Lecture 14

ECE 278 Mathematics for MS Comp Exam

UCSD Jacobs | Electrical and Computer Engineering

- Standard statistics is used to collect, organize, and describe data
  - Mean, variance, covariance etc.
- Statistical **inference** uses a set of observations to infer general characteristics about a population
- Widely used across all fields in science and engineering
- Additional reference for notes: "Fundamentals of Statistical Signal Processing" Estimation Theory Steven Kay and note from MIT on class web site

- The first step in making a statistical inference is to model the population(s) by a probability distribution characteristized by one or more unknown parameters
  - eg. a zero-mean gaussian for which the variance is unknown.
- Goal is to use data to infer the parameters so the the resulting probability distribution "best" describes the population as a whole given an incomplete set of data
- For this introduction are interested in only one class of inference problems: **estimation** of parameters
- Other classes of inference include:
  - hypotheses testing—providing a yes/no answer as to whether the parameter lies in a specified region of values.
  - statistical classification— deciding which set a new observation belongs to based on a previously determined model generated from a data set.

Frameworks of Estimation

- There are two primary methods of inference:, **frequentist** and **Bayesian**, for making statistical inferences.
- In a frequentist approach, the parameters are treated as **fixed** but unknown quantities in the distribution which governs variation in the sample data.
- A Bayesian approach, the unknown parameters are taken to be **random variables** and, prior to sampling, assign a **prior distribution** for each parameters
- After the data are collected, Bayes rule is used and the **likelihood** is multiplied by the prior distribution to obtain the posterior distribution of the parameter. Characteristic statistics (mean, variance etc.) of the posterior distribution are then used to make inferences.
- More generally, a distribution is assigned for anything that is unknown or uncertain.
- As more data become available, the prior distribution is updated using the conditional probabilities.
- Note that Bayesian estimation can deal with situations where the sequence of observations are not necessarily independent.

Types of Estimators

- There are **many** different techniques for parameter estimation
- A specific method is called an **estimator**
- An estimator is applied to a set of data to construct an estimate.
- Desirable properties of an estimator (consistency, bias, and variance) are covered in more advanced classes.
- The **bias** of an estimator is the difference between the expected value of the parameter and the true value of the parameter.
- Look at several widely-used estimators : maximum posterior estimator (MAP) estimator, maximum likelihood estimator (MLE), minimum mean-square estimator (MMSE) and the least-squares estimator (LSE), which is a generalization of simple regression.
- Other methods of estimation, which will not be discussed, rely on estimating the moments of the distribution.

- Given an random variable $\underline{y}$ and an observation $\underline{x}$ of that random variable, which itself is treated as a random variable, we would like to derive an estimate $\widehat{y}$ of the the random variable $\underline{y}$ given the joint probability distribution $p_{\underline{xy}}(x,y)$
- A natural (but not unique) function to derive the estimate is to minimize the expectation of the square of the difference between the random variable $\underline{y}$ and the estimate $\widehat{y}$given by

$$\left\langle (\underline{y}-\widehat{y})^2 \right\rangle = \int (\underline{y}-\widehat{y})^2 f_{\underline{y}}(y)\mathrm{d}y \tag{1}$$

  where $f_{\underline{y}}(y)$ is the pdf.

- The minimum is obtained by setting the derivative with respect to the estimate $\widehat{y}$ equal to zero, which gives

$$-2\int \left(\underline{y}-\widehat{y}\right) f_{\underline{y}}(y)\mathrm{d}y = 0$$

  or

$$\int y f_{\underline{y}}(y)\mathrm{d}y = \int \widehat{y} f_{\underline{y}}(y)\mathrm{d}y$$

- Therefore

$$\widehat{y} = \left\langle \underline{y} \right\rangle \tag{2}$$

  as one might expect. The second derivative is positive, so that the estimate is indeed the minimum.

