Mismatched Estimation and Relative Entropy

Sergio Verdú, *Fellow, IEEE*

***Abstract*—A random variable with distribution P is observed in Gaussian noise and is estimated by a mismatched minimum mean-square estimator that assumes that the distribution is Q, instead of P. This paper shows that the integral over all signal-to-noise ratios (SNRs) of the excess mean-square estimation error incurred by the mismatched estimator is twice the relative entropy $D(P\|Q)$ (in nats). This representation of relative entropy can be generalized to nonreal-valued random variables, and can be particularized to give new general representations of mutual information in terms of conditional means. Inspired by the new representation, we also propose a definition of free relative entropy which fills a gap in, and is consistent with, the literature on free probability.**

***Index Terms*—Divergence, free probability, minimum mean-square error (MMSE) estimation, mutual information, relative entropy, Shannon theory, statistics.**

I. Introduction

A LINK between the central notions in estimation theory [minimum mean-square error (MMSE)] and information theory (entropy) was given in [1]: The entropy[1] of a discrete random variable X can be expressed as [1, Th. 13]

$$H(X) = \frac{1}{2}\int_0^\infty \mathsf{E}[(X - \mathsf{E}[X|\sqrt{\gamma}\,X + N])^2]\,\mathrm{d}\gamma \tag{1}$$

where N is a standard Gaussian random variable independent of X. If X does not have a finite second moment or it takes values on $\mathcal{A} \subset \mathbb{R}$, then X can be replaced on the right-hand side of (1) by $g(X)$ for any one-to-one function $g : \mathcal{A} \mapsto \mathbb{R}$, such that $\mathsf{E}[g^2(X)] < \infty$. Thus, the uncertainty in a random variable is equal to the integral over signal-to-noise ratio (SNR) of its MMSE when observed in Gaussian noise.

More generally, Guo *et al.* [1] show that for a real-valued random variable X

$$I(X; \sqrt{\mathsf{snr}}\,X + N) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma \tag{2}$$

where we have denoted the integrand in (1) by

$$\mathsf{mmse}(\gamma) = \mathsf{E}[(X - \mathsf{E}[X|\sqrt{\gamma}\,X + N])^2]. \tag{3}$$

Manuscript received July 21, 2009; revised March 21, 2010. Date of current version July 14, 2010. This work was supported in part by the National Science Foundation under Grants CCF-0635154 and CCF-0728445. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Seoul, South Korea, June 2009.

The author is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: Verdu@Princeton.edu).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2050800

[1]For brevity, throughout the paper, information units are nats, and logarithms are natural.

As shown in [1] (see also [2]), (2) can be generalized to complex-valued random variables, random vectors, and continuous-time random processes.

Furthermore, the non-Gaussianness (relative entropy with respect to a Gaussian distribution with the same mean and variance) of a continuous random variable is equal to [1, Th. 14]

$$\begin{aligned} D(X) &= D\left(P_X \| \mathcal{N}\left(\mathsf{E}[X], \sigma_X^2\right)\right) & (4)\\ &= \frac{1}{2}\int_0^\infty \frac{\sigma_X^2}{1+\gamma\,\sigma_X^2} - \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma & (5)\end{aligned}$$

where the relative entropy (or divergence) of probability distributions $P \ll Q$ is

$$D(P\,\|\,Q) = \int \log\frac{\mathrm{d}P}{\mathrm{d}Q}\,\mathrm{d}P. \tag{6}$$

In the special case of Gaussian X, the MMSE estimator is linear and

$$\mathsf{mmse}(\gamma) = \frac{\sigma_X^2}{1+\gamma\,\sigma_X^2}. \tag{7}$$

Therefore, the non-Gaussianness of a random variable is equal to half the integral (over SNR) of the excess mean-square error (MSE) incurred by the linear MMSE estimator over the (nonlinear) MMSE estimator. Note that the representation in (5) gives an estimation theoretic representation for the differential entropy of continuous random variables in view of

$$h(X) = \frac{1}{2}\log\left(2\pi e\sigma_X^2\right) - D(X). \tag{8}$$

The various estimation-theoretic representations of information-theoretic quantities are not more advantageous than their conventional definitions from a computational standpoint. However, they have proved surprisingly useful in a variety of applications: (5) is instrumental to obtain particularly simple proofs of Shannon's entropy-power inequality [3] and of the monotonic decrease with n of $D(X_1 + \cdots + X_n)$ where X_i are independent identically distributed (i.i.d.) [4]; for other applications and generalizations of (1), (2), and (5), see [1], [5] [2], [6], [7], [9], [10], [11], and [12].

The purpose of this paper is to give a new estimation-theoretic representation for relative entropy. Since entropy, mutual information, differential entropy, and non-Gaussianness can be obtained through particularizations of relative entropy, the formula in this paper can be used to recover the known relationships (1), (2), and (5).

The new formula for $D(P\,\|\,Q)$ is presented in Section II, along with several properties and examples; the proof is given in Section III. The multivariate version of the formula is given in Section IV, which is particularized to give new estimation theoretic representations of mutual information in Section V.

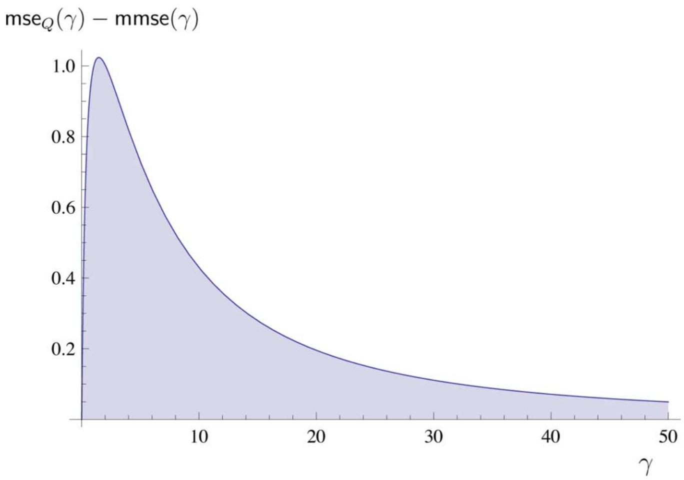

Fig. 1. Excess MSE for $P = \mathcal{N}(0,2)$,$Q = \mathcal{N}(0,0.1)$.

