Presentation on the term paper:
A Simple Proof of the EPI via Properties of the Mutual Information
by
Olivier Rioul

Saurabh Deshpande

20 April 2011

Outline

- What the paper is about?
- Relationship between FI and MMSE
- Dual versions of de Bruijn's Identity
- Equivalent Integral Representations of $h(X)$
- Proofs of EPI via Integral Representations of $h(X)$
- New Proof of EPI

About the paper: EPI & its Two Different Proofs

- Entropy-Power Inequality (EPI): $X \perp Y$ (r.v. s)

$$\begin{gathered} \exp(2h(X+Y)) \geq \exp(2h(X)) + \exp(2h(Y)) \\ \Updownarrow \\ h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1-\lambda)h(Y) \end{gathered} \tag{1}$$

- Proof 1 (Stam-1959, Blachman-1965): $X \perp Z \sim \mathcal{N}(0,1)$
  Based on Fisher Information (FI) and de Bruijn's identity

$$\frac{d}{dt} h(X + \sqrt{t}Z)\Big|_{t=0} = \frac{1}{2} J(X), \tag{2}$$

About the paper

- Proof 2 (Verdú-Guo-2006): $X \perp Z \sim \mathcal{N}(0,1)$
  Based on Minimum Mean-Square Error (MMSE) estimator in Gaussian Channels and the identity

$$\begin{aligned} \frac{d}{dt} I(X; X + \sqrt{t}Z) &= \frac{1}{2}\mathrm{mmse}(X|\sqrt{t}X + Z), \\ I(X;Y) &= h(Y) - h(Y|X), \\ \mathrm{mmse}(X|Y) &= \mathbb{E}[(X - \mathbb{E}[X|Y])^2]. \end{aligned} \tag{3}$$

- What does the paper do? (Note: h in (2) and I in (3) are related.)
  – Addresses the question: (So,) How are the J in (2) and mmse in (3) related?
  – Derives *equivalent integral representations* of $h(X)$.
  – Gives alternatives to the two proofs via the integral representations of $h(X)$.
  – Gives a new proof of EPI based on the data processing inequality for I.

Relationship between FI and MMSE

- Recall FI: $p(x)$: probability density function of X

$$J(X) := \mathbb{E}[\, S^2(X)\,], \quad S(X) := \frac{\dot{p}(X)}{p(X)}. \quad (\mathbb{E}[\, S(X)\,] = 0)$$

- (Eq. after (10) in) Blachman-1965:

$$S(X+Z) = \mathbb{E}[\, S(Z)|X+Z\,] \qquad \{*****\}$$

- $\implies \quad J(X+Z) = \mathrm{Var}[\, S(X+Z)\,] = \mathrm{Var}[\mathbb{E}[\, S(Z)|X+Z\,]]$

- Law of total variance:

$$\mathrm{Var}[\, U\,] = \mathrm{Var}[\, U|V\,] + \mathrm{Var}[\mathbb{E}[\, U|V\,]],$$

where $\quad \mathrm{Var}[\, U|V\,] := \mathbb{E}[\,(U - \mathbb{E}[\, U|V\,])^2\,]$.
$\mathbb{E}[\;]$ over joint distribution of U and $V \implies \mathrm{Var}[\, U|V\,]$ is a no.

- $\implies \quad \mathrm{Var}[\mathbb{E}[\, S(Z)|X+Z\,]] = \mathrm{Var}[\, S(Z)\,] - \mathrm{Var}[\, S(Z)|X+Z\,]$