UCSD Jacobs | Electrical and Computer Engineering

- The error in the estimate is evaluated by substituting (2) into (1) and gives

$$\left\langle (\underline{y} - \widehat{y})^2 \right\rangle = \sigma_y^2,$$

which is simply the variance of the random variable $\underline{y}$.

- Now suppose that we have additional information from a sample x of a random variable $\underline{x}$.

- The random variables $\underline{x}$and $\underline{y}$ are related by a joint pdf $p_{\underline{xy}}(x, y)$.

- Define for the error as

$$\left\langle (\underline{y} - \widehat{y}(x))^2 \middle| \underline{x} = x \right\rangle = \int \left(\underline{y} - \widehat{y}(x)\right)^2 f_{\underline{y}|x}(y|x) \mathrm{d}y$$

where now the estimate $\widehat{y}(x)$ depends on the specific realization x of the data.

- The minimum is obtained the same way as before with

$$\widehat{y}(x) = \left\langle \underline{y} | \underline{x} = x \right\rangle, \tag{3}$$

which is a the **conditional expectation** of the random variable $\underline{y}$ given a specific realization x of the random variable $\underline{x}$.

- The error is the **conditional variance** given by

$$\left\langle (\underline{y} - \widehat{y}(x))^2 \right\rangle = \sigma^2_{y|x},$$

where the variance $\sigma^2_{y|x}$ is now conditioned on the data value.

- When a set of observations is used, which is typically the case, the single value is replaced by a vector $\mathbf{x}$ of values with would like to derive an estimate $\widehat{y}$ of the the random variable $\underline{y}$ given the joint probability distribution $p_{\underline{x}\underline{y}}(x, y)$.

- A natural (but not unique) function to derive the estimate is to minimize the expectation of the square of the difference between the random variable $\underline{y}$ and the estimate $\widehat{y}$given by

$$\widehat{y}(\mathbf{x}) = \left\langle \underline{y} | \underline{\mathbf{x}} = \mathbf{x} \right\rangle,$$

with the error given by

$$\left\langle (\underline{y} - \widehat{y}(\mathbf{x}))^2 \right\rangle = \sigma^2_{y|\mathbf{x}},$$

where the variance for estimate is now conditioned on the complete set of data $\mathbf{x}$.

- An estimate is based on a single of a sample or a set of samples
- An **estimator** is the method by which the estimate is obtained.
- An estimator can be derived from an estimate by simply replacing the realization x by the random variable $\underline{x}$ to give

$$\widehat{y}(\underline{\mathbf{x}}) = \left\langle \underline{y} | \underline{\mathbf{x}} \right\rangle$$

or

$$\widehat{\underline{y}} = \left\langle \underline{y} | \underline{\mathbf{x}} \right\rangle$$

where $\widehat{\underline{y}} \doteq \widehat{y}(\underline{\mathbf{x}})$.

- The error in the estimate is the variance of the conditional distribution $f_{\underline{y}|\mathbf{x}}(y|\mathbf{x})$.

- Let a joint distribution be given as joint distribution (note that this is corrected)

$$f_{\underline{x,y}}(x,y) = \begin{cases} kxy & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the normalized constant and the MMSE estimator for $\underline{y}$ in terms of $\underline{x}$

- **Solution**

The area under the curve is

$$\begin{aligned} A &= \int_0^1 \mathrm{dx} \int_x^1 xy\mathrm{d}y \\ &= \frac{1}{2}\int_0^1 (\mathrm{x} - \mathrm{x}^3)\mathrm{dx} \\ &= \frac{1}{8} \end{aligned}$$

so that $k = 8$.

