10 - MMSE Estimation

- Estimation given a pdf
- Minimizing the mean square error
- The minimum mean square error (MMSE) estimator
- The MMSE and the mean-variance decomposition
- Example: uniform pdf on the triangle
- Example: uniform pdf on an L-shaped region
- Example: Gaussian
- Posterior covariance
- Bias
- Estimating a linear function of the unknown
- MMSE and MAP estimation

Estimation given a PDF

Suppose $x : \Omega \to \mathbb{R}^n$ is a random variable with pdf p^x.

One can *predict* or *estimate* the outcome as follows

- Given *cost function* $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$
- pick *estimate* $\hat{x}$ to minimize $\mathbf{E}\, c(x, \hat{x})$

We will look at the cost function

$$c(x, \hat{x}) = \|x - \hat{x}\|^2$$

Then the *mean square error* (MSE) is

$$\mathbf{E}\big(\|x - \hat{x}\|^2\big) = \int \|x - \hat{x}\|^2 p^x(x)\, dx$$

Minimizing the MSE

Let's find the *minimum mean-square error* (MMSE) estimate of x; we need to solve

$$\min_{\hat{x}} \mathbf{E}\big(\|x - \hat{x}\|^2\big)$$

We have

$$\begin{aligned} \mathbf{E}\big(\|x - \hat{x}\|^2\big) &= \mathbf{E}\big((x - \hat{x})^T(x - \hat{x})\big) \\ &= \mathbf{E}\big(x^T x - 2\hat{x}^T x + \hat{x}^T \hat{x}\big) \\ &= \mathbf{E}\|x\|^2 - 2\hat{x}^T \mathbf{E}\, x + \hat{x}^T \hat{x} \end{aligned}$$

Differentiating with respect to $\hat{x}$ gives the optimal estimate

$$\hat{x}_{\mathsf{mmse}} = \mathbf{E}\, x$$

The MMSE estimate

The minimum mean-square error estimate of x is

$$\boxed{\hat{x}_{\mathsf{mmse}} = \mathbf{E}\, x}$$

Its mean square error is

$$\boxed{\mathbf{E}\big(\|x - \hat{x}_{\mathsf{mmse}}\|^2\big) = \mathbf{trace}\,\mathbf{cov}(x)}$$

since $\mathbf{E}\big(\|x - \hat{x}_{\mathsf{mmse}}\|^2\big) = \mathbf{E}\big(\|x - \mathbf{E}\, x\|^2\big)$

The mean-variance decomposition

We can interpret this via the MVD. For any random variable z, we have

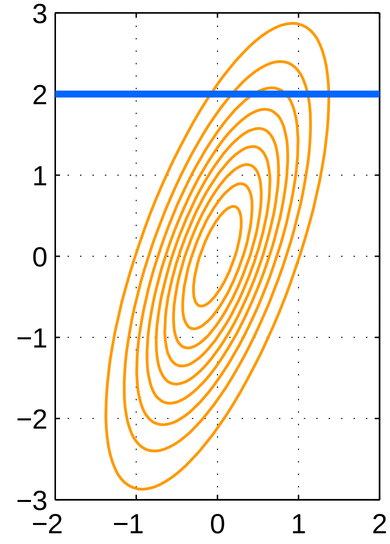
$$\mathbf{E}\big(\|z\|^2\big) = \mathbf{E}\big(\|z - \mathbf{E}\,z\|^2\big) + \|\mathbf{E}\,z\|^2$$

Applying this to $z = x - \hat{x}$ gives

$$\mathbf{E}\big(\|x - \hat{x}\|^2\big) = \mathbf{E}\big(\|x - \mathbf{E}\,x\|^2\big) + \|\hat{x} - \mathbf{E}\,x\|^2$$

- The first term is the *variance* of x; it is the error we *cannot remove*
- The second term is the *bias* of $\hat{x}$; the best we can do is make this zero.

The estimation problem



Suppose x, y are random variables, with joint pdf $p(x, y)$.

- We measure $y = y_{\text{meas}}$.
- We would like to find the MMSE estimate of x given $y = y_{\text{meas}}$.

The *estimator* is a function $\phi : \mathbb{R}^m \to \mathbb{R}^n$.

We measure $y = y_{\text{meas}}$, and estimate x by $\hat{x}_{\text{est}} = \phi(y_{\text{meas}})$

We would like to find the function ϕ which minimizes the cost function

$$J = \mathbf{E}\left(\|\phi(y) - x\|^2\right)$$

Notation

We'll use the following notation.

