Bimodal distributions consisting of two uniform densities
5. Bimodal distributions consisting of one Gaussian with a gap in the centre part
6. Gaussian distributions with truncated tails

The learning procedure consists of adapting these parameters to the data at hand.

The nonparametric models often represent the density using some kind of interpolation. Possible methods are:

7. Parzen estimation (with different kernels)
8. k-NNR estimation

Model selection

In any case, the model that we hypothesize depends on parameters that must be selected appropriately (even the nonparametric models have parameters to tune). The best parameters are found by minimizing the error criterion. This error criterion is numerically obtained using the samples from training data. With the optimal parameters thus found we evaluate the model once again, but this time with evaluation/validation data. The model with minimal error criterion will be selected.

The training data and the evaluation/validation data should be independent. Otherwise, the risk of overfitting occurs. The best procedure for model selection and performance evaluation is cross-validation. See Section 5.4 and Section 9.2. Once the best model has been selected, the final estimator is formed by using the full dataset as training data. The final performance is established by averaging the results obtained during the cross-validation.

Synopsis

```
For all type of estimators (i.e. MMSE, MAP, MMAE, etc):
1.    For all considered distribution models:
      1.1.  Apply cross-validation using an error criterion matching the type of estimator.
      1.2.  Calculate the average of the performances found in 1.1.
2.    Compare the averaged performance (calculated in 1.2.), and select the best model.
3.    For the model selected in 2., determine the best parameters, but now by using the full
      dataset.
4.    Generate a graph of the estimator corresponding to the parameters found in 3.
```