- To find the MMSE, we need conditional distribution $f_{\underline{y|x}}(y|x)$. which can be written as

$$f_{\underline{y|x}}(y|x) = \frac{f_{\underline{x,y}}(x,y)}{f_{\underline{x}}(x)}$$

- The marginal distribution is

$$\begin{aligned} f_{\underline{x}}(x) &= \int_x^1 xy\mathrm{d}y \\ &= 4x(1-x^2) \qquad 0 < x < 1 \end{aligned}$$

so that

$$f_{\underline{y|x}}(y|x) = \frac{2y}{(1-x^2)} \qquad 0 < y < x < 1$$

- The MMSE estimator is the conditional expectation, which is given by

$$\begin{aligned}\widehat{\underline{y}} &= \left\langle \underline{y} | \underline{x} \right\rangle \\ &= \int_x^1 y \, f_{\underline{y}|x}(y|x) \mathrm{d}y \\ &= \frac{2}{1-x^2} \int_x^1 y^2 \mathrm{d}y \\ \widehat{\underline{y}} &= \frac{2}{3} \frac{1-x^3}{1-x^2}\end{aligned}$$

Note that this estimator is nonlinear in x.

- Now consider the bivariate gaussian distribution considered earlier

$$f_{\underline{x},\underline{y}}(x,y) \quad = \quad \frac{1}{2\pi\sigma^2\sqrt{1-\rho_{xy}^2}} \exp\left(-\frac{x^2-2\rho_{xy}xy+y^2}{2\sigma^2(1-\rho_{xy}^2)}\right). \tag{4}$$

with the correlation coefficient given by (See Lec. 12)

$$\rho_{xy} \doteq \frac{\left\langle \underline{xy} \right\rangle - \left\langle \underline{x} \right\rangle \left\langle \underline{y} \right\rangle}{\sigma_x \sigma_y} = \frac{\mathrm{cov}(\underline{xy})}{\sigma_x \sigma_y}$$

where the means need not be zero as was the case for the previous example.

- The MMSE estimator is the conditional pdf

$$\widehat{\underline{y}} = \left\langle \underline{y} | \underline{x} \right\rangle$$

- Skipping the details of the derivation, we can write

$$\widehat{\underline{y}} = \left\langle \underline{y} \right\rangle + \rho_{xy} \frac{\sigma_y}{\sigma_x} \left(\underline{x} - \left\langle \underline{x} \right\rangle \right)$$

- Note that this is a **linear** estimator! (Actually linear with a constant offset, which is called *affine*.)

- The MMSE error is the variance in the conditional pdf, which is given by

$$\text{MMSE} = \sigma_y^2 (1 - \rho_{xy}).$$

- Note that with no correlation the error is just the variance of $\underline{y}$ because knowing a realization of $\underline{x}$ does not help.

- As the random variables become more correlated, the error decreases. For $\rho_{xy} = 0$, there is only a single random variable and the error is zero.

- In general, except for the special case of a gaussian distribution, the MMSE estimator is nonlinear, and can be hard to calculate.
- This motivates minimizing the error for a *linear model* of the data instead of a general model .
- This linear model is given by

$$\widehat{\underline{y}} = a\underline{x} + b \tag{5}$$

where a and b are parameters that minimize the mean square error

$$\left\langle \left(\underline{y} - \widehat{\underline{y}}\right)^2 \right\rangle = \left\langle \left(\underline{y} - a\underline{x} + b\right)^2 \right\rangle \tag{6}$$

where the expectation is over the joint distribution $f_{\underline{x},\underline{y}}(x, y)$.

- To find the optimal values of a and b, we take derivatives of (6) with respect to a and b.
- Because the expectation operation is linear, we can take the derivative inside the expectation to give

$$\frac{\mathrm{d}}{\mathrm{d}b}\left\langle \left(\underline{y} - a\underline{x} + b\right)^2 \right\rangle = 0$$

or

$$\left\langle \underline{y} - (a\underline{x} + b) \right\rangle = 0$$

or

$$\left\langle \underline{y} \right\rangle = \left\langle a\underline{x} + b \right\rangle.$$

- Solving for b gives

$$b = \left\langle \underline{y} \right\rangle - a\left\langle \underline{x} \right\rangle.$$