- p^y is the *marginal* or *induced* pdf of y

$$p^y(y) = \int p(x,y)\,dx$$

- $p^{|y}$ is the pdf *conditioned* on y

$$p^{|y}(x,y) = \frac{p(x,y)}{p^y(y)}$$

The MMSE estimator

The *mean-square-error conditioned on y* is $e_{\text{cond}}(y)$, given by

$$e_{\text{cond}}(y) = \int \|\phi(y) - x\|^2 \, p^{|y}(x, y) \, dx$$

Then the mean square error J is given by

$$J = \mathbf{E}\big(e_{\text{cond}}(y)\big)$$

because

$$\begin{aligned} J &= \int \int \|\phi(y) - x\|^2 \, p(x, y) \, dx \, dy \\ &= \int p^{y}(y) \, e_{\text{cond}}(y) \, dy \end{aligned}$$

The MMSE estimator

We can write the MSE conditioned on y as

$$e_{\text{cond}}(y_{\text{meas}}) = \mathbf{E}\Big(\|\phi(y) - x\|^2 \,|\, y = y_{\text{meas}}\Big)$$

- For each y_{meas}, we can pick a value for $\phi(y_{\text{meas}})$
- So we have an MMSE prediction problem for each y_{meas}

The MMSE estimator

Apply the MVD to $z = \phi(y) - x$ conditioned on $y = w$ to give

$$e_{\text{cond}}(w) = \mathbf{E}\big(\|\phi(y) - x\|^2 \,|\, y = w\big)$$

$$= \mathbf{E}\big(\|x - h(w)\|^2 \,|\, y = w\big) + \|\phi(w) - h(w)\|^2$$

where $h(w)$ is the mean of x conditioned on $y = w$

$$h(w) = \mathbf{E}(x \,|\, y = w)$$

To minimize $e_{\text{cond}}(w)$ we therefore pick

$$\phi(w) = h(w)$$

The error of the MMSE estimator

With this choice of estimator

$$e_{\mathsf{cond}}(w) = \mathbf{E}\big(\|x - h(w)\|^2 \,|\, y = w\big)$$

$$= \mathbf{trace}\,\mathbf{cov}(x \,|\, y = w)$$

Summary: the MMSE estimator

The MMSE estimator is

$$\phi_{\text{mmse}}(y_{\text{meas}}) = \mathbf{E}(x \mid y = y_{\text{meas}})$$

$$e_{\text{cond}}(y_{\text{meas}}) = \mathbf{trace}\,\mathbf{cov}(x \mid y = y_{\text{meas}})$$

- We often write

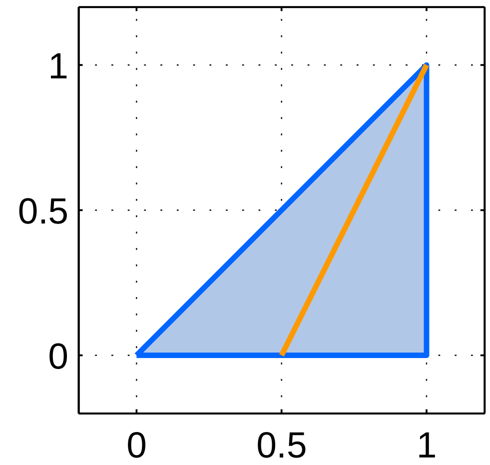
$$\hat{x}_{\text{mmse}} = \phi_{\text{mmse}}(y_{\text{meas}}) \qquad e_{\text{cond}}(y_{\text{meas}}) = \mathbf{E}\Big(\|x - \hat{x}_{\text{mmse}}\|^2 \;\Big|\; y = y_{\text{meas}}\Big)$$

- The estimate only depends on the *conditional pdf* of $x \mid y = y_{\text{meas}}$
- The means and covariance are those of the *conditional pdf*
- The above formulae give the MMSE estimate for *any pdf* on x and y

Example: MMSE estimation

(x, y) are uniformly distributed on the triangle A.
i.e., the pdf is

$$p(x,y) = \begin{cases} 2 & \text{if } (x,y) \in A \\ 0 & \text{otherwise} \end{cases}$$



