ESE 524: Detection and Estimation Theory

Recitation 0

Washington University in St. Louis

Topics for Today

- Description of a model
- Probability distribution of a model

But Why Male Models?

- The most important concept from I0 is the Model.
- Data: $x[n]$ are our measurements. They are numbers. For example:

$$\boldsymbol{x} = [5, 5, 5, 5, 5]^T$$

  has $x[0] = 5$, $x[1] = 5$, and so on.
- $x[n]$ do not represent variables, think of them as constants.
- This class does not deal with the process of measurement, we assume we have $x[n]$ already.
- Philosophically, we want an equation that describes $x[n]$.
- Therefore we pick a family of functions $\mathcal{F}$ -e.g., Linear models, physics equations, sine waves.
- Each $f \in \mathcal{F}$ has the same functional form, but some parameters $\boldsymbol{\theta}$ let us describe the data.
- This leads us to the model:

$$x[n] \approx f(n; \boldsymbol{\theta})$$

Adding noise

- I used an approximation in the last slide because it is impossible to perfectly model anything.
- To get perfect equality, define an error term:

$$e[n] = x[n] - f(n; \boldsymbol{\theta})$$

- Now I can express a "perfect" model:

$$x[n] = f(n; \boldsymbol{\theta}) + e[n] \tag{1}$$

- The only problem is that we don't know the true parameters $\boldsymbol{\theta}$, so we don't have a formula for $e[n]$.
- Therefore, $e[n]$ is represented by a random variable.
- This means that $x[n]$ are realizations or samples of a random variable.

Estimators

- An Estimator, $\hat{\boldsymbol{\theta}}$, is a function of $x[n]$ which attempts to solve Eq. 1 for $\boldsymbol{\theta}$..
- Since $x[n]$ is a random variable, the estimator is also a random variable.
- We need $p(\boldsymbol{x};\theta)$ to both build estimators and calculate the distribution of estimators.
- But we also have $x[n]$ as a constant number, which means that the estimators are also a number.
- So the estimator value is a realization or sample, which means we do not know it's true value either!
- To characterize performance of the estimators, we look at the error term:
$$\theta - \hat{\theta}(\boldsymbol{x})$$
- Bias is the expected value of this term, $\mathrm{E}[\theta - \hat{\theta}]$.
- Mean-square-error is the variance, $\mathrm{E}[(\theta - \hat{\theta})^2]$

Example : Student's Homework Grades

- Let $x[n]$ represent your homework grades, for $n = 0, 1, 2..., 7$.
- Say you get:

$$\boldsymbol{x} = [100, 50, 95, 89, 93, 92, 96, 88]^T$$

- From this data, I would like to determine your:
  - Time put into the class, θ_T.
  - Understanding of the material, θ_U.
  - Stress level, θ_S.
  - Amount of pizza consumed, θ_P.
- From the first principals of education, I use the ACME-Education-Equation-3000 -

$$x[n] \approx f(n; \boldsymbol{\theta}) = \underbrace{\theta_T}_{\text{Time}} + \underbrace{\cos(\theta_U n)}_{\text{Understanding}} - \underbrace{\exp(\theta_S n - \theta_P)}_{\text{Stress } vs. \text{ Pizza}}$$

- My estimators $\hat{\theta}_T$, $\hat{\theta}_S$, $\hat{\theta}_U$, and $\hat{\theta}_P$ will give the curve that best fits the grades.

Example Continued: What's with all the Noise?

- However, I understand that these are not the only factors impacting your grade.
- To account for this, define the error term:

$$e[n] = x[n] - f(n; \boldsymbol{\theta}) = x[n] - (\theta_T + \cos(\theta_U n) - \exp((\theta_S n - \theta_P)/10)$$

- I don't know what the error is, since I don't know the true vale of the parameters.
- Let's model the error with a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 20$.
- I update my model to give an exact equality:

$$x[n] = f(n; \boldsymbol{\theta}) + e[n] = \theta_T + \cos(\theta_U n) - \exp(\theta_S n - \theta_P) + e[n]$$

- Now $x[n]$ is a normal random variable with mean $\theta_T + \cos(\theta_U n) - \exp(\theta_S n - \theta_P)$ and variance 20, and I can use this distribution to figure out maximum-likelihood estimators (calculations not shown).

Example Continued: Plots

- It turns out that the maximum likelihood estimates given this data are:

$$\hat{\theta}_T = 87.81 \quad \hat{\theta}_U = 5.02 \quad \hat{\theta}_S = -4.91 \quad \hat{\theta}_P = 8.69$$

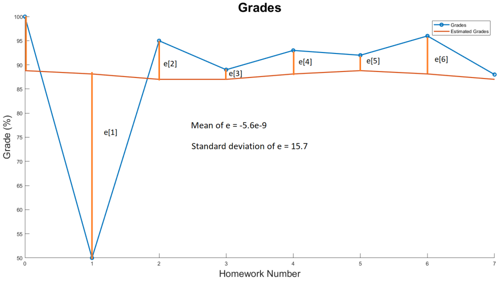


Figure 1: Plot of the grades, grades estimated by the model, and the model error. Note that the model roughly follows the trend, but is dropped by the student's outlier grade in HW 1.