FI and MMSE

- Identify $\quad \mathrm{Var}[\,S(Z)|X+Z\,] = \mathrm{mmse}(S(Z)|X+Z) \quad \{*****\}$
- $\Longrightarrow$
$$\begin{aligned} J(X+Z) &= \mathrm{Var}[\,S(Z)\,] - \mathrm{mmse}(S(Z)|X+Z) \\ &= J(Z) - \mathrm{mmse}(S(Z)|X+Z) \end{aligned}$$
- $Z \sim \mathcal{N}(0,\sigma^2) \qquad \Longrightarrow \qquad S(Z) = -\frac{Z}{\mathrm{Var}[\,Z\,]}, \quad J(Z) = \frac{1}{\mathrm{Var}[\,Z\,]}$
- $\Longrightarrow$
$$\mathrm{Var}[\,Z\,]J(X+Z) + J(Z)\mathrm{mmse}(X|X+Z) = 1, \quad Z \sim \mathcal{N}(0,\sigma^2). \tag{4}$$
- $\Longrightarrow$
$$\boxed{J(X+Z) + \mathrm{mmse}(X|X+Z) = 1, \quad Z \sim \mathcal{N}(0,1).} \tag{5}$$

First Dual version of De Bruijn's Identity

- Begin from de Bruijn's Identity (2)

$$
\begin{aligned}
\frac{d}{dt}h(Y+\sqrt{t}Z)\Big|_{t=0} &= \frac{1}{2}J(Y), \quad Y \perp Z \sim \mathcal{N}(0,1) \\
&\Downarrow\ Y = X+\sqrt{t'}Z', \quad X \perp Z' \perp Z, \\
&\Downarrow\quad Z' \sim \mathcal{N}(0,1), \quad t' \neq 0 \\
\frac{d}{dt}h(X+\sqrt{t'}Z'+\sqrt{t}Z)\Big|_{t=0} &= \frac{1}{2}J(X+\sqrt{t'}Z') \\
&\Downarrow\ \sqrt{t'}Z'+\sqrt{t}Z \sim \sqrt{t'+t}Z' \\
\frac{d}{dt}h(X+\sqrt{t'+t}Z')\Big|_{t=0} &= \frac{1}{2}J(X+\sqrt{t'}Z') \\
&\Downarrow\ t'+t = u
\end{aligned}
$$

$$
\boxed{\frac{d}{du}h(X+\sqrt{u}Z')\Big|_{u=t'} = \frac{1}{2}J(X+\sqrt{u}Z')\Big|_{u=t'}, \quad \forall\, t' > 0.} \tag{6}
$$

Second Dual version of De Bruijn's Identity

- $Z \sim \mathcal{N}(0,1)$

$$
\begin{aligned}
\frac{d}{dt} h(Z + \sqrt{t}X) &= \frac{d}{dt}\left\{ h\left(\frac{Z}{\sqrt{t}} + X\right) + \frac{1}{2}\log t \right\} \\
&\Downarrow u = 1/t \\
&= -\frac{1}{t^2}\frac{d}{du} h(X + \sqrt{u}Z) + \frac{1}{2t} \\
&\Downarrow (6) \\
&= -\frac{1}{2t^2} J(X + \sqrt{u}Z) + \frac{1}{2t} \\
&\Downarrow J(\sqrt{t}X) = J(X)/t \\
&= -\frac{1}{2t} J(\sqrt{t}X + Z) + \frac{1}{2t} \qquad (7)
\end{aligned}
$$

Second Dual version ...

- Recall relation between FI & MMSE (5) (+ replace X by $\sqrt{t}X$):

$$J(\sqrt{t}X + Z) + t\,\mathrm{mmse}(X|\sqrt{t}X + Z) = 1, \quad Z \sim \mathcal{N}(0,1). \tag{8}$$

- Use $J(\sqrt{t}X + Z)$ from (8) in earlier (7):

$$\frac{d}{dt}h(Z + \sqrt{t}X) = -\frac{1}{2t}\left\{1 - t\,\mathrm{mmse}(X|\sqrt{t}X + Z)\right\} + \frac{1}{2t}$$

$$\implies \quad \boxed{\frac{d}{du}h(\sqrt{t}X + Z) = \frac{1}{2}\mathrm{mmse}(X|\sqrt{t}X + Z).} \tag{9}$$