- Substituting b into (5) gives

$$\widehat{\underline{y}} = \left\langle \underline{y} \right\rangle + a\left(\underline{x} - \left\langle \underline{x} \right\rangle\right)$$

- This estimate is simply the mean value $\left\langle \underline{y} \right\rangle$ of $\underline{y}$ wighted by the deviation of $\underline{x}$ from its mean value with the weighting term given by a
- The optimal value of a can be understood by view the random variables with the means removed. Call these variables $\underline{X} = \underline{x} - \left\langle \underline{x} \right\rangle$ and $\underline{Y} = \underline{y} - \left\langle \underline{y} \right\rangle$.
- Then the value of a satisfies

$$\left\langle (\underline{Y} - a\underline{X})\underline{X} \right\rangle = 0$$

or

$$\left\langle \underline{YX} \right\rangle - a\left\langle \underline{X}^2 \right\rangle = 0.$$

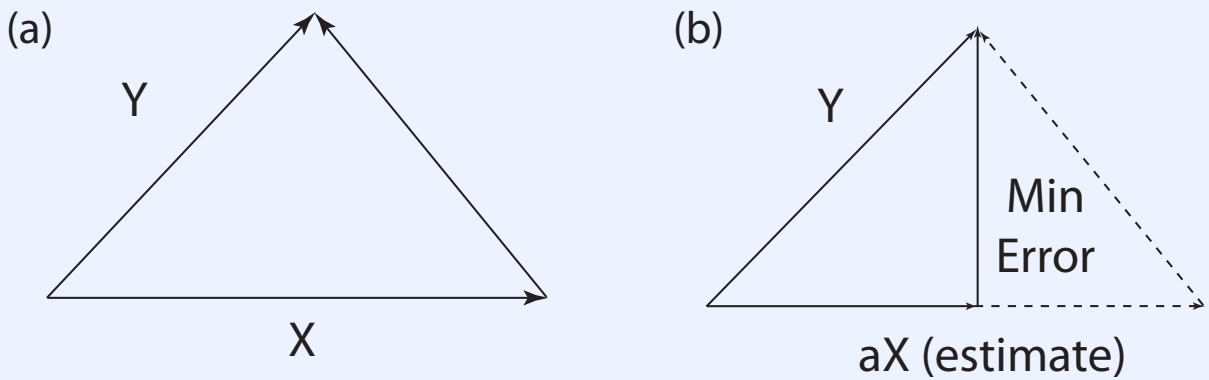
- The first term is $\left\langle \underline{YX} \right\rangle$ is the correlation, which is equal to $\mathrm{cov}(\underline{x}, \underline{y})$ when the means are included.
- The second term is the mean-square value of $\underline{X}$ or the variance σ_x^2 when the mean is included.
- Therefore, the weighting factor of a is given by

$$a = \frac{\mathrm{cov}(\underline{x}, \underline{y})}{\sigma_x^2}.$$

- When the means removed and treating the random variable as vectors, we see that the expression

$$\left\langle (\underline{Y} - a\underline{X})\underline{X} \right\rangle = 0$$

- implies that the vector $\underline{Y} - a\underline{X} = \underline{Y} - \left\langle \underline{Y} \right\rangle$ is **orthogonal** to the vector $\underline{X}$ where orthogonal random variables are uncorrelated.



Orthogonality cont.

- Error $\underline{Y} - \widehat{\underline{Y}} = \underline{Y} - a\widehat{\underline{X}}$ is orthogonal to the observation $\underline{X} = \underline{x} - \left\langle \underline{x} \right\rangle$.
- This means that the LMMSE estimator is unbiased with the estimation error orthogonal to the random variable used to form the estimate.
- A more general statement applies to MMSE for which the error is orthogonal to any function of the random variable used to form the estimate (not just the linear function used for the LMMSE estimator.
- Note that estimation error has the same form as the general case with

$$\text{MMSE} = \sigma_y^2(1 - \rho_{xy}).$$

- Let a joint distribution be given as before with

$$f_{\underline{x},\underline{y}}(x,y) = \begin{cases} 8xy & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

  Find LMMSE estimator for $\underline{y}$ in terms of $\underline{x}$.

