MODEL QUALITY EVALUATION

ADVANCED TOPICS

- **Decay rate of the C-R bound: An example**
- **Barankin Bound and SNR-Threshold Effect**
- **Superefficiency**
- **Nuisance Parameters**

Decay rate of the C-R bound: An example

In general, it holds that $\mathrm{cov}\{\hat{\theta}_N\} = O(N^{-1})$, but not always...
Consider a signal of the type

$$y_t = \sum_{k=1}^{M} A_k \cos(\omega_k t - \phi_k) + w_t$$

where $\{w_t\}$ is zero mean Gaussian white noise of variance σ^2, and the parameters to be determined are $\{\sigma^2, A_1, \ldots, A_M, \phi_1, \ldots, \phi_M, \omega_1, \ldots, \omega_M\}$. It can be shown that

$$\boxed{\mathrm{var}\{\hat{\omega}_k\} \geq \frac{24\sigma^2}{N^3 A_k^2}\left[1 + \frac{3\cos(2\phi_k)\sin(N\omega_k)}{N\sin(\omega_k)} + o(N^{-1})\right]}$$

In this case, the C-R bound of $\hat{\omega}_k$ decays like $1/N^3$!!

BARANKIN BOUND AND SNR-THRESHOLD EFFECT

- Under regularity conditions, ML estimators achieve the C-R bound for large N and SNR (signal-to-noise ratio)
- However, when the SNR goes below a *threshold*, the performance of ML greatly deteriorates, both in bias and variance
- This is caused by *outliers*, i.e., by large errors in the likelihood function
- The C-R bound does not exhibit this phenomenon, and below the SNR threshold, the mean square error of ML does not match this bound
- One possibility to study this threshold is to consider a better bound than the C-R
- The greatest lower bound on the MSE of unbiased estimators is due to Barankin

BARANKIN BOUND AND SNR-THRESHOLD EFFECT (CONT.)

Barankin Bound:

$$\operatorname{var}\{\hat{\theta}(Y)\} \geq \sup_{P_{H_1}\{h\}, P_{H_2}\{h\}} \frac{[E_{H_1}\{H\} - E_{H_1}\{H\}]^2}{\int \left[\frac{\left(\int [P_{H_1}\{h\} - P_{H_2}\{h\}] P_Y\{y;\theta+h\}\, dh \right)^2}{P_Y\{y;\theta\}} \right] dy}$$

This bound is the best possible for unbiased estimators, since it is possible to construct an estimator that achieves it, and it doesn't impose any regularity conditions. However, the "optimal" estimator sometimes depends on θ_0, so it is only "locally" optimal

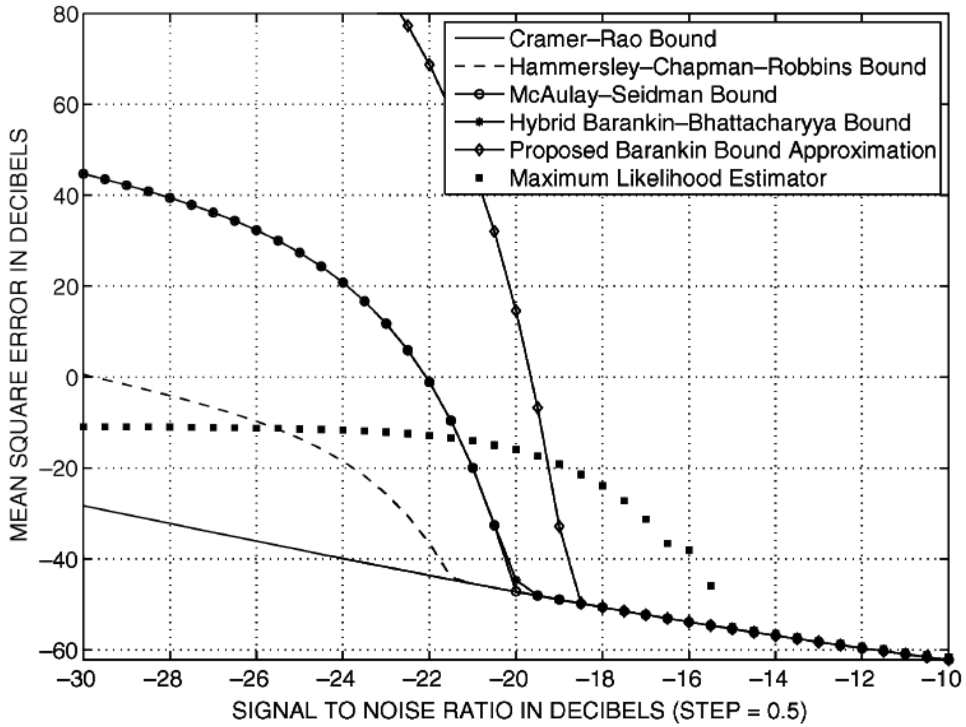
Unfortunately, it is very difficult to compute the Barankin bound, so there is a vast literature on this problem

