Probability in the Engineering and Informational Sciences, **30**, 2016, 125–140.
doi:10.1017/S0269964815000315

A DE BRUIJN'S IDENTITY FOR DEPENDENT RANDOM VARIABLES BASED ON COPULA THEORY

NAYEREH BAGHERI KHOOLENJANI AND MOHAMMAD HOSSEIN ALAMATSAZ
Department of Statistics, University of Isfahan, Isfahan 81746-73441, Iran
E-mail: n.b.khoolenjani@gmail.com; alamatho@sci.ui.ac.ir.

De Bruijn's identity shows a link between two fundamental concepts in information theory: entropy and Fisher information. In the literature, De Bruijn's identity has been stated under the assumption of independence between input signal and an additive noise. However, in the real world, the noise could be highly dependent on the main signal. The main aim of this paper is, firstly, to extend De bruijn's identity for signal-dependent noise channels and, secondly, to study how Stein and heat identities are related to De bruijn's identity. Thus, new versions of De Bruijn's identity are introduced in the case when input signal and additive noise are dependent and are jointly distributed according to Archimedean and Gaussian copulas. It is shown that in this generalized model, the derivatives of the differential entropy can be expressed in terms of a function of Fisher information. Our finding enfolds the conventional De Bruijn's identity as some remarks. Then, the equivalence among the new De Bruijn-type identity, Stein's identity and heat equation identity is established. The paper concludes with an application of the developed results in information theory.

1. INTRODUCTION

De Bruijn's identity was first discussed by Stam [25], and since then it has been widely used by many researchers. De Bruijn's identity shows a link between two fundamental concepts in information theory: entropy and Fisher information [11]. Due to its applications in estimation and turbo (iterative) decoding schemes, De Bruijn's identity has attracted much of interest. Verdu and his collaborators conducted a series of studies [9,19,21] to analyze the relationship between the input–output mutual information and the minimum mean-square error (MMSE), a result referred to as the I-MMSE identity for additive Gaussian noise channels.

Consider additive noise channels of the form

$$Y = X + \sqrt{t}Z, \tag{1}$$

where input signal X and additive noise Z are arbitrary random variables and the parameter t is assumed non-negative. When additive noise Z is an independent Gaussian random variable with zero mean and unit variance, the elegant algebraic connection between differential entropy of output signal Y and Fisher information is established through a relation

© Cambridge University Press 2015 0269-9648/15 $25.00

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

known as De Bruijn's identity. If we further assume that input signal X is also Gaussian, then both random variables X and Z are Gaussian and, thus, the output signal Y is also Gaussian. In this case, Stein's identity and De Bruijn's identity are equivalent, and they are also equivalent to the heat equation identity. Recently, an extension of De Bruijn's identity to non-Gaussian random variables is provided by Park, Serpedin, and Qaraqe [20].

However, all aforementioned results are based on the assumption of independence between input random variable X and additive noise Z. In the real world, there are signal-dependent noise channels in which the noise characteristics highly depend on the transmitted signal [8,17]. Application for such dependency appears also in radar and sonar systems [13]. Copula theory is one of the best methods used for modeling the dependency in conventional works. Copula is an elegant concept introduced by Sklar in order to find the link between a joint probability density function (pdf) and its marginal pdfs [24]. Copula functions have been extensively studied and a comprehensive discussion of their mathematical properties has been presented [10,18]. In this paper, similar forms of De Bruijn's identity will be derived in the case when input signal X and additive noise Z are jointly distributed according to Archimedean and Gaussian copulas. Considering dependent signal and noise, the derivatives of the differential entropy of output signal Y will be presented by a function of Fisher information. Then, the equivalence among De Bruijn-type identity mentioned in this paper, Stein's identity and heat equation identity will be proved.

The rest of this paper is organized as follows. In Section 2, we recall basic facts and definitions on copulas and the dependency problem. New versions of De Bruijn's identity for dependent random variables are introduced in Section 3. Relationships among new De Bruijn-type identity, Stein's identity and heat equation identity are also established in this section. In Section 4, an application of the proposed results in information theory is presented. Finally, conclusions are made in Section 5.

Let us first review the fundamental notions and basic definitions used in the paper.

Fisher information of a deterministic parameter t is defined as

$$J_t(Y) = \int_{-\infty}^{\infty} f_Y(y;t) \left(\frac{\partial}{\partial t} \log f_Y(y;t) \right)^2 dy. \tag{2}$$

This is a general definition of Fisher information in signal processing. Fisher information provides a lower bound, called the Cramér–Rao lower bound, for MSE of any unbiased estimator. Like other concepts, such as entropy and mutual information, in information theory, Fisher information also shows information about uncertainty. However, it is difficult to adopt the definition of Fisher information directly in information theory despite of the fact that it has been commonly used in Statistics. Instead, a more specific definition of Fisher information is proposed as follows.

If t is assumed to be a location parameter, differentiation with respect to y is equivalent to differentiation with respect to t. Therefore, the definition of Fisher information in (2) is changed as follows:

$$\begin{aligned} J(Y) &= \int_{-\infty}^{\infty} f_Y(y;t) \left(\frac{\partial}{\partial y} \log f_Y(y;t) \right)^2 dy, \\ &= E_Y \left[\left(\frac{\partial}{\partial Y} \log f_Y(Y;t) \right)^2 \right]. \end{aligned} \tag{3}$$

The random variable $(\partial/\partial Y) \log f_Y(Y;t)$ is known as the score of the distribution f_Y.

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

If $\log f_Y(y;t)$ is twice differentiable w.r.t y, under certain regularity conditions (see [14]), then the Fisher information may also be written as

$$J(Y) = -E_Y\left[\frac{\partial^2}{\partial Y^2}\log f_Y(Y;t)\right]. \tag{4}$$

Differential entropy of a random variable Y, $h(Y)$, is defined as

$$h(Y) = -\int_{-\infty}^{\infty} f_Y(y;t)\log f_Y(y;t)dy, \tag{5}$$

where t is a deterministic parameter in the pdf f_Y (see [15]).

Entropy power of a random variable Y, $N(Y)$, is defined as

$$N(Y) = \frac{1}{2\pi e}\exp(2h(Y)). \tag{6}$$

Based on the definitions mentioned previously, three preliminary theorems – De Bruijn's, Stein's, and heat equation identities – are recalled next.

THEOREM 1.1 (De Bruijn's Identity [7]): *Let X be an arbitrary random variable with a finite second-order moment, and Z be independent normally distributed with zero mean and unit variance. Then, for the additive noise channel $Y = X + \sqrt{t}Z$ we have*

$$\frac{\partial}{\partial t}h(Y) = \frac{1}{2}J(Y). \tag{7}$$

PROOF: See [7]. ■

THEOREM 1.2 (Stein's Identity [12]): *Let a random variable Y be normally distributed with mean μ_y and variance σ_y^2. Then*

$$E_Y[r(Y)(Y-\mu_Y)] = \sigma_y^2 E_Y\left[\frac{d}{dY}r(Y)\right], \tag{8}$$

holds for any function $r(Y)$ for which the expectations in (8) exist.

PROOF: See [12]. ■

THEOREM 1.3 (Heat equation Identity [5]): *Let Y be normally distributed with mean μ and variance $1+t$. Assume that $g(y)$ is a twice continuously differentiable function, and both $g(y)$ and $|(d/dy)g(y)|$ are $O\left(e^{c|y|}\right)$ for some $0 \le c < \infty$. ($O(.)$ denotes the limiting behavior of a function, that is, $g(y) = O(q(y))$ if, and only if, there exist positive real numbers K and y^* such that $g(y) \le K|q(y)|$ for any y which is greater than y^*.) Then*

$$\frac{d}{dt}E_Y[g(Y)] = \frac{1}{2}E_Y\left[\frac{d^2}{dY^2}g(Y)\right]. \tag{9}$$

PROOF: See [5]. ■

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

2. COPULA THEORY

Copula theory is a fully developed mathematical theory for multivariate distribution analysis. Copula is a function that links univariate distribution functions to generate a multivariate distribution function and, thus, represents the dependency structure of random variables. In other words, copulas enable us to extract the dependence structure from joint distribution function of a set of random variables and, at the same time, to isolate the dependence structure from the univariate marginal behavior. This means that copulas allow us to study the dependence structure apart from the marginal distribution of each variable. In recent years, there has been a revival of copula in applications where the matter of dependency between random variables is of importance [3].

The foundation theorem for copula was introduced by Sklar [24] which states that for a given joint multivariate pdf and the relevant marginal pdfs, there exists a copula function that relates them. In a bivariate case, Sklar's Theorem states that: *If $F_{X,Y}$ is a joint cumulative distribution function (cdf) with margins F_X and F_Y, then there exists a function $C : [0,1]^2 \longrightarrow [0,1]$ such that:*

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)). \tag{10}$$

If F_X and F_Y are continuous, then C is unique; otherwise, C is uniquely determined on the (range of F_X) $\times$ (range of F_Y). Conversely, if C is a copula and F_X and F_Y are cdfs, then the function $F_{X,Y}$ defined by (10) is a joint cdf with margins F_X and F_Y. A thorough description of copula properties is found in [18].

For any copula function, there is a corresponding copula density function, which is the mixed partial derivative of function C, given by:

$$c(u,v) = \frac{f_{X,Y}(u,v)}{f_X(u) f_Y(v)}, \tag{11}$$

where $f_{X,Y}$, f_X and f_Y are the joint and marginal pdfs of X and Y, respectively. Equation (11) can be expressed in an equivalent and more suitable form:

$$f_{X,Y}(x,y) = c(u,v) f_X(x) f_Y(y), \tag{12}$$

where u, v are related to x, y through the marginal cdfs:

$$u = F_X(x), \quad v = F_Y(y). \tag{13}$$

There are two main families of copulas; namely the family of elliptical copulas (i.e. copulas following Elliptical distributions such as Laplacian, t-Student, etc.) and the family of Archimedean copulas. The most common elliptical copulas is Gaussian. The key advantage of an elliptical copula is that one can specify different levels of dependency between the margins. Archimedean copulas are popular because they are constructed easily and allow modeling the dependence in arbitrarily high dimensions with only one parameter, governing the strength of dependence. In this paper, we use Archimedean and Gaussian copulas to describe the dependencies between input signal and additive noise in the channel model (1).

The Gaussian copula is defined as

$$C_\rho(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)), \tag{14}$$

where ρ is the Gaussian copula parameter, Φ_ρ is the bivariate standard Gaussian distribution function and Φ^{-1} is the inverse of the univariate standard Gaussian distribution function.

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

The corresponding Gaussian copula density is

$$c(u,v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right] \times \exp\left(\frac{x^2}{2}\right) \exp\left(\frac{y^2}{2}\right), \tag{15}$$

where $-1 < \rho < 1$, $0 \le u, v \le 1$, $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$.

