COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2020.1849722

Deconvolution of $\mathbb{P}(X < Y)$ with unknown error distributions

Cao Xuan Phuong[a] and Le Thi Hong Thuy[b]

[a]Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam; [b]Faculty of Basic Sciences, Van Lang University, Ho Chi Minh City, Vietnam

ABSTRACT
This paper is devoted to a nonparametric estimation of the probability $\theta := \mathbb{P}(X < Y)$, where X, Y are continuous univariate random variables of interest and observed with additional random errors. We focus on the case where the distributions of the random errors are unknown but symmetric around zero and can be estimated from some additional samples. Using deconvolution techniques, we propose an estimator of θ which depends on a regularization parameter. We then establish upper and lower bounds on convergence rate of the estimator under mean squared error when error densities are assumed to be supersmooth.

ARTICLE HISTORY
Received 9 May 2020
Accepted 4 November 2020

KEYWORDS
Deconvolution; characteristic function; supersmooth errors

MSC 2010
62G05; 62G20

1. Introduction

Let X, Y be continuous univariate random variables of interest defined on a same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose we have two independent samples $(X'_1, ..., X'_n)$ and $(Y'_1, ..., Y'_m)$ of independent and identically distributed (i.i.d.) observations drawn respectively from the distributions of X' and Y', where

$$X' = X + \zeta, \quad Y' = Y + \eta. \tag{1.1}$$

Here X, Y, ζ and η are assumed to be mutually independent random variables. The variables X' and Y' play the role of noisy versions of X and Y respectively, whereas ζ and η represent respectively random measurement errors with densities f_ζ and f_η, called error densities. We aim to estimate nonparametrically the probability

$$\theta := \mathbb{P}(X < Y)$$

on this basis of the samples.

The problem of estimating θ has been risen from many scientific fields, such as reliability and medicine. In stress-strength models of reliability theory (see, e.g., Kotz and Pensky (2003)), X represents stress impacting on a component and Y represents strength of the component. The component fails if the stress exceeds the strength, and θ is then defined as reliability of the component. In medicine (see, e.g., Zhou (2008)), θ is a global index used to evaluate the accuracy of a diagnostic test. For instance, if X and Y represent respectively diagnostic test measurements for a healthy subject and

CONTACT Cao Xuan Phuong caoxuanphuong@tdtu.edu.vn Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam