# W14 Notes

## Mean square error

### 06 Theory - Minimum mean square error

Suppose our problem is to *estimate* or *guess* or *predict* the value of a random variable $X$ in a particular run of our experiment. Assume we have the distribution of $X$. Which value do we choose as our guess?

There is no single best answer to this question. The "best guess" number depends on additional factors in the problem context.

One method is to pick a value where the PMF or PDF of $X$ is *maximal*. This is a value of highest probability. There may be more than one!

Another method is to pick the expected value $E[X]$. This value may be impossible!

For the normal distribution, or any symmetrical distribution, these are the same value. For most distributions, though, they are not the same.

---

**Mean square error (MSE)**

Given some estimate $\hat{x} \in \mathbb{R}$ for a random variable $X$, the **mean square error (MSE)** of $\hat{x}$ is:

$$E\left[(X - \hat{x})^2\right] \quad = \text{"cost function"}$$

(one example; we study this unit)

The MSE quantifies the typical (square of the) error. Error here means the difference between the true value $X$ and the estimate $\hat{x}$.

Other error estimates are reasonable and useful in niche contexts. For example, $E\left[\,|X - \hat{x}|\,\right]$ or $\mathrm{Max}\,|X - \hat{x}|$. They are not frequently used, so we do not consider their theory further.

Minimizer of MSE:

Find $\hat{x}$:

$$E[\underbrace{(X-\hat{x})^2}_{X^2 - 2\hat{x}X + \hat{x}^2}] = \overset{\text{number}}{E[X^2]} - 2\hat{x}\,E[X] + (\hat{x})^2$$

variable

Quadratic function. Minimize with calculus.

$$\frac{d}{d\hat{x}}\left(E[X^2] - 2\hat{x}E[X] + \hat{x}^2\right)$$

$$= -2E[X] + 2\hat{x} \overset{\text{to find max}}{=} 0$$

$\rightsquigarrow$ $\boxed{\hat{x} = E[X]}$

Then what is that error?

MSE @ $\hat{x} = E[X]$

is $E[(X - E[X])^2]$

i.e. $= Var[X]$

---

In problem contexts where large errors are more costly than small errors (i.e. many real problems), the most likely value of $X$ (the point with maximal PDF) may fare poorly as an estimate.

It turns out that the *expected value* $E[X]$ also happens to be the value that *minimizes the MSE*.

**Expected value minimizes MSE**

Given a random variable $X$, its expected value $\hat{x} = E[X]$ is the estimate of $X$ with **minimal mean square error**.

The MSE error for $\hat{x} = E[X]$ is:

$$E[\,(X - \hat{x})^2\,] \;=\; \mathrm{Var}[X]$$

**Proof that $E[X]$ minimizes MSE**

Expand the MSE error:

$$E[(X - \hat{x})^2] \quad \gg\gg \quad E[X^2] - 2\hat{x}\,E[X] + \hat{x}^2$$

Now minimize this parabola. Differentiate:

$$\frac{d}{d\hat{x}}\,E[(X - \hat{x})^2] \quad \gg\gg \quad 0 - 2E[X] + 2\hat{x}$$

Find zeros:

$$0 - 2E[X] + 2\hat{x} \;=\; 0$$

$$\gg\gg \quad 2\hat{x} \;=\; 2E[X] \quad \gg\gg \quad \hat{x} \;=\; E[x]$$

---

When the estimate $\hat{x}$ is made in the absence of information (besides the distribution of $X$), it is called a **blind estimate**. Therefore, $\hat{x}_B = E[X]$ is the blind minimal MSE estimate, and $e_B = \mathrm{Var}[X]$ is the error of this estimate.

In the presence of additional information, for example that event $A$ is known, then the MSE estimate is $\hat{x}_A = E[X \mid A]$ and the error of this estimate is $e_{X|A} = \mathrm{Var}[X \mid A]$.