If a copula C_ψ has the form

$$C_\psi(u,v) = \psi(\psi^{-1}(u) + \psi^{-1}(v)), \tag{16}$$

where $\psi : R^+ \to [0,1]$ is a 2-monotone function such that $\psi(0) = 1$ and $\lim_{\alpha\to\infty} \psi(\alpha) = 0$, then it is called an Archimedean copula and ψ is said to be the generator function of this Archimedean copula. In this work, we concentrate on Archimedean copulas whose generator functions ψ are completely monotone. Then, an equivalent representation for (16) becomes

$$C_\psi(u,v) = \int_0^\infty G^\alpha(u) G^\alpha(v) dM_\psi(\alpha), \tag{17}$$

where $G(t) = \exp\{-\psi^{-1}(t)\}$ and M_ψ is a distribution function with Laplace Transform ψ. The corresponding Archimedean copula density is

$$c_\psi(u,v) = \int_0^\infty \left[\alpha g(u) G^{\alpha-1}(u)\right] . \left[\alpha g(v) G^{\alpha-1}(v)\right] dM_\psi(\alpha), \tag{18}$$

in which

$$g(t) = G'(t) = -\frac{\exp(-\psi^{-1}(t))}{\psi'(\psi^{-1}(t))}, \quad t \in (0,1).$$

3. MAIN RESULTS

In this section, the channel model (1) will be extended to allow for dependent random variables. In particular, the dependence between input signal X and additive noise Z is modeled by assuming Archimedean and Gaussian copula functions for the joint distribution of signal and noise. In each case, the relationships between the derivatives of the differential entropy of output signal Y and Fisher information are established.

THEOREM 3.1: *In channel model (1), let X be a r.v. with a finite second-order moment depending on the r.v. Z. Assume that Z is normally distributed with zero mean and unit variance and the dependent structure of X and Z can be described by the Archimedean copula (18). Then, under the following conditions:*

$$\lim_{y\to\pm\infty} y\sqrt{f_Y(y;t)} = 0, \tag{19}$$

$$\left|\frac{E_X[f_{Y|X}(y \mid X;t)]}{\sqrt{f_Y(y;t)}}\right| < \infty, \tag{20}$$

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

the first derivative of the differential entropy of output signal Y can be expressed as

$$\frac{\partial}{\partial t}h(Y) = \frac{1}{2}J(Y) - \frac{1}{2}E_Y\left\{E_{X|Y}\left[\frac{\partial}{\partial Y}\log f_Y(Y;t).\left(\frac{\partial}{\partial Y}\log g\left(\Phi\left(\frac{Y-X}{\sqrt{t}}\right)\right)\right.\right.\right.$$
$$\left.\left.\left. - \frac{\partial}{\partial Y}\log G\left(\Phi\left(\frac{Y-X}{\sqrt{t}}\right)\right)\right) \mid Y\right]\right\}$$
$$- \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty}\left(\frac{\partial}{\partial y}\log f_Y(y;t)\right).\frac{\partial}{\partial y}\log G\left(\Phi\left(\frac{y-x}{\sqrt{t}}\right)\right).$$
$$\times \alpha g_1(x;\alpha)g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right)dM_\psi(\alpha)dx\,dy. \tag{21}$$

with $g_1(.;\alpha) = \alpha.g(F_X(.)).f_X(.).G^{\alpha-1}(F_X(.))$ *and* $g_2(.;\alpha) = \alpha.g(\Phi(.)).\phi_Y(.).G^{\alpha-1}(\Phi(.))$.

Proof: When Z is normally distributed with zero mean and unit variance, the density of Y is given by

$$f_Y(y;t) = \int_{-\infty}^{\infty}\frac{1}{\sqrt{t}}f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right)dx, \tag{22}$$

where

$$f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right) = \int_{0}^{\infty}g_1(x;\alpha)g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right)dM_\psi(\alpha),$$

Therefore, we have

$$-\left(\frac{y-x}{t}\right)f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right)$$
$$= \frac{\partial}{\partial y}f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right) - \frac{\partial}{\partial y}\log G\left(\Phi\left(\frac{y-x}{\sqrt{t}}\right)\right)\int_{0}^{\infty}(\alpha-1).g_1(x;\alpha)g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right)$$
$$\times dM_\psi(\alpha) - \frac{\partial}{\partial y}\log g\left(\Phi\left(\frac{y-x}{\sqrt{t}}\right)\right)\int_{0}^{\infty}g_1(x;\alpha)g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right)dM_\psi(\alpha). \tag{23}$$

Also, between differentials of the pdf $f_{Y|X}(y \mid x;t)$ with respect to y and t, there exists a relation that can be established as follows: Since

$$\frac{\partial}{\partial t}g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right) = \left(-\frac{y-x}{2t}\right)\frac{\partial}{\partial y}g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right),$$

we have

$$\frac{\partial}{\partial t}f_{Y|X}(y \mid x;t) = \frac{\partial}{\partial t}\left(\frac{f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right)}{\sqrt{t}f_X(x)}\right),$$
$$= -\frac{1}{2t\sqrt{t}f_X(x)}f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right) + \frac{1}{\sqrt{t}f_X(x)}\frac{\partial}{\partial t}f_{X,Z}\left(x,\frac{y-x}{\sqrt{t}}\right),$$
$$= -\frac{1}{2t}\left\{\left(\frac{\partial}{\partial y}(y-x)\right)f_{Y|X}(y \mid x;t) + (y-x)\frac{\partial}{\partial y}f_{Y|X}(y \mid x;t)\right\},$$

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

and therefore,

$$\frac{\partial}{\partial t} f_{Y|X}(y \mid x; t) = -\frac{1}{2t} \frac{\partial}{\partial y} \left((y - x) f_{Y|X}(y \mid x; t) \right). \tag{24}$$

We will use the relationships (23) and (24) to calculate the derivative of the entropy of Y.