Section VI uses the mismatched estimation approach to give a new relationship between relative entropy and relative Fisher information. Section VII gives a definition of free relative entropy and explores its relationship with existing notions in free probability.

II. New Formula

For clarity, we focus attention first on real-valued random variables with finite second moments. The conditional mean (conditioned on a particular value of the Gaussian-contaminated observation) assuming that X is distributed according to Q is equal to

$$\mathsf{E}_Q[X \,|\, \alpha X + N = y] = \frac{\int z\, e^{-(y-\alpha z)^2/2}\, \mathrm{d}Q(z)}{\int e^{-(y-\alpha v)^2/2}\, \mathrm{d}\, Q(v)} \tag{9}$$

where N is standard Gaussian independent of X and the integrals are over the whole real line. Define now the *mismatched MSE* as

$$\mathsf{mse}_Q(\gamma) = \mathsf{E}[(X - \mathsf{E}_Q[X \,|\, \sqrt{\gamma}X + N])^2]. \tag{10}$$

In (10) and henceforth, we follow the convention that an unsubscripted expectation is with respect to X distributed according to P and independent standard Gaussian N, i.e., $\mathsf{E}_P = \mathsf{E}$. Although, for brevity, $\mathsf{mse}_Q(\gamma)$ only shows explicitly its dependence on Q, it also depends on P: it is equal to the mismatched MSE attained by an estimator that minimizes the MSE if X were distributed according to Q, when the actual distribution is P. Therefore, the MMSE in (3) can be written as

$$\mathsf{mmse}(\gamma) = \mathsf{mse}_P(\gamma). \tag{11}$$

The main result is the following.

Theorem 1:

$$D(P \,\|\, Q) = \frac{1}{2}\int_0^\infty \mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma \tag{12}$$

The nonnegativity of relative entropy follows immediately from (12) since the integrand is nonnegative; the uniqueness of the minimum mean square estimator yields that relative entropy can be 0 only if $P = Q$.

A variational representation for relative entropy can be obtained writing

$$\begin{aligned} D(P \,\|\, Q) &\\ &= \sup_{P \ll R \ll Q} \frac{1}{2}\int_0^\infty \mathsf{mse}_Q(\gamma) - \mathsf{mse}_R(\gamma)\, \mathrm{d}\gamma \qquad (13)\\ &= \sup_{P \ll R \ll Q} \mathsf{E}\left[\log \frac{R(X)}{Q(X)}\right]. \qquad (14) \end{aligned}$$

where X is distributed according to P. It can be shown that (14) is equivalent to the Legendre–Fenchel representation of relative entropy known as the Donsker–Varadhan formula [13].

Note that in the special case where the reference measure is Gaussian with the same mean and variance as the first measure, (12) reduces to (5) since in that case the mismatched MSE is the linear MMSE

$$\mathsf{mse}_Q(\gamma) = \frac{\sigma_X^2}{1 + \gamma\, \sigma_X^2}. \tag{15}$$

In the special case $P = \mathcal{N}(\mu_1, \sigma_1^2)$ and $Q = \mathcal{N}(\mu_2, \sigma_2^2)$, we have

$$\mathsf{mse}_Q(\gamma) = \frac{(\mu_1 - \mu_2)^2 + \sigma_1^2 + \gamma\sigma_2^4}{(1 + \gamma\sigma_2^2)^2} \tag{16}$$

$$\mathsf{mmse}(\gamma) = \frac{\sigma_1^2}{1 + \gamma\sigma_1^2} \tag{17}$$

and (cf., Fig. 1)

$$\begin{aligned} \int_0^\infty &\mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma \\ &= \frac{(\mu_1 - \mu_2)^2 + \sigma_1^2 - \sigma_2^2}{\sigma_2^2} + \log \frac{\sigma_2^2}{\sigma_1^2} \qquad (18)\\ &= 2D\left(\mathcal{N}\left(\mu_1, \sigma_1^2\right) \|\, \mathcal{N}\left(\mu_2, \sigma_2^2\right)\right). \qquad (19) \end{aligned}$$

For a random variable equally likely to take M values, (1) states that the integral of the MMSE based on a Gaussian-contaminated observation is equal to the natural logarithm of M^2. Interestingly, Theorem 1 enables us to prove a robustness property of the MMSE estimator for the equiprobable distribution when used for a nonequiprobable random variable (distributed on the same M values): the integral of the ensuing mismatched MSE

$$\int_0^\infty \mathsf{mse}_U(\gamma)\, \mathrm{d}\gamma = 2\log M \tag{20}$$

does not depend on the true distribution even though the mismatched MSE $\mathsf{mse}_U(\gamma)$ does depend on the true distribution.

The relative-entropy processing theorem states that the distributions at the output of a system are closer than at the input

$$D(P \,\|\, Q) \geq D(\bar{P} \,\|\, \bar{Q}) \tag{21}$$

where

$$\bar{P}(y) = \int W(y \,|\, x) dP(x) \tag{22}$$

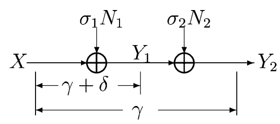

Fig. 2. SNR-incremental Gaussian channel.

$$\bar{Q}(y) = \int W(y \,|\, x) dQ(x). \tag{23}$$

Theorem 1 suggests exploring a possible refinement of the relative-entropy processing theorem whereby the excess MSE $\mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)$ decreases with processing at any γ, and not just its integral. A counterexample to such refinement is easy to construct since a one-to-one transformation can indeed change the excess MSE but not its integral. For example, multiplication by a scalar α changes the excess square error to $\alpha^2 \mathsf{mse}_Q(\alpha^2\gamma) - \alpha^2 \mathsf{mmse}(\alpha^2\gamma)$.

III. Proof

Proof: The idea is to show that

$$D(P * \mathcal{N}(0,\sigma^2) \,\|\, Q * \mathcal{N}(0,\sigma^2)) = \frac{1}{2}\int_0^{1/\sigma^2} \mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma \tag{24}$$

and therefore the relative entropy between two distributions observed in Gaussian noise decreases, as the noise variance increases infinitesimally, proportionally to the penalty incurred by a mismatched estimator.

First, we give two auxiliary results (whose proofs follow that of Theorem 1). In keeping with longstanding tradition in information theory, it is convenient, and not subject to confusion, to denote

$$D(U \| V) = D(P_U \| P_V) \tag{25}$$

whenever U and V are random variables defined on the same measurable space with respective distributions P_U and P_V.