First Integral Representation of $h(X)$

- Define $D_h(X) := h(X_G) - h(X), \quad X_G \sim \mathcal{N}(\mathbb{E}[X], \mathrm{Var}[X])$
- $D_h(X)$: a measure of *non-Gaussianness* of X
- Properties of $D_h(X)$:
  - $D_h(tX) = D_h(X)$
  - $D_h(X + \sqrt{t}Z)\big|_{t=0} = D_h(X)$
  - $Z \sim \mathcal{N}(0, \sigma^2) \implies$
    $\lim_{t\to\infty} D_h(X + \sqrt{t}Z) = \lim_{t\to\infty} D_h(\frac{X}{\sqrt{t}} + Z) = D_h(Z) = 0$
-
$$\int_0^\infty \frac{d}{dt} D_h(X + \sqrt{t}Z)dt = -D_h(X) \tag{10}$$
- Some Details

First Integral Representation of $h(X)$

- Define $D_J(X) := J(X) - J(X_G), \quad X_G \sim \mathcal{N}(\mathbb{E}[X], \mathrm{Var}[X])$
- $D_J(X)$: another measure of *non-Gaussianness* of X
- First dual version of de Bruijn's inequality (6) $\implies$

$$\frac{d}{dt} D_h(X + \sqrt{t}Z)dt = -\frac{1}{2} D_J(X + \sqrt{t}Z)$$

- Integrating above from $t = 0$ to ∞ and using (10) $\implies$

$$\boxed{D_h(X) = \frac{1}{2}\int_0^\infty D_J(X + \sqrt{t}Z)dt} \tag{11}$$

or

$$h(X) = h(X_G) - \frac{1}{2}\int_0^\infty D_J(X + \sqrt{t}Z)dt$$

- Some Details

Second Integral Representation of $h(X)$

- Define $D_V(X|Y) := \mathrm{mmse}(X_G|Y_G) - \mathrm{mmse}(X|Y)$, $X_G \sim \mathcal{N}(\mathbb{E}[X], \mathrm{Var}[X])$, $Y_G \sim \mathcal{N}(\mathbb{E}[Y], \mathrm{Var}[Y])$
- $D_V(X)$: still another measure of *non-Gaussianness* of X
-
$$\begin{aligned} D_h(\sqrt{t}X + Z)\Big|_{t=0} &= 0, \\ \lim_{t\to\infty} D_h(\sqrt{t}X + Z) = \lim_{t\to\infty} D_h(X + \frac{Z}{\sqrt{t}}) &= D_h(X) \end{aligned} \tag{12}$$
- Second dual version of de Bruijn's inequality (9) $\Longrightarrow$
$$\frac{d}{dt} D_h(\sqrt{t}X + Z)dt = \frac{1}{2} D_V(X|\sqrt{t}X + Z)$$
- Integrating above from $t = 0$ to ∞ and using (12) $\Longrightarrow$
$$\boxed{D_h(X) = \frac{1}{2}\int_0^{\infty} D_V(X|\sqrt{t}X + Z)dt} \tag{13}$$
- Some Details

Relationship between the 2 Two Integral Representations & EPI

- (4) $\implies D_J(X+Z) = D_V(X|X+Z)$
- $u = 1/t$ in (13) $\implies$ (11)
- Equality in EPI holds for Gaussians $\implies$ EPI (1) in terms of $D_h(X)$: $D_h(W) \le \lambda D_h(X) + (1-\lambda) D_h(Y), \quad W := \sqrt{\lambda} X + \sqrt{1-\lambda} Y$
- Integral representation (11) $\implies$

$$D_J(W+\sqrt{t}Z) \le \lambda D_J(X+\sqrt{t}Z) + (1-\lambda) D_J(Y+\sqrt{t}Z) \implies \text{EPI}$$

$$\Updownarrow$$

$$J(W+\sqrt{t}Z) \le \lambda J(X+\sqrt{t}Z) + (1-\lambda) J(Y+\sqrt{t}Z) \tag{14}$$

