Optimum Estimation via Partition Functions and Information Measures

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

ISIT 2010, Austin TX, June 2010

Background – Relating Estimation and Info Measures

Early work:

Duncan (1968,1970); Kailath (1968,1969,1970); Kadota, Ziv & Zakai (1971); Bucy (1979); Mayer–Wolf & Zakai (1983).

More recent research activity:

Forney (2004); Guo, Shamai & Verdú (2005, 2008); Palomar & Verdú (2006, 2007); Mayer–Wolf & Zakai (2007); Guo (2009); Verdú (2009); Raginsky & Coleman (2009); Weissman (2010); Merhav, Guo & Shamai (2010).

Context of This Work

In [MGS, 2010], stat–mech methods where applied on the I–MMSE relation

$$\frac{\mathrm{d}I(\boldsymbol{X};\sqrt{snr}\boldsymbol{X}+\boldsymbol{N})}{\mathrm{d}\; snr}=\frac{1}{2}\mathrm{mmse}(\boldsymbol{X}|\sqrt{snr}\boldsymbol{X}+\boldsymbol{N}),\quad \boldsymbol{N}\sim\mathcal{N}(0,I)$$

to compute MMSE and to relate threshold effects (in estimation) to phase transitions (in physics).

Main theme of this work: For the purpose of evaluating the MMSE (using stat-mech methods), more direct relations can be used: Given $P(\boldsymbol{x},\boldsymbol{y})$, $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{y}\in\mathbb{R}^m$:

$$Z(\boldsymbol{y},\boldsymbol{\lambda})=\sum_{\boldsymbol{x}}\exp\{\boldsymbol{\lambda}^T\boldsymbol{x}\}P(\boldsymbol{x},\boldsymbol{y})$$

$$\hat{\boldsymbol{x}}=\mathbf{E}\{\boldsymbol{X}|\boldsymbol{y}\}=\nabla_{\boldsymbol{\lambda}}\ln Z(\boldsymbol{y},\boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=0};\quad \mathrm{Cov}\{(\boldsymbol{X}-\hat{\boldsymbol{X}})\}=\mathbf{E}\left\{\nabla^2_{\boldsymbol{\lambda}}\ln Z(\boldsymbol{Y},\boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=0}\right\}$$

Advantages of the Proposed Approach

- More direct and easier to use than the I–MMSE relation.
- Applies to a general $P(\boldsymbol{x}, \boldsymbol{y})$; $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$.
- In addition to the MMSE – provides also the conditional mean estimator.
- Several variants can be used.
- Easy to extend to the mismatched case.
- Allows mismatch in the full joint distribution, not just the source.

Some Comments

$$Z(\boldsymbol{y}, \boldsymbol{\lambda}) = \sum_{\boldsymbol{x}} \exp\{\boldsymbol{\lambda}^T \boldsymbol{x}\} P(\boldsymbol{x}, \boldsymbol{y})$$

$$\mathbf{E}\{\boldsymbol{X}|\boldsymbol{y}\} = \nabla_{\boldsymbol{\lambda}} \ln Z(\boldsymbol{y}, \boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=0}; \quad E \stackrel{\triangle}{=} \text{Cov}\{(\boldsymbol{X} - \hat{\boldsymbol{X}})\} = \mathbf{E}\left\{\nabla^2_{\boldsymbol{\lambda}} \ln Z(\boldsymbol{Y}, \boldsymbol{\lambda})\Big|_{\boldsymbol{\lambda}=0}\right\}$$

- $\ln\left[\sum_{\boldsymbol{x}} \exp\{\boldsymbol{\lambda}^T \boldsymbol{x}\} P(\boldsymbol{x}|\boldsymbol{y})\right]$ generates conditional cumulants.
- OK and easier to replace $P(\boldsymbol{x}|\boldsymbol{y})$ by $P(\boldsymbol{x}, \boldsymbol{y})$.
- Physics interpretation:
  - $Z(\boldsymbol{y}, \boldsymbol{\lambda}) \Longleftrightarrow$ partition function.
  - $\{\lambda_i\} \Longleftrightarrow$ ‘forces’ acting on $\{y_i\}$.
  - Error covariance relation $\Longleftrightarrow$ fluctuation–dissipation thm.