The local behavior of relative entropy is [15]

$$D(P_\theta \,\|\, P_0) = \frac{\theta^2}{2} I(0) + o(\theta^2) \tag{26}$$

where $I(\theta)$ is Fisher's information of the parametric family P_θ. For the purposes of the next result, we do not need to deal with the full generality of Fisher's information since the perturbations are around a Gaussian random variable; however, (26) is not sufficient since the reference measure is also changing with the parameter.

Lemma 1: Let N be standard Gaussian and let U and V have finite second moments and be independent of N. Then

$$D(N+\theta U \| N + \theta V) = \frac{\theta^2}{2}(\mathsf{E}U - \mathsf{E}V)^2 + o(\theta^2). \tag{27}$$

Another auxiliary result gives an alternative expression for the integrand in (12).

Lemma 2:

$$\mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma) = \mathsf{E}[(\mathsf{E}_Q[X \,|\, Y] - \mathsf{E}[X \,|\, Y])^2] \tag{28}$$

where the outer expectation is with respect to $Y = \sqrt{\gamma}X + N$ with X distributed according to P. Lemma 2 is reminiscent of the well-known result that if the sigma-fields satisfy $\mathcal{G} \subset \mathcal{F}$, then

$$\mathsf{E}[(X - \mathsf{E}[X \,|\, \mathcal{G}])^2] - \mathsf{E}[(X - \mathsf{E}[X \,|\, \mathcal{F}])^2] = \mathsf{E}[(\mathsf{E}[X \,|\, \mathcal{G}] - \mathsf{E}[X \,|\, \mathcal{F}])^2]. \tag{29}$$

As in the incremental-channel proof of [1, Th. 1], consider the setup in Fig. 2 where N_1 and N_2 are independent standard Gaussian random variables and

$$\gamma + \delta = \frac{1}{\sigma_1^2} \tag{30a}$$

$$\gamma = \frac{1}{\sigma_1^2 + \sigma_2^2}. \tag{30b}$$

When X is distributed according to P (resp., Q), we denote the joint distribution of Y_1 and Y_2 by $P_{Y_1Y_2}$ (resp., $Q_{Y_1Y_2}$). The relative entropy between P_{Y_2} and Q_{Y_2} is smaller than the relative entropy between P_{Y_1} and Q_{Y_1}. Moreover, when δ (and thus, σ_2) is small, the decrease is linear with δ as the following result shows.

Lemma 3:

$$D(P_{Y_1} \,\|\, Q_{Y_1}) - D(P_{Y_2} \,\|\, Q_{Y_2}) = \frac{\delta}{2}\mathsf{E}[(X - \mathsf{E}_Q[X \,|\, Y_2])^2 - (X - \mathsf{E}[X \,|\, Y_2])^2] + o(\delta). \tag{31}$$

In view of Lemma 2 and (30), Lemma 3 is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} D\left(P * \mathcal{N}\left(0, \frac{1}{\gamma}\right) \middle\| Q * \mathcal{N}\left(0, \frac{1}{\gamma}\right)\right) = \frac{1}{2}(\mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)) \tag{32}$$

from which (24) follows since

$$\lim_{\gamma \to 0} D\left(P * \mathcal{N}\left(0, \frac{1}{\gamma}\right) \middle\| Q * \mathcal{N}\left(0, \frac{1}{\gamma}\right)\right) = 0 \tag{33}$$

for any P and Q. □

Proof (Lemma 1): We will assume $\mathsf{E}V = 0$ and will replace $\mathsf{E}U$ by $\mathsf{E}U - \mathsf{E}V$ at the end, since

$$D(N + \theta U \,\|\, N + \theta V) = D(N + \theta(U - \mathsf{E}V) \,\|\, N + \theta(V - \mathsf{E}V)). \tag{34}$$

Let $\bar{U}$ be a Gaussian random variable with the same mean and variance as U, and also independent of N. Denoting the standard normal distribution by Φ, it is immediate to check that

$$D(N + \theta U \,\|\, N + \theta V) = D(N + \theta U) + D(N + \theta \bar{U} \,\|\, N) - \mathsf{E}\left[\log \frac{P_{N+\theta V}(N+\theta U)}{\Phi(N + \theta U)}\right]. \tag{35}$$

According to [16, Th. 1], the first term on the right-hand side of (35) satisfies

$$D(N+\theta U)=o(\theta^2). \tag{36}$$

The second term is equal to

$$D(N+\theta\bar{U}\,\|\,N)=\frac{\theta^2}{2}\mathsf{E}[U^2]-\frac{1}{2}\log\left(1+\theta^2\sigma_U^2\right) \tag{37}$$
$$=\frac{\theta^2}{2}\mathsf{E}^2[U]+o(\theta^2). \tag{38}$$

The third term can be taken care analogously to (36) [16]

$$\mathsf{E}\left[\log\frac{P_{N+\theta V}(N+\theta U)}{\Phi(N+\theta U)}\right] \tag{39}$$
$$=\mathsf{E}\left[\log\mathsf{E}\left[\exp\left(-\frac{\theta^2V^2}{2}+(N+\theta U)\theta V\right)\middle|\,N,U\right]\right]$$
$$=\mathsf{E}[(N+\theta U)\theta V]+o(\theta^2) \tag{40}$$
$$=o(\theta^2) \tag{41}$$

where V is independent of U with distribution P_V (and zero mean), and (39) follows by writing the density of $N+\theta V$ as an expectation with respect to V. □

Proof (Lemma 2):

$$\mathsf{mse}_Q(\gamma)-\mathsf{mmse}(\gamma)$$
$$=\mathsf{E}[(X-\mathsf{E}_Q[X\,|\,Y])^2-(X-\mathsf{E}[X\,|\,Y])^2] \tag{42}$$
$$=\mathsf{E}[(2X-\mathsf{E}_Q[X\,|\,Y]-\mathsf{E}[X\,|\,Y])(\mathsf{E}[X\,|\,Y]-\mathsf{E}_Q[X\,|\,Y])] \tag{43}$$
$$=\mathsf{E}[(\mathsf{E}_Q[X\,|\,Y]-\mathsf{E}[X\,|\,Y])^2]+2\,\mathsf{E}[(X-\mathsf{E}[X\,|\,Y])(\mathsf{E}[X\,|\,Y]-\mathsf{E}_Q[X\,|\,Y])] \tag{44}$$
$$=\mathsf{E}[(\mathsf{E}_Q[X\,|\,Y]-\mathsf{E}[X\,|\,Y])^2] \tag{45}$$

where (45) holds for any Q on account of the orthogonality principle. □

Proof (Lemma 3): As in [1], we write Y_1 as a function of Y_2, X, N_1, and N_2

$$(\gamma+\delta)\,Y_1=\gamma\,(Y_2-\sigma_2N_2)+\delta\,(X+\sigma_1N_1) \tag{46}$$
$$=\gamma\,Y_2+\delta\,X+\sqrt{\delta}\,N \tag{47}$$

where

$$N=\frac{1}{\sqrt{\delta}}\,(\delta\,\sigma_1\,N_1-\gamma\,\sigma_2\,N_2). \tag{48}$$

