

Application of transformations in parametric inference.

Naomi Brownstein and Marianna Pensky
Department of Mathematics
University of Central Florida
Orlando, FL 32816

Abstract

The objective of the present paper is to provide a simple approach to statistical inference using the method of transformations of variables. We demonstrate performance of this powerful tool on examples of constructions of various estimation procedures, hypothesis testing, Bayes analysis and statistical inference for the stress-strength systems. We argue that the tool of transformations not only should be used more widely in statistical research but should become a routine part of calculus-based courses of statistics. Finally, we provide sample problems for such a course as well as possible undergraduate reserach projects which utilize transformations of variables.

Key words: Transformations of variables, estimation, testing, stress-strength model, Bayesian inference.

1 Introduction

Transformations of random variables have been a standard tool in statistical inference. They have been used for solutions of a variety of statistical problems such as nonparametric density estimation, nonparametric regression, analysis of time series, construction of equivariant estimators and de-noising (see, e.g., [9], [11], [12], [14], [18], [19] and [21]). However, with a number of complex applications of transformation tools, one very simple application of transformations has been largely overlooked by the statistical community. This paper discusses utilization of a well-known fact that, when some of the parameters are treated as known, the majority of familiar probability distributions are just transformations of one another. Therefore, results of parametric statistical inference for one family of pdfs can be reproduced without much work for another family.

How many distributions are there in statistics? There are several dozens of them in the two volumes of “Continuous Univariate Distributions ” [6] and [7]. One will find similar lists in many other books, the goal of which is statistical inference of a certain type such as, for example, in [20]. In the majority of texts, estimators, tests and other statistical procedures are usually constructed for each distribution family separately which leads to a great deal of calculations and sometimes errors. Unifying those procedures will result in saving hours of work.

The objective of the present paper is to provide a simple approach to statistical inferences using the method of transformations of variables. We demonstrate performance of this powerful tool on examples of constructions of various estimation procedures, hypothesis testing, Bayes analysis and

statistical inference for the stress-strength systems. To the best of authors' knowledge, standard courses in statistics usually mention transformations only in relation to the maximum likelihood estimation (MLE) procedure. We argue that the tool of transformations not only should be used more widely in statistical research but should become a routine part of calculus-based courses of statistics. The material of the present paper can be used as such a supplement. For this purpose, we provide several sample homework problems and undergraduate research projects. Note that, although in what follows we consider only the case of a one-dimensional random variable, the theory has obvious extension to the case of random vectors. However, making this generalization in the present paper will unnecessarily complicate the presentation.

Consider a random variable X with the pdf $f(x|\boldsymbol{\theta})$, where parameter $\boldsymbol{\theta}$ is a scalar or a vector. Suppose also that there exist a random variable ξ , a monotone function u and a one-to-one transformation ν such that $X = u(\xi)$, where the pdf $g(\xi|\boldsymbol{\tau})$ of ξ has a different parameterization from X , namely,

$$g(\xi|\boldsymbol{\tau}) = f(u(\xi)|\nu(\boldsymbol{\tau})) |u'(\xi)|, \quad \boldsymbol{\theta} = \nu(\boldsymbol{\tau}). \quad (1)$$

Denoting $u^{-1} = v$ and $\nu^{-1} = \eta$, we rewrite (1) as

$$f(x|\boldsymbol{\theta}) = g(v(x)|\eta(\boldsymbol{\theta})) |v'(x)|, \quad \boldsymbol{\tau} = \eta(\boldsymbol{\theta}). \quad (2)$$

Now, let $g(\xi|\boldsymbol{\tau})$ be a popular distribution family, so that all sorts of statistical results are available. The objective of the present paper is to show how these results can be re-formulated for $f(x|\boldsymbol{\theta})$. Notice that the correspondence (1) is quite common but, however, is not used to a full extent. For example, Table 1 contains distributions which can be obtained from one or two parameter exponential distributions by appropriate re-parameterization. Our goal is however not to explore all possible correspondences of this sort but to provide few examples which will illustrate the general idea. The theory presented herein can be easily extended to many more kinds of statistical procedures and various other families of distributions.

Critics of this paper can adequately remark that some of the results listed here can be obtained in a general form for, say, one or two-parameter exponential families. The goal, however, is not to provide such a generalization but to supply a simple and yet powerful methodological tool to modify statistical procedures. The scale and location-scale families of exponential distributions are used here only as an example. In fact, techniques described below can be used for a distribution family which does not have a sufficient statistic.