- Integral representation (13) $\implies$ $D_V(W|\sqrt{t}W+Z) \le$ $\lambda D_V(X|\sqrt{t}X+Z) + (1-\lambda) D_V(Y|\sqrt{t}Y+Z) \implies$ EPI

$$\Updownarrow$$

$$\text{mmse}(W|\sqrt{t}W+Z) \ge$$
$$\lambda\,\text{mmse}(X|\sqrt{t}X+Z) + (1-\lambda)\text{mmse}(Y|\sqrt{t}Y+Z) \tag{15}$$

Alternative Proofs of EPI based on FI & MMSE

- $\tilde{X} := X + \sqrt{t}Z, \tilde{Y} := Y + \sqrt{t}Z \quad \Longrightarrow$
  (14) $\quad \Leftrightarrow \quad J(\sqrt{\lambda}\tilde{X} + \sqrt{1-\lambda}\tilde{Y}) \leq \lambda J(\tilde{X}) + (1-\lambda)J(\tilde{Y})$
  : FI Property 4 seen in class!
- Interpretation: In terms of Cramér-Rao lower bound, *estimation* from individual measurements of independent r.v. s is better than that from the sum of the measurements.
- FI Property 4 seen in class $\Longrightarrow$ (14) $\Longrightarrow$ EPI. $\square$
- (9)-(10) in Verdú-Guo-2006 $\Longrightarrow$
  $\mathrm{mmse}(W|\sqrt{t}W + \sqrt{\lambda}Z' + \sqrt{1-\lambda}Z'') \geq$
  $\mathrm{mmse}(W|\sqrt{t}X + Z', \sqrt{t}Y + Z'')$
  $\Longrightarrow$ (15)
- Interpretation: In terms of MMSE, *estimation* from individual measurements of independent r.v. s is better than that from the sum of the measurements.
- (9)-(10) in Verdú-Guo-2006 $\Longrightarrow$ (15) $\Longrightarrow$ EPI. $\square$

New Proof of EPI based on Mutual Information

- Abstract key *ingredients* in the proofs of EPI:
  - 2 (independent) variables jointly bring more *information* than their sum
  - Gaussian perturbation argument using an auxilliary variable Z
- Recall:

$$\begin{tikzcd}
h(X) \arrow[rr, leftrightarrow] \arrow[d, leftrightarrow, "\text{de Bruijn (2)}"'] & & I(X;Y) \arrow[d, leftrightarrow, "\text{(3) Verdú-Guo}"] \\
J(X) \arrow[rr, leftrightarrow, dashed, "(5)"'] & & \mathrm{mmse}(X|Y)
\end{tikzcd}$$

- Why not use *Mutual Information directly* to prove EPI?

New Proof of EPI

- First key ingredient in terms of Mutual Information :
  well-known Data Processing Inequality for Mutual Information:

$$I(W + \sqrt{t}Z; Z) \leq I(X + \sqrt{\lambda t}Z, Y + \sqrt{(1-\lambda)t}Z; Z) \tag{16}$$

- Recall: $W + \sqrt{t}Z = \sqrt{\lambda}(X + \sqrt{\lambda t}Z) + \sqrt{1-\lambda}(Y + \sqrt{(1-\lambda)t}Z)$
- Apply Chain Rule of Mutual Information to (16). (Details) $\Longrightarrow$

$$I(W + \sqrt{t}Z; Z) \leq I(X + \sqrt{\lambda t}Z; Z) + I(Y + \sqrt{(1-\lambda)t}Z; Z) \tag{17}$$

- Gaussian perturbations $\Longrightarrow$ smooth densities $\Longrightarrow$
  $I(X + \sqrt{t}Z; Z)$ and $I(Y + \sqrt{t}Z; Z)$ are differentiable w.r.t. t.
- It can be proved that (Details)

$$\begin{aligned} I(X + \sqrt{\lambda t}Z; Z) &= \lambda I(X + \sqrt{t}Z; Z) + o(t), \\ I(Y + \sqrt{(1-\lambda)t}Z; Z) &= (1-\lambda) I(Y + \sqrt{t}Z; Z) + o(t) \end{aligned} \tag{18}$$