From (5), the first derivative of the differential entropy of output signal Y is

$$\frac{\partial}{\partial t} h(Y) = -\int_{-\infty}^{\infty} \log f_Y(y; t) \frac{\partial}{\partial t} f_Y(y; t) dy, \tag{25}$$

$$= -\int_{-\infty}^{\infty} \log f_Y(y; t) \frac{\partial}{\partial t} E_X[f_{Y|X}(y \mid X; t)] dy,$$

$$= -\int_{-\infty}^{\infty} \log f_Y(y; t) E_X \left[\frac{\partial}{\partial t} f_{Y|X}(y \mid X; t) \right] dy. \tag{26}$$

Using (24), (26) is expressed as

$$\frac{\partial}{\partial t} h(Y) = \frac{1}{2t} \int_{-\infty}^{\infty} \log f_Y(y; t) E_X \left[\frac{\partial}{\partial y} (y - X) f_{Y|X}(y \mid X; t) \right] dy,$$

$$= \frac{1}{2t} \int_{-\infty}^{\infty} \log f_Y(y; t) \frac{\partial}{\partial y} E_X[(y - X) f_{Y|X}(y \mid X; t)] dy. \tag{27}$$

Integrating by parts with respect to y and using (23), (27) can be rewritten as

$$\frac{\partial}{\partial t} h(Y) = -\frac{1}{2t} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial y} \log f_Y(y; t) \right) E_X[(y - X) f_{Y|X}(y \mid X; t)] dy,$$

$$= \frac{1}{2\sqrt{t}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial y} \log f_Y(y; t) \right) \left(-\frac{y - x}{t} \right) f_{X,Z} \left(x, \frac{y - x}{\sqrt{t}} \right) dx\, dy,$$

$$= \frac{1}{2\sqrt{t}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial y} \log f_Y(y; t) \right) \frac{\partial}{\partial y} f_{X,Z} \left(x, \frac{y - x}{\sqrt{t}} \right) dx\, dy$$

$$- \frac{1}{2\sqrt{t}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial y} \log f_Y(y; t) \right) \frac{\partial}{\partial y} \log g \left(\Phi \left(\frac{y - x}{\sqrt{t}} \right) \right) f_{X,Z} \left(x, \frac{y - x}{\sqrt{t}} \right) dx\, dy$$

$$+ \frac{1}{2\sqrt{t}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial y} \log f_Y(y; t) \right) \frac{\partial}{\partial y} \log G \left(\Phi \left(\frac{y - x}{\sqrt{t}} \right) \right) f_{X,Z} \left(x, \frac{y - x}{\sqrt{t}} \right) dx\, dy$$

$$- \frac{1}{2\sqrt{t}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{0}^{\infty} \left(\frac{\partial}{\partial y} \log f_Y(y; t) \right) . \frac{\partial}{\partial y} \log G \left(\Phi \left(\frac{y - x}{\sqrt{t}} \right) \right)$$

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

$$\times \alpha g_1(x;\alpha) g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right) dM_\psi(\alpha)dx\,dy,$$

$$= \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial y}\log f_Y(y;t)\right)\frac{\partial}{\partial y}E_X[f_{Y|X}(y\mid X;t)]dy \tag{28}$$

$$-\frac{1}{2}E_Y\left\{E_{X|Y}\left[\frac{\partial}{\partial Y}\log f_Y(y;t)\left(\frac{\partial}{\partial Y}\log g\left(\Phi\left(\frac{Y-X}{\sqrt{t}}\right)\right)\right.\right.\right.$$

$$\left.\left.\left.-\frac{\partial}{\partial Y}\log G\left(\Phi\left(\frac{Y-X}{\sqrt{t}}\right)\right)\right)\mid Y\right]\right\}$$

$$-\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty}\left(\frac{\partial}{\partial y}\log f_Y(y;t)\right).\frac{\partial}{\partial y}\log G\left(\Phi\left(\frac{y-x}{\sqrt{t}}\right)\right)$$

$$\times \alpha g_1(x;\alpha) g_2\left(\frac{y-x}{\sqrt{t}};\alpha\right) dM_\psi(\alpha)dx\,dy. \tag{29}$$

Again, integrating by parts with respect to y and using (20), the first term on the right-hand side of (28) can be expressed as

$$\frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial y}\log f_Y(y;t)\right)\frac{\partial}{\partial y}E_X[f_{Y|X}(y\mid X;t)]dy$$

$$= \frac{1}{2}\left(\frac{\partial}{\partial y}\log f_Y(y;t)\right)E_X[f_{Y|X}(y\mid X;t)]\Big|_{y=-\infty}^{\infty}$$

$$-\frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\partial^2\log f_Y(y;t)}{\partial y^2}\right)E_X[f_{Y|X}(y\mid X;t)]dy,$$

$$= \frac{1}{2}J(Y). \tag{30}$$

and the proof is completed. ■

Remark 3.2: Suppose the underlying Archimedean copula is generated by $\psi(\alpha) = \exp(-\alpha)$. In this case, the copula reduces to the independent copula and $\psi'(\psi^{-1}(\alpha)) = -\exp(-\psi^{-1}(\alpha))$. Thus, $g(\Phi((y-x)/\sqrt{t})) = 1$ and the third term in (23) becomes zero. Based on the definition, $M_\psi(\alpha)$ is the distribution function of a degenerate random variable α at 1. Hence, the second term in (23) is also equal to zero. Therefore, Theorem 3.1 reduces to the conventional De Bruijn's identity, that is, the renown Theorem 1.1 is a corollary of our Theorem 3.1.