The rest of the paper is organized as follows. In Section 2 we consider basic statistical inference for $f(x|\boldsymbol{\theta})$ based on our knowledge of inference for $g(\xi|\boldsymbol{\tau})$: sufficient statistics, maximum likelihood estimators (MLE) and uniform minimum variance unbiased estimators (UMVUE), Bayes estimators, interval estimators and likelihood ratio tests. Section 3 deals with more sophisticated statistical inference such as analysis of stress-strength systems and elicitation of noninformative priors. In both Sections 2 and 3, we provide examples of statistical procedures which can be obtained with no effort by using known results and applying transformation of variables suggested in this paper. Section 4 presents several sample homework problems and undergraduate research projects. Section 5 concludes the paper with the discussion.

Table 1. Transformations of random variables.

$f(x \boldsymbol{\theta})$	Transforms	$g(\xi \boldsymbol{\tau})$
Weibull distribution: $f(x \sigma) = \frac{\alpha}{\sigma^\alpha} x^{\alpha-1} \exp\left\{-\left(\frac{x}{\sigma}\right)^\alpha\right\}$, α known, $x > 0$.	$v(x) = x^\alpha$ $\boldsymbol{\theta} = \sigma$ $\boldsymbol{\tau} = \lambda$ $\eta(\sigma) = \sigma^{-\alpha}$	One-parameter exponential distribution: $g(\xi \lambda) = \lambda \exp\{-\lambda\xi\}$, $\xi > 0$.
Rayleigh distribution: $f(x \sigma) = \frac{x}{\sigma} \exp\left\{-\frac{x^2}{2\sigma}\right\}$, $x > 0$.	$v(x) = x^2/2$ $\boldsymbol{\theta} = \sigma$ $\boldsymbol{\tau} = \lambda$ $\eta(\sigma) = \sigma^{-1}$	One-parameter exponential distribution: $g(\xi \lambda) = \lambda \exp\{-\lambda\xi\}$, $\xi > 0$.
Burr type X distribution: $f(x \sigma) = 2\sigma x e^{-x^2} (1 - e^{-x^2})^{\sigma-1}$, $x > 0$.	$v(x) = -\ln(1 - e^{-x^2})$ $\boldsymbol{\theta} = \sigma$ $\boldsymbol{\tau} = \lambda$ $\eta(\sigma) = \sigma$	One-parameter exponential distribution: $g(\xi \lambda) = \lambda \exp\{-\lambda\xi\}$, $\xi > 0$.
Burr type XII distribution: $f(x \sigma) = \frac{\alpha x^{\alpha-1}}{\sigma(1+x^\alpha)^{\frac{1+\alpha}{\sigma}}}$, α known, $x > 0$.	$v(x) = \ln(1 + x^\alpha)$ $\boldsymbol{\theta} = \sigma$ $\boldsymbol{\tau} = \lambda$ $\eta(\sigma) = \sigma^{-1}$	One-parameter exponential distribution: $g(\xi \lambda) = \lambda \exp\{-\lambda\xi\}$, $\xi > 0$.
Extreme Value distribution: $f(x \sigma) = \frac{1}{\sigma} \exp\left\{x - \frac{\exp x - 1}{\sigma}\right\}$, $x > 0$.	$v(x) = \exp\{x\} - 1$ $\boldsymbol{\theta} = \sigma$ $\boldsymbol{\tau} = \lambda$ $\eta(\sigma) = \sigma^{-1}$	One-parameter exponential distribution: $g(\xi \lambda) = \lambda \exp\{-\lambda\xi\}$, $\xi > 0$.
One-parameter Beta distribution: $f(x \sigma) = \sigma x^{\sigma-1} I(0 < x < 1)$	$v(x) = -\ln(x)$ $\boldsymbol{\theta} = \sigma$ $\boldsymbol{\tau} = \lambda$ $\eta(\sigma) = \sigma$	One-parameter exponential distribution: $g(\xi \lambda) = \lambda \exp\{-\lambda\xi\}$, $\xi > 0$.
Pareto distribution: $f(x \sigma, \rho) = \frac{\sigma \rho^\sigma}{x^{\sigma+1}} I(x \geq \rho)$.	$v(x) = \ln x$ $\boldsymbol{\theta} = (\sigma, \rho)$ $\boldsymbol{\tau} = (\lambda, \mu)$ $\eta(\sigma, \rho) = (\sigma, \ln \rho)$	Two-parameter exponential distribution: $g(\xi \mu, \lambda) = \lambda \exp\{-\lambda(\xi - \mu)\}$, $\xi \geq \mu$.
Power distribution: $f(x \sigma, \rho) = \frac{\sigma}{\rho^\sigma} x^{\sigma-1} I(0 < x < \rho)$.	$v(x) = \ln(1/x)$ $\boldsymbol{\theta} = (\sigma, \rho)$ $\boldsymbol{\tau} = (\lambda, \mu)$ $\eta(\sigma, \rho) = (\sigma, \ln(1/\rho))$	Two-parameter exponential distribution: $g(\xi \mu, \lambda) = \lambda \exp\{-\lambda(\xi - \mu)\}$, $\xi \geq \mu$.

