On Transformations of Random Vectors

Jeffrey A. Fessler

COMMUNICATIONS & SIGNAL PROCESSING LABORATORY
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, Michigan 48109-2122

Aug. 1998, Revised September 3, 1998

Technical Report No. 314
Approved for public release; distribution unlimited.

On Transformations of Random Vectors

Jeffrey A. Fessler

4240 EECS, University of Michigan, Ann Arbor, MI 48109-2122

email: fessler@umich.edu, phone: 734-763-1434, fax: 734-764-8041

Technical Report # 314
Communications and Signal Processing Laboratory
Dept. of Electrical Engineering and Computer Science
The University of Michigan

Abstract

This technical report treats some technical considerations related to the probability density function of a function of a random vector.

I. Introduction

Let $X \in R^n$ be a continuous random vector with known pdf $f_X(x)$. In many problems it is necessary to find the pdf $f_Y(y)$ of a new random vector Y defined as a function $Y = g(X)$ of X, where $g : R^n \to R^n$.

Many textbooks on probability and random variables state the following equality:

$$f_Y(y) = f_X(g^{-1}(y))\,|J(y)|, \tag{1}$$

where J is the Jacobian of $g^{-1}(y)$.

There is considerable variation in how precisely the textbook authors state the conditions for the above equality. Most books do state the condition that g be one-to-one (and hence invertible). However, the stated conditions on differentiability vary widely.

Many engineering books make *no mention* of the need for g^{-1} to be differentiable, *e.g.* [1–7]. Many books assume that g^{-1} is *globally* differentiable *e.g.* [8–13], but this condition is too restrictive in some applications. Some books [14–16] assume that $g : \mathcal{S} \to g(\mathcal{S})$ is one-to-one and differentiable on some open set $\mathcal{S} \subseteq R^n$, and that the pdf of X vanishes (is zero) outside of $\mathcal{S}$. This is reasonably general, but still inapplicable to problems where, for example, X has a Gaussian pdf and $\mathcal{S}$ is a proper subset of R^n.

A more general requirement is to assume that $P[X \in \mathcal{S}] = 1$, for which the condition that $f_X(x)$ vanishes outside $\mathcal{S}$ is a special case. Hoel, Port, and Stone [17] provide such a theorem without proof. Bickel and Doksum [18] provide a proof of the transformation formula under the condition $P[X \in \mathcal{S}] = 1$, but the proof is not entirely rigorous since the integrals given can cover points outside $\mathcal{S}$ where the Jacobian need not exist. This technical report provides a rigorous proof of (1), properly handling the technical details of the set $\mathcal{S}$.

This work was motivated by [19], in which a transformation function arises that is differentiable except on a set of hyperplanes of Lebesgue measure zero.

II. Theory

The following is simply Theorem 17.2 of [14], included for convenience.

Theorem 1: Let $h : \mathcal{V} \to h(\mathcal{V})$ be a one-to-one mapping of an open set $\mathcal{V}$ onto an open set $h(\mathcal{V})$. Suppose that (on $\mathcal{V}$) h is continuous and that h has continuous partial derivatives h_{ij} with Jacobian $J(y) \triangleq \det[h_{ij}(y)]$. Then for $\mathcal{A} \subseteq \mathcal{V}$, for any nonnegative function f

$$\int_{\mathcal{A}} f(h(y))|J(y)|\,dy = \int_{h(\mathcal{A})} f(x)\,dx. \tag{2}$$

Work supported in part by NIH grants CA-60711 and CA-54362 and the Whitaker Foundation.

The following Theorem is a generalization of (20.20) in [14]. Standard treatments *e.g.* [13, p. 143] assume that the transformation function is globally differentiable. Our generalization allows for a (measure zero) set where the Jacobian is undefined.

Theorem 2: Let $g : R^n \to R^n$ be one-to-one and assume that $h = g^{-1}$ is continuous. Assume that on an open set $\mathcal{V} \subseteq R^n$ h is continuously differentiable with Jacobian $J(y)$. Define $J_0 : R^n \to \mathbb{R}$ by

$$J_0(y) = \begin{cases} J(y), & y \in \mathcal{V} \\ 0, & y \in \mathcal{V}^c, \end{cases} \tag{3}$$

where $\mathcal{V}^c$ is the set complement (in R^n) of $\mathcal{V}$.

