Parameter estimation

Ex Know observation Y depends on some parameter(s), but we don't know the values of these parameters.
So we take some observations, and estimate the parameters

Ex Cannot measure a RV X directly
(for example, temperature of the sun)
(or because it's distorted by noise)
(or because it will happen in the future),
So we observe Y (which is hopefully correlated with X) and estimate X

Example Y is Gaussian, mean 0, unknown variance

Let variance by $X+1$ and X be an exponential RV

Example Y is binomial w/ parameters n, p, but p is unknown.

Let $X = p$ and define a prior $f_X(x)$

How to estimate?

a) what information is available to help?
   - observations?
   - prior knowledge?

b) how do we define "goodness" of the estimate?
   - small error
     - mean squared error
     - maximum absolute error
   - maximally likely (most probable)
   - easy to implement
     - example: linear

This gives rise to several popular estimators which are summarized on the next page

Estimator	Available information		goodness
	prior	observations	
$\hat{x} = E(X)$	yes	no	minimum mean-squared error (MMSE)
$\hat{x} = E(X\|A)$	yes	range	
$\hat{x} = E(X\|y)$	yes	$Y=y$	
$\hat{x} = a^* y + b^*$	yes	$Y=y$	Linear MMSE
$\underset{x}{\operatorname{argmax}}\ f_X(x\|y)$ $= \underset{x}{\operatorname{argmax}}\ f_{XY}(x,y)$	yes	$Y=y$	Maximum A posteriori Probability (uses Bayes Rule) (MAP)
$\underset{x}{\operatorname{argmax}}\ f_Y(y\|x)$	no	$Y = y$	Maximum Likelihood (ML)

Minimum Mean Square Error (MMSE) estimators

a) no observations

"blind estimation"

$$\hat{x} = E(X)$$

b) observation is in set A $\qquad X \in A$

$$\hat{x} = E(X \mid A)$$

c) observation $Y=y$ and prior knowledge

$$\hat{x} = E(x \mid y)$$

<u>why?</u> Find the $\hat{x}$ that minimizes the MSE:

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

where B indicates our knowledge, example $B = \{Y = y\}$

$$E((X - \hat{x})^2 \mid B) = E(X^2 \mid B) - 2E(X \mid B)\hat{x} + \hat{x}^2$$

set derivative wrt $\hat{x}$ to 0 and solve: $2E(X \mid B) = 2\hat{x}$

$\Rightarrow$ if B is nothing, $\quad \hat{x} = E(X)$

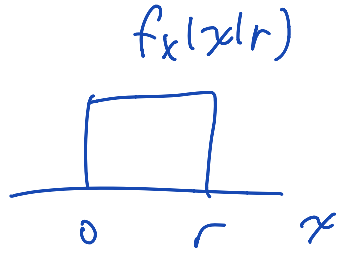
if B is $X \in A$ $\quad \hat{x} = E(X \mid A)$

if B is $\{Y = y\}$ $\quad \hat{x} = E(X \mid Y = y)$

Example MMSE estimators

Suppose R is uniform on $[0,1]$,
and X is uniform on $(0,r)$.
What is MMSE estimator of X given R?

$$f_X(x \mid r) = \begin{cases} 1/r & 0 < x < r \\ 0 & \text{else} \end{cases}$$



$$E(X \mid r) = \frac{r}{2}$$

Example

Suppose R is uniform on $[0,1]$,
and X is uniform on $(0,r)$.
What is MMSE estimator of R given X?

$$f_R(r \mid x) = \frac{f_X(x \mid r)\, f_R(r)}{f_X(x)}$$

⋮ details omitted

$$f_R(r \mid x) = \begin{cases} \dfrac{1}{-r \ln x} & 0 \le x \le r \le 1 \\ 0 & \text{else} \end{cases}$$

$$E(R \mid x) = \int_x^1 -r \frac{1}{r \ln x}\, dr = \frac{x-1}{\ln x}$$

Note: this estimate may be computationally difficult to obtain, and may only have slightly lower error.
Next estimator is constrained to be a linear function.

Linear MMSE estimator

Assume $\hat{x} = ay + b$ and find the best values a^* and b^*

Let X and Y have means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation coefficient ρ_{XY}.

The optimal LMMSE of X is

$$\hat{X}_L = a^* Y + b^*$$

where $a^* = \rho_{XY} \frac{\sigma_X}{\sigma_Y}$ $\quad b^* = \mu_X - a^* \mu_Y$

The minimum mean squared error is

$$\sigma_X^2 (1 - \rho_{XY}^2)$$

The estimation error is uncorrelated with Y.
(the orthogonality principle of the LMMSE.)

