
Model: A *known* model of system or process with *unknown* parameter a.
Data: An observation R of a *random* variable r whose pdf depends on a.
Model$\rightarrow f_{r|a}(R|A)$: If knew $a=A$, would know pdf of observation r.
Goal: Estimate a from R and conditional pdf $f_{r|a}(R|A)$: Compute $\hat{a}(R)$.
Example: Flip coin 10 times. **Data:** #heads in 10 independent flips.
Model: Binomial pmf for r. **Unknown parameter:** a=Pr[heads].

---

1. **Non-Bayesian:** a is an unknown *constant* (do not know $f_a(A)$).
Given: $f_{r|a}(R|A)$ from model; observation (data) R of rv r; nothing more.
Advantage: Need very little; no (possibly wrong) prior information.
Soln: *Maximum Likelihood Estimator:* max likelihood of what happened:r=R.
MLE: $\hat{a}_{MLE}(R) = \underset{A}{argmax}[f_{r|a}(R|A)]$. **Compute:** $\frac{\partial}{\partial A}[\log f_{r|a}(R|A)] = 0$.

BLUE: *Best (minimum variance) Linear Unbiased Estimator* of *constant* x from $y = Hx + v, E[v] = 0$ is $\hat{x}(Y) = (H'H)^{-1}H'Y$. **Proof:** p. 290.

---

2. **Bayesian:** a is *itself* random with known *a priori* pdf $f_a(A)$.
Given: $f_{r|a}(R|A)$ from model; $f_a(A)$=*a priori* info; observation R of r.
Advantage: Incorporate *a priori* in estimate, but this better be right!
Soln: min $E[c(e)]$ where $e = a - \hat{a}(r)$=error and $c(\cdot)$=cost=MEP or LSE:

2a. MEP: *Min Error Prob:* $c(e) = \begin{cases} 0 & \text{if } |e| < \epsilon; \\ 1 & \text{if } |e| > \epsilon. \end{cases}$ "close only counts in horseshoes" "a miss is as good as a mile"

$E[c(e)] = 1 - \int_{-\infty}^{\infty} dR \int_{\hat{a}(R)-\epsilon}^{\hat{a}(R)+\epsilon} dA\, f_{r,a}(R,A) = 1 - 2\epsilon \int_{-\infty}^{\infty} f_{r,a}(R, \hat{a}(R))dR$.
This is minimized when $f_{r,a}(R, \hat{a}(R))$ maximized for each R.

MAP: *Max A Posteriori:* $\hat{a}_{MAP}(R) = \underset{A}{argmax}[f_{r|a}(R|A)f_a(A)]$ (compare MLE).
Compute: $\frac{\partial}{\partial A}[\log f_{r|a}(R|A) + \log f_a(A)] = 0$. MEP criterion$\rightarrow$MAP solution.

2b. LSE: *Least Squares Estimation* criterion: $c(e) = e^2$. Penalize big errors.
LSE: $\hat{a}_{LS}(R) = E[a|r=R] = \frac{\int A f_{r|a}(R|A) f_a(A)dA}{\int f_{r|a}(R|A')f_a(A')dA'}$ Denominator just $f_r(R)$: no effect on argmax of A
Proof: Page 298. Moment of inertia minimized around center of mass.

Bias: Let a be an unknown constant A_{act} so that $f_a(A) = \delta(A - A_{act})$.
DEF: $\hat{a}(R)$ is *unbiased* if $E[\hat{a}(r)] = A_{act} \leftrightarrow E[e] = 0$. How to compute:
$E[\hat{a}(r)] = \int\int \hat{a}(R) f_{r|a}(R|A)\delta(A - A_{act})dR\,dA = \int \hat{a}(R) f_{r|a}(R|A_{act})dR$.
MSE: $\hat{a}(R)$ unbiased$\rightarrow E[(\hat{a}(r) - A_{act})^2] = \sigma^2_{\hat{a}(r)} \rightarrow$MSE =variance of $\hat{a}(R)$.

Given: Flip coin with Pr[heads]=a. **Data:** #heads in 10 independent flips.
Model: pmf $p_{r|a}(R|A) = \binom{10}{R} A^R(1-A)^{10-R}, R = 0, 1 \ldots 10; 0 \leq A \leq 1$.
Goal: Estimate a=Pr[heads] from r=#heads in 10 flips and *a priori* $f_a(A)$.