Recalling (30), it is straightforward to verify that:

- N is a standard Gaussian random variable;
- N is independent of the Gaussian noise $\sigma_1N_1+\sigma_2N_2$ affecting the observation

$$Y_2=X+\sigma_1N_1+\sigma_2N_2. \tag{49}$$

This is important because it allows us to view (47), upon conditioning on Y_2, as a simple single-input channel (with distribution X conditioned on Y_2) subject to additive independent Gaussian noise and an irrelevant additive deterministic component.

For every scalar y

$$D\left(P_{Y_1\,|\,Y_2=y}\middle\|\,Q_{Y_1\,|\,Y_2=y}\right)$$
$$=D\left(P_{\sqrt{\delta}X+N\,|\,Y_2=y}\middle\|\,Q_{\sqrt{\delta}X+N\,|\,Y_2=y}\right) \tag{50}$$
$$=D\left(\mathcal{N}(0,1)*P_{\sqrt{\delta}X\,|\,Y_2=y}\middle\|\,\mathcal{N}(0,1)*Q_{\sqrt{\delta}X\,|\,Y_2=y}\right) \tag{51}$$
$$=\frac{\delta}{2}(\mathsf{E}[X|Y_2=y]-\mathsf{E}_Q[X\,|\,Y_2=y])^2+o(\delta) \tag{52}$$

where (52) follows from Lemma 1. Averaging (52) over $Y_2=y$ given by (49) where X is distributed according to P, we obtain[2]

$$\frac{\delta}{2}\mathsf{E}[(\mathsf{E}[X|Y_2]-\mathsf{E}_Q[X\,|\,Y_2])^2]+o(\delta)$$
$$=D\left(P_{Y_1\,|\,Y_2}\middle\|\,Q_{Y_1\,|\,Y_2}|P_{Y_2}\right) \tag{53}$$
$$=D\left(P_{Y_1}\,\|\,Q_{Y_1}\right)-D\left(P_{Y_2}\,\|\,Q_{Y_2}\right) \tag{54}$$

where (54) follows from the fact that $P_{Y_2\,|\,Y_1}=Q_{Y_2\,|\,Y_1}$. Applying Lemma 2 (where Y takes the role of Y_2) to the left-hand side of (53), we obtain the desired result. □

IV. Multivariate Version

We now consider the case of random vectors with finite second moments. The vector $\mathbf{X}$ can have multivariate distributions P or Q. The Gaussian vector $\mathbf{N}$ is independent of $\mathbf{X}$, has zero mean, and its covariance matrix is the identity. Theorem 1 holds verbatim upon defining

$$\mathsf{mse}_Q(\gamma)=\mathsf{E}[\|\,\mathbf{X}-\mathsf{E}_Q[\mathbf{X}\,|\,\sqrt{\gamma}\mathbf{X}+\mathbf{N}]\|^2]. \tag{55}$$

To verify that generalization of Theorem 1, we can follow the incremental channel approach in Section III along with the more general auxiliary results:

Lemma 4:

$$D(\mathbf{N}+\theta\mathbf{U}\,\|\,\mathbf{N}+\theta\mathbf{V})=\frac{\theta^2}{2}\,\|\,\mathsf{E}[\mathbf{U}-\mathbf{V}]\|^2+o(\theta^2). \tag{56}$$

Lemma 5:

$$\mathsf{mse}_Q(\gamma)-\mathsf{mmse}(\gamma)=\mathsf{E}[\,\|\,\mathsf{E}_Q[\mathbf{X}\,|\,\mathbf{Y}]-\mathsf{E}[\mathbf{X}\,|\,\mathbf{Y}]\|^2]. \tag{57}$$

As for the entropy formula (1), when P and Q are not defined on $(\mathbb{R}^n,\mathcal{B}^n)$ but on some measurable space $(\mathcal{A},\mathcal{F})$ (or are not second order), the main formula still holds provided that $\mathbf{X}$ in (55) is replaced by $g(\mathsf{X})$ where X is the random variable that can have distributions P or Q and g is a measurable one-to-one function $g:\mathcal{A}\mapsto\mathbb{R}^n$, such that $\mathsf{E}[\,\|\,g(\mathsf{X})\|^2]<\infty$ under either distribution.

V. Mutual Information

We can recover (2) from Theorem 1 noting that

$$I(X;\sqrt{\mathsf{snr}}X+N)$$
$$=\int D(N+\sqrt{\mathsf{snr}}\,x\,||\,N+\sqrt{\mathsf{snr}}X)\,\mathrm{d}P_X(x) \tag{58}$$
$$=\int D(\delta_x*\mathcal{N}(0,\mathsf{snr}^{-1})\;||\;P_X*\mathcal{N}(0,\mathsf{snr}^{-1}))\,\mathrm{d}P_X(x) \tag{59}$$

[2]The issue of interchanging the expectation and limit with respect to δ can be handled as in the incremental-channel proof of (2) [18].

$$= \frac{1}{2}\int\int_0^{\mathsf{snr}} \mathsf{mse}_{P_X,\delta_x}(\gamma) - \mathsf{mmse}_{\delta_x}(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}P_X(x) \quad (60)$$

$$= \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma \quad (61)$$

where (60) follows from (24) at $\sigma^2 = 1/\mathsf{snr}$ and the mismatched MSE assuming distribution P_X when the true distribution is δ_x is denoted by

$$\mathsf{mse}_{P_X,\delta_x}(\gamma) = \mathsf{E}[(x - \mathsf{E}[X|\sqrt{\gamma}x + N])^2] \quad (62)$$

$$\mathsf{mmse}(\gamma) = \int \mathsf{mse}_{P_X,\delta_x}(\gamma)\,\mathrm{d}\,P_X(x) \quad (63)$$

$$\mathsf{mmse}_{\delta_x}(\gamma) = 0 \quad (64)$$

since the MMSE is zero when the true distribution is an atom.