Define $\mathcal{U} = h(\mathcal{V})$. Suppose random vector X has pdf $f_X(x)$ (with respect to Lebesgue measure) with nonzero mass in $\mathcal{U}^c$, *i.e.* $P[X \in \mathcal{U}^c] = \int_{\mathcal{U}^c} f_X(x)\,dx = 0$. Then the pdf of $Y = g(X)$ is given by

$$\boxed{f_Y(y) = f_X(g^{-1}(y))\,|J_0(y)| = \begin{cases} f_X(g^{-1}(y))\,|J(y)|, & y \in \mathcal{V} \\ 0, & y \in \mathcal{V}^c. \end{cases}} \tag{4}$$

Proof:

For (measurable) $\mathcal{B} \subseteq R^n$

$$0 \le P[g(X) \in \mathcal{B} \cap \mathcal{V}^c] \le P[g(X) \in \mathcal{V}^c] = P[X \in g^{-1}(\mathcal{V}^c)] = P[X \in \mathcal{U}^c] = 0.$$

Thus $P[g(X) \in \mathcal{B} \cap \mathcal{V}^c] = 0$, so

$$\begin{aligned} P[g(X) \in \mathcal{B}] &= P[g(X) \in \mathcal{B} \cap \mathcal{V}] + P[g(X) \in \mathcal{B} \cap \mathcal{V}^c] \\ &= P[X \in h(\mathcal{B} \cap \mathcal{V})] = \int_{h(\mathcal{B}\cap\mathcal{V})} f_X(x)\,dx = \int_{\mathcal{B}\cap\mathcal{V}} f_X(h(y))\,|J(y)|\;dy \end{aligned}$$

by Theorem 1, which applies since $\mathcal{B} \cap \mathcal{V} \subseteq \mathcal{V}$. (The set $\mathcal{U}$ is open since by assumption $\mathcal{V}$ is open and h is continuous.) Thus by (3):

$$P[g(X) \in \mathcal{B}] = \int_{\mathcal{B}\cap\mathcal{V}} f_X(h(y))\,|J_0(y)|\;dy = \int_{\mathcal{B}} f_X(h(y))\,|J_0(y)|\;dy - \int_{\mathcal{B}\cap\mathcal{V}^c} f_X(h(y))\,|J_0(y)|\;dy,$$

since $\mathcal{B}$ is the union of the disjoint sets $\mathcal{B} \cap \mathcal{V}$ and $\mathcal{B} \cap \mathcal{V}^c$. The second integral above is zero since $|J_0(y)|$ is zero for $y \in \mathcal{V}^c$ by (3). Thus

$$P[g(X) \in \mathcal{B}] = \int_{\mathcal{B}} f_X(h(y))\,|J_0(y)|\;dy,$$

for $\mathcal{B} \subseteq R^n$, proving that (4) is a pdf of $g(X)$. □

REFERENCES

[1] R. G. Brown and P. Y. C. Hwang, *Introduction to random signals and applied Kalman filtering*, Wiley, New York, 3 edition, 1997.
[2] C. W. Helstrom, *Probability and stochastic processes for engineers*, Macmillan, New York, 1991.
[3] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, Wiley, New York, 1994.
[4] A. Papoulis, *Probability, random variables, and stochastic processes*, McGraw-Hill, New York, 2 edition, 1984.
[5] Y. Viniotis, *Probability and random processes for electrical engineers*, McGraw-Hill, Boston, 1998.
[6] R. Walpole and R. H. Myers, *Probability and statistics for engineers and scientists*, Prentice Hall, New York, 5 edition, 1993.
[7] R. E. Ziemer, *Elements of engineering probability and statistics*, Prentice Hall, NJ, 1997.
[8] J. A. nón and V. Chandrasekar, *Introduction to probability and random processes*, McGraw-Hill, New York, 1997.
[9] A. Leon-Garcia, *Probability and random processes for electrical engineering*, Addison Wesley, New York, 2 edition, 1994.
[10] P. Z. Peebles and Jr., *Probability, random variables, and random signal principles*, McGraw-Hill, New York, 3 edition, 1993.
[11] J. A. Rice, *Mathematical statistics and data analysis*, Brooks/Cole, Monterey, Calif, 1988.
[12] S. Ross, *A first course in probability*, Macmillan, Englewood Cliffs, NJ, 4 edition, 1994.
[13] H. Stark and J. W. Woods, *Probability, random processes, and estimation theory for engineers*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
[14] R. Billingsley, *Probability and measure*, Wiley, New York, 2 edition, 1986.
[15] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the theory of statistics*, McGraw Hill, New York, 3 edition, 1974.
[16] D. Stirzaker, *Elementary probability*, Cambridge University Press, Cambridge, 1994.
[17] P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to probability theory*, Houghton Mifflin, Boston, 1971.
[18] M. statistics, *P J Bickel K A Doksum*, Holden-Day, Oakland, CA, 1977.
[19] J. A. Fessler and H. Erdoğan, "Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise," *IEEE Tr. Im. Proc.*, 1998. almost Submitted.