THEOREM 3.3: *Consider the channel model (1) and let X be an arbitrary random variable with a finite second-order moment with cdf $F_X(x)$, and Z be a dependent normally distributed random variable with zero mean and unit variance. Assume that the dependence between X and Z is modeled by a Gaussian copula with the copula density (15). If*

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

$E_X[X\Phi^{-1}(F_X(X))]$ and $E_X[\Phi^{-1}(F_X(X))]^2$ exist, then:

$$\frac{\partial}{\partial t}h(Y) = \left(\frac{1-\rho^2}{2}\right) J(Y) + A_t, \tag{31}$$

where

$$A_t = \frac{\rho}{2t(1-\rho^2)} E_Y \left\{ \log f_Y(Y;t).E_{X|Y}\left[\Phi^{-1}(F_X(X))\left(\rho\Phi^{-1}(F_X(X)) - \frac{(Y-X)}{\sqrt{t}}\right) \mid Y\right]\right\}.$$

Proof: The density of Y is given by

$$f_Y(y;t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right) dx.$$

By using (12) and (15), we have

$$\begin{aligned} &f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right) \\ &\quad = \exp\left\{-\frac{1}{2(1-\rho^2)}\left[[\Phi^{-1}(F_X(x))]^2 - 2\rho\Phi^{-1}(F_X(x))\left(\frac{y-x}{\sqrt{t}}\right) + \left(\frac{y-x}{\sqrt{t}}\right)^2\right]\right\} \\ &\qquad \times \frac{1}{\sqrt{2\pi(1-\rho^2)}} f_X(x) \exp\left\{\frac{[\Phi^{-1}(F_X(x))]^2}{2}\right\}. \end{aligned} \tag{32}$$

Differentiating of $f_Y(y;t)$ with respect to t yields

$$\frac{\partial}{\partial t} f_Y(y;t) = \int_{-\infty}^{\infty} \frac{\partial}{\partial t}\left[\frac{1}{\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right)\right] dx, \tag{33}$$

where

$$\begin{aligned} &\frac{\partial}{\partial t}\left[\frac{1}{\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right)\right] \\ &\quad = \frac{1}{2t\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right)\left[\frac{(y-x)^2}{t(1-\rho^2)} - \frac{\rho\Phi^{-1}(F_X(x))(y-x)}{\sqrt{t}(1-\rho^2)} - 1\right]. \end{aligned}$$

On the other hand, we have

$$\frac{\partial}{\partial y} f_Y(y;t) = \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} \frac{\partial}{\partial y}\left[f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right)\right] dx, \tag{34}$$

$$= \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right)\left[\frac{\rho\Phi^{-1}(F_X(x))}{\sqrt{t}(1-\rho^2)} - \frac{y-x}{t(1-\rho^2)}\right] dx \tag{35}$$

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

and

$$\frac{\partial^2}{\partial y^2} f_Y(y;t) = \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} \frac{\partial^2}{\partial y^2} \left[f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right)\right] dx, \tag{36}$$

$$= \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right) \left[\left(\frac{\rho\Phi^{-1}(F_X(x))}{\sqrt{t}(1-\rho^2)} - \frac{y-x}{t(1-\rho^2)}\right)^2 - \frac{1}{t(1-\rho^2)}\right] dx. \tag{37}$$

Therefore,

$$\frac{\partial}{\partial t} f_Y(y;t) = \left(\frac{1-\rho^2}{2}\right) \frac{\partial^2}{\partial y^2} f_Y(y;t) - \int_{-\infty}^{\infty} \frac{1}{2\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right) \times \left[\frac{\rho^2[\Phi^{-1}(F_X(x))]^2}{t(1-\rho^2)} - \frac{\rho\Phi^{-1}(F_X(x))(y-x)}{t\sqrt{t}(1-\rho^2)}\right] dx. \tag{38}$$

Now, using (25) and (38), we obtain

$$\frac{\partial}{\partial t} h(Y) = -\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial t} f_Y(y;t)\right) \log f_Y(y;t) dy, \tag{39}$$

$$= -\left(\frac{1-\rho^2}{2}\right) \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial y^2} f_Y(y;t)\right) . \log f_Y(y;t) dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right) \times \left[\frac{\rho^2[\Phi^{-1}(F_X(x))]^2}{t(1-\rho^2)} - \frac{\rho\Phi^{-1}(F_X(x))(y-x)}{t\sqrt{t}(1-\rho^2)}\right] \log f_Y(y;t) dx\, dy. \tag{40}$$

The second term in (40) may be rewritten as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\sqrt{t}} f_{X,Z}\left(x, \frac{y-x}{\sqrt{t}}\right) \left[\frac{\rho^2[\Phi^{-1}(F_X(x))]^2}{t(1-\rho^2)} - \frac{\rho\Phi^{-1}(F_X(x))(y-x)}{t\sqrt{t}(1-\rho^2)}\right] \log f_Y(y;t) dx\, dy,$$

$$= \frac{\rho}{2t(1-\rho^2)} E_Y \left\{ \log f_Y(Y;t) . E_{X|Y}\left[\Phi^{-1}(F_X(X)) \left(\rho\Phi^{-1} F_X(X) - \frac{(Y-X)}{\sqrt{t}}\right) \mid Y\right]\right\}.$$