New Proof of EPI

- (17) & (18) $\Longrightarrow$

$$I(W+\sqrt{t}Z;Z) \leq \lambda I(X+\sqrt{t}Z;Z)+(1-\lambda)I(Y+\sqrt{t}Z;Z)+o(t) \quad (19)$$

- $X' := X + \sqrt{t'}Z_1$, $Y' := Y + \sqrt{t'}Z_2$, $\quad Z_1, Z_2, Z$ i.i.d.
- Gaussian perturbations $\Longrightarrow$ smooth densities $\Longrightarrow$ (19) applicable
- Suitable rearrangement (Details) $\Longrightarrow$

$$f(t'+t) \leq f(t') + o(t), \quad \text{(i.e.,} f(t) \text{ is } \textit{non-increasing}) \quad (20)$$

$$f(t) := I(W+\sqrt{t}Z;Z) - \lambda I(X+\sqrt{t}Z;Z) - (1-\lambda)I(Y+\sqrt{t}Z;Z)$$

- $\perp$ of r.v. s $\Longrightarrow$ $f(0) = 0$ $\Longrightarrow$ $f(t) \leq 0$ $\Longrightarrow$

$$I(W+\sqrt{t}Z;Z) \leq \lambda I(X+\sqrt{t}Z;Z) + (1-\lambda)I(Y+\sqrt{t}Z;Z) \quad (21)$$

Final Steps in the New Proof of EPI

- Use identity $I(U+\sqrt{t}Z;Z) = I(U+\sqrt{t}Z;U) - h(U) + h(Z)$
- Rearrange (21): $\forall\, t > 0$,

$$\begin{aligned} & h(W) - \lambda h(X) - (1-\lambda)h(Y) \geq \\ & I(W+\sqrt{t}Z;W) - \lambda I(X+\sqrt{t}Z;X) - (1-\lambda)I(Y+\sqrt{t}Z;Y) \end{aligned} \tag{22}$$

- For any $U \perp Z$,

$$\begin{aligned} I(U+\sqrt{t}Z;U) &= I(\frac{U}{\sqrt{t}}+Z;U), \quad \text{due to scale-invariance of } I \\ &= h(\frac{U}{\sqrt{t}}+Z) - h(Z), \quad \text{due to} \quad U \perp Z \\ &\leq h(\frac{U_G}{\sqrt{t}}+Z) - h(Z), \quad \text{due to } D_h(U) \geq 0 \\ &= \frac{1}{2}\log\left(1+\frac{\mathrm{Var}[\,U\,]}{t\mathrm{Var}[\,Z\,]}\right) \end{aligned}$$

The New Proof completed . . . :-)

- $$\lim_{t\to\infty} I(U+\sqrt{t}Z;U) \leq \lim_{t\to\infty} \frac{1}{2}\log\left(1+\frac{\mathrm{Var}[U]}{t\mathrm{Var}[Z]}\right) = 0$$

- Mutual Information ($I(U+\sqrt{t}Z;U)$) is *non-negative*

- $\implies \lim_{t\to\infty} I(U+\sqrt{t}Z;U) = 0$

- As $t \to \infty$, (22) $\implies$

$$h(W) - \lambda h(X) - (1-\lambda)h(Y) \geq 0. \quad \square$$

: EPI (1)

Questions?