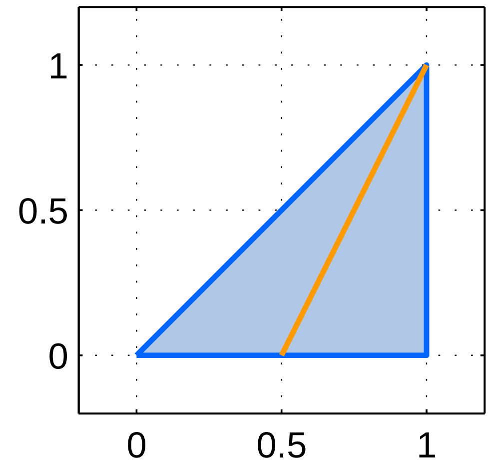
- the conditional distribution of $x \mid y = y_{\text{meas}}$ is uniform on $[y_{\text{meas}}, 1]$
- the MMSE estimator of x given $y = y_{\text{meas}}$ is

$$\hat{x}_{\text{mmse}} = \mathbf{E}(x \mid y = y_{\text{meas}}) = \frac{1 + y_{\text{meas}}}{2}$$

- the conditional mean square error of this estimate is

$$\mathbf{E}\left(\|x - \hat{x}_{\text{mmse}}\|^2 \mid y = y_{\text{meas}}\right) = \frac{1}{12}(y_{\text{meas}} - 1)^2$$

Example: MMSE estimation



The mean square error is

$$\mathbf{E}\big(\|\phi_{\mathsf{mmse}}(y) - x\|^2\big) = \mathbf{E}\big(e_{\mathsf{cond}}(y)\big)$$

$$= \int_{x=0}^{1}\int_{y=0}^{x} p(x,y)\, e_{\mathsf{cond}}(y)\, dy\, dx$$

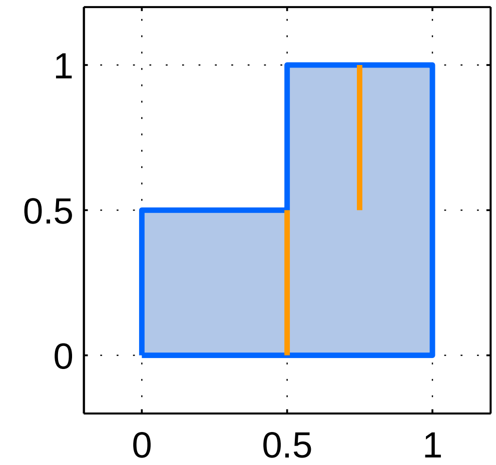
$$= \int_{x=0}^{1}\int_{y=0}^{x} \frac{1}{6}(y-1)^2\, dy\, dx$$

$$= \frac{1}{24}$$

Example: MMSE estimation

(x, y) are uniformly distributed on the L-shaped region A, i.e., the pdf is

$$p(x,y) = \begin{cases} \frac{4}{3} & \text{if } (x,y) \in A \\ 0 & \text{otherwise} \end{cases}$$



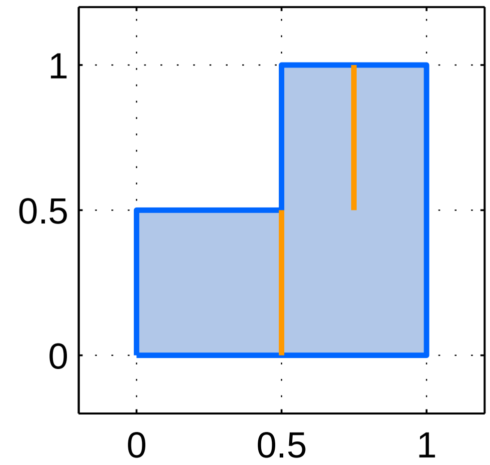
- the MMSE estimator of x given $y = y_{\text{meas}}$ is

$$\hat{x}_{\text{mmse}} = \mathbf{E}(x \mid y = y_{\text{meas}}) = \begin{cases} \frac{1}{2} & \text{if } 0 \le y_{\text{meas}} \le \frac{1}{2} \\ \frac{3}{4} & \text{if } \frac{1}{2} < y_{\text{meas}} \le 1 \end{cases}$$

- the mean square error of this estimate is

$$\mathbf{E}\left(\|x - \hat{x}_{\text{mmse}}\|^2 \mid y = y_{\text{meas}}\right) = \begin{cases} \frac{1}{12} & \text{if } 0 \le y_{\text{meas}} \le \frac{1}{2} \\ \frac{1}{48} & \text{if } \frac{1}{2} < y_{\text{meas}} \le 1 \end{cases}$$