In addition to (2) giving the mutual information between a random variable and its Gaussian contaminated version, a relationship between mutual information and estimation had been found in [19]

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(X;\sqrt{\gamma}Y + N)\,|_{\gamma=0} = \frac{1}{2}\mathsf{E}[(\mathsf{E}[Y\,|\,X] - \mathsf{E}[Y])^2]. \quad (65)$$

In the remainder of this section, we give two general expressions for mutual information between arbitrary random variables. To motivate the first expression, note that if X is discrete, then (1) leads to

$$H(X) - H(X\,|\,Y) = \frac{1}{2}\int_0^\infty \mathsf{mmse}(\gamma) - \mathsf{E}[(X - \mathsf{E}[X\,|\,\sqrt{\gamma}X + N, Y])^2]\mathrm{d}\gamma. \quad (66)$$

More generally, applying Theorem 1 to the conditional relative entropy $D(P_{Y\,|\,X}\,\|\,P_Y\,|\,P_X)$ and using (29), it is easy to obtain the following result.

Theorem 2: If Y has finite second-order moment

$$I(X;Y) = \frac{1}{2}\int_0^\infty \mathsf{E}[(\mathsf{E}[Y\,|\,\sqrt{\gamma}Y + N] - \mathsf{E}[Y\,|\,\sqrt{\gamma}Y + N, X])^2]\mathrm{d}\gamma \quad (67)$$

where N is standard Gaussian independent of X and Y.

Therefore, the mutual information between X and Y is the integrated (over SNR) decrease in MMSE in the estimation of Y thanks to the availability of X.

By specializing the multivariate version of Theorem 1 where $P = P_{XY}$ and $Q = P_X \times P_Y$ and X and Y are real-valued with finite second moments[3] we obtain the following symmetric representation.

Theorem 3: The mutual information between real-valued random variables with finite second moments is given by

$$I(X;Y) = \frac{1}{2}\int_0^\infty \mathsf{E}[(\mathsf{E}[X\,|\,Z_1] - \mathsf{E}[X\,|\,Z_1,Z_2])^2]\,\mathrm{d}\gamma + \frac{1}{2}\int_0^\infty \mathsf{E}[(\mathsf{E}[Y\,|\,Z_2] - \mathsf{E}[Y\,|\,Z_1,Z_2])^2]\,\mathrm{d}\gamma \quad (68)$$

where

$$Z_1 = \sqrt{\gamma}X + N_1 \quad (69)$$

$$Z_2 = \sqrt{\gamma}Y + N_2 \quad (70)$$

and N_1 and N_2 are standard Gaussian independent of each other and of (X,Y).

Proof: Under $Q = P_X \times P_Y$, (55) becomes

$$\mathsf{mse}_Q(\gamma) = \mathsf{E}[(X - \mathsf{E}[X\,|\,Z_1])^2] + \mathsf{E}[(Y - \mathsf{E}[Y\,|\,Z_2])^2] \quad (71)$$

while under the true joint distribution P_{XY}

$$\mathsf{mmse}(\gamma) = \mathsf{E}[(X - \mathsf{E}[X\,|\,Z_1,Z_2])^2] + \mathsf{E}[(Y - \mathsf{E}[Y\,|\,Z_1,Z_2])^2]. \quad (72)$$

Subtracting (72) from (71), it is easy to verify that we obtain

$$\begin{aligned}\mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma) &= \mathsf{E}[\mathsf{E}^2[X\,|\,Z_1,Z_2] - \mathsf{E}^2[X\,|\,Z_1]] + \mathsf{E}[\mathsf{E}^2[Y\,|\,Z_1,Z_2] - \mathsf{E}^2[Y\,|\,Z_1]] &(73)\\ &= \mathsf{E}[(\mathsf{E}[X\,|\,Z_1] - \mathsf{E}[X\,|\,Z_1,Z_2])^2] + \mathsf{E}[(\mathsf{E}[Y\,|\,Z_2] - \mathsf{E}[Y\,|\,Z_1,Z_2])^2] &(74)\end{aligned}$$

□

In the discrete case, (68) readily follows from (1) since under the (mismatched) product distribution

$$H(X) + H(Y) = \frac{1}{2}\int_0^\infty \mathsf{mse}_{P_X\times P_Y}(\gamma)\,\mathrm{d}\gamma. \quad (75)$$

and

$$H(X,Y) = \frac{1}{2}\int_0^\infty \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma \quad (76)$$

with the integrands in (75) and (76) given in (71) and (72), respectively.

Recalling (29), note that

$$\mathsf{E}[(\mathsf{E}[X\,|\,Z_1] - \mathsf{E}[X\,|\,Z_1,Z_2])^2]$$

is the increase in MSE incurred by neglecting the availability of Z_2 when estimating X (which goes to 0 as $\gamma \to 0$ or $\gamma \to \infty$).

Theorem 3 suggests the dissection of mutual information into

$$I(X;Y) = I(X\,|\,Y) + I(Y\,|\,X) \quad (77)$$

with the partial asymmetrical information of X provided by Y defined as the nonnegative quantity

$$I(X\,|\,Y) = \frac{1}{2}\int_0^\infty \mathsf{E}[(\mathsf{E}[X\,|\,\sqrt{\gamma}X + N_1] - \mathsf{E}[X\,|\,\sqrt{\gamma}X + N_1, \sqrt{\gamma}Y + N_2])^2]\,\mathrm{d}\gamma \quad (78)$$

$$= \frac{1}{2}\int_0^\infty \mathsf{mmse}(\gamma) - \mathsf{E}[(X - \mathsf{E}[X\,|\,\sqrt{\gamma}X + N_1, \sqrt{\gamma}Y + N_2])^2]\,\mathrm{d}\gamma. \quad (79)$$

[3] As usual, this restriction can be dropped appealing to one-to-one transformations.

