Week 14 notes

Statistical testing cont'd

01 Theory - Binary testing, MAP and ML

Binary hypothesis test

Ingredients of a binary hypothesis test:

- Complementary hypotheses H_0 and H_1
  - Maybe also know the **prior probabilities** $P[H_0]$ and $P[H_1]$
  - Goal: determine which case we are in, H_0 or H_1
- Decision rule made of complementary events A_0 and A_1
  - A_0 is likely given H_0, while A_1 is likely given H_1
  - Decision rule: outcome A_0, accept H_0; outcome A_1, accept H_1
  - Usually: A_i written in terms of **decision statistic** X using a **design**
  - We cover three **designs**:
    - MAP and ML (minimize 'error probability')
    - MC (minimizes 'error cost')
  - Designs use $P_{X|H_0}$ and $P_{X|H_1}$ (or $f_{X|H_0}$, $f_{X|H_1}$) to construct A_0 and A_1

MAP design

Suppose we know:

- Both prior probabilities $P[H_0]$ and $P[H_1]$
- Both conditional distributions $P_{X|H_0}(x)$ and $P_{X|H_1}(x)$ (or $f_{X|H_0}(x)$ and $f_{X|H_1}(x)$)

The **maximum a posteriori probability (MAP)** design for a decision statistic X:

$$A_0 \quad = \quad \text{set of } x \text{ for which:}$$

Discrete case:

$$P_{X|H_0}(x) \cdot P[H_0] \quad \geq \quad P_{X|H_1}(x) \cdot P[H_1]$$

Continuous case:

$$f_{X|H_0}(x) \cdot P[H_0] \quad \geq \quad f_{X|H_1}(x) \cdot P[H_1]$$

Then $A_1 \;=\; \{x \in \mathbb{R} \mid x \notin A_0\}$.

The MAP design minimizes the total probability of error.

ML design

Suppose we know only:

- Both conditional distributions

The **maximum likelihood (ML)** design for X:

$$A_0 \quad = \quad \text{set of } x \text{ for which:} \quad \begin{matrix} P_{X|H_0}(x) \ \geq \ P_{X|H_1}(x) & \text{(discrete)} \\ f_{X|H_0}(x) \ \geq \ f_{X|H_1}(x) & \text{(continuous)} \end{matrix}$$

ML is a simplified version of MAP. (Set $P[H_0]$ and $P[H_1]$ to 0.5.)

The probability of a *false alarm*, a Type I error, is called P_{FA}.

The probability of a *miss*, a Type II error, is called P_{Miss}.

$$P_{FA} \quad = \quad P[A_1 \mid H_0]$$

$$P_{\text{Miss}} \quad = \quad P[A_0 \mid H_1]$$

Total probability of error:

$$P_{\text{ERR}} \quad = \quad P[A_1 \mid H_0] \cdot P[H_0] + P[A_0 \mid H_1] \cdot P[H_1]$$

⚠ False alarm $\neq$ false alarm

Suppose A_1 sets off a smoke alarm, and H_0 is 'no fire' and H_1 is 'yes fire'.

Then P_{FA} is the odds that we get an alarm *assuming there is no fire*.

This is *not* the odds of *experiencing* a false alarm (no context). That would be $P[A_1 H_0]$.

This is *not* the odds of a *given* alarm being a false one. That would be $P[H_0 \mid A_1]$.

02 Illustration

Example - ML test: Smoke detector

Suppose that a smoke detector sensor is configured to produce 8 V when there is smoke, and 0 V otherwise. But there is background noise with distribution $\mathcal{N}(0, 3^2\,\text{V})$.

Design an ML test for the detector electronics to decide whether to activate the alarm.

What are the three error probabilities? (Type I, Type II, Total.)

Solution

First, establish the conditional distributions:

$$X \mid H_0 \ \sim \mathcal{N}(0, 3^2) \qquad X \mid H_1 \ \sim \mathcal{N}(8, 3^2)$$

Density functions:

$$f_{X|H_0} \ = \ \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \qquad f_{X|H_1} \ = \ \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2}$$

The ML condition becomes:

$$\frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \quad \overset{?}{\geq} \quad \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2}$$

$$\gg\gg \quad -\frac{1}{2}\left(\frac{x-0}{3}\right)^2 \quad \overset{?}{\geq} \quad -\frac{1}{2}\left(\frac{x-8}{3}\right)^2$$

$$\gg\gg \quad x^2 \quad \overset{?}{\leq} \quad (x-8)^2$$

$$\gg\gg \quad x \leq 4$$

Therefore, A_0 is $x \leq 4$, while A_1 is $x > 4$.

The decision rule is: activate alarm when $x > 4$.