The MSE estimate can also be conditioned on another variable, say $Y$:

**Minimal MSE of $X$ given $Y$**

The minimal MSE estimate of $X$ given another variable $Y$:

$$\hat{x}_M(y) \;=\; E[X \mid Y = y]$$

The error of this estimate is $\mathrm{Var}[X \mid Y = y]$, which equals $E\big[\,(X - \hat{x}_M(y))^2 \mid Y = y\,\big]$.

Notice that the minimal MSE of $X$ given $Y$ can be used to define a random variable:

$$\hat{X}_M(Y) \;=\; E[X \mid Y] \;=\; \hat{x}_M(Y)$$

This is a derived variable from $Y$ given by composing $Y$ with the function $\hat{x}_M$.

The variable $\hat{X}_M(Y)$ provides the minimal MSE estimates of $X$ when experimental outcomes are viewed as providing the information of $Y$ only, and the model is used to derive estimates of $X$ from this information.

## 07 Illustration

**Example - Minimal MSE estimate given PMF, given fixed event**

Suppose $X$ has the following PMF:

| $k$ | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| $P_X(k)$ | 0.15 | 0.28 | 0.26 | 0.19 | 0.13 |

Find the minimal MSE estimate of $X$, given that $X$ is even. What is the error of this estimate?

**Solution**

The minimal MSE given $A$ is just $E[X \mid A]$ where $A = \{2, 4\}$.

First compute the conditional PMF:

$$P_{X|A}(k) \quad = \quad \begin{cases} 0.19/0.47 & k = 4 \\ 0.28/0.47 & k = 2 \\ 0 & k \neq 2, 4 \end{cases}$$

Therefore:

$$\hat{x}_A \quad = \quad 2\frac{0.28}{0.47} + 4\frac{0.19}{0.47} \quad \approx \quad 2.80851$$

The error is:

$$e_{X|A} \quad = \quad (2 - 2.81)^2 \frac{0.28}{0.47} + (4 - 2.81)^2 \frac{0.19}{0.47}$$

$$\gg\gg \quad \approx 0.9633$$

### Exercise - Minimal MSE estimate from joint PDF

Here is the joint PDF of $X$ and $Y$:

$$f_{X,Y} \quad = \quad \begin{cases} 8xy & 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the minimal MSE estimate of $X$ in terms of $Y$.

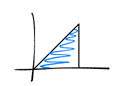
What is the estimate of $X$ when $Y = 0.2$? When $Y = 0.8$?

**Answer**

$$\hat{x}_M(y) \quad = \quad \frac{2}{3} \cdot \frac{1 - y^3}{1 - y^2}$$

$\hat{x}_M(0.2) \;=\; 0.6889 \qquad \hat{x}_M(0.8) \;=\; 0.9037$

$$f_{X,Y} = \begin{cases} 8xy & 0 \le y \le x \le 1 \\ 0 & \text{else} \end{cases}$$



$$\hat{x}_M(y) = E[X \mid Y = y]$$

$$= \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

$$f_{X|Y} = \frac{f_{X,Y}}{f_Y} \quad \text{(in domain)}$$

$$f_Y = \int_{-\infty}^{\infty} f_{X,Y}\, dx = \int_{x=y}^{1} 8xy\, dx$$

$$= 4x^2 y \Big|_y^1 = 4y(1 - y^2)$$

Thus $f_{X|Y} = \frac{8xy}{4y(1-y^2)} = \frac{2x}{1-y^2} \quad 0 \le y \le x \le 1$

$$\int_{x=y}^{1} x \frac{2x}{1-y^2}\, dx = \frac{2x^3}{3(1-y^2)} \Big|_{x=y}^{1}$$

$$= \boxed{\frac{2}{3} \cdot \frac{1-y^3}{1-y^2}}$$

## 08 Theory - Line of minimal MSE

Linear approximation is very common in applied math.

One *could* consider the linearization of $\hat{x}_M(y)$ (its tangent line at some *point*) instead of the exact function $\hat{x}_M(y)$.

One could instead minimize the MSE over all linear functions of $Y$ as an RV. The line with minimal MSE is called the **linear estimator**.

The difference here is:

- line of best fit *at a point* vs.
- line of best fit *over an interval*

### Linear estimator: Line of minimal MSE

Let $L(y)$ be an arbitrary line $L(y) = ay + b$. Let $\hat{X}_L(Y) = L(Y) = aY + b$.