- **Solution**

  The linear model for the estimate is

$$\widehat{\underline{y}} = a\underline{x} + b$$

  with

$$b = \langle \underline{y} \rangle - a\langle \underline{x} \rangle$$

  and

$$a = \frac{\mathrm{cov}(\underline{x},\underline{y})}{\sigma_x^2}.$$

- From earlier, the marginal distribution is

$$f_{\underline{x}}(x) = \int_x^1 xy\mathrm{d}y$$
$$= 4x(1-x^2) \qquad 0 < x < 1$$

so that the mean value is

$$\left\langle \underline{x} \right\rangle = \int_0^1 x f_{\underline{x}}(x)\mathrm{d}x$$
$$= \int_0^1 4x^2(1-x^2)\mathrm{d}x$$
$$= \frac{8}{15}$$

- The mean value $\left\langle \underline{x} \right\rangle$ can also be obtained directly from the joint pdf as

$$\left\langle \underline{x} \right\rangle = \int_0^1 \int_x^1 x f_{\underline{x},\underline{y}}(x,y)\mathrm{d}y\mathrm{d}x$$
$$= \int_0^1 \mathrm{dx} \int_x^1 x^2 y\mathrm{d}y$$
$$= \frac{8}{15}$$

- The mean of $\langle y \rangle$ is

$$\begin{aligned}\langle \underline{y} \rangle &= \int_0^1 \int_x^1 y f_{\underline{x,y}}(x,y) \mathrm{d}y \mathrm{d}x \\ &= \int_0^1 \mathrm{d}x \int_x^1 x y^2 \mathrm{d}y \\ &= \frac{4}{5}\end{aligned}$$

- The mean square value is

$$\begin{aligned}\langle \underline{x^2} \rangle &= \int_0^1 x^2 f_{\underline{x}}(x) \mathrm{d}x \\ &= \int_0^1 4x^3(1-x^2) \mathrm{d}x \\ &= \frac{1}{3}\end{aligned}$$

so that the variance σ_x^2 is

$$\sigma_x^2 = \langle \underline{x^2} \rangle - \langle \underline{x} \rangle^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}$$

- The correlation $\left\langle \underline{x}\underline{y} \right\rangle$ is

$$\begin{aligned}\left\langle \underline{x}\underline{y} \right\rangle &= \int_0^1 \int_x^1 xy f_{\underline{x},\underline{y}}(x,y)\mathrm{dxdy} \\ &= \int_0^1 \mathrm{dx} \int_x^1 x^2 y^2 \mathrm{dy} \\ &= \frac{4}{9}\end{aligned}$$

- The covariance is

$$\begin{aligned}\mathrm{cov}(\underline{x},\underline{y}) &= \left\langle \underline{x}\underline{y} \right\rangle - \left\langle \underline{x} \right\rangle \left\langle \underline{y} \right\rangle \\ &= \frac{4}{9} - \left(\frac{8}{15}\right)\left(\frac{4}{5}\right) = \frac{4}{225}\end{aligned}$$

- Using these values

$$a = \frac{\mathrm{cov}(\underline{x}, \underline{y})}{\sigma_x^2} = \frac{\frac{4}{225}}{\frac{11}{225}} = \frac{4}{11}$$

and

$$b = \langle \underline{y} \rangle - a \langle \underline{x} \rangle = \frac{4}{5} - \frac{4}{11}\frac{8}{15} = \frac{20}{33}$$

so that

$$\widehat{\underline{y}} = \frac{4}{11}\underline{x} + \frac{20}{33}.$$

Plot of MMSE and LMMSE Estimators