Example Continued: Bias and Variance

- To approximate the bias and variance of my estimators, I generated many (1000) random sets of grades from some "true" parameters (not possible in real life for one student).
- I computed the estimators for each sample, and generated estimates of the parameters.
- Then I approximated the expected value operation for the bias and variance using the mean() command in MATLAB.
- Any ideas on how I could improve my estimators?

Table 1: Results of the Student Grade Estimator

Parameter	True Parameter	Expected Value of Parameter	Bias	Variance
θ_T	90	89.31	-0.69	3.61
θ_U	5	6.28	1.28	1.65
θ_S	-5	-5.84	-0.84	1.68
θ_P	10	4.03	-5.97	43.87

Getting a Probability Distribution from a Model

- The general method for solving problems in this class will proceed as follows
  - Given a set of known samples, $\boldsymbol{x} = [x[0], x[1], ..., x[n-1]]^T$, and unknown parameter (or vector of parameters) θ, assume a model $x[n] = f(\theta, n) + e[n]$.
  - Use the model and the transformation of variables formula to find the likelihood function, $p(x|\theta)$.
  - Use the likelihood to construct an estimator $\hat{\theta}(\boldsymbol{x})$. Note this is another transformation of random variables, so it has a probability distribution, $p_{\hat{\theta}}(\hat{\theta})$.
  - Examine the performance of an estimator, usually by computing an expected value such as the Mean-Squared-Error.

Example 2: Estimating a DC Level with Additive Noise

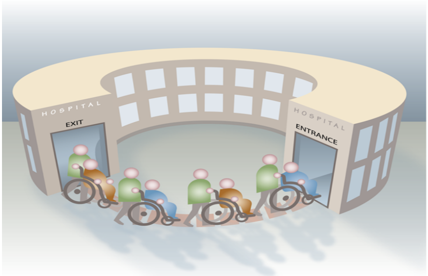
- Model: $x[n] = \theta + w[n]$ where each $w[n]$ is a sample from white noise, and is normally distributed with mean 0 and known variance σ^2.
- Likelihood Function: Assuming the samples $x[n]$ are independent, we can make the likelihood function as a multivariate gaussian.
- Estimator: Intuitively, the average, $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$, is the best estimator here.
- Performance: We saw in example 1 from Lecture 1 that the MSE of this estimator is σ^2/N.

Examples of Detection and Estimation Research Applications

- The material taught in this class has a very wide range of applications.
- It lays foundations for some of the most actively researched fields in industry and academy.
- A few examples taken from the research conducted by Dr. Nehorai's research group are shown in the next slides.

Example 3: Hospital Readmissions

- Hospital Readmission: patients are admitted to a hospital within 30 days of discharge - accrue significant additional hospital costs.
- Cost: The Agency for Healthcare Research and Quality reports that in 2011 an additional $41 billion dollars in hospital costs were caused by readmissions across the country.
- Overall Goal: Identify patients at high risk of readmission and reduce Barnes Jewish's annual readmissions rates to under 17.5%.



Graphic credit: Greg Cross/The Bulletin. Republished with permission.

Quick Data Overview

- Observations: $\boldsymbol{x}$ consists of the results of 776 patients who were either readmitted ($x[n] = 1$) or not-readmitted ($x[n] = 0$).
- Data: The matrix $\boldsymbol{H}$ has rows representing patients and columns representing:
  - LACE Score - current hospital risk metric
  - Whether the patient has diabetes
  - Principal Diagnoses - Heart Failure, Chronic Obstructory Pulminary Disease, Myocardial Infarction, or Pneumonia
  - Ethnicity
  - Gender
  - Month patient was treated
  - Length of Stay
  - Access to Primary Care Provider
  - Age
  - Discharge Status - back to home, to rehab, etc.
  - Zipcode - very rough proxy for income/demographic data

Parameters

- In a logistic regression model, the paramters $\theta_1, ..., \theta_{11}$ will represent the impact of each of these factors on the readmissions rate.
- The Odds Ratio gives the relationship between variables and the observation, i.e.:

$$\mathrm{OR} = \frac{p(\text{readmission; patient has diabetes})}{p(\text{readmission; patient does not have diabetes})}$$

- In the case of Linear Models, the odds ratio for each variable is simply θ_i.
- For Logistic Regression, the odds ratio for each variable is $\exp(\theta_i)$.

