Computing Bayesian Cramer-Rao Bounds

Chandra R. Murthy

26 March 2011

Outline

- Part 1: Cramer-Rao Bounds
- Part 2: Factor Graphs
- Part 3: Computing CRBs from Factor Graphs

PART 1: CRAMER RAO BOUNDS

- Recap of Cramer-Rao Bounds
- Bayesian Cramer-Rao Bounds

Cramer-Rao Bound

- Need for bounds:
  - Many practical estimation problems: computing *optimal* (e.g., MMSE, MAP, ML) estimators is infeasible
  - Typically, we use *suboptimal* techniques
    - EM
    - Belief propagation
  - So, want to know how good these estimators are!
- So, alternative strategy:
  - Find lower bound on MSE among all (unbiased) estimators
  - Check how close we can get to the lower bound

Schur Complement

- Consider $$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$
- Can diagonalize:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta}_{11} \end{bmatrix}$$

- Where Δ_{11} is the Schur complement of A_{11}

$$\mathbf{\Delta}_{11} \triangleq \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

Decorrelation

- Let the covariance of $[X_1, X_2]^T$ be $\mathbf{A} \geq \mathbf{0}$
- Can decorrelate X:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1 \end{bmatrix}$$

$$\mathsf{Cov}(\mathbf{Y}) = \mathsf{diag}\{\mathbf{A}_{11}, \mathbf{\Delta}_{11}\}$$

$$\mathbf{\Delta}_{11} \triangleq \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \geq 0$$

- Equality iff $\mathbf{X}_2 = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1$ a.s.

Scalar CRLB Theorem

- Let Y ~ f(y|θ). Let $\hat{\theta}$ be an unbiased est. of θ
- Under regularity conditions

$$\mathsf{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

- I(θ) is the *Fisher Information*
- Equality iff

$$s(y;\theta) \triangleq \frac{\partial}{\partial\theta} \ln f(y|\theta) = I(\theta)(\hat{\theta}(y) - \theta)$$

Proof

- Consider $\mathbf{z} \triangleq \begin{bmatrix} \mathbf{s}(\mathbf{y};\theta) \\ \hat{\theta}(y) - \theta \end{bmatrix}$
- Note Z is zero-mean, with covariance

$$\mathsf{Cov}(\mathbf{z}) = \begin{bmatrix} I(\theta) & 1 \\ 1 & \mathsf{Var}(\hat{\theta}) \end{bmatrix}$$

- Taking the schur complement of I(θ):

$$\mathsf{Var}(\hat{\theta}) - I^{-1}(\theta) \geq 0$$

- Equality iff

$$\hat{\theta}(y) - \theta = I^{-1}(\theta) s(\mathbf{y};\theta) \quad \text{almost surely}$$

Vector Parameter CRB

Theorem Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$. Then

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathsf{T}}\} \geq \mathbf{I}^{-1}(\boldsymbol{\theta})$$

where $I(\boldsymbol{\theta})$ is the Fisher Information Matrix

$$\begin{aligned} [I(\theta)]_{ij} &= \mathbb{E}\{\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \ln f(\mathbf{Y}|\boldsymbol{\theta})}{\partial \theta_j}\} = -\mathbb{E}\{\frac{\partial^2 \ln f(\mathbf{Y}|\theta)}{\partial \theta_i \partial \theta_j}\} \\ \mathbf{I}(\boldsymbol{\theta}) &= \mathbb{E}\{[\nabla_\theta \ln f(\mathbf{Y}|\boldsymbol{\theta})][\nabla_\theta \ln f(\mathbf{Y}|\boldsymbol{\theta})]^{\mathsf{T}}\} = -\mathbb{E}\{\nabla^2 \ln f(\mathbf{Y}|\boldsymbol{\theta})\} \end{aligned}$$

The equality holds iff

$$\nabla_\theta \ln f(\mathbf{Y}|\boldsymbol{\theta}) = \mathbf{I}(\theta)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