BARANKIN BOUND AND SNR-THRESHOLD EFFECT (CONT.)

Example: *Single-Tone Estimation from Discrete-Time Observations*

$$z = a[1 \quad e^{j2\pi\theta} \quad \cdots \quad e^{j31\cdot 2\pi\theta}]^T + n \in \mathbb{C}^{32}, \qquad \theta \in (-0.5, 0.5)$$

where $n \sim \mathcal{CN}(0, I)$ and $a \sim \mathcal{CN}(0, \sigma_a^2)$. We take 100 measurements of z



SUPEREFFICIENCY

We have seen that there are biased estimators with $MSE < C\text{-}R$. Since for consistent estimators, typically $\|\text{bias}\|^2 = O(1/N^2) < O(1/N) = \text{var}$, we would expect that as $N \to \infty$,

$$\lim_{N\to\infty} MSE(\hat{\theta}_N) \geq C\text{-}R(\theta)$$

However, this is not correct...

Example: Let $\hat{\theta}_N$ be an estimator of θ such that $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, \Sigma_{C\text{-}R})$. Then, take a fixed θ_g, and define

$$\hat{\theta}'_N = \begin{cases} \hat{\theta}_N, & \text{if } \left\|\hat{\theta}_N - \theta_g\right\| > N^{-1/4} \\ \theta_g, & \text{if } \left\|\hat{\theta}_N - \theta_g\right\| \leq N^{-1/4} \end{cases}$$

If $\theta \neq \theta_g$, then $\sqrt{N}(\hat{\theta}'_N - \theta) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, \Sigma_{C\text{-}R})$, but if $\theta = \theta_g$, then $N \operatorname{cov} \hat{\theta}'_N \to 0$!!

SUPEREFFICIENCY(CONT.)

Interestingly, Le Cam (1952) showed that the set of θ's for which the C-R bound is asymptotically violated has at most measure zero

NUISANCE PARAMETERS

Let a statistical model $P\{y;\theta\}$ be parameterized by $\theta = (\psi, \lambda)$, where we are actually interested only on ψ (*parameter of interest*). λ is then called a *nuisance parameter*
By the parsimony principle, the covariance of ψ increases because of λ

Orthogonal parameters:
ψ and λ are *orthogonal* if the information matrix of θ has the form

$$M_\theta = \begin{bmatrix} M_\psi & 0 \\ 0 & M_\lambda \end{bmatrix}$$

Properties:
1. The ML estimators $\hat{\psi}$ and $\hat{\lambda}$ are asymptotically independent
2. The asymptotic covariance of $\hat{\psi}$ is the same whether λ is known or estimated by $\hat{\lambda}$
3. $\hat{\psi}_\lambda = \hat{\psi}(\lambda)$, the ML estimator of ψ when λ is given, varies only slowly with λ

Therefore, the ideal situation is when ψ and λ are “orthogonalizable”

NUISANCE PARAMETERS (CONT.)

Alternatives:

1. *Profile likelihood:* Work with $L_P(\psi) = \max_\lambda P\{y;\psi,\lambda\}$

2. *Marginal likelihood:* If there is a statistic T whose distribution only depends on ψ, work with $L_M(\psi) = P_T\{T;\psi\}$

3. *Conditional likelihood:* If there is a statistic S such that $P\{y \mid S;\psi,\lambda\}$ only depends on ψ, work with $L_C(\psi) = P\{y \mid S;\psi\}$

4. *Integrated likelihood:* Work with $L_I(\psi;\pi) = \int P\{y;\psi,\lambda\}\pi(\lambda)d\lambda$

5. *Bayesian approach:* From $P\{\psi,\lambda \mid y\}$, directly obtain $P\{\psi \mid y\} = \int P\{\psi,\lambda \mid y\}d\lambda$

A final alternative, currently under intense research, is the *modified profile likelihood*