Proofs. of optimal coefficients for Linear MMSE

$$E\left[(X-\hat{X}_L)^2\right] = E\left[(X-aY-b)^2\right]$$

$$= E(X^2) - 2E\left[X(aY+b)\right] + E\left[(aY+b)^2\right]$$

$$= E(X^2) - 2aE(XY) - 2bE(X) + a^2E(Y^2) + 2abE(Y) + b^2$$

find best a, b by setting partial derivatives to 0.

$$0 = \frac{\partial}{\partial a} = -2E(XY) + 2aE(Y^2) + 2bE(Y)$$

$$0 = \frac{\partial}{\partial b} = -2E(X) + 2aE(Y) + 2b$$

solving,

$$a^* = \frac{Cov(X,Y)}{Var(Y)} = \rho_{XY}\frac{\sigma_X}{\sigma_Y}$$

$$b^* = E(X) - a^*E(Y)$$

Proofs. $E\left[(X-\hat{X}_2)^2\right]$

$$= E\left[(X-aY-b)^2\right]$$

$$= E(X^2) - 2E\left[X(aY+b)\right] + E\left[(aY+b)^2\right]$$

$$= E(X^2) - 2aE(XY) - 2bE(X) + a^2E(Y^2) + 2abE(Y) + b^2$$

find best a, b by setting partial derivatives ~~to~~ 0.

$$0 = \frac{\partial}{\partial a} = -2E(XY) + 2aE(Y^2) + 2bE(Y)$$

$$0 = \frac{\partial}{\partial b} = -2E(X) + 2aE(Y) + 2b$$

solving,

$$a^* = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} = \rho_{XY}\frac{\sigma_X}{\sigma_Y}$$

$$b^* = E(X) - a^*E(Y)$$

MAP estimator and ML estimator

previous methods minimized the mean squared error. Maximum A posteriori estimators and Maximum Likelihood estimators do not. But they may be easier to implement, and often perform quite well.

MAP: maximize the a posteriori probability

$$\hat{x}_{MAP} = \arg\max_{x} f_X(x/y)$$

(the x that maximizes $f_X(x/y)$.

(Note: this is equivalent to finding the x that maximizes $f_{XY}(x,y)$, because

$$f_{XY}(x,y) = f_X(x/y)\, f_Y(y)$$

and the $f_Y(y)$ does not depend on x.)

So

$$\hat{x}_{MAP} = \arg\max_{x} f_Y(y/x)\, f_X(x)$$

which means we need prior probability $f_X(x)$.

Maximum Likelihood estimator

$$\hat{x}_{ML} = \arg\max_{x} f_{Y}(y|x)$$

This does <u>NOT</u> need prior probability, $f_X(x)$

$\Rightarrow$ ML rule is same as MAP rule if all values of X are equally likely

Example comparing and contrasting estimators

Bernoulli RV, where parameter p is unknown.

Perform n independent experiments and use the observed # successes X to estimate the unknown $p = Y$

X is binomial w/ parameters (n, p)

Y is beta RV

$f_Y(y)$ represents our prior knowledge of the value of $p = y$

$$f_Y(y) = \begin{cases} 6y(1-y) & 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

symmetric function of y

a) blind estimate $E(Y) = 1/2$

b) ML estimate:

$$P_X(x \mid y) = \binom{n}{x} y^x (1-y)^{n-x}$$

to find the y that maximizes this differentiate wrt y and set to zero

$$\frac{d}{dy} P_X(x|y) = \binom{n}{x}\Big[x\, y^{x-1}(1-y)^{n-x} + y^{x}(-1)(n-x)(1-y)^{n-x-1}\Big]$$

$$= \binom{n}{x} y^{x-1}(1-y)^{n-x-1}\big[x(1-y) - y(n-x)\big]$$

$$= 0 \quad \Rightarrow \quad y=0,\ y=1,\ \boxed{y = \frac{x}{n}}$$ ML estimator

c) MAP estimator

find y to maximize $\dfrac{P_X(x|y)\, f_Y(y)}{P_X(x)}$

Can ignore denominator, since its constant in y.

So maximize $6\binom{n}{x} y^{x+1}(1-y)^{n-x+1}$

$$\frac{d}{dy} = 0 = 6\binom{n}{x} y^{x}(1-y)^{n-x}\big[(x+1)(1-y) - y(n-x+1)\big]$$

solving for y, $y=0$, $y=1$,

or $\boxed{y = \dfrac{x+1}{n+2}}$ MAP estimator

d) mmSE estimator

$$\hat{y} = E(Y \mid X)$$

(a bit trickier. details omitted)

$$= \frac{x+2}{n+4}$$

e) Linear mmSE estimator

$$\hat{y} = ax + b$$

can show best (a, b) lead to same answer as above (details omitted)

$$= \frac{x+2}{n+4}$$

Comparison :

Blind : $\hat{y} = \frac{1}{2}$ regardless of any observations

ML : $\hat{y} = \frac{x}{n}$ ← the fraction of observed "success" regardless of prior knowledge

MAP : $\hat{y} = \frac{x+1}{n+2}$

This is actually a famous result

→ Laplace ?

if the sun has risen the last n days, what is the probability it will rise again tomorrow ?

$x = n \Rightarrow \boxed{\hat{y} = \frac{n+1}{n+2}}$

More practically, you can see the balance here between the prior knowledge and the experimental evidence.

If we do no experiments, $n = 0$,

and $\hat{y}_{MAP} = \frac{1}{2} = \hat{y}_{blind}$

(Estimate based on prior info only)

If we do many experiments $n \to \infty$

and $\hat{y}_{MAP} \longrightarrow \hat{y}_{ML}$

(negligible contribution of prior information)