Example: MMSE estimation



The mean square error is

$$\mathbf{E}\left(\|\phi_{\text{mmse}}(y) - x\|^2\right) = \mathbf{E}\left(e_{\text{cond}}(y)\right)$$

$$= \int_A p(x,y)\, e_{\text{cond}}(y)\, dy\, dx$$

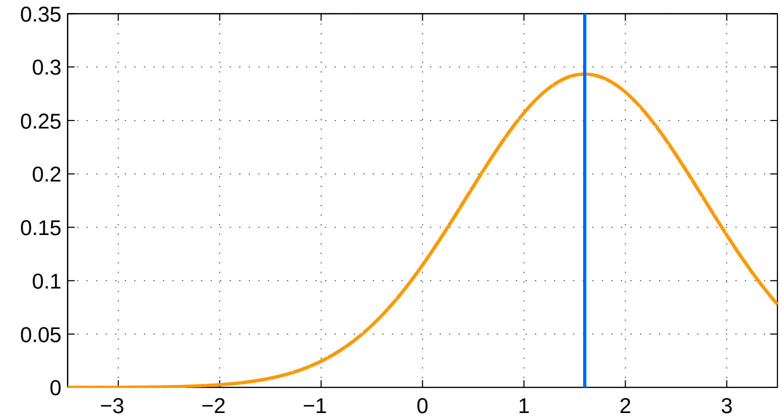
$$= \int_{x=0}^{1} \int_{y=0}^{\frac{1}{2}} \frac{1}{12} p(x,y)\, dy\, dx + \int_{x=\frac{1}{2}}^{1} \int_{y=\frac{1}{2}}^{\frac{1}{2}} \frac{1}{48} p(x,y)\, dy\, dx$$

$$= \frac{1}{16}$$

MMSE estimation for Gaussians

Suppose $\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$ where

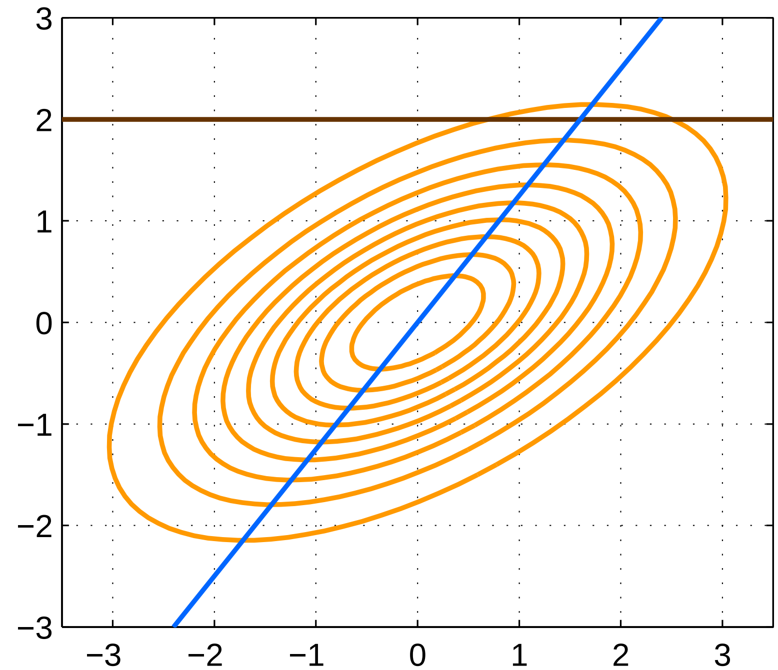
$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix}$$



We know that the conditional density of $x \mid y = y_{\text{meas}}$ is $\mathcal{N}(\mu_1, \Sigma_1)$ where

$$\mu_1 = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y_{\text{meas}} - \mu_y)$$

$$\Sigma_1 = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T$$



MMSE estimation for Gaussians

The MMSE estimator when x, y are jointly Gaussian is

$$\phi_{\text{mmse}}(y_{\text{meas}}) = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y_{\text{meas}} - \mu_y)$$

$$e_{\text{cond}}(y_{\text{meas}}) = \mathbf{trace}\big(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T\big)$$

- The conditional MSE $e_{\text{cond}}(y)$ is independent of y; a special property of Gaussians

  Hence the optimum MSE achieved is

$$\mathbf{E}\big(\|\phi_{\text{mmse}}(y) - x\|^2\big) = \mathbf{trace}\big(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T\big)$$