Type I error:

$$P_{FA} \ = \ P[A_1 \mid H_0] \quad \gg\gg \quad P[X > 4 \mid H_0]$$

$$\gg\gg \quad 1 - P\left[\frac{X-0}{3} \leq \frac{4}{3} \,\middle|\, H_0\right]$$

$$\gg\gg \quad 1 - P[Z \leq 1.3333] \quad \gg\gg \quad \approx 0.0912$$

Type II error:

$$P_{\text{Miss}} \quad = \quad P[A_0 \mid H_1] \quad \gg\gg \quad P[X \leq 4 \mid H_1]$$

$$\gg\gg \quad P\left[\frac{X-8}{3} \leq \frac{4-8}{3} \,\middle|\, H_1\right]$$

$$\gg\gg \quad P[Z \leq -1.3333] \quad \gg\gg \quad \approx 0.0912$$

Total error:

$$P_{\text{ERR}} \ = \ P_{FA} \cdot 0.5 + P_{\text{Miss}} \cdot 0.5 \quad \approx \quad 0.0912$$

☰ Example - MAP test: Smoke detector

Suppose that a smoke detector sensor is configured to produce 8 V when there is smoke, and 0 V otherwise. But there is background noise with distribution $\mathcal{N}(0, 3^2\,\mathrm{V})$.

Suppose that the background chance of smoke is 5%. Design a MAP test for the alarm.

What are the three error probabilities? (Type I, Type II, Total.)

Solution

First, establish priors:

$$P[H_0] \;=\; 0.95 \qquad P[H_1] \;=\; 0.05$$

The MAP condition becomes:

$$\frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \cdot 0.95 \quad \overset{?}{\geq} \quad \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2} \cdot 0.05$$

$$\gg\gg \quad e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \quad \overset{?}{\geq} \quad e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2} \cdot \frac{0.05}{0.95}$$

$$\gg\gg \quad -\frac{1}{2}\left(\frac{x-0}{3}\right)^2 \quad \overset{?}{\geq} \quad -\frac{1}{2}\left(\frac{x-8}{3}\right)^2 + \ln\left(\frac{0.05}{0.95}\right)$$

$$\gg\gg \quad x^2 \quad \overset{?}{\leq} \quad (x-8)^2 - 18\ln\left(\frac{0.05}{0.95}\right)$$

$$\gg\gg \quad x \leq 7.31$$

Therefore, A_0 is $x \leq 7.31$, while A_1 is $x > 7.31$.

The decision rule is: activate alarm when $x > 7.31$.

Type I error:

$$P_{FA} \;=\; P[A_1 \mid H_0] \quad \gg\gg \quad P[X > 7.31 \mid H_0]$$

$$\gg\gg \quad 1 - P[Z \leq 2.4367] \quad \gg\gg \quad \approx 0.007411$$

Type II error:

$$P_{\text{Miss}} \quad = \quad P[A_0 \mid H_1] \quad \gg\gg \quad P[X \leq 7.31 \mid H_1]$$

$$\gg\gg \quad P[Z \leq -0.23] \quad \gg\gg \quad \approx 0.4090$$

Total error:

$$P_{\text{ERR}} \;=\; P_{FA} \cdot 0.95 + P_{\text{Miss}} \cdot 0.05 \quad \approx \quad 0.02749$$

03 Theory - MAP criterion proof

Explanation of MAP criterion - discrete case

First, we show that the MAP design selects for A_0 all those x which render H_0 more likely than H_1.

Observe this Calculation:

$$P[H_i \mid X = x] \;=\; P[X = x \mid H_i] \cdot \frac{P[H_i]}{P[X]} \qquad \text{(Bayes' Rule)}$$

$$= \; P_{X|H_i}(x) \cdot \frac{P[H_i]}{P[X]} \qquad \text{(Conditional PMF)}$$

Now, take the condition for A_0, and cross-multiply:

$$\gg\gg \quad P_{X|H_0}(x) \cdot P[H_0] \;\geq\; P_{X|H_1}(x) \cdot P[H_1]$$

Divide both sides by $P[X]$ and apply the above Calculation in reverse:

$$\gg\gg \quad P[H_0 \mid X = x] \;\geq\; P[H_1 \mid X = x]$$

This is what we sought to prove.

Next, we verify that the MAP design minimizes the total probability of error.

The total probability of error is:

$$P_{\text{ERR}} \quad = \quad P[A_1 \mid H_0] \cdot P[H_0] + P[A_0 \mid H_1] \cdot P[H_1]$$

Expand this with summation notation (assuming the discrete case):

$$\gg\gg \quad \sum_{x \in A_1} P_{X|H_0}(x) \cdot P[H_0] + \sum_{x \in A_0} P_{X|H_1}(x) \cdot P[H_1]$$

Now, how do we choose the set $A_0 \subset \mathbb{R}$ (and thus $A_1 = A_0^c$) in such a way that this sum is minimized?