Comparing (79) with (76), we see that if N_2 in (79) had zero variance instead of unit variance we would obtain $I(X;Y)$. Note that for discrete random variables

$$I(X\,|\,X) = \frac{1}{2}H(X). \tag{80}$$

The generalization of Theorem 3 to the multivariate case where $\mathbf{X}$ and $\mathbf{Y}$ are vectors of dimensions n and m, respectively, is straightforward

$$I(\mathbf{X};\mathbf{Y}) = \frac{1}{2}\int_0^\infty \mathsf{E}[\,\|\,\mathsf{E}[\mathbf{X}\,|\,\mathbf{Z}_1] - \mathsf{E}[\mathbf{X}\,|\,\mathbf{Z}_1,\mathbf{Z}_2]\|^2]\,\mathrm{d}\gamma + \frac{1}{2}\int_0^\infty \mathsf{E}[\,\|\,\mathsf{E}[\mathbf{Y}\,|\,\mathbf{Z}_2] - \mathsf{E}[\mathbf{Y}\,|\,\mathbf{Z}_1,\mathbf{Z}_2]\|^2]\mathrm{d}\gamma \tag{81}$$

where

$$\mathbf{Z}_1 = \sqrt{\gamma}\,\mathbf{X} + \mathbf{N}_1 \tag{82}$$

$$\mathbf{Z}_2 = \sqrt{\gamma}\,\mathbf{Y} + \mathbf{N}_2 \tag{83}$$

and $\mathbf{N}_1$ and $\mathbf{N}_2$ are Gaussian vectors of dimensions n and m, respectively, with unit covariance matrices, independent of each other and of $(\mathbf{X},\mathbf{Y})$.

VI. Relative Fisher Information

Although it does not appear to have received widespread attention, it is natural to define the relative Fisher information[4] (e.g., [20]) as

$$I(U\,\|\,V) = \mathsf{E}\left[\left(\nabla \log \frac{\mathrm{d}P_U}{\mathrm{d}P_V}(U)\right)^2\right]. \tag{84}$$

In view of de Bruijn's identity, we might expect that there is a connection between the relative Fisher information and relative entropy. One such connection, dual to Lemma 1, was unveiled in [17, Th. 1]

$$\frac{\mathrm{d}}{\mathrm{d}\delta}D(U+\sqrt{\delta}Z\,\|\,V+\sqrt{\delta}Z)\,|_{\,\delta=0} = -\frac{1}{2}I(U\,\|\,V) \tag{85}$$

where Z is a (not necessarily Gaussian) random variable independent of U and V.

Instead, using mismatched estimation as a bridge, here we establish the relationship

$$I(\sqrt{\gamma}U+N\,\|\,\sqrt{\gamma}V+N) = 2\gamma\frac{\mathrm{d}}{\mathrm{d}\gamma}D(\sqrt{\gamma}U+N\,\|\,\sqrt{\gamma}V+N) \tag{86}$$

where N is standard normal independent of U and V. Integrating (86), we obtain the following representation of relative entropy in terms of relative Fisher information:

$$D(U\,\|\,V) = \frac{1}{2}\int_0^\infty I(U+\sqrt{\delta}N\,\|\,V+\sqrt{\delta}N)\,\mathrm{d}\delta. \tag{87}$$

In view of (32), to verify (86), it suffices to show that

$$I(\sqrt{\gamma}U+N\,\|\,\sqrt{\gamma}V+N) = \gamma(\mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)) \tag{88}$$

[4] In this brief treatment, we restrict attention to the scalar case.

where P and Q take the role of P_U and P_V, respectively. In turn, (88) follows by representing its right-hand side as in Lemma 2 and using the identity

$$\frac{1}{\sqrt{\gamma}}\nabla \log \frac{\mathrm{d}P_{\sqrt{\gamma}U+N}}{\mathrm{d}P_{\sqrt{\gamma}V+N}}(z) = \mathsf{E}[U\,|\,\sqrt{\gamma}U+N=z] - \mathsf{E}[V\,|\,\sqrt{\gamma}V+N=z] \tag{89}$$

which can be obtained from (9).

VII. Free Relative Entropy

Free probability has emerged in the last two decades as a major application of noncommutative operator algebras, and has seen its most well-known success in the study of the spectral distribution of large random matrices. Introductions to free probability with the bare minimum required degree of abstraction can be found in [21] and [22]. A number of free counterparts of distributions and transforms are known; for example, the free counterpart of the log-moment generating transform is Voiculescu's R-transform (the inverse R-transform of the sum of R-transforms is the free convolution); the free counterpart of the Gaussian distribution is the semicircle law

$$w_a(x) = \begin{cases} \frac{2}{a^2\pi}\sqrt{a^2-x^2}, & |\,x\,| \le a \\ 0, & |\,x\,| > a \end{cases} \tag{90}$$

whose second moment is $a^2/4$. The free counterpart of the central limit theorem states that repeated free convolutions of properly normalized distributions converge to the semicircle law. The free counterpart of the Cauchy distribution is the Cauchy distribution itself. The free counterpart of the Poisson distribution is the Marčenko–Pastur distribution, which is the asymptotic distribution of the squared singular values of matrices with i.i.d. entries [21]. Voiculescu has also introduced the free entropy (to be more precise, the free differential entropy); in the case of a univariate density $f(t)$, it takes the form [23], [24]

$$\chi(X) = \int\int \log|\,t-s\,|\,f(t)f(s)\,\mathrm{d}t\,\mathrm{d}s + \frac{3}{4} + \frac{1}{2}\log 2\pi \tag{91}$$

while the free Fisher information is defined as

$$\mathcal{I}(X) = \frac{4\pi^2}{3}\int f^3(t)\,\mathrm{d}t \tag{92}$$

which is equal to the reciprocal of the variance in the case of the semicircle law. Among all distributions with second moment equal to σ^2 (91) is maximized by the semicircular law $w_{2\sigma}$ which attains the value

$$\chi(W) = \frac{1}{2}\log 2\pi e\sigma^2. \tag{93}$$

The free counterpart of the I-MMSE relationship (2) is given in [12]. Also known [26] are free counterparts of the entropy power inequality and of de Bruijn's identity

$$\chi(X+\sqrt{\alpha}W) - \chi(X) = \frac{1}{2}\int_0^\alpha \mathcal{I}(X+\sqrt{\lambda}W)\,\mathrm{d}\lambda \tag{94}$$

where W and X are free and W has the standard semicircle law. (Recall that the distribution of the sum of free random variables is obtained through the free convolution of their distributions.) A particularly simple proof of the free entropy power inequality (free analogue of that in [3]) is given in [12]. The monotonic decrease of nonsemicircularity is shown in [27]. However, a fully satisfactory free counterpart of relative entropy has yet to be found (cf., [28] and [29]).