2 Basic Statistical Inference Procedures

In this section we discuss construction of the basic statistical procedure for $f(x|\boldsymbol{\theta})$ based on the relevant knowledge for $g(\xi|\boldsymbol{\tau})$. We provide the statements followed by few examples of their applications. The proofs of all statements are very elementary and can be obtained by simple change of variables. Note that, while Theorem 2 on the use of transformations for construction of the MLEs is the common knowledge, the rest of the statements, though very useful, rarely appear in statistical texts.

Let the relation between $f(x|\boldsymbol{\theta})$ and $g(\xi|\boldsymbol{\tau})$ be given by (1) and (2) and $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ be the independent and identically distributed (i.i.d.) samples from those pdfs with $X_i = u(\xi_i), i = 1, \dots, n$. Then the parametric family $f(x|\boldsymbol{\theta})$ inherits all useful properties of the family $g(\xi|\boldsymbol{\tau})$. To simplify the notations, in what follows $v(\mathbf{X}) = (v(X_1), v(X_2), \dots, v(X_n))$.

Theorem 1 (Sufficient statistics) *Let $T(\boldsymbol{\xi}) = T(\xi_1, \xi_2, \dots, \xi_n)$ be a scalar or vector valued sufficient statistic for the family of pdfs $g(\xi|\boldsymbol{\tau})$. Then $T^*(\mathbf{X}) = T(v(\mathbf{X}))$ is a sufficient statistic for the family of pdfs $f(x|\boldsymbol{\theta})$.*

Theorem 2 (MLE) *Let $U(\boldsymbol{\xi}) = U(\xi_1, \xi_2, \dots, \xi_n)$ be a MLE of $\boldsymbol{\tau}$ based on the sample $\boldsymbol{\xi}$. Then $U^*(\mathbf{X}) = \nu(U(v(\mathbf{X})))$ is the MLE of $\boldsymbol{\theta}$ based on observations \mathbf{X} . If, moreover, $T = T(\boldsymbol{\xi})$ is a sufficient statistic for $\boldsymbol{\tau}$, so that the MLE of $\boldsymbol{\tau}$ has the form $U(\boldsymbol{\xi}) = V(T)$, then the MLE of $\boldsymbol{\theta}$ is $U^*(\mathbf{X}) = \nu(V(T^*))$ where $T^* = T^*(\mathbf{X})$ is defined in Theorem 1.*

Theorem 3 (UMVUE) *Let $h(\boldsymbol{\tau})$ be a function of $\boldsymbol{\tau}$ and let $V(T)$ be the UMVUE of $h(\boldsymbol{\tau})$ based on the sufficient statistic $T = T(\boldsymbol{\xi})$. Then $V(T^*)$ is the UMVUE of $h^*(\boldsymbol{\theta}) = h(\eta(\boldsymbol{\theta}))$ based on the sufficient statistic $T^*(\mathbf{X})$.*

Corollary 1 (UMVUE) *Let $\Psi(\xi_1, \dots, \xi_s)$ be a function of observations ξ_1, \dots, ξ_s with expectation $E_{\boldsymbol{\tau}}\Psi(\xi_1, \dots, \xi_s)$ over the pdf $\prod_{j=1}^s g(\xi_j|\boldsymbol{\tau})$. If $V(T)$ is the UMVUE of $E_{\boldsymbol{\tau}}\Psi(\xi_1, \dots, \xi_s)$, then $V(T^*)$ is the UMVUE of $E_{\boldsymbol{\theta}}\Psi(v(X_1), \dots, v(X_s))$.*