- The estimate $\hat{x}_{\text{mmse}}$ is an *affine* function of y_{meas}

Posterior covariance

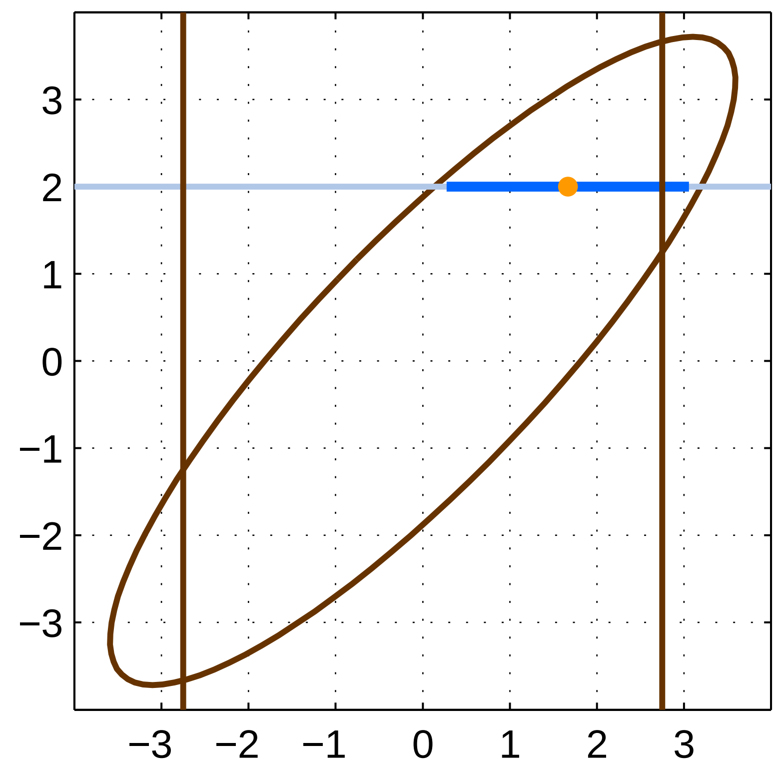
Let's look at the error $z = \phi(y) - x$. We have

$$\mathbf{cov}(z \mid y = y_{\mathsf{meas}}) = \mathbf{cov}\big(x \mid y = y_{\mathsf{meas}}\big)$$

- We use this to find a *confidence region* for the estimate
- $\mathbf{cov}\big(x \mid y = y_{\mathsf{meas}}\big)$ is called the *posterior covariance*

Example: MMSE for Gaussians

Here $\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$ with $\Sigma = \begin{bmatrix} 2.8 & 2.4 \\ 2.4 & 3 \end{bmatrix}$



We measure

$$y_{\text{meas}} = 2$$

and the MMSE estimator is

$$\begin{aligned} \phi_{\text{mmse}}(y_{\text{meas}}) &= \Sigma_{12}\Sigma_{22}^{-1} y_{\text{meas}} \\ &= 0.8 y_{\text{meas}} \\ &= 1.6 \end{aligned}$$

The posterior covariance is $Q = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx} = 0.88$

Let $\alpha^2 = Q F_{\chi^2}^{-1}(0.9)$, then $\alpha \approx 1.54$ and the confidence interval is

$$\mathbf{Prob}\Big(|x - 1.6| \leq \alpha \;\Big|\; y = y_{\text{meas}} \Big) = 0.9$$

Bias

The MMSE estimator is *unbiased*; that is the *mean error* is zero.

$$\mathbf{E}\big(\phi_{\mathsf{mmse}}(y) - x\big) = 0$$

Estimating a linear function of the unknown

Suppose

- $p(x, y)$ is a pdf on x, y
- $q = Cx$ is a random variable
- we measure y and would like to estimate q

The optimal estimator is

$$q_{\text{mmse}} = C\,\mathbf{E}\big(x \mid y = y_{\text{meas}}\big)$$

- Because $\mathbf{E}\big(q \mid y = y_{\text{meas}}\big) = C\,\mathbf{E}\big(x \mid y = y_{\text{meas}}\big)$
- The optimal estimate of Cx is C multiplied by the optimal estimate of x
- Only works for *linear functions* of x

MMSE and MAP estimation

Suppose x and y are random variables, with joint pdf $f(x,y)$

The *maximum a posterior (MAP)* estimate is the x that maximizes

$$h(x, y_{\text{meas}}) = \text{ conditional pdf of } x \,|\, (y = y_{\text{meas}})$$

- The MAP estimate also maximizes the *joint pdf*

$$x_{\text{map}} = \arg\max_x f(x, y_{\text{meas}})$$

- When x, y are jointly Gaussian, then the peak of the conditional pdf is the *conditional mean*.
- i.e., for Gaussians, the *MMSE estimate is equal to the MAP estimate.*