In the remainder of this section, we put forward a definition of free relative entropy. Consider a noncommutative von Neumann algebra (an algebra of bounded operators on a Hilbert space that is closed in the weak operator topology and contains the identity operator) of self-adjoint operators endowed with a *state* (unit-norm positive linear functional) τ which is also a trace [30]. The probability measure μ_X with compact support on the real line associated with an element X of the algebra is uniquely determined by its moments

$$\int_{-\infty}^{\infty} x^n \,\mathrm{d}\mu_X(x) = \tau(X^n). \tag{95}$$

We define the free counterpart to (10) by

$$\mathsf{mse}_Q(\gamma) = \tau((X - \mathsf{E}_Q[X \,|\, Y])^2) \tag{96}$$

where $Y = \sqrt{\gamma}X + W$; X has measure P; W is standard semicircular; W and X are free; and $\mathsf{E}_Q[\cdot \,|\, Y]$ is the conditional expectation (orthogonal projection with respect to the scalar product derived from the trace) onto the von Neumann subalgebra generated by Y when X is distributed according to Q.

When $Q = P$, $\mathsf{E}_P[\cdot \,|\, Y] = \mathsf{E}[\cdot \,|\, Y]$ is the projection onto the W^*-subalgebra generated by Y, and we write

$$\mathsf{mse}_P(\gamma) = \mathsf{mmse}(\gamma) \tag{97}$$

$$= \frac{1}{\gamma}(1 - \mathcal{I}(\sqrt{\gamma}X + W)) \tag{98}$$

where (98) is shown in [12] and is identical to the relationship in [1, eq. (58)].

We now put forth the definition of the free relative entropy between P and Q in parallel to the result given in Theorem 1

$$\mathcal{D}(P \,\|\, Q) = \frac{1}{2}\int_0^{\infty} \mathsf{mse}_Q(\gamma) - \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma. \tag{99}$$

An alternative to Voiculescu's liberation-based quantification of the degree of freeness [25] is the free mutual information defined in parallel to (68)

$$\mathcal{I}(X;Y) = \frac{1}{2}\int_0^{\infty} \tau[(\mathsf{E}[X \,|\, Z_1] - \mathsf{E}[X \,|\, Z_1, Z_2])^2]\,\mathrm{d}\gamma + \frac{1}{2}\int_0^{\infty} \tau[(\mathsf{E}[Y \,|\, Z_2] - \mathsf{E}[Y \,|\, Z_1, Z_2])^2]\,\mathrm{d}\gamma \tag{100}$$

where

$$Z_1 = \sqrt{\gamma}\,X + W_1 \tag{101}$$

$$Z_2 = \sqrt{\gamma}\,Y + W_2 \tag{102}$$

and W_1 and W_2 are standard semicircular freely independent of each other and of (X, Y). With this definition of free mutual information, the free counterpart of (2) holds. Moreover, if X and W are freely independent semicircular with variances σ_X^2 and 1, respectively, then we obtain the sweet formula

$$\mathcal{I}(X; \sqrt{\mathsf{snr}}\,X + W) = \frac{1}{2}\log\left(1 + \sigma_X^2\right). \tag{103}$$

The "nonsemicircularity" of a measure P is then its free relative entropy with respect to the semicircle law with identical first and second moments. But it is easy to check that $\mathsf{E}_Q[X \,|\, Y]$ is a scalar multiple of Y when X is semicircular under Q. Appealing to the orthogonality principle, it can be checked that for any unit-variance measure absolutely continuous with respect to Lebesgue measure on $[-2, 2]$, we obtain

$$\mathsf{mse}_w(\gamma) = \frac{1}{1+\gamma} \tag{104}$$

and therefore, the nonsemicircularity of a unit variance X (the unnormalized case readily follows by scaling that does not affect the relative entropy) is given by

$$\mathcal{D}(X) = \frac{1}{2}\int_0^{\infty} \frac{1}{1+\gamma} - \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma. \tag{105}$$

This is consistent with Voiculescu's definition of free differential entropy (91) since

$$\chi(X) = \frac{1}{2}\log 2\pi e - \frac{1}{2}\int_0^{\infty} \mathcal{I}(X + \sqrt{t}W) - \frac{1}{1+t}\,\mathrm{d}t \tag{106}$$

$$= \frac{1}{2}\log 2\pi e - \frac{1}{2}\int_0^{\infty} \frac{1}{1+\gamma} - \mathsf{mmse}(\gamma)\,\mathrm{d}\gamma \tag{107}$$

$$= \frac{1}{2}\log 2\pi e - \mathcal{D}(X) \tag{108}$$

where (106) follows from (94) (e.g., [12]) and (107) follows from (98). And so (108) is indeed what it should be.

As an application of (99), we proceed to obtain the free relative entropy between two semicircular laws $D(w_a \,\|\, w_b)$. Using the orthogonality principle, it is straightforward to check that if $Y = \sqrt{\gamma}X + W$, then

$$\mathsf{E}_{w_b}[X \,|\, Y] = \frac{\sqrt{\gamma}}{\frac{4}{b^2} + \gamma}Y \tag{109}$$

and therefore

$$\mathsf{mse}_{w_b}(\gamma) = \frac{\frac{4a^2}{b^4} + \gamma}{\left(\frac{4}{b^2} + \gamma\right)^2}. \tag{110}$$

Then, the free relative entropy in nats is

$$\mathcal{D}(w_a \,\|\, w_b) = \frac{1}{2}\int_0^{\infty} \frac{\frac{4a^2}{b^4} + \gamma}{\left(\frac{4}{b^2} + \gamma\right)^2} - \frac{1}{\frac{4}{a^2} + \gamma}\,\mathrm{d}\gamma \tag{111}$$

$$= \frac{1}{2}\left(\frac{a^2}{b^2} - 1\right) + \log\frac{b}{a} \tag{112}$$

$$= D(\mathcal{N}(0, a) \,\|\, \mathcal{N}(0, b)) \tag{113}$$

where (113) is the (classical) relative entropy in (19).

A representation for the free relative entropy (99) in terms of the R-transforms of P and Q would be useful. However, this appears to be a challenging problem if only because a direct representation in terms of characteristic functions is unknown for the classical relative entropy.