The mean square error (MSE) of this line $L$ is:

$$e_L(a,b) \;=\; E\left[\left(X - \hat{X}_L(Y)\right)^2\right]$$

The **linear estimator** of $X$ in terms of $Y$ is the line $L_{\min}$ with minimal MSE, and it is:

$$L_{\min}(y) \quad=\quad \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) + \mu_X$$

The error value at the (best) linear estimator, $e_{L_{\min}}$, is:

$$e_{L_{\min}} \;=\; \sigma_X^2(1 - \rho_{X,Y}^2)$$

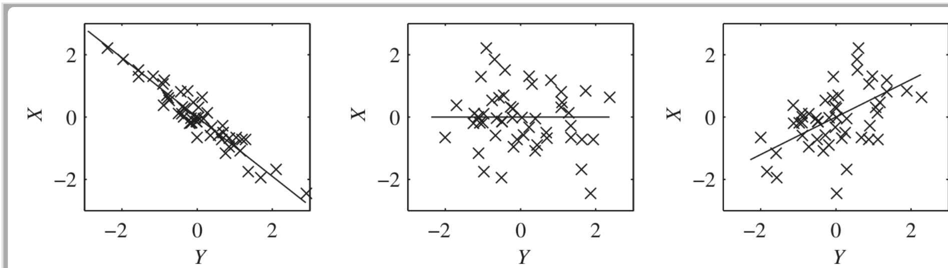
Theorem: The error variable of the linear estimator, $X - \hat{X}_{L_{\min}}(Y)$, is perfectly uncorrelated with $Y$.

### Slope and $\rho_{X,Y}$

Notice:

$$\frac{\hat{X}_{L_{\min}}(Y) - \mu_X}{\sigma_X} \quad=\quad \rho_{X,Y}\cdot\left(\frac{Y-\mu_Y}{\sigma_Y}\right)$$

Thus, for *standardized* variables $X$ and $Y$, it turns out $\rho_{X,Y}$ is the *slope* of the linear estimator.



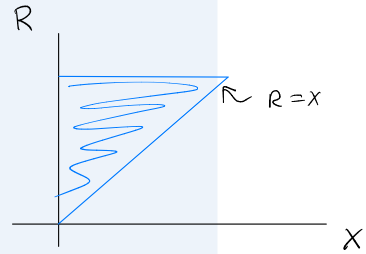

(a) $\rho_{X,Y} = -0.95$ (b) $\rho_{X,Y} = 0$ (c) $\rho_{X,Y} = 0.60$

In each graph, $E[X] = E[Y] = 0$ and $\mathrm{Var}[X] = \mathrm{Var}[Y] = 1$.

The line of minimal MSE is the "best fit" line, $\hat{X}_{L_{\min}}(Y) = \rho_{X,Y}\,Y$.

## 09 Illustration

### ☰ Example - Estimating on a variable interval




Suppose that $R \sim \text{Unif}((0,1))$ and suppose $X \sim \text{Unif}(0, R)$.

(a) Find $\hat{x}_M(r)$ (b) Find $\hat{r}_M(x)$ (c) Find $\hat{R}_{L_{\min}}(X)$

**Solution**

(a) Find $\hat{x}_M(r)$:

$\hat{x}_M(r) = \frac{r}{2}$

We know $\hat{x}_M(r) = E[X \mid R = r]$.

Given $R = r$, so $X$ is uniform on $(0, r)$, we have $E[X \mid R = r] = \frac{r}{2}$.

---

(b) Find $\hat{r}_M(x)$:

We know $\hat{r}_M(x) = E[R \mid X = x]$.