Theorem 4 (Bayes) *Let $U(\boldsymbol{\xi}) = U(\xi_1, \xi_2, \dots, \xi_n)$ be the Bayes estimator of $h(\boldsymbol{\tau})$ based on the sample $\boldsymbol{\xi}$ and the prior pdf $\pi(\boldsymbol{\tau})$. Then, $U^*(\mathbf{X}) = U(v(\mathbf{X}))$ is the Bayes estimator of $h^*(\boldsymbol{\theta}) = h(\eta(\boldsymbol{\theta}))$ based on the sample \mathbf{X} and prior pdf $\pi(\eta(\boldsymbol{\theta})) |J_{\eta}(\boldsymbol{\theta})|$ where $|J_{\eta}(\boldsymbol{\theta})|$ is the Jacobian of the transformation $\eta(\boldsymbol{\theta})$. Moreover, if $g(\boldsymbol{\tau}|\boldsymbol{\xi})$ is the posterior pdf of $\boldsymbol{\tau}$ given $\boldsymbol{\xi}$ based on the prior pdf $\pi(\boldsymbol{\tau})$ then $g^*(\boldsymbol{\theta}|\mathbf{X}) = g(\eta(\boldsymbol{\theta})|v(\mathbf{X})) |J_{\eta}(\boldsymbol{\theta})|$ is the posterior pdf of $\boldsymbol{\theta}$ corresponding to the prior pdf $\pi(\eta(\boldsymbol{\theta})) |J_{\eta}(\boldsymbol{\theta})|$.*

Theorems 2–4 and Corollary 1 refer to construction of various types of point estimators. However, interval estimation and hypothesis testing procedures can be modified in a similar way. We give few examples below.

Theorem 5 (Interval estimation) *Let $\varphi(\boldsymbol{\tau})$ be a parametric function of interest and let $(\underline{U}(\boldsymbol{\xi}), \overline{U}(\boldsymbol{\xi}))$ be an interval estimator of $\varphi(\boldsymbol{\tau})$ corresponding to the confidence level $(1 - \gamma)$, i.e.*

$$P(\underline{U}(\boldsymbol{\xi}) < \varphi(\boldsymbol{\tau}) < \overline{U}(\boldsymbol{\xi})) \geq 1 - \gamma.$$

Let $\underline{U}^(\mathbf{X}) = \underline{U}(v(\mathbf{X}))$ and $\overline{U}^*(\mathbf{X}) = \overline{U}(v(\mathbf{X}))$. Then $(\underline{U}^*(\mathbf{X}), \overline{U}^*(\mathbf{X}))$ is an interval estimator of $\varphi(\eta(\boldsymbol{\theta}))$ corresponding to the confidence level $(1 - \gamma)$.*

Theorem 6 (Likelihood ratio test) Let $U(\boldsymbol{\xi}) = U(\xi_1, \xi_2, \dots, \xi_n)$ be a likelihood ratio test (LRT) statistic for testing hypothesis $H_0 : \boldsymbol{\tau} \in \Omega_0$ versus $H_1 : \boldsymbol{\tau} \in \Omega_0^C$. Consider sets Θ_0 and Θ_0^C such that

$$\boldsymbol{\tau} \in \Omega_0 \Leftrightarrow \boldsymbol{\theta} = \nu(\boldsymbol{\tau}) \in \Theta_0, \quad \boldsymbol{\tau} \in \Omega_0^C \Leftrightarrow \boldsymbol{\theta} = \nu(\boldsymbol{\tau}) \in \Theta_0^C.$$

Then $U^*(\mathbf{X}) = U(\nu(\mathbf{X}))$ is a LRT statistic for testing hypothesis $H_0^* : \boldsymbol{\theta} \in \Theta_0$ versus $H_1^* : \boldsymbol{\theta} \in \Theta_0^C$. Moreover, if $U(\boldsymbol{\xi}) > C_\gamma$ is a level γ test for H_0 , then $U^*(\mathbf{X}) > C_\gamma$ is a level γ test for H_0^* .

Example 1. Consider the task of finding the UMVUE of $\ln(\sigma)$ given a sample \mathbf{X} from the Weibull distribution (see Table 1). Using the fact that $\sigma = \nu(\lambda) = \lambda^{-1/\alpha}$ we obtain $\ln(\sigma) = -1/\alpha \ln(\lambda)$. Then, the UMVUE of $\ln(\lambda)$ based on the sample $\boldsymbol{\xi}$ from the one-parameter exponential distribution is of the form (see [20], page 359)

$$V(T) = \ln(T) - \Psi(n) \tag{3}$$

where $T = \bar{\boldsymbol{\xi}}$ and $\Psi(n) =$ is the Euler's psi-function $\Psi(x) = \frac{d}{dx} \ln \Gamma(x)$. Theorems 1 and 3 with $v(x) = x^\alpha$ imply that the UMVUE for $\ln(\sigma)$ is $V(T^*) = -1/\alpha [\ln(T^*) - \Psi(n)]$ where $T^* = n^{-1} \sum_{i=1}^n X_i^\alpha$.