VIII. Conclusion

The relative entropy $D(P \parallel Q)$ is known to play various important operational roles (see, e.g., [31]):

- the reliability of rejecting Q when P is true in Bayesian hypothesis testing (via the law of large numbers);
- the difficulty of impersonating P by Q (large deviations);
- the asymptotic excess rate for a lossless data compressor optimal for source Q, when the true source is P (discrete case);
- the capacity per unit cost of a binary-input channel with conditional output distributions P and Q, where the input letter that produces Q has zero cost.

To those roles we can add:

- the excess MSE (integrated over SNR) of an optimal estimator for Q when the true distribution is P and the desired random variable is observed in Gaussian noise.

In view of the well-known connections of relative entropy with other measures of distance between probability distributions, the new formula links those measures with mismatched estimation. For example, the Csiszár–Kemperman–Pinsker inequality [14, p. 58], together with (12) and Lemma 2 leads to the inequality

$$V^2(P,Q) \leq \int_0^\infty \mathsf{E}[(\mathsf{E}_Q[X \mid \sqrt{\gamma}X + N] - \mathsf{E}[X \mid \sqrt{\gamma}X + N])^2]\, \mathrm{d}\gamma \quad (114)$$

where $V(P,Q)$ is the variational distance between P and Q, and the bound is tight in the sense that the ratio of the left-hand side to the right-hand side cannot be upper bounded by any constant smaller than 1.

Another application of the main result in this paper has been found in [32] which shows that the relationship found in [1] between the errors in causal and noncausal continuous-time filtering carries over to the case of mismatched filters.

Estimation-theoretic representations of information-theoretic quantities have proven useful both showing new results and providing the simplest known proofs for old results. The fundamental role of those representations is further hinted by the fact that the projection representation of free relative entropy given in this paper is the only known general definition of this quantity.

Acknowledgment

The author is grateful for suggestions by an anonymous reviewer.

References

[1] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1283, Apr. 2005.

[2] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[3] S. Verdú and D. Guo, "A simple proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2165–2166, May 2006.

[4] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4295–4297, Sep. 2006.

[5] M. Zakai, "On mutual information, likelihood ratios, and estimation error for the additive Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3017–3024, Sep. 2005.

[6] J. Binia, "Divergence and minimum mean-square error in continuous-time additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1160–1163, Mar. 2006.

[7] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[8] D. Palomar and S. Verdú, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 453–470, Feb. 2007.

[9] E. Mayer-Wolf and M. Zakai, "Some relations between mutual information and estimation error in Wiener space," *Ann. Appl. Probab.*, vol. 17, no. 3, pp. 1102–1116, Jun. 2007.

[10] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.

[11] D. Guo, S. Shamai, and S. Verdú, "Estimation of non-Gaussian random variables in Gaussian noise: Properties of the MMSE," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 6–11, 2008, pp. 1083–1087.

[12] H. Yoshida, "Remarks on a semicircular perturbation of the free Fisher information," *Infinite Dimensional Anal., Quantum Probab. Related Topics*, vol. 11, no. 1, pp. 97–108, 2008.

[13] P. Dupuis and R. E. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: Wiley-Interscience, 1997.

[14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[15] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, pp. 79–86, 1951.

[16] M. Pinsker, S. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1877–1888, Nov. 1995.

[17] D. Guo, "Relative entropy and score function: New information estimation relationships through arbitrary additive perturbation," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun. 28–Jul. 3 2009, pp. 814–818.

[18] Y. Wu, D. Guo, and S. Verdu, "Low-SNR behavior of mutual information with Gaussian noise," 2010, preprint.

[19] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[20] G. Blower, *Random Matrices: High Dimensional Phenomena*, ser. London Mathematical Society Lecture Notes. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[21] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communication," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[22] P. Biane, "Free probability for probabilists," MSRI Rep. 1998-040, 1998 [Online]. Available: arXiv:math/9809193v1

[23] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory I," *Commun. Math. Phys.*, vol. 155, pp. 71–92, 1993.

[24] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory II," *Invent. Math.*, vol. 118, pp. 411–440, 1994.

[25] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory VI: Liberation and mutual free information," *Adv. Math.*, vol. 146, pp. 101–166, 1999.

[26] S. V. Szarek and D. Voiculescu, "Volumes of restricted Minkowski sums and free analogue of the entropy power inequality," *Commun. Math. Phys.*, vol. 178, pp. 563–570, 1996.

[27] D. Shlyakhtenko and H. Schultz, "Shannon's monotonicity problem for free and classical entropy," in *Proc. Nat. Acad. Sci.*, Sep. 19, 2007, vol. 104, pp. 15254–15258.

[28] P. Biane and R. Speicher, "Free diffusions, free entropy and free Fisher information," *Ann. Inst. H. Poincaré Probab. Stat.*, vol. 37, pp. 581–606, 2001.

[29] F. Hiai, M. Mizuo, and D. Petz, "Free relative entropy for measures and a corresponding perturbation theory," *J. Math. Soc. Jpn.*, vol. 54, pp. 679–718, 2002.

[30] B. Blackadar, *Operator Algebras: Theory of C*-Algebras and Von Neumann Algebras*. London, U.K.: Taylor & Francis, 2006.

[31] S. Verdú, "teaching it," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 28, 2007, DOI: 10.1109/ISIT.2007.4557074.

[32] T. Weissman, "The relationship between causal and non-causal mismatched estimation in continuous-time AWGN channels," *IEEE Trans. Inf. Theory*, 2010, to be published.

Sergio Verdú (S'80–M'84–SM'88–F'93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona, Barcelona, Spain, in 1980 and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1984.

Since 1984, he has been a member of the faculty of Princeton University, Princeton, NJ, where he is the Eugene Higgins Professor of Electrical Engineering.

Dr. Verdú is the recipient of the 2007 Claude E. Shannon Award and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from the IEEE Communications Society. He has also received paper awards from the Japanese Telecommunications Advancement Foundation and from Eurasip. He received the 2000 Frederick E. Terman Award from the American Society for Engineering Education for his book *Multiuser Detection* (Cambridge, U.K.: Cambridge Univ. Press, 1998). He served as President of the IEEE Information Theory Society in 1997. He is currently Editor-in-Chief of *Foundations and Trends in Communications and Information Theory.*