The first term in (40) can be integrated by parts to obtain

$$-\left(\frac{1-\rho^2}{2}\right) \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial y^2} f_Y(y;t)\right) . \log f_Y(y;t) dy$$

$$= -\left(\frac{1-\rho^2}{2}\right) \left[\left(\frac{\partial}{\partial y} f_Y(y;t)\right) . \log f_Y(y;t)\right]_{-\infty}^{\infty} + \left(\frac{1-\rho^2}{2}\right) \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial y} f_Y(y;t)\right]^2 \frac{1}{f_Y(y;t)} dy. \tag{41}$$

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

So the proof will be completed if we show that the first term in (41) vanishes. We can rewrite the first term as

$$\left(\frac{\partial}{\partial y} f_Y(y;t)\right) \log f_Y(y;t) = \left[\frac{\frac{\partial}{\partial y} f_Y(y;t)}{\sqrt{f_Y(y;t)}}\right] \left[2\sqrt{f_Y(y;t)} \log \sqrt{f_Y(y;t)}\right]. \tag{42}$$

The square of the first factor integrates to the Fisher information and hence must be bounded as $y \to \pm\infty$. The second factor goes to zero since $x \log x \to 0$ as $x \to 0$ and $f_Y(y;t) \to 0$ as $y \to \pm\infty$. Hence, the first term in (41) goes to 0 and, thus, we have the assertion. ■

Remark 3.4: In the above proofs, we have exchanged integration and differentiation in (23), (25)–(27), (33), (34), (36) and (39). Strict justification of these exchanges requires application of the Bounded Convergence Theorem.

Remark 3.5: Setting $\rho = 0$ in (31), we shall arrive at the conventional De Bruijn's identity (Theorem 1.1) for the case where input signal X and additive noise Z are two independent random variables, as a corollary.

Remark 3.6: Most signal processing algorithms are designed and based on prior knowledge of signal and noise characteristics. It is therefore natural (and useful) to view them as Bayesian inference strategies. In this case, A_t may be viewed as a posterior mean.

As was shown in Theorem 3.3, the first derivative of the differential entropy of output signal Y is expressed in terms of a function of the Fisher information. In what follows, we consider special cases of this result with additive normally distributed noise.

COROLLARY 3.7: *Consider the channel model (1) and let X and Z be two dependent random variables distributed according to a Gaussian copula with standard normal marginals. Then, the following identity holds:*

$$\frac{\partial}{\partial t} h(Y) = \frac{1}{2}\left(1 + \frac{\rho}{\sqrt{t}}\right) J(Y). \tag{43}$$

PROOF: Since the random variables X and Z are white Gaussian, the output signal Y is also Gaussian with zero mean and variance $1 + t + 2\rho\sqrt{t}$. Thus, by using (39), $(\partial/\partial t)h(Y)$ can be expressed as

$$\begin{aligned}
\frac{\partial}{\partial t} h(Y) &= -\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial t} f_Y(y;t)\right) \log f_Y(y;t) dy, \\
&= -\int_{-\infty}^{\infty} \left[\frac{-\left(1 + \frac{\rho}{\sqrt{t}}\right)}{2(1 + t + 2\rho\sqrt{t})} + \frac{y^2\left(1 + \frac{\rho}{\sqrt{t}}\right)}{2(1 + t + 2\rho\sqrt{t})^2}\right] f_Y(y;t) \\
&\quad \times \left\{-\frac{1}{2}\log(2\pi(1 + t + 2\rho\sqrt{t})) - \frac{y^2}{2(1 + t + 2\rho\sqrt{t})}\right\} dy,
\end{aligned}$$

which simplifies to

$$\frac{\partial}{\partial t} h(Y) = \frac{(1 + \frac{\rho}{\sqrt{t}})}{2(1 + t + 2\rho\sqrt{t})}.$$

Since for the normal distribution, Fisher information is the inverse of its variance, the proof is completed. ■

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

Now, using Relation (43), we consider the second derivative of the differential entropy of output signal Y. One interesting property of the second derivative of the differential entropy is that it can also be expressed as a function of the Fisher information.

COROLLARY 3.8: *Consider the channel model (1) and let X and Z be two dependent random variables distributed according to a Gaussian copula with standard normal marginals. Then, the second derivative of the differential entropy is expressed as*

$$\frac{\partial^2}{\partial t^2}h(Y) = -\frac{1}{2}J(Y)\left[\frac{\rho}{2t\sqrt{t}} + \left(1+\frac{\rho}{\sqrt{t}}\right)^2 J(Y)\right]. \tag{44}$$

In Section 1, Theorems 1.1–1.3 share the analogy of presenting an identity between expectations of certain functions and expectation of their derivatives. If De Bruijn's identity happen to be equivalent to Stein's identity, it would also be equivalent to the heat equation identity, because the equivalence between the heat equation identity and Stein's identity has been already proved to hold in [5]. Given the channel (1), Park et al. [20] showed the equivalence among the three identities, provided that the signal X and additive noise Z are two independent normal random variables. However, as mentioned in Section 1, the noise characteristics could be highly dependent on the main signal. Thus, signal-dependent noise is a more realistic assumption in practical applications such as in image transmission [1], radar [2] , and wireless communications [17]. Now, in what follows, we shall establish the existence of similar relationships among Stein's identity, heat equation identity and our extended De Bruijn-type identity mentioned in Relation (43). A corollary of this result is the fact that the whole set of applications established via Stein's identity could be transferred and proved into the realm of this new version of De Bruijn's identity.