Note that the formula for UMVUE of $\ln(\sigma)$ is not listed in the most comprehensive existing collection of UMVUEs [20] in the table for the Weibull distribution nor are estimators for many other functions of σ that can be easily obtained by using our simple technique. Moreover, the method of transformations of random variables is not even listed among half a dozen techniques suggested for derivation of UMVUEs. If the authors introduced this very simple idea in the book, the tables would be more comprehensive, and the book would be much shorter.

Example 2. The tool of transformations is very useful for solution of simple problems in various statistics courses. For example, consider problem 8.17 in [4] which we re-formulate in notations of the present paper. When (X_1, \dots, X_n) and (Y_1, \dots, Y_m) are samples from one-parameter beta distributions (see Table 1) with parameters σ_1 and σ_2 , respectively, the objective is to construct the LRT for the hypothesis $H_0^* : \sigma_1 = \sigma_2 = \sigma$ versus $H_1^* : \sigma_1 \neq \sigma_2$, to show that this test can be based on statistics

$$T^* = \frac{\sum \log X_i}{\sum \log X_i + \sum \log Y_i} \tag{4}$$

and to find the distribution of T^* when H_0 is true. For this purpose, turn to problem 8.6 in the same section of [4] where samples (X_1, \dots, X_n) and (Y_1, \dots, Y_m) have one-parameter exponential distributions with parameters λ_1 and λ_2 , respectively, and the goal is to construct the same LRT for the hypothesis $H_0 : \lambda_1 = \lambda_2 = \lambda$ versus $H_1 : \lambda_1 \neq \lambda_2$, to show that this test can be based on statistics

$$T = \frac{\sum X_i}{\sum X_i + \sum Y_i}$$

and to find the distribution of T when H_0 is true. Solving problem 8.6 we discover that the rejection region for H_0 is of the form $T^n(1-T)^m \leq c$ and, since when H_0 is true, $\sum X_i$ and $\sum Y_i$ have Gamma (n, λ) and Gamma (m, λ) distributions respectively, T has the Beta (n, m) distribution. Now, use the fact (see Table 1) that whenever X has a one-parameter Beta distribution, $\ln X$ has the one-parameter exponential distribution with identical parameterization. Hence, T^* given by (4) has the same Beta (n, m) distribution as T and the rejection region for H_0^* is $(T^*)^n [1 - (T^*)]^m \leq c$. Similarly, the LRTs can be constructed for all distributions which can be obtained by transformations of variables from the one-parameter exponential distribution.

3 More Sophisticated Applications

Applications of transformations are not limited to elementary statistical inference procedures. In what follows, we consider two models where transformations of variables can provide a final result with very little effort: the stress-strength system and derivation of noninformative priors. However, we want to point out that applications of transformation of variables are by no means limited to these two cases.

3.1 The Stress-Strength Problem

Consider estimation of probability $R = P(X_1 < X_2)$ on the basis of observations $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$, $i = 1, 2$. If X_1 is the stress applied to the device and X_2 is its strength, then R is the probability of successful operation of the device, and the problem of drawing statistical inference about R is called the *stress-strength problem*.

Let, as before, random variables X_i , $i = 1, 2$, have pdfs $f_i(x_i|\boldsymbol{\theta}_i)$ and $X_i = u(\xi_i)$ where u is a monotone function. Then, similarly to (1), ξ_i , $i = 1, 2$, have the pdfs $g_i(\xi_i|\boldsymbol{\tau}_i)$ of the forms

$$g_i(\xi_i|\boldsymbol{\tau}_i) = f_i(u(\xi_i)|\nu(\boldsymbol{\tau}_i)) |u'(\xi_i)|, \quad \boldsymbol{\theta}_i = \nu(\boldsymbol{\tau}_i). \quad (5)$$

where ν is a one-to-one transformation. As before, denote $u^{-1} = v$ and $\nu^{-1} = \eta$ and observe that if the function u is monotonically increasing, then $R^* = P(X_1 < X_2) = P(u(\xi_1) < u(\xi_2)) = P(\xi_1 < \xi_2) = R$, so that R remains invariant. If u is a decreasing function then, similarly, $R^* = P(\xi_1 > \xi_2) = 1 - R$. Hence, if $g_i(\xi_i|\boldsymbol{\tau}_i)$, $i = 1, 2$, are familiar distribution families, and statistical inference procedures are available for these families, results can be easily reformulated for the families $f_i(x_i|\boldsymbol{\theta}_i)$, $i = 1, 2$. The example below is one of many where authors waste a lot of effort instead of simply writing the final result.

