GOAL: Estimate a from observation R of random variable r.

NEED: Conditional pdf $p_{r|a}(R|A)$ and *a priori* pdf $p_a(A)$ for a.
If we knew a, we would know pdf for random variable r.

WANT: min $E[c(e)]$ where random variable $e = a - \hat{a}(r)$=error.

COST: Different $c(\cdot)$ →different estimators:

MEP: $c(e) = \begin{cases} 0 & \text{if } |e| < \epsilon; \\ 1 & \text{if } |e| > \epsilon. \end{cases}$ compare to detection criterion "a miss is as good as a mile"

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

MAP: $\hat{a}_{MAP}(R) = \underset{A}{ARGMAX}\,[p_{r|a}(R|A)p_a(A)] \quad (p_{r,a} = p_{r|a}p_a)$.
Often use: $\frac{\partial}{\partial A}[\log p_{r|a}(R|A) + \log p_a(A)] = 0$.
As in detection, MEP criterion→MAP solution.

LSE: $c(e) = e^2$; Least-Squares Estimation criterion.

LSE: $\hat{a}_{LS}(R) = E[a|r = R] = \frac{\int A p_{r|a}(R|A)p_a(A)dA}{\int p_{r|a}(R|A')p_a(A')dA'}$ denominator is just $p_r(R)$

PROOF: Stark and Woods, page 298.
↔the moment of inertia of a body is minimized around its center of mass (parallel-axis theorem of mechanics).

What if we don't have *a priori* $p_a(A)$? (non-Bayesian)
Use Maximum-Likelihood Estimator (MLE):

MLE: $\hat{a}_{MLE}(R) = \underset{A}{ARGMAX}\,[p_{r|a}(R|A)]$.
Often use: $\frac{\partial}{\partial A}[\log p_{r|a}(R|A)] = 0$.
Maximizes likelihood of what actually happened (r=R).

BIAS: Let a be a *parameter*: $p_a(A) = \delta(A - A_{act})$.
$\hat{a}(R)$ is *unbiased* if $E[\hat{a}(r)] = A_{act} \leftrightarrow E[e] = 0$.

$E[\hat{a}(r)] = \int\int \hat{a}(R)p_{r|a}(R|A)\delta(A - A_{act})\,dR\,dA = \int \hat{a}(R)p_{r|a}(R|A_{act})\,dR$

MSE: $\hat{a}(R)$ unbiased→ $E[(\hat{a}(r) - A_{act})^2] = \sigma^2_{\hat{a}(r)}$: MSE=variance.

EXAMPLE: Flip a coin with $Pr[heads] = a$ 100 times.
OBSERVE: r=#heads in 100 independent flips of the coin.
ESTIMATE: $a = Pr[heads]$ from observation R of RV r.
MODEL: pmf $p_{r|a}(R|A) = \binom{100}{R} A^R (1-A)^{100-R}$,
$R = 0, 1, \ldots 100, \quad 0 \leq A \leq 1.$

MLE: $\frac{\partial}{\partial A}[\log \binom{100}{R} + R \log A + (100-R)\log(1-A)]$
$= \frac{R}{A} - \frac{100-R}{1-A} = 0 \rightarrow \hat{a}_{MLE}(R) = \frac{R}{100}.$

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

MEP: *a priori* distribution: $p_a(A) = 1$ for $0 \leq A \leq 1$.
MAP: Clearly $\hat{a}_{MAP}(R) = \hat{a}_{MLE}(R) = \frac{R}{100}$.
Uniform *a priori* distribution of $a \rightarrow \hat{a}_{MAP}(R) = \hat{a}_{MLE}(R)$.
Uniform pdf: $a \sim N(0, \sigma^2)$ with $\sigma^2 \rightarrow \infty$.

MEP: *a priori* distribution: $p_a(A) = 2A$ for $0 \leq A \leq 1$.
MAP: $\frac{\partial}{\partial A}[\log \binom{100}{R} + R\log A + (100-R)\log(1-A) + \log 2 + \log A]$
$= \frac{R}{A} - \frac{100-R}{1-A} + \frac{1}{A} = 0 \rightarrow \hat{a}_{MAP}(R) = \frac{R+1}{101}.$
Nonuniform *a priori* pdf has slanted MAP estimator!

LSE: *a priori* distribution: $p_a(A) = 1$ for $0 \leq A \leq 1$.
LSE: $\hat{a}_{LS}(R) = E[a|r=R] = \frac{\int_0^1 A\binom{100}{R}A^R(1-A)^{100-R}dA}{\int_0^1 \binom{100}{R}A^R(1-A)^{100-R}dA} = \frac{R+1}{102}.$
(Schaum's Outline *Math. Handbook*, (15.24) on p. 95)
Note even with a uniform *a priori* distribution for a,
least-squares estimator *still* slanted! ($\hat{a}_{LS}(50) = \frac{51}{102} = \frac{1}{2}$)

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

$\rightarrow \hat{a}_{LS}(R) = \hat{a}_{LLSE}(R) = E[a] + \frac{\lambda_{ar}}{\sigma_r^2}(R - E[r])$ (Linear LSE).