THEOREM 3.9: *Consider the channel model (1) and let X and Z be two dependent random variables distributed according to a Gaussian copula with the copula density (15) with standard normal marginals. If we define the function $r(Y)$ in (8) as*

$$r(y;t) = -\frac{\partial}{\partial y}\log f_Y(y;t), \tag{45}$$

then Stein's identity is equivalent to De Bruijn-type identity in (43). Moreover, if we define $g(Y)$ in (9) as

$$g(y;t) = -\log f_Y(y;t), \tag{46}$$

then De Bruijn-type identity in (43) is also equivalent to the heat equation identity.

PROOF: Similar to the proof of Theorem 3 in [5], the equivalence is proved by showing that each identity is derived from the other one. First, in Stein's identity, the required function is defined as:

$$r(y;t) = -\frac{\partial}{\partial y}\log f_Y(y;t).$$

Then, the identity in (43) is derived from Stein's identity as follows:

$$\begin{aligned}\frac{1}{2}\left(1+\frac{\rho}{\sqrt{t}}\right)J(Y) &= \frac{1}{2}\left(1+\frac{\rho}{\sqrt{t}}\right)E_Y\left[\frac{\partial}{\partial Y}r(Y;t)\right],\\ &= \frac{1}{2\sigma_Y^2}\left(1+\frac{\rho}{\sqrt{t}}\right)E_Y\left[r(Y;t)(Y-\mu_Y)\right].\end{aligned} \tag{47}$$

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

Therefore, since $\mu_Y = 0$, the right-hand side of (47) takes the form

$$\frac{1}{2\sigma_Y^2}\left(1+\frac{\rho}{\sqrt{t}}\right)E_Y\left[r(Y;t)(Y-\mu_Y)\right] = \frac{\left(1+\frac{\rho}{\sqrt{t}}\right)}{2(1+t+2\rho\sqrt{t})}E_Y\left[Yr(Y;t)\right],$$
$$= \frac{\partial}{\partial t}h(Y)E_Y\left[Yr(Y;t)\right]. \tag{48}$$

But by integration by parts, we have $E_Y\left[Yr(Y;t)\right] = 1$, which yields the result.

Second, we show that Stein's identity is derived from De Bruijn-type identity in (43). In our case, as shown above, the left-hand side of (8) is equal to 1. So, we ought to prove that $E_Y[(\partial/\partial Y)r(Y,t)] = 1/\sigma_Y^2$. This is true using the identity in (43), because we have

$$E_Y\left[\frac{\partial}{\partial Y}r(Y;t)\right] = J(Y),$$
$$= \frac{2\frac{\partial}{\partial t}h(Y)}{1+\frac{\rho}{\sqrt{t}}},$$
$$= \frac{1}{1+t+2\rho\sqrt{t}},$$
$$= \frac{1}{\sigma_Y^2}. \tag{49}$$

Therefore, the equivalence between the new version of De Bruijn's identity and Stein's identity is proved. Since the equivalence between Stein's identity and the heat equation identity is proved in [5] in general, the proof is completed. ■

4. APPLICATION

As mentioned in [9] and [22], De Bruijn's identity has wide applications in a variety of areas such as information theory, estimation theory and so on. Similarly, De Bruijn-type identities mentioned in this paper can be adopted in many applications. Here, we reveal one of its applications in information theory.

The entropy power inequality (EPI) is a basic and powerful tool in information theory. It has been used very effectively in proofs of converse results for various channel coding theorems; indeed this was Shannon's original motivation in introducing it in [23]. Another famous application to Gaussian broadcast channels was developed by Bergmans [4], and this was significantly generalized in an important (and much more recent) paper of Weingarten, Steinberg, and Shamai, [26] to the Gaussian multi-input multi-output (MIMO) setting. Numerous variants and generalizations of the EPI such as a classical EPI [25], Costa's EPI [6], and an extremal inequality [16] were proposed by several different authors. Here, we will present a new version of the classical EPI in the case when the summands are dependent. The EPI for dependent variables finds its application in entropy-based blind source separation of dependent components.

THEOREM 4.1: *Consider the channel model (1) and let X and Z be two dependent random variables distributed according to a Gaussian copula with the copula density (15). Then,*

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

under the following condition:

$$(J(Y)+A_t^*)^2 \leq -\frac{\frac{\partial}{\partial t}(J(Y)+A_t^*)}{1-\rho^2}, \tag{50}$$

where $A_t^* = 2A_t/(1-\rho^2)$, *we have*

$$\frac{\partial^2}{\partial t^2} N(X+\sqrt{t}Z) \leq 0, \tag{51}$$

that is, $N(X+\sqrt{t}Z)$ *is a concave function of* t.

Proof: The second derivative of the entropy power $N(Y)$ is

$$\begin{aligned}\frac{\partial^2}{\partial t^2}N(Y) &= 2\frac{\partial}{\partial t}N(Y)\frac{\partial}{\partial t}h(Y) + 2N(Y)\frac{\partial^2}{\partial t^2}h(Y),\\ &= 2N(Y)\left\{2\left(\frac{\partial}{\partial t}h(Y)\right)^2 + \frac{\partial^2}{\partial t^2}h(Y)\right\},\end{aligned} \tag{52}$$

where $Y = X+\sqrt{t}Z$. Since $N(Y) \geq 0$, proving the inequality in (51) is equivalent to proving that:

$$2\left(\frac{\partial}{\partial t}h(Y)\right)^2 \leq -\frac{\partial^2}{\partial t^2}h(Y). \tag{53}$$