Example 3. Consider, for example, the situation described in [15] where both X_1 and X_2 have the Burr type X distributions (see Table 1). The objective of the authors is to find the MLE and the UMVUE of R and also to develop lower confidence bounds for R . The authors carry out all details of statistical inference while the only thing they need is to exploit known results for estimation of R in the case when ξ_i , $i = 1, 2$, have one-parameter exponential distributions. In particular, they could just copy the estimators of Tong [17] with the only difference that sufficient statistics $T_i = \sum_{j=1}^{n_i} \xi_j^{(i)}$ should be replaced by $T_i^* = -\sum_{j=1}^{n_i} \ln(1 - \exp(-X_j^{(i)}))$, $i = 1, 2$, and $R^* = 1 - R$ (see Theorems 1 and 3). Similarly, the confidence intervals for R^* can be obtained using known results by Enis and Geisser [5].

3.2 Elicitation of Noninformative Priors

Bayesian inference can be carried out with various kinds of prior distributions. However, in order to minimize subjectivity contributed by the choice of prior, one may choose *objective* Bayesian analysis, which is based on *noninformative* prior pdfs. There are various kinds of noninformative priors, e.g., Jeffreys' prior, reference prior and matching prior (for the review of noninformative priors see, e.g., [2], [3] or [13]). Evaluation of those priors is usually not an easy task, therefore many authors were involved in their derivation. To help researchers in their use of objective Bayesian methods, Yang and Berger [22] compiled the catalog of noninformative priors. It is evident from the catalog, that if

the pdfs of two distributions are related to each other as (1) and (2), and $\pi(\boldsymbol{\tau})$ is the Jeffreys or reference noninformative prior for $g(\xi|\boldsymbol{\tau})$, then $\pi^*(\boldsymbol{\theta}) = \pi(\eta(\boldsymbol{\theta}) | J_\eta(\boldsymbol{\theta}))$ is the noninformative prior of the same kind for $f(x|\boldsymbol{\theta})$ (compare with Theorem 4). Hence, one can find noninformative priors for less common distribution families easily using transformations without re-doing all the calculations.

Example 4. Consider the problem of finding noninformative priors for parameter σ of Burr type X distribution and parameters (σ, ϱ) of the Pareto distribution (see Table 1). Note that neither of the families is a location-scale parameter family, and the catalog does not list the expressions for the Jeffreys and reference priors in this case. However, it states that in the case of the one-parameter exponential distribution the Jeffreys and the reference priors are of the form $1/\lambda$ while in the case of the two-parameter exponential distribution the Jeffreys prior is 1 while the reference prior is $1/\lambda$. Hence, the Jeffreys and the reference priors for the parameter σ of Burr type X distribution are both equal to $1/\sigma$, while in the case of the Pareto distribution the Jeffreys prior is equal to 1 while the reference prior is $1/\sigma$.

Example 5. In [8], Kim et al. developed the Jeffreys' and the reference priors for the stress-strength system in the case of the Burr type X distribution. However, since the Burr type X distribution can be obtained from the one-parameter exponential distribution by setting $\xi = -\ln(1 - e^{-X^2})$, the results could have just been copied from [16] where noninformative priors are obtained in the case of the one-parameter exponential distribution. Namely, the Jeffreys' and the reference priors which both are equal to $1/(\lambda_1\lambda_2)$ where λ_1 and λ_2 are the scale parameters of the exponential distributions. Thompson and Basu [16] also derived the corresponding posterior pdf. Hence, the Jeffreys' and the reference priors coincide in the case of the Burr type X distribution and are equal to $1/(\sigma_1\sigma_2)$. The corresponding posterior can be found from [16] by replacing x with $-\ln(1 - e^{-x^2})$ in the posterior and multiplying it by xe^{-x^2} .

4 Incorporating Transformations into Statistics Courses

Even in the absence of specific materials in the current textbooks, transformations of variables can be incorporated into an advanced course on statistical inference. While teaching topics in probability, an instructor should explain how one distribution can be obtained from another via transformations of variables (using, for example, Table 1 of this paper). Then each of the standard topics in statistical inference can be supplemented by examples and problems where transformations are applied. For this purpose, of course, Theorems 1–6 of this paper should be presented in class at appropriate moments. Below we list few sample problems which can be assigned for homework and more advanced projects that are suitable for undergraduate research.