We know $f_R$ and $f_{X|R}$. From these we derive the joint distribution $f_{X,R}$:

$$f_R(r) = \begin{cases} 1 & r \in (0,1) \\ 0 & \text{otherwise} \end{cases} \qquad f_{X|R}(x|r) = \begin{cases} 1/r & x \in (0,r) \\ 0 & \text{otherwise} \end{cases}$$

$$\gg\gg \quad f_{X,R}(x,r) \quad = \quad f_{X|R} \cdot f_R \;=\; \begin{cases} 1/r & 0 < x < r < 1 \\ 0 & \text{else} \end{cases}$$

Now extract the marginal $f_X$:

$$\gg\gg \quad f_X(x) \quad = \quad \int_{-\infty}^{\infty} f_{X,R}(x,r)\, dr$$

$$\gg\gg \quad \int_x^1 \frac{1}{r}\, dr \quad \gg\gg \quad -\ln x \qquad (0 < x < 1)$$

Now deduce the conditional $f_{R|X}$:

$$f_{R|X} \quad = \quad \frac{f_{X,R}}{f_X} \;=\; \begin{cases} \frac{-1}{r \ln x} & 0 < x < r < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then:

$$E[R \mid X = x] \quad \gg\gg \quad \int_x^1 r \frac{-1}{r \ln x}\, dr$$

$$\gg\gg \quad \frac{x-1}{\ln x}$$

So $\hat{r}_M(x) = \frac{x-1}{\ln x}$.

---

(c) Find $\hat{R}_{L_{\min}}(X)$:

We need all the basic statistics.

$E[R] \;=\; 1/2$ because $R \sim \text{Unif}((0,1))$.

$\sigma_R^2 \;=\; \frac{(b-a)^2}{12} \;=\; 1/12$.

$E[X] \;=\; 1/4$ using the marginal PDF $f_X(x) = -\ln x$ on $x \in (0,1)$. (IBP and L'Hopital are needed.)

$\sigma_X \;=\; \sqrt{7}/12$ also using the marginal $f_X(x) = -\ln x$.

$E[XR] = 1/6$ using $f_{X,R}(x,r)$, namely:

$$E[XR] \;=\; \int_{r=0}^{1} \int_{x=0}^{r} xr \frac{1}{r}\, dx\, dr$$

$$\gg\gg \quad \int_0^1 \frac{x^2}{2}\, dx \quad \gg\gg \quad \frac{1}{6}$$

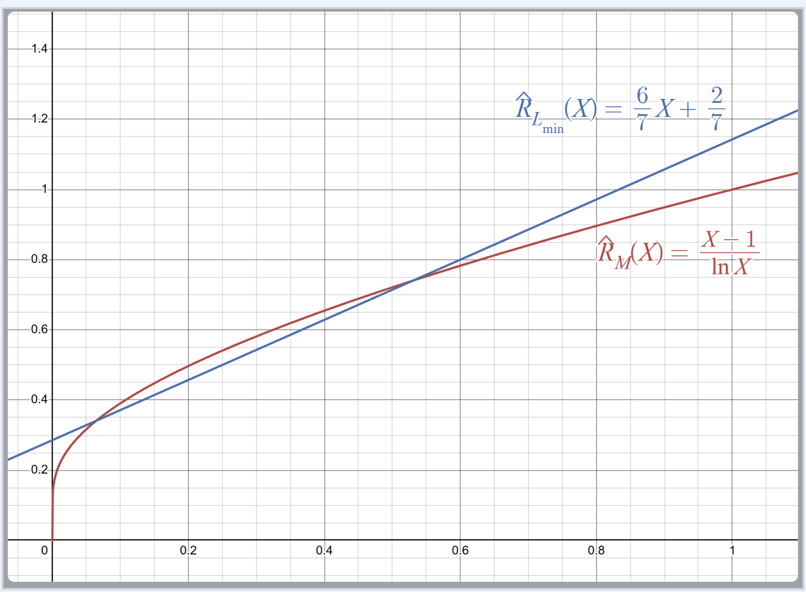
From all this we infer $\text{Cov}[X,R] = 1/24$ and $\rho_{X,R} = \sqrt{3/7}$.

Hence:

$$L_{\min}(x) \;=\; \frac{6}{7}x + \frac{2}{7}$$

Thus:

$$\hat{R}_{L_{\min}}(X) \;=\; \frac{6}{7}X + \frac{2}{7}$$




### Exercise - Line of minimal MSE given joint PDF

Here is the joint PDF of $X$ and $Y$:

$$f_{X,Y} \;=\; \begin{cases} 8xy & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the line giving the linear MSE estimate of $X$ in terms of $Y$.