Since all terms are positive, and any $x \in \mathbb{R}$ may be placed in A_1 or in A_0 freely and independently of all other choices, the total sum is minimized when we minimize the impact of placing each x.

So, for each x, we place it in A_0 if:

$$P_{X|H_0}(x) \cdot P[H_0] \;\geq\; P_{X|H_1}(x) \cdot P[H_1]$$

That is equivalent to the MAP condition.

04 Theory - MC design

- Write C_{10} for cost of false alarm, i.e. cost when H_0 is true but decided H_1.
    - Probability of incurring cost C_{10} is $P_{FA} \cdot P[H_0]$.
- Write C_{01} for cost of miss, i.e. cost when H_1 is true but decided H_0.

- Probability of incurring cost C_{01} is $P_{\text{Miss}} \cdot P[H_1]$.

Expected value of cost incurred

$$E[C] \quad = \quad P[A_1 \mid H_1] \cdot P[H_0] \cdot C_{10} + P[A_0 \mid H_1] \cdot P[H_1] \cdot C_{01}$$

MC design

Suppose we know:

- Both prior probabilities $P[H_0]$ and $P[H_1]$
- Both conditional distributions $P_{X|H_0}(x)$ and $P_{X|H_1}(x)$ (or $f_{X|H_0}(x)$ and $f_{X|H_1}(x)$)

The **minimum cost (MC)** design for a decision statistic X:

$$A_0 \quad = \quad \text{set of } x \text{ for which:}$$

Discrete case:

$$P_{X|H_0}(x) \cdot P[H_0] \cdot C_{10} \quad \geq \quad P_{X|H_1}(x) \cdot P[H_1] \cdot C_{01}$$

Continuous case:

$$f_{X|H_0}(x) \cdot P[H_0] \cdot C_{10} \quad \geq \quad f_{X|H_1}(x) \cdot P[H_1] \cdot C_{01}$$

Then $A_1 \;=\; \{x \in \mathbb{R} \mid x \notin A_0\}$.

The MC design minimizes the expected value of the cost of error.

MC minimizes expected cost

Inside the argument that MAP minimizes total probability of error, we have this summation:

$$P_{\text{ERR}} \quad = \quad \sum_{x \in A_1} P_{X|H_0}(x) \cdot P[H_0] + \sum_{x \in A_0} P_{X|H_1}(x) \cdot P[H_1]$$

The expected value of the cost has a similar summation:

$$E[C] \quad = \quad \sum_{x \in A_1} P_{X|H_0}(x) \cdot P[H_0] \cdot C_{10} + \sum_{x \in A_0} P_{X|H_1}(x) \cdot P[H_1] \cdot C_{01}$$

Following the same reasoning, we see that the cost is minimized if each x is placed into A_0 precisely when the MC design condition is satisfied, and otherwise it is placed into A_1.

05 Illustration

Example - MC Test: Smoke detector

Suppose that a smoke detector sensor is configured to produce 8 V when there is smoke, and 0 V otherwise. But there is background noise with distribution

$\mathcal{N}(0, 3\,\mathrm{V})$.

Suppose that the background chance of smoke is 5%. Suppose the cost of a miss is 50× the cost of a false alarm. Design an MC test for the alarm.

Compute the expected cost.

Solution

We have priors:

$$P[H_0] \;=\; 0.95 \qquad P[H_1] \;=\; 0.05$$

And we have costs:

$$C_{10} = 1 \qquad C_{01} = 50$$

(The ratio of these numbers is all that matters in the inequalities of the condition.)

The MC condition becomes:

$$\frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \cdot 0.95 \cdot 1 \quad \overset{?}{\geq} \quad \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2} \cdot 0.05 \cdot 50$$

$$\gg\gg \quad e^{-\frac{1}{2}\left(\frac{x-0}{3}\right)^2} \quad \overset{?}{\geq} \quad e^{-\frac{1}{2}\left(\frac{x-8}{3}\right)^2} \cdot \frac{2.5}{0.95}$$

$$\gg\gg \quad -\frac{1}{2}\left(\frac{x-0}{3}\right)^2 \quad \overset{?}{\geq} \quad -\frac{1}{2}\left(\frac{x-8}{3}\right)^2 + \ln\left(\frac{2.5}{0.95}\right)$$

$$\gg\gg \quad x^2 \quad \overset{?}{\leq} \quad (x-8)^2 - 18\ln\left(\frac{2.5}{0.95}\right)$$

$$\gg\gg \quad x \leq 2.91$$

Therefore, A_0 is $x \leq 2.91$, while A_1 is $x > 2.91$.