A Few Variants

$$P_{\boldsymbol{\lambda}}(\boldsymbol{y}) \triangleq \frac{\sum_{\boldsymbol{x}} \exp\{\boldsymbol{\lambda}^T \boldsymbol{x}\} P(\boldsymbol{x}, \boldsymbol{y})}{\sum_{\boldsymbol{x}} \exp\{\boldsymbol{\lambda}^T \boldsymbol{x}\} P(\boldsymbol{x})} \triangleq \frac{Z(\boldsymbol{y}, \lambda)}{\Theta(\boldsymbol{\lambda})}$$

$$J(\boldsymbol{Y}) = -\mathbf{E}\left\{ \nabla^2_{\boldsymbol{\lambda}} \ln P_{\boldsymbol{\lambda}}(\boldsymbol{Y}) \Big|_{\boldsymbol{\lambda}=0} \right\} \quad \text{(Fisher info)}$$

$$\begin{aligned}
\text{tr}\{E\} &= \sum_{i=1}^{n} \mathbf{E}\left\{ \frac{\partial^2 \ln Z(\boldsymbol{Y}, \boldsymbol{\lambda})}{\partial \lambda_i^2} \Big|_{\boldsymbol{\lambda}=0} \right\} \\
&= \sum_{i=1}^{n} \left[\text{Var}\{X_i\} + \mathbf{E}\left\{ \frac{\partial^2 \ln P_{\boldsymbol{\lambda}}(\boldsymbol{Y})}{\partial \lambda_i^2} \Big|_{\boldsymbol{\lambda}=0} \right\} \right] \\
&= \sum_{i=1}^{n} \left[\text{Var}\{X_i\} - \mathbf{E}\left\{ \left[\frac{\partial \ln P_{\boldsymbol{\lambda}}(\boldsymbol{Y})}{\partial \lambda_i} \right]^2 \Big|_{\boldsymbol{\lambda}=0} \right\} \right] \\
&= \sum_{i=1}^{n} \left[\mathbf{E}\{X_i^2\} - \mathbf{E}\left\{ \left[\frac{\partial \ln Z(\boldsymbol{Y}, \boldsymbol{\lambda})}{\partial \lambda_i} \right]^2 \Big|_{\boldsymbol{\lambda}=0} \right\} \right]
\end{aligned}$$

2nd & 3rd lines: $\ln P_{\boldsymbol{\lambda}}(\boldsymbol{Y})$ can be replaced by $i(\boldsymbol{X}; \boldsymbol{Y}) = \ln[P(\boldsymbol{y}|\boldsymbol{x})/P_{\boldsymbol{\lambda}}(\boldsymbol{y})]$.

Example 1: Codeword Sent Over an AWGN Channel

Channel input: $M = e^{nR}; \quad \mathcal{C} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{M-1}\};$
$\boldsymbol{x}_i \sim$ Surf{sphere of radius $\sqrt{nP}$}.

AWGN channel:

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}; \quad \boldsymbol{N} \sim \mathcal{N}(0, \sigma^2 I)$$

Partition function:

$$Z(\boldsymbol{y}, \boldsymbol{\lambda}) = \sum_{\boldsymbol{x} \in \mathcal{C}} e^{-nR} \cdot \exp\{-\|\boldsymbol{y} - \boldsymbol{x}\|^2/(2\sigma^2) + \boldsymbol{\lambda}^T \boldsymbol{x}\}.$$

Can be analyzed using techniques borrowed from the random energy model (REM) of spin glasses:

$$\text{mmse}(\boldsymbol{X}|\boldsymbol{Y}) = \begin{cases} 0 & R < C \\ \frac{P\sigma^2}{P+\sigma^2} & R > C \end{cases}$$

Example 2: Curie–Weiss Model

Let $\boldsymbol{x} \in \{-1, +1\}^n$ and

$$P(\boldsymbol{x}) \propto \exp\left\{n\left[\frac{a \cdot m^2(\boldsymbol{x})}{2n} + b \cdot m(\boldsymbol{x})\right]\right\}$$

where

$$m(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and let the channel be a BSC:

$$P(y|x) = \frac{e^{\beta x y}}{2\cosh(\beta)} \quad y \in \{-1, +1\}.$$

Then

$$Z(\boldsymbol{y}, \boldsymbol{\lambda}) \propto \sum_{\boldsymbol{x}} \exp\left\{\sum_i x_i(b + \lambda_i + \beta y_i)\right\} \cdot \exp\left\{\frac{a}{2n}\left(\sum_i x_i\right)^2\right\}$$

Example 2 (Cont'd)

Using the identity

$$\exp\left\{\frac{a}{2n}\left(\sum_{i=1}^{n} x_i\right)^2\right\} = \int_{-\infty}^{+\infty} \mathrm{d}t \exp\left\{-\frac{nt^2}{2a} + t\sum_{i=1}^{n} x_i\right\}$$

$$Z(\boldsymbol{y}, \boldsymbol{\lambda}) \propto \int_{-\infty}^{+\infty} \mathrm{d}t \exp\left\{-\frac{nt^2}{2a} + \sum_{i=1}^{n} \ln\cosh(\beta y_i + \lambda_i + b + t)\right\}$$

$$\hat{x}_i = \left.\frac{\partial \ln Z}{\partial \lambda_i}\right|_{\boldsymbol{\lambda}=0} = \langle \tanh(\beta y_i + b + t)\rangle$$

where the averaging is w.r.t. a weight func. proportional to

$$\exp\left\{-\frac{nt^2}{2a} + \sum_{i=1}^{n} \ln\cosh(\beta y_i + b + t)\right\}.$$

As $n \to \infty$, this is dominated by t^* that maximizes this weight.