Problem 1. The MLE of the parameters λ and μ in the two-parameter exponential distribution are \bar{X}^{-1} and $X_{(1)}$, respectively. Using transformations of variables and Table 1, a) find the MLE's of the parameters ϱ and σ of the Pareto distribution; b) find the MLE's of the parameters ϱ and σ of the Power distribution.

Problem 2. It is well known that \bar{X} is the UMVUE of $1/\lambda$ where λ is the scale parameter in one-parameter exponential distribution (see Table 1). Using transformations of variables and Table 1, a) find the UMVUE of the parameter σ of the Rayleigh distribution; b) find the UMVUE

of the parameter σ of the Extreme Value distribution.

Can the UMVUE for the parameter σ of the Weibull distribution be constructed using transformation techniques if parameter α of the Weibull distribution is known and $\alpha \neq 1$?

Problem 3. a) Find the LRT of

$$H_0 : \mu \leq 0 \quad \text{versus} \quad H_1 : \mu > 0$$

based on sample X_1, \dots, X_n from a population with the pdf $f(x|\lambda, \mu) = \lambda \exp(-\lambda(x - \mu)) I(X \geq \mu)$ where both μ and λ are unknown (problem 8.7(a) of [4]).

b) Using transformations of variables and Table 1 find the LRT of

$$H_0 : \varrho \leq 0 \quad \text{versus} \quad H_1 : \varrho > 0$$

based on a sample from a two-parameter Pareto distribution where both parameters ϱ and σ are unknown.

Sample project 1. Monograph [20] contains UMVUEs for a variety of distribution families and parametric functions. However, as Example 1 of this paper shows, the lists of UMVUEs are not complete. For example, Section A5 contains unbiased estimators of 105 functions of λ in one-parameter exponential family (some of those estimators are also suitable for functions of the scale parameter of the gamma distribution with the known shape parameter). However, Section A7 presents unbiased estimators for only 31 functions of the scale parameter of the Weibull distribution and Section A8 lists unbiased estimators for only 32 functions of the parameter of the Rayleigh distribution. Using Table 1 and technique of transformations, one can expand the collection of UMVUEs in Sections A7 and A8 using the UMVUEs in Section A5 of [20]. Some of these estimators will be totally new and can be published.

Sample project 2. Monograph [10] discusses a variety of techniques for point and interval estimation of $R = P(X_1 < X_2)$ on the basis of observations $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$, $i = 1, 2$, when X_1 and X_2 belong to familiar distribution families. In the case when X_1 and X_2 are independent normally distributed random variables, the authors describe construction of the MLEs, the UMVUEs, the Bayes estimators and exact and asymptotic confidence intervals for R (see Sections 3.1.1, 3.2.1, 3.3.1, 4.1.2 and 4.2.1, respectively). However, for some distribution families the inference for R has never been pursued. One of the examples is the alpha distribution $\alpha(\mu, \tau)$ with the pdf

$$f(x|\mu, \tau) = \frac{1}{\sqrt{2\pi\tau}x^2} \exp \left\{ -\frac{1}{2\tau} \left(\frac{1}{\mu} - \frac{1}{x} \right)^2 \right\}.$$

Let $X_1 \sim \alpha(\mu_1, \tau_1)$ and $X_2 \sim \alpha(\mu_2, \tau_2)$ be independent random variables. After a careful examination one can notice that the alpha distribution can be obtained from the normal distribution by mere transformation of variables and parameters. Hence, using transformation of variables, one can derive the MLE, the UMVUE, the Bayes estimator and an exact and an asymptotic confidence intervals for R in the case of the alpha distribution family (see also Theorems 2.7–2.9 in [10] which describe applications of transformations in the case of the stress-strength problem). The resulting estimators will be entirely new and of interest to researchers in the field of reliability.

5 Discussion

In the present paper, we exploit the tool of transformations of random variables to obtain various statistical inference procedures with minimal effort. We discuss a few very straightforward examples of applications of the transformations which can be appreciated even by a student who took an upper level undergraduate/lower level graduate statistics course. Moreover, using simple techniques suggested above, one can easily expand the list of UMVUEs provided in [20] and construct statistical procedures for less familiar distribution families using known results. In addition, we discuss few more sophisticated applications of the transformation techniques in stress-strength model and elicitation of prior distributions.

The utility of the method, however, is not limited to just the procedures and models discussed in Sections 2 and 3. It can be applied to virtually any area where one has to deal with a variety of parametric families of distributions. It can be argued that since transformation of variables is the routine part of almost any calculus-based statistics course, it would be worth introducing the methodology developed above into standard textbooks, thereby saving hundreds of hours which are spent on re-deriving inferential procedures for various distribution families.

Acknowledgments

Marianna Pensky was supported in part by the NSF grants DMS-0505133 and DMS-0652524. Naomi Brownstein was supported by a Student Mentor Academic Research Teams (SMART) grant from the Burnett Honors College at the University of Central Florida.

References

- [1] Awad, A.M., Gharraf, M.K. (1986) Estimation of $P(Y < X)$ in the Burr case: a comparative study. *Commun. Statist. – Simul. Comp.*, **15**, 389–403.
- [2] Berger, J. O. (1993) *Statistical Decision Theory and Bayesian Analysis*. Corrected reprint of the second (1985) edition. Springer-Verlag, New York.
- [3] Bernardo, J. M., Smith, A. F. M. (1994) *Bayesian Theory*. Wiley, Chichester.
- [4] G. Casella, R. L. Berger (2002) *Statistical Inference*. Second Ed., Duxbury, California.
- [5] Enis, P., Geisser, S. (1971) Estimation of the probability that $Y > X$. *J. Amer. Statist. Assoc.*, **66**, 162–168.
- [6] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994) *Continuous Univariate Distributions.*, Vol. 1. Second ed. Wiley, New York.
- [7] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995) *Continuous Univariate Distributions.*, Vol. 2. Second ed. Wiley, New York.
- [8] Kim, D.H., Sang, G.H., Jang S.C. (2000) Noninformative priors for stress-strength system in Burr-type X model. *Journ. Korean Stat. Soc.*, **29**, 17 – 27.

- [9] Konishi, S. (1991) Normalizing Transformations and Bootstrap Confidence Intervals, *Ann. Statist.* **19**, 2209–2225.
- [10] Kotz, S., Lumelskii, Y., Pensky, M. (2003) *The Stress-Strength Model and Its Generalizations. Theory and Applications.* World Scientific Co., Singapore.
- [11] Linton, O. B. , Chen, R., Wang, N., Hrdle, W. (1997) An Analysis of Transformations for Additive Nonparametric Regression, *J. Amer. Statist. Assoc.* **92**, 1512–1521.
- [12] Marron, J.S., Ruppert, D. (1994) Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **56**, 653–671.
- [13] Robert, C. P. (2001) *The Bayesian Choice. From Decision-theoretic Foundations to Computational Implementation.* Second ed. Springer-Verlag, New York.
- [14] Ruppert, D., Cline, D.B.H. (1994) Bias reduction in kernel density estimation by smoothed empirical transformations. *Ann. Statist.* **22**, 185–210.
- [15] Surles, J.G., Padgett, W.J. (1998) Inference for $P(Y < X)$ in the Burr type X model. *J. Appl. Statist. Sci.*, **7**, 225–238.
- [16] Thompson, R.D., Basu, A.P. (1993) Bayesian reliability of stress-strength systems. In *Advances in Reliability*, ed. Basu, A.P., Elsevier Science Publishers, Amsterdam, pp. 411–421.
- [17] Tong, H. (1974) A note on the estimation of $P(Y < X)$ in the exponential case. *Technometrics*, **16**, 625. Errata: *Technometrics*, **17**, 395.
- [18] Van Buuren, S. (1997) Optimal transformations for categorical autoregressive time series. *Statist. Neerlandica* **51**, 90–106.
- [19] Vidakovic, B. (2004) Transforms in Statistics, In: *Handbook of Computational Statistics Concepts and Methods, Chapter II.7.* Eds. Gentle, J., Härdle, W., and Mori, Y., Springer-Verlag, Heidelberg pp. 199–236.
- [20] Voinov, V.G., Nikulin, M.S. (1993) *Unbiased Estimators and Their Applications. Volume 1: Univariate Case.* Kluwer Academic Publishers, Dordrecht, Netherland.
- [21] Yang, L., Marron, J.S. (1999) Iterated transformation-kernel density estimation. *J. Amer. Statist. Assoc.* **94**, 580–589.
- [22] Yang, R. and Berger, J. (1997) A catalog of noninformative priors. ISDS Discussion Paper 97-42.