What is the expected error of this line, $e_{L_{\min}}$?

What is the estimate of $X$ when $Y = 0.2$? When $Y = 0.8$?

**Answer**

$$\hat{X}_{L_{\min}}(Y) \ = \ 0.3637\,Y + 0.6060$$

$$e_{L_{\min}} \ = \ 0.02020$$

$$\hat{x}_{L_{\min}}(0.2) \ = \ 0.67874 \qquad \hat{x}_{L_{\min}}(0.8) \ = \ 0.89696$$

Here is the joint PDF of $X$ and $Y$:

$$f_{X,Y} = \begin{cases} 8xy & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

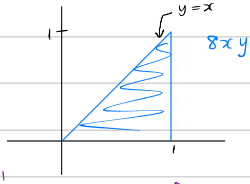
Find the line giving the linear MSE estimate of $X$ in terms of $Y$.

What is the expected error of this line, $e_{L_{\min}}$?

What is the estimate of $X$ when $Y = 0.2$? When $Y = 0.8$?

$$L_{min}(y) = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) + \mu_X$$

$$\frac{Cov[X,Y]}{\sigma_X \sigma_Y} \cdot \frac{\sigma_X}{\sigma_Y} = \frac{Cov[X,Y]}{Var[Y]}$$




$$E[X] = \int_{y=0}^{1}\int_{x=y}^{1} x(8xy)\,dx\,dy$$

$$= \int_0^1 \frac{8}{3}y(1-y^3) = \frac{4}{3}y^2 - \frac{8}{15}y^4\Big|_0^1 = \frac{4}{3} - \frac{8}{15} = \frac{20-8}{15} = \frac{12}{15} = \frac{4}{5}$$

have $f_Y = 4y(1-y^2)$

$$E[Y] = \int_{y=0}^{1} y(4y(1-y^2))\,dy = \frac{4}{3}y^3 - \frac{4}{5}y^5\Big|_0^1 = \frac{20-12}{15} = \frac{8}{15}$$

$$E[Y^2] = \int_0^1 y^2(4y(1-y^2))\,dy = y^4 - \frac{2}{3}y^6\Big|_0^1 = \frac{1}{3}$$

$$E[XY] = \int_0^1\int_{x=y}^{1} xy\,(8xy)\,dx\,dy = \int_0^1 \frac{8}{3}y^2(1-y^3)\,dy = \frac{8}{9}y^3 - \frac{8}{18}y^6\Big|_0^1 = \frac{16-8}{18} = \frac{4}{9}$$

$\rightarrow$ $$Cov[X,Y] = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}$$

$\rightarrow$ $$Var[Y] = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{75-64}{9 \cdot 25} = \frac{11}{225}$$

$$\rightarrow a = \frac{Cov}{\sigma_Y^2} = \frac{4}{11}$$

Thus: $$L_{min}(y) = \frac{4}{11}\left(y - \frac{8}{15}\right) + \frac{4}{5} = \boxed{\frac{4}{11}y + \frac{20}{33} = L_{min}(y)}$$

$$\rho_{X,Y}^2 = \frac{Cov[X,Y]^2}{\sigma_X^2\sigma_Y^2}$$

$$e_{L_{min}} = \sigma_X^2(1 - \rho_{X,Y}^2)$$

$$E[X^2] = \int_0^1\int_{x=y}^{1} x^2(8xy)\,dx\,dy = \int_0^1 2y(1-y^4)\,dy = y^2 - \frac{y^6}{3}\Big|_0^1 = \frac{2}{3}$$

$\rightarrow$ $$Var[X] = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{6}{225}$$

$$e_{L_{min}} = \frac{6}{225}\left(1 - \frac{(4/225)^2}{\frac{6}{225}\cdot\frac{11}{225}}\right) = \boxed{\frac{2}{99} = e_{L_{min}}}$$