Transformation of Random Vectors: Continued

White Random Vectors

A random vector $\mathbf{X}$ in an underlying sample space $\mathbf{S} \subseteq \mathbf{R}^n$ is said to be *white* if the components of the random vector are statistically independent of each other, i.e., the joint n^{th}–order PDF of the components is separable, i.e,

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

As a consequence, the components are also pairwise independent of each other, i.e.,

$$f_{X_i,X_j}(x_i, x_j) = f_{X_i}(x_i) f_{X_j}(x_j), \quad i \neq j.$$

Since the components are pairwise independent they are also pairwise uncorrelated. This implies that the covariance matrix associated with a white random vector is diagonal, i.e.,

$$\mathbf{C}_x = \mathbf{\Lambda} \equiv diag(\lambda_1, \lambda_2, \ldots \lambda_n).$$

If the uncorrelated random variables are further identically distributed, i.e., possess identical statistical characteristics then the random vector is said to be a *i.i.d* random vector. In this case the covariance matrix becomes an identity matrix:

$$\mathbf{C}_x = \sigma^2 \mathbf{I},$$

where σ is the common standard deviation of the components. A random vector is said to be *weakly white* if the components are just statistically uncorrelated. Henceforth when we refer to a white random vector it will mean white in the weak sense.

In certain digital communication applications such as Vitterbi decoding, used to remove *inter symbol interference* (ISI) introduced by a channel with finite memory, a white channel noise model is required. In some cases this may not be true and it will be necessary to whiten the noise before the application of the Vitterbi algorithm.

Whitening of a Random Vector:

If we transform a random vector $\mathbf{X}$ using a linear transformation $\mathbf{A}$ then the mean vector and covariance matrix of the transformed random vector $\mathbf{Y}$ are given by:

$$\mathbf{m}_y = \mathbf{A}\mathbf{m}_x, \quad \mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T$$

In linear algebra terminology we seek a transformation $\mathbf{A}$ that transforms the random vector $\mathbf{X}$ into another random vector $\mathbf{Y} = \mathbf{A}\mathbf{X}$ that has a diagonal covariance matrix $\mathbf{C}_y$, i.e., we require:

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T = \mathbf{D},$$

where $\mathbf{D}$ is some diagonal, positive semi–definite matrix. The solution to this whitening problem is to choose the linear transformation $\mathbf{A}$ to be equal to $\mathbf{V}^T$, where $\mathbf{V}$ is the unitary matrix of eigenvectors of the covariance matrix $\mathbf{C}_x$, i.e.,

$$\mathbf{Y} = \mathbf{V}^T\mathbf{X} \Longleftrightarrow \mathbf{C}_y = \mathbf{V}^T\mathbf{C}_x\mathbf{V} = \mathbf{\Lambda}.$$

This process of whitening the random vector $\mathbf{X}$ using the eigenvectors of its covariance matrix is also called as the *Karhunen Loeve Transform* (KLT).

IID Whitening:

If we require that the transformed vector be not only white but also i.i.d then we require:

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T = \sigma^2\mathbf{I}.$$

The solution to this i.i.d whitening problem is given by:

$$\mathbf{A} = \sigma\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T,$$

where $\mathbf{\Lambda}^{\frac{1}{2}}$ refers to the self–adjoint matrix square–root of $\mathbf{\Lambda}$, $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues of $\mathbf{C}_x$ and $\mathbf{V}$ is the unitary matrix of eigenvectors of $\mathbf{C}_x$. The fact that the covariance of the transformed vector $\mathbf{Y}$ is identity can be verified via:

$$\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T = \sigma^2\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{C}_x\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}} = \sigma^2\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Lambda}\mathbf{\Lambda}^{-\frac{1}{2}} = \sigma^2\mathbf{I}.$$

Geometrical Interpretation of Whitening

A unitary transformation $\mathbf{A}$ is a rotation in $\mathbf{R}^n$ if the coloums are orthonormal, i.e .,

$$\mathbf{A}^T\mathbf{A} = \mathbf{I} \Longleftrightarrow \mathbf{A}^{-1} = \mathbf{A}^T.$$

For a deterministic vector, the length of a vector remains invariant under a unitary operation, i.e.,

$$||\mathbf{y}|| = ||A\mathbf{x}|| = ||\mathbf{x}||.$$

In the space of random vectors $\mathbf{X} \in \mathbf{S} \subseteq \mathbf{R}^n$ this implies that the average power is preserved, i.e.,

$$< \mathbf{Y}, \mathbf{Y} >=< \mathbf{A}\mathbf{X}, \mathbf{A}\mathbf{X} >= E\{\mathbf{X}^T\mathbf{A}^T\mathbf{A}\mathbf{X}\} =< \mathbf{X}, \mathbf{X} >= P_{\text{ave}}.$$

The KLT in the first case, i.e., the weakly white case therefore corresponds to a rotation of the random vector $\mathbf{X}$. In the second case, i.e., the i.i.d white case the KLT corresponds to first a rotation by $\mathbf{V}^T$ followed by inverse scaling of the axes by $\sigma\mathbf{\Lambda}^{-\frac{1}{2}}$. In a sense this is analogous to the process of reducing a general quadratic form to the standard quadratic form that we encounter in coordinate geometry.