Extension to Random Parameters

- Let

$$(\mathbf{Y}, \boldsymbol{\Theta}) \sim f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = f(\mathbf{y}, \boldsymbol{\theta})$$

- $\Theta_i \in (-\infty, \infty)$ $\qquad f(\mathbf{y}, \boldsymbol{\theta}) > 0$
- Let $\hat{\boldsymbol{\Theta}}$ be a Bayesian estimator of $\boldsymbol{\Theta}$

regularity conditions

1. $f(\mathbf{y}, \boldsymbol{\theta})$ is absolutely continuous with respect to $\boldsymbol{\theta}$;
2. $\lim_{\theta_i \to \pm\infty} \theta_i f(\mathbf{y}, \theta_i) = 0 \quad \forall i$, or

2' the conditional bias satisfies

$$\lim_{\theta_i \to \pm\infty} \mathbb{E}(\hat{\Theta}_i - \theta_i | \boldsymbol{\Theta} = \boldsymbol{\theta}) f(\theta_i) = 0, \quad \forall i$$

Random Parameter CRLB

- The BCRLB is given by

$$\mathcal{M}(\hat{\mathbf{\Theta}}) \stackrel{\Delta}{=} \mathbb{E}(\hat{\mathbf{\Theta}} - \mathbf{\Theta})(\hat{\mathbf{\Theta}} - \mathbf{\Theta})^T \geq \mathbf{J}^{-1}$$

- Where

$$\mathbf{J} = \mathbb{E}\{[\nabla_\theta \ln f(\mathbf{Y}, \mathbf{\Theta})][\nabla_\theta \ln f(\mathbf{Y}, \mathbf{\Theta})]^T\}$$

- (assuming expectations are finite and inverses exist)

BCRB: Some Remarks

- Also holds for biased estimators
- Need the "weak unbiasedness" condition:

$$\int_x \nabla_{x_j} [p(x) B(x)] = 0$$

- Where:

$$B(x) \triangleq \int_y [\hat{x}(y) - x] p(y|x) dy$$

BCRB: More Remarks

- In practice, more interested in BCRB of a particular component X_k:

$$\begin{aligned}
&\mathrm{E}_{X_k Y}[(\hat{x}_k(Y) - X_k)(\hat{x}_k(Y) - X_k)^T] \\
&\qquad = \mathrm{E}_{XY}[(\hat{x}_k(Y) - X_k)(\hat{x}_k(Y) - X_k)^T] \\
&\qquad \triangleq \mathbf{E}_{kk},
\end{aligned}$$

- The CRB is given by

$$\mathbf{E}_{kk} \succeq [\mathbf{J}^{-1}]_{kk}$$

PART 2: FACTOR GRAPHS

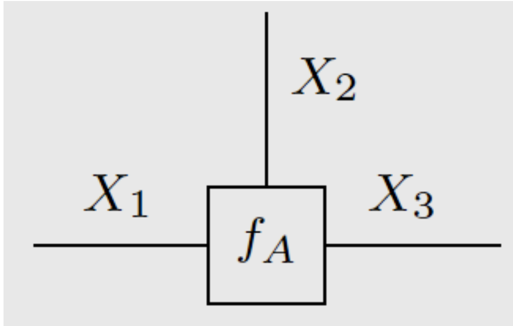
A brief introduction

What is a Factor Graph?

- It is a graphical model of a function
  - Next question: *what is a graphical model?*
- Graphical model: helps to visualize interactions between variables
- Examples:
  - Error control codes
  - Communication channel representations
- Types of graphical models *other* than factor graphs:
  - Markov random fields, neural networks, etc