The decision rule is: activate alarm when $x > 2.91$.

Type I error:

$$P_{FA} \;=\; P[A_1 \mid H_0]$$

$$\gg\gg \quad P[X > 2.91 \mid H_0] \quad \gg\gg \quad \approx 0.1660$$

Type II error:

$$P_{\text{Miss}} \;=\; P[A_0 \mid H_1]$$

$$\gg\gg \quad P[X \leq 2.91] \quad \gg\gg \quad \approx 0.04488$$

Total error:

$$P_{\text{ERR}} \;=\; P_{FA} \cdot 0.95 + P_{\text{Miss}} \cdot 0.05 \quad \approx \quad 0.1599$$

PMF of total cost:

$$P_C(c) \quad = \quad \begin{cases} 0.002244 & c = 50 \\ 0.1577 & c = 1 \\ 0.840056 & c = 0 \end{cases}$$

Therefore $E[C] \;=\; 0.27$.

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 one run of the experiment. Assume we have the distribution of X. Which value do we choose?

There is no single best answer to this question. The best answer is a function of 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 $E[X]$.

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

Mean square error

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

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

The MSE quantifies the typical (square of the) error, meaning the difference between the true value X and the estimate $\hat{x}$. The expected value calculates the typical value of this error.

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

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

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

Minimal mean square error

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

The MSE error itself of $\hat{x} = E[X]$:

$$\text{MSE error when } \hat{x} = E[X]: \qquad E[(X - \hat{x})^2] \quad = \quad \text{Var}[X]$$

Proof that $E[X]$ gives minimal MSE

Expand the MSE error:

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

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:

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

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

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

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 $e_B = \text{Var}[X]$ is the error of this estimate.

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

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:

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

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

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

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

This variable is a derived variable of Y given by post-composition 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:

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

$$\ggg \quad \approx 0.9633$$

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

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

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

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) instead of the exact function $\hat{x}_M(y)$.

Instead, one can minimize the MSE over all possible linear functions of Y. The line with minimal MSE is called the **linear estimator**.

Line of minimal MSE

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

The mean square error (MSE) of L is:

$$e_L(a,b) \quad = \quad E\Big[\big(X - \hat{X}_L(Y)\big)^2\Big]$$

The **linear estimator** 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 minimal error value $e_{L_{\min}}$ is:

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

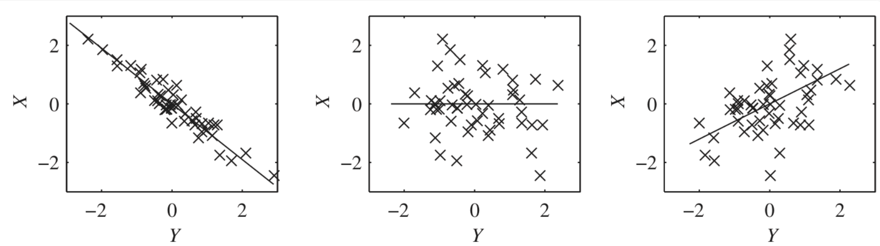
The variable of minimal error, $X - \hat{X}_{L_{\min}}(Y)$, is uncorrelated with Y.

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, $\rho_{X,Y}$ is the *slope* of the minimal MSE line for *standardized* variables X and Y.



(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 $\text{Var}[X] = \text{Var}[Y] = 1$.

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 $\hat{x}_M(r)$.

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

To compute this function, we calculate a sequence of densities.

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:

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

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

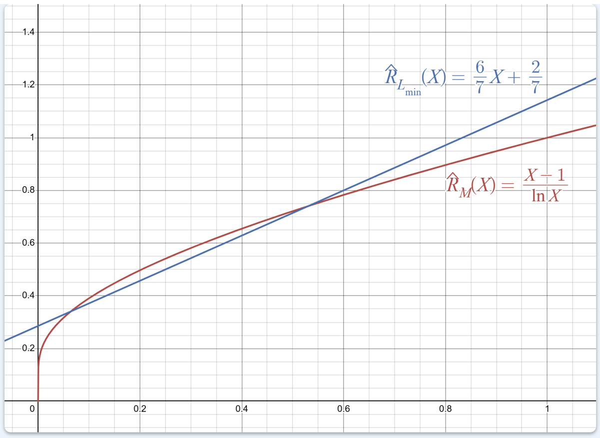
From this we infer $\text{Cov}[X,R] = 1/24$ and $\rho_{X,R} = \sqrt{3/7}$.

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:

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

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

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

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

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