Example 2 (Cont'd)

By taking a 2nd order derivative:

$$\lim_{n\to\infty} \frac{\text{mmse}(\boldsymbol{X}|\boldsymbol{Y})}{n} = 1 - \mathbf{E}\{\tanh^2(\beta Y + b + t_0)\},$$

where t_0 is the solution to the equation

$$t = a\mathbf{E}\{\tanh(\beta Y + b + t)\},$$

and where Y is a binary $\{\pm 1\}$ RV, with mean $m^* \tanh(\beta)$, m^* being the dominant solution to the equation

$$m = \tanh(am + b),$$

i.e., the maximizer of

$$h_2((1+m)/2) + am^2/2 + bm,$$

where $h_2(\cdot)$ is the binary entropy function.

Generalized Spherical Symmetry

Suppose $m = n$ and

$$P(\boldsymbol{x}, \boldsymbol{y}) = F_n\left(\sum_i \phi(x_i, y_i)\right)$$

and let $f_n(t) = \mathcal{L}^{-1}\{F_n(s)\}$. Then,

$$\begin{aligned} Z(\boldsymbol{y}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}^n} \mathrm{d}\boldsymbol{x} e^{\boldsymbol{\lambda}^T \boldsymbol{x}} \int_0^\infty \mathrm{d}t f_n(t) \exp\left\{-t \sum_i \phi(x_i, y_i)\right\} \\ &= \int_0^\infty \mathrm{d}t f_n(t) \int_{\mathbb{R}^n} \mathrm{d}\boldsymbol{x} e^{\boldsymbol{\lambda}^T \boldsymbol{x}} \exp\left\{-t \sum_i \phi(x_i, y_i)\right\} \\ &= \int_0^\infty \mathrm{d}t f_n(t) \prod_i \int_{\mathbb{R}} \mathrm{d}x_i e^{\lambda_i x_i} \exp\{-t\phi(x_i, y_i)\}. \end{aligned}$$

and we have a product form, which can be handled easily.

Generalized Spherical Symmetry (Cont'd)

$$\rho(\lambda, y, t) \triangleq \ln\left[\int_{-\infty}^{\infty} \mathsf{d}x e^{\lambda x - t\phi(x,y)}\right],$$

$$\rho_0(y, t) \triangleq \rho(0, y, t) = \ln\left[\int_{-\infty}^{\infty} \mathsf{d}x e^{-t\phi(x,y)}\right],$$

$$\zeta(y, t) \triangleq \left.\frac{\partial \rho(\lambda, y, t)}{\partial \lambda}\right|_{\lambda=0} = \frac{\int_{\mathbb{R}} \mathsf{d}x \cdot x e^{-t\phi(x,y)}}{\int_{\mathbb{R}} \mathsf{d}x \cdot e^{-t\phi(x,y)}}.$$

Then,

$$\mathbf{E}\{X_i|\boldsymbol{y}\} = \frac{\int_0^\infty \mathsf{d}t f_n(t)\zeta(y_i, t)e^{\sum_i \rho_0(y_i,t)}}{\int_0^\infty \mathsf{d}t f_n(t)e^{\sum_i \rho_0(y_i,t)}}$$

which is approximated by $\zeta(y_i, \hat{t})$, where $\hat{t}$ is the maximizer of the expression

$$\ln|f_n(t)| + \sum_i \rho_0(y_i, t).$$

Similar ideas are applied to the MMSE analysis.

Extensions

- The range of t may not necessarily be $[0, \infty)$.
- For

$$P(\boldsymbol{x}, \boldsymbol{y}) = F_n\left(\sum_{i=1}^{n} \phi_1(x_i, y_i), \ldots, \sum_{i=1}^{n} \phi_k(x_i, y_i)\right)$$

  apply a multidimensional Laplace transform.

- $\sum_i \phi(x_i, y_i) \Longrightarrow \sum_i \phi(x_i, y_i, y_{i-1}, \ldots, y_{i-k})$ for some k.
- Can handle $P(\boldsymbol{x}, \boldsymbol{y}) = F_n[(\boldsymbol{x}, \boldsymbol{y})^T S(\boldsymbol{x}, \boldsymbol{y})]$ for a positive matrix S.

Conclusion

- We proposed a simple “partition function” for derivations pertaining to MMSE estimation.
- There are several advantages relative to I-MMSE relations.
- The relations connect also to some info measures, e.g., information density, Fisher information.
- A few examples demonstrated.
- Defined a class of joint pmf’s for which MMSE calculations are easy.