Logistic Regression Results

Predictor	Odds ratio	95% Confidence Interval	p value
LACE	1.22	(1.14, 1.31)	<0.001
COPD (vs CHF)	0.21	(0.11, 0.45)	<0.001
MI (vs CHF)	0.66	(0.45, 0.99)	<0.001
Discharged to Skilled Nursing Facility (vs Home)	0.50	(0.29, 0.86)	0.01
Discharged with Home Health (vs Home)	2.34	(1.57, 3.49)	<0.001
Male (vs Female)	1.97	(1.36, 2.86)	<0.001
LOS	0.97	(0.95, 0.99)	0.03
Age70-74yrs (vs 65-69 yrs)	0.69	(0.49, 0.98)	0.04
Age75-79 (vs 65-69 yrs)	1.78	(1.2, 2.6)	<0.001
Has PCP (vs no PCP)	1.77	(1.15, 2.72)	0.01

Figure 2: Odds Ratios, with confidence intervals and p-values for each estimated parameter. Variables with high p-values, bad confidence intervals are assigned $\theta_i = 0$ as they don't help the model. Results were originally reported at the 2018 GSA conference.

Connection to Machine Learning: Predicting Readmissions

- Once θ_i have been estimated, we want to predict whether or not patients will be readmitted, so define a Decision Threshold:

$$D(\boldsymbol{x}, T) = \begin{cases} 1 & \text{if } p(\text{readmission}; \theta) > T \\ 0 & \text{otherwise} \end{cases}$$

- Then we choose T by trying to maximize:
  - True Positives: Correctly predicted readmissions.
  - True Negatives: Correctly predicted safe patients.
- and minimize:
  - False Positive: Predicting that a safe patient will be readmitted.
  - False Negative: Predicting that a readmitted patient would not return.
- This is a Binary Classification Problem, and there is usually a tradeoff between the metrics we want to find.

Comments

- The best logistic regression models with decision thresholds predict roughly 60% of patients correctly.
- This is ... non-ideal.
- We tested many different models and were able to come up with a model of the form:

$$p(\boldsymbol{x};\theta) = \sum_{i=1}^{k} a_i f_i(\boldsymbol{H},\theta)$$

  where a_i are constants, f_i represents logistic regression models trained on cleverly selected sub-sets of $\boldsymbol{H}$, and $\boldsymbol{\theta}$ is a vector containing all the parameters for each f_i.
- While we have better classification performance, this model loses the easy interpretation of the GLM approach.

Example 4: Detecting Birth

- We can predict (detection) if a woman will give birth soon, based on recordings of the uterine electrical activity
- Birth related problems are a major health concern for the newborn babies and their mothers.
- Having an accurate prediction of the due date helps prepare for delivery, and therefore reduces complications.

Detecting Birth (Cont.)

- First we measure the magnetic (or electrical potentials) fields from the mother's abdomen

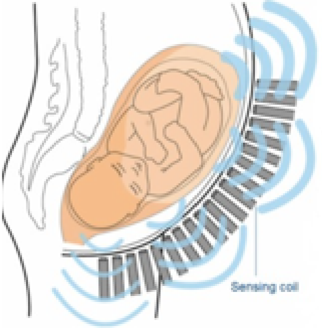


Figure 3: Magnetic field recordings of the uterine activity using the SARA system.

Detecting Birth (Cont.)

- Estimate the internal currents that created those fields (inverse problem)

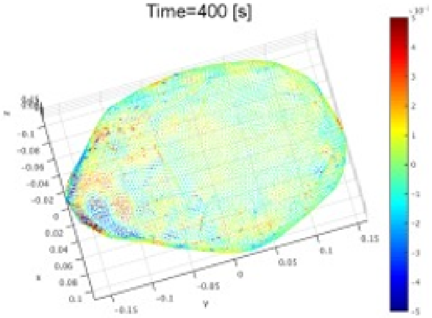


Figure 4: Solution to the inverse problem of uterine activity

Detecting Birth (Cont.)

- Make a model and determine the null hypothesis (H0)

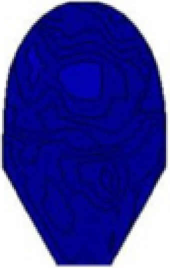


Figure 5: H_0: The uterine electrical currents are unsynchronized

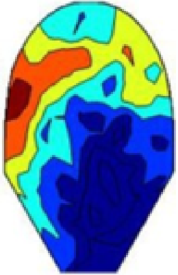


Figure 6: H_1: The uterine electrical currents are synchronized

Detecting Birth (Cont.)

- Use the existing model to accept or reject the null hypothesis on a new measurement
  - Accept H_0: Birth will **probably not** happen soon.
  - Reject H_0: Birth will **probably** happen soon

Conclusions

- Understanding what models are and how they relate to the data is key to doing well in this course.
- The first example was a situation of additive noise, which we will use for most examples in this class.
- Remember that $x[n]$ is a realization of a random number, so it has a probability distribution.
- Remember the standard workflow for estimation problem. Keep in mind that problems can vary a lot, but their solution usually follows this procedure.