Thank You

Extra Details 1

- $h(X_G) = \frac{1}{2}\log(2\pi e \mathrm{Var}[\,X\,])$
- scale invariance of $D_h(X)$:

$$D_h(tX) = h(Y) - h(tX), \quad Y \sim \mathcal{N}(\mathbb{E}[\,tX\,], \mathrm{Var}[\,tX\,]).$$

Hence, $Y \sim t\mathcal{N}(\mathbb{E}[\,X\,], \mathrm{Var}[\,X\,]) \sim tX_G$.
Hence,

$$D_h(tX) = h(tX_G) - h(tX) = h(X_G) + \log|t| - h(X) - \log|t| = D_h(X). \quad \square$$

- $D_J(X) \geq 0$: Property 2 of $J(X)$ seen in class (special case of Cramér-Rao Inequality)
- $D_V(X) \geq 0$: Property of the MMSE (sub-optimal for non-Gaussians)
- The above two & (11), (13) $\implies h(X_G) \geq h(X)$

Application of Chain Rule for Mutual Information to (16)

- $U := X + \sqrt{\lambda t} Z, \quad V := Y + \sqrt{(1-\lambda)t} Z$
- $$\begin{aligned} I(U,V;Z) = I(U;Z) + I(V;Z|U), & \\ \leq I(U;Z) + I(V;Z|U) + I(U;V), & \quad \because I(U;V) \geq 0 \\ = I(U;Z) + I(V;Z,U), & \\ = I(U;Z) + I(V;Z) + I(U;V|Z). & \end{aligned}$$
- $X \perp Y \implies U \perp V$ given $Z \implies I(U;V|Z) = 0 \implies$

$$I(U,V;Z) \leq I(U;Z) + I(V;Z). \quad \square$$

Extra Details 2

- $I(X+\sqrt{t}Z;Z)\big|_{t=0} = I(X;Z) = 0$, since $X \perp Z$
- Differentiability of $I(X+\sqrt{t}Z;Z)$ w.r.t. $t \implies$

$$\begin{aligned}
\exists \quad \lim_{t\to 0} \frac{I(X+\sqrt{t}Z;Z) - I(X;Z)}{t} &= \lim_{t\to 0} \frac{I(X+\sqrt{t}Z;Z)}{t} \\
&\Downarrow t \to \lambda t \\
&= \lim_{\lambda t\to 0} \frac{I(X+\sqrt{\lambda t}Z;Z)}{\lambda t}
\end{aligned} \tag{23}$$

- Hence

$$\frac{I(X+\sqrt{t}Z;Z)}{t} = \frac{I(X+\sqrt{\lambda t}Z;Z)}{\lambda t} + \frac{o(t)}{t}, \quad \lim_{t\to 0} \frac{o(t)}{t} = 0$$

- Rearrangement $\implies$

$$I(X+\sqrt{\lambda t}Z;Z) = \lambda I(X+\sqrt{t}Z;Z) + o(t). \quad \square$$

Rearrangement

- Identity $I(X+\sqrt{t'}Z_1+\sqrt{t}Z;Z) =$ $I(X+\sqrt{t'}Z_1+\sqrt{t}Z;\sqrt{t'}Z_1+\sqrt{t}Z) - I(X+\sqrt{t'}Z_1;Z_1)$
- Stability of Gaussian Distribution & $Z_1 \sim Z \implies \sqrt{t'}Z_1+\sqrt{t}Z \sim \sqrt{t'+t}Z$
- Hence, Identity becomes

$$I(X+\sqrt{t'}Z_1+\sqrt{t}Z;Z) = f_X(t'+t) - f_X(t') \tag{24}$$

$$f_X(t) := I(X+\sqrt{t}Z;Z)$$

- Similarly, we have

$$I(Y+\sqrt{t'}Z_2+\sqrt{t}Z;Z) = f_Y(t'+t) - f_Y(t'), \tag{25}$$

$$I(W+\sqrt{t'}\hat{Z}+\sqrt{t}Z;Z) = f_W(t'+t) - f_W(t'), \tag{26}$$

$$\text{where} \quad \hat{Z} = \sqrt{\lambda}Z_1 + \sqrt{1-\lambda}Z_2 \implies \hat{Z} \sim Z$$

- (26) - (24) - (25) $\implies$ (20). $\square$