MLE: $\frac{\partial}{\partial A}[\log \binom{10}{R} + R \log A + (10-R)\log(1-A)] = \frac{R}{A} - \frac{10-R}{1-A} = 0$
$\rightarrow \hat{a}_{MLE}(R) = \frac{R}{10}$. Easy to interpret! **Note:** No *a priori* pdf for a.

Bias: $E[\hat{a}_{MLE}(r)] = E[\frac{r}{10}] = \frac{10A_{act}}{10} = A_{act} \rightarrow \hat{a}_{MLE}(r)$ unbiased.
MSE: $E[(\hat{a}_{MLE}(r) - A_{act})^2] = \sigma^2_{\frac{r}{10}}$ (since unbiased) $= \frac{10A_{act}(1-A_{act})}{100}$.

EX2: Now suppose have $f_a(A) = 1$ for $0 \leq A \leq 1$ (Bayesian problem).
MAP: $\log f_a(A) = 0 \rightarrow$same algebra$\rightarrow \hat{a}_{MAP}(R) = \hat{a}_{MLE}(R) = \frac{R}{10}$.
Have: Uniform *a priori* pdf $a \sim N(0, \sigma^2 \rightarrow \infty) \rightarrow \hat{a}_{MAP}(R) = \hat{a}_{MLE}(R)$.

EX3: Now suppose have $f_a(A) = 2A$ for $0 \leq A \leq 1$ (Bayesian problem).
MAP: $\frac{\partial}{\partial A}[\log \binom{10}{R} + R \log A + (10-R)\log(1-A) + \log 2 + \log A]$
$= \frac{R}{A} - \frac{10-R}{1-A} + \frac{1}{A} = 0 \rightarrow \hat{a}_{MAP}(R) = \frac{R+1}{11}$. A slanted estimator!

EX4: Now suppose have $f_a(A) = 1$ for $0 \leq A \leq 1$ (Bayesian problem).
LSE: $\hat{a}_{LS}(R) = E[a|r = R] = \frac{\int_0^1 A\binom{10}{R}A^R(1-A)^{10-R}dA}{\int_0^1 \binom{10}{R}A^R(1-A)^{10-R}dA} = \frac{R+1}{12}$.
Ref: Schaum's Outline *Math. Handbook*, (15.24) on p. 95. $\hat{a}_{LS}(5) = \frac{1}{2}$.
Note: Even with a uniform *a priori* distribution for a, $\hat{a}_{LS}$ still slanted!

LLSE: min $E[(a - \hat{a}(r))^2]$ such that $\hat{a}(R) = cR + b$ for some constants b, c.
Soln: $\frac{\partial}{\partial c}E[(a - cr - b)^2] = 0 \rightarrow \hat{a}_{LLSE}(R) = E[a] + \frac{\lambda_{ar}}{\sigma_r^2}(R - E[r])$.
& $\frac{\partial}{\partial b}E[(a - cr - b)^2] = 0$. This is *Linear Least Squares Estimator*.

LSE: r, a *jointly Gaussian*$\rightarrow \begin{bmatrix} r \\ a \end{bmatrix} \sim N\left(\begin{bmatrix} E[r] \\ E[a] \end{bmatrix}, \begin{bmatrix} \sigma_r^2 & \lambda_{ra} \\ \lambda_{ra} & \sigma_a^2 \end{bmatrix}\right)$

$\rightarrow \hat{a}_{LS}(R) = E[a|r = R] = E[a] + \frac{\lambda_{ar}}{\sigma_r^2}(R - E[r]) = \hat{a}_{LLSE}(R)$!
Fact: Two very *different* problems have the *same* solution!

Norm: *Normalized form:* $(\hat{a}(R) - E[a])/\sigma_a = \rho_{ar}(R - E[r])/\sigma_r$.
MSE: $E[(a - \hat{a}(r))^2] = \sigma_a^2 - \frac{\lambda_{ar}^2}{\sigma_r^2} \rightarrow E\left[\left(\frac{\hat{a}(r)-a}{\sigma_a}\right)^2\right] = 1 - \rho_{ar}^2$.