Using the identity (31), the inequality in (53) is expressed as

$$\left(J(Y)+\frac{2A_t}{1-\rho^2}\right)^2 \leq -\frac{\partial}{\partial t}\left(J(Y)+\frac{2A_t}{1-\rho^2}\right)\cdot\frac{1}{1-\rho^2}, \tag{54}$$

and thus the proof is completed. ■

Remark 4.2: Consider an additive white Gaussian noise channel. Given the channel model (1), let X and Z be two dependent random variables jointly distributed according to a Gaussian copula with standard normal marginals. Then,

$$J(Y) = \frac{1}{1+t+2\rho\sqrt{t}},$$

and

$$A_t^* = \frac{\rho}{1-\rho^2}\left(\rho+\frac{1}{\sqrt{t}}\right),$$

and the condition (50) is satisfied. Therefore, for $\rho \geq 0$, $N(X+\sqrt{t}Z)$ is a concave function of t.

5. CONCLUSION

In this paper, we have introduced new versions of De Bruijn's identity considering the statistical dependence of the input signal and additive noise based on the copula theory. First, it was proved that the derivatives of the differential entropy can be expressed in terms of Fisher information. Second, the equivalence among Stein's identity, heat equation

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

identity and De Bruijn-type identity was established under certain conditions. Then, an application of our main results in information theory was provided. It is worth mentioning that we believe that a broad range of problems (in probability, decision theory, Bayesian statistics and graph theory) could be considered as potential applications of the proposed results.

Acknowledgement

The first author (NBK) is grateful to the office of Graduate Studies of the University of Isfahan for their support. The authors are also grateful to anonymous reviewers whose valuable comments and suggestions led to improvement of the manuscript.

References

1. Acito, N., Diani, M. & Corsini, G. (2011). Signal-dependent noise modeling and model parameter estimation in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing* 49(8): 2957–2971.
2. Aiazzi, B., Alparone, B., Baronti, S. & Garzelli, A. (2003). Coherence estimation from multilook incoherent SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 41(11): 2531–2539.
3. Arias-Nicolas, J.P., Fernandez-Ponce, J.M., Luque-Calvo, P. & Suarez-Llorens, A. (2005). Multivariate dispersion order and the notion of copula applied to the multivariate t-distribution. *Probability in the Engineering and Informational Sciences* 19: 363–375.
4. Bergmans, P.P. (1974). A simple converse for broadcast channels with additive white Gaussian noise, *IEEE Transactions on Information Theory* IT-20: 279–280.
5. Brown, L., DasGupta, A., Haff, L.R. & Strawderman, W.E. (2006). The heat equation and Stein's identity: Connections, applications. *Journal of Statistical Planning and Inference* 136: 2254–2278.
6. Costa, M.H. (1985). A new entropy power inequality. *IEEE Transactions on Information Theory* 31(6): 751–760.
7. Cover, T.M. & Thomas, J.A. (2006). *Elements of Information Theory*. 2nd edn., New York: Wiley.
8. Gholizadeh, M.H., Amindavar, H. & Ritcey, J.A. (2013). Analytic Nakagami fading parameter estimation in dependent noise channel using copula. *EURASIP Journal on Advances in Signal Processing* 2013(1):129.
9. Guo, D., Shamai (Shitz), S. & Verdu, S., (2005). Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory* 51(4): 1261–1282.
10. Joe, H. (1997). Multivariate Models and Dependence Concepts. *In Monographs on Statistics and Applied Probability*. vol. 73. London: Chapman & Hall.
11. Johnson, O. (2004). *Information Theory and the Central Limit Theorem*. London, UK: Imperial College Press.
12. Kattumannil, S.K. (2009). On Stein's identity and its applications. *Statistics and Probability Letters* 79: 1444–1449.
13. Kay, S. (2009). Waveform design for multistatic radar detection. *IEEE Transactions on Aerospace and Electronic Systems* 45(3): 1153–1166.
14. Lehmann, E.L. & Casella, G. (1998). *Theory of Point Estimation*. 2nd edn., New York: Springer.
15. Li, X. & Zhang, S. (2011). Some new results on Renyi entropy of residual life and inactivity time. *Probability in the Engineering and Informational Sciences* 25: 237–250.
16. Liu, T. & Viswanath, P. (2007). An extremal inequality motivated by multiterminal information-theoretic problems. *IEEE Transactions on Information Theory* 53(5): 1839–1851.
17. Moon, J. & Park, J. (2001). Pattern-dependent noise prediction in signal-dependent noise. *IEEE Journal on Selected Areas in Communications* 19(4): 730–743.
18. Nelsen, R.B. (2006). *An Introduction to Copulas*. 2nd edn., New York: Springer.
19. Palomar D.P. & Verdu, S. (2006). Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory* 52(1): 141–154.
20. Park, S., Serpedin, E. & Qaraqe, K. (2012). On the equivalence between stein and De Bruijn identities. *IEEE Transactions on Information Theory* 58(12): 7045–7067.
21. Payaro M. & Palomar D.P. (2009). Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels. *IEEE Transactions on Information Theory* 55(8): 3613–3628.

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press

22. Rioul, O. (2011). Information theoretic proofs of entropy power inequalities. *IEEE Transactions on Information Theory* 57(1): 33–55.
23. Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656.
24. Sklar, A. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8: 229–231.
25. Stam, A.J. (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* 2(2): 101–112.
26. Weingarten, H., Steinberg, Y. & Shamai, S. (2006). The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Transactions on Information Theory* 52(9): 3936–3964.

https://doi.org/10.1017/S0269964815000315 Published online by Cambridge University Press