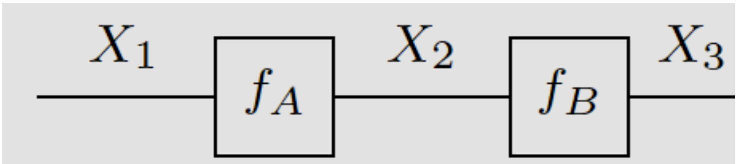
Basic Setup

- Functions = nodes, Edges = variables
- No structure: $f_A(X_1, X_2, X_3)$



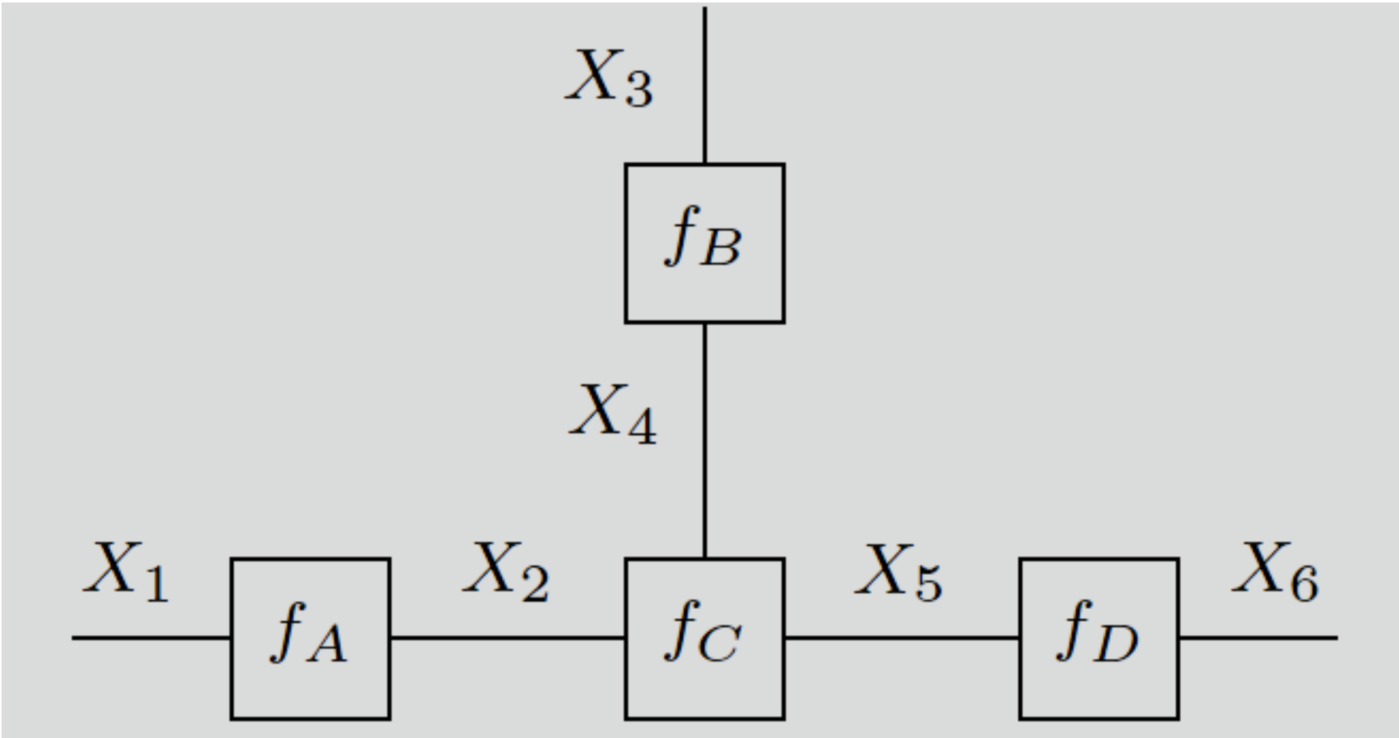
- Function with structure:

$$f(x_1, x_2, x_3) \triangleq f(x_1, x_2) f(x_2, x_3)$$



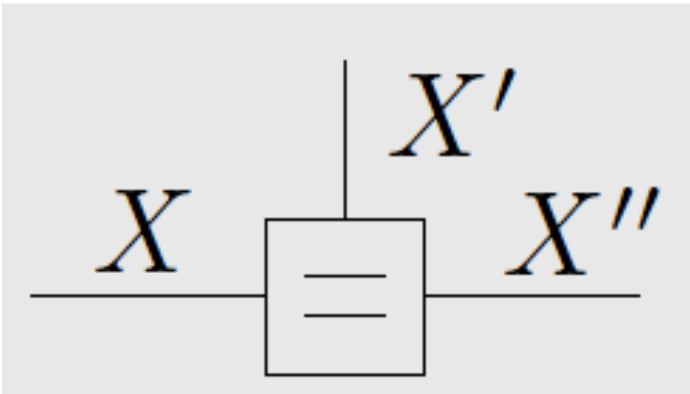
One More Example

$$f(x_1, x_2, x_3, x_4, x_5, x_6) \triangleq f_A(x_1, x_2) f_B(x_3, x_4) f_C(x_2, x_4, x_5) f_D(x_5, x_6)$$



Equality Constraint

- Sometimes, the same variable is input to multiple function
- But if we represent a variable by an edge, it can only be input to two functions at max!
- So, use *equality constraint nodes*



Summary Propagation Algo

- Suppose we want to compute the marginalization

$$f(x_5) \triangleq \sum_{x_1,x_2,x_3,x_4,x_6} f(x_1, x_2, x_3, x_4, x_5, x_6)$$

- Suppose, in particular, we are given that

$$f(x_5) = \sum_{x_1,x_2,x_3,x_4,x_6} f_A(x_1, x_2) \cdot f_B(x_3, x_4) \cdot f_C(x_2, x_4, x_5) \cdot f_D(x_5, x_6)$$

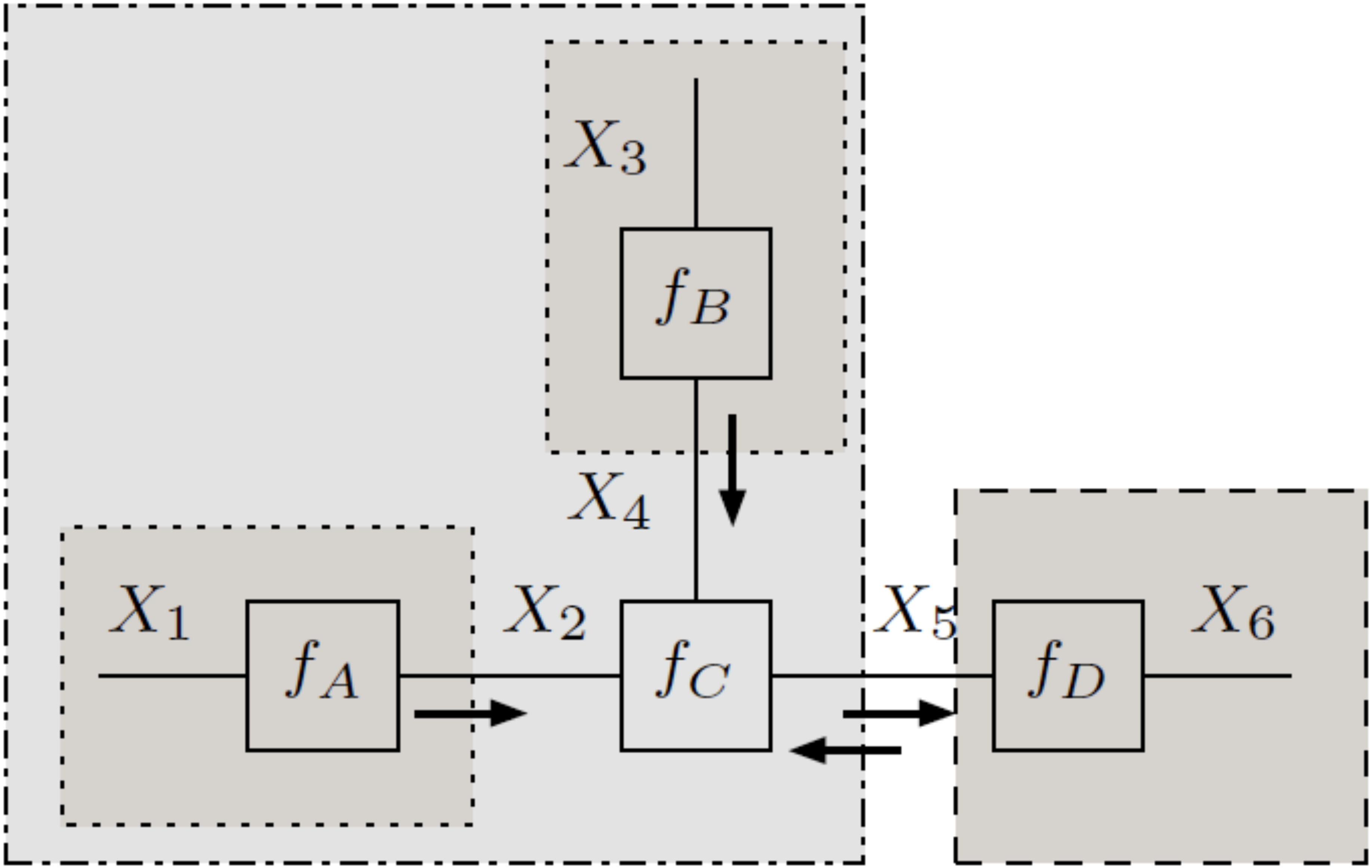
Key Step: Use "Brackets"

$$= \sum_{x_2,x_4} f_C(x_2,x_4,x_5) \underbrace{\left(\sum_{x_1} f_A(x_1,x_2) \right) \cdot \left(\sum_{x_3} f_B(x_3,x_4) \right)}_{\mu_{f_A \to x_2}(x_2) \qquad \mu_{f_B \to x_4}(x_4)} \cdot$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\mu_{f_C \to x_5}(x_5)}$$

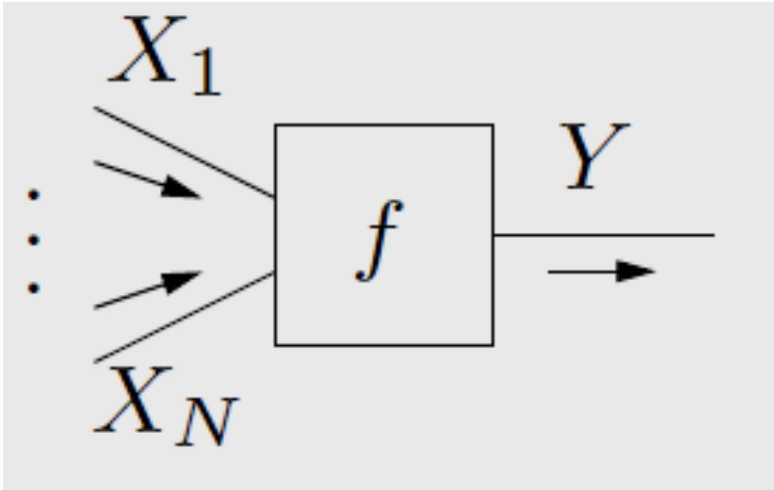
$$\underbrace{\left(\sum_{x_6} f_D(x_5,x_6) \right)}_{\mu_{f_D \to x_5}(x_5)} \cdot$$

SPA Illustrated



Sum Product Rule

- Message out of node *f* along edge *Y* is the product of the function *f* and all messages towards node *f* summed over all other edges (variables) except *Y*



- This rule is the central building block of factor graph based computations

PART 3: COMPUTING BCRBS

Switch to paper/thesis

Page numbers: 210-215 in the manuscript

Corresponds to: 238-244 of PDF file

Reference material: Justin Dauwels, “On Graphical Models for Communications and Machine Learning: Algorithms, Bounds, and Analog Implementation” Ph.